413

90th Congress
2d Session

JOINT COMMITTEE PRINT

FEDERAL PROGRAMS FOR THE DEVELOPMENT OF HUMAN RESOURCES

A COMPENDIUM OF PAPERS

SUBMITTED TO THE

SUBCOMMITTEE ON ECONOMIC PROGRESS

OF THE

JOINT ECONOMIC COMMITTEE
CONGRESS OF THE UNITED STATES

VOLUME 2

Part III. Income Maintenance and Family Support
Part IV. Health Care and Improvement
Part V. Housing and the Quality of Man's Environment

Printed for the use of the Joint Economic Committee

U.S. GOVERNMENT PRINTING OFFICE
88-744 WASHINGTON : 1968

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $1.25

JOINT ECONOMIC COMMITTEE

(Created pursuant to sec. 5(a) of Public Law 304, 79th Cong.)

WILLIAM PROXMIRE, Wisconsin, *Chairman*
WRIGHT PATMAN, Texas, *Vice Chairman*

SENATE

JOHN SPARKMAN, Alabama
J. W. FULBRIGHT, Arkansas
HERMAN E. TALMADGE, Georgia
STUART SYMINGTON, Missouri
ABRAHAM RIBICOFF, Connecticut
JACOB K. JAVITS, New York
JACK MILLER, Iowa
LEN B. JORDAN, Idaho
CHARLES H. PERCY, Illinois

HOUSE OF REPRESENTATIVES

RICHARD BOLLING, Missouri
HALE BOGGS, Louisiana
HENRY S. REUSS, Wisconsin
MARTHA W. GRIFFITHS, Michigan
WILLIAM S. MOORHEAD, Pennsylvania
THOMAS B. CURTIS, Missouri
WILLIAM B. WIDNALL, New Jersey
DONALD RUMSFELD, Illinois
W. E. BROCK 3D, Tennessee

JOHN R. STARK, *Executive Director*
JAMES W. KNOWLES, *Director of Research*

ECONOMISTS

WILLIAM H. MOORE — JOHN B. HENDERSON — GEORGE R. IDEN
DONALD A. WEBSTER (Minority)

SUBCOMMITTEE ON ECONOMIC PROGRESS

WRIGHT PATMAN, Texas, *Chairman*

HOUSE OF REPRESENTATIVES

MARTHA W. GRIFFITHS, Michigan
WILLIAM S. MOORHEAD, Pennsylvania
THOMAS B. CURTIS, Missouri
W. E. BROCK 3D, Tennessee

SENATE

WILLIAM PROXMIRE, Wisconsin
J. W. FULBRIGHT, Arkansas
HERMAN E. TALMADGE, Georgia
LEN B. JORDAN, Idaho
CHARLES H. PERCY, Illinois

GEORGE R. IDEN, *Economist*

CONTENTS

VOLUME 1

VOLUME 2

PART III. INCOME MAINTENANCE AND FAMILY SUPPORT

PART IV: HEALTH CARE AND IMPROVEMENT

PART V: HOUSING AND THE QUALITY OF MAN'S ENVIRONMENT

AUTHORS LISTED ALPHABETICALLY

Part III

Income Maintenance and Family Support

PROVIDING BASIC INCOME SECURITY FOR THE POOREST OF THE POOR

BY ELLEN WINSTON*

In a money economy, the most crucial need of the poor, and indeed the near-poor, is the possession of money with which to purchase essential goods and services. Moreover, some 30 million or more of the total population of the United States, distributed throughout the Nation and found not just in urban ghettos or on wornout farmland, face the daily predicament of little or no money. They do so in a period of unparalleled affluence. This has led to the widening discussion of how to provide an income floor for the economically deprived in our affluent society.

The concept of such a floor is not new. Almost 20 years ago, the then-Senator Robert A. Taft said in a policy address delivered in New York, "I believe that the American people feel that with the high production of which we are now capable, there is enough left over to prevent extreme hardship and maintain a minimum floor under subsistence, education, medical care, and housing, to give to all a minimum standard of decent living and to all children a fair opportunity to get a start in life." This was wishful, prospective thinking at the time. Today the critical questions are related not to whether we can afford to provide such security but rather to our willingness to do so and the most feasible mechanisms.

In 1964 in his message on poverty, President Lyndon B. Johnson substantially restated the goal:

> Today, for the first time in our history, we have the power to strike away the barriers to full participation in our society. Having the power, we have the duty.

We cannot actually eliminate poverty because poverty is a relative term. We can, however, eliminate the basic factors in extreme destitution—hunger, inadequate clothing, housing unfit for human habitation, untreated health conditions, the hopelessness and apathy of individuals that result from the daily struggle for survival with no real opportunity for either setting goals or moving however slowly toward their attainment.

We already have many ways besides public assistance, some more effective than others, of dealing with lack of or inadequate current income. Among them are minimum wage laws, unemployment insurance, social security benefits, veterans' benefits, public work and training programs, various types of housing subsidies, and surplus food programs. To the extent that these and other programs may be ex-

*Former U.S. Commissioner of Welfare and currently consultant to the Department of Health, Education, and Welfare and to the National Study Service.

NOTE.—The author alone is responsible for the point of view expressed in this article.

panded they affect the potentials for other approaches to basic economic security. If such programs are substantially expanded, the need for a floor under income involves fewer people. The effect of the expansion of old-age insurance in reducing the need for old-age assistance is a case in point.

In the United States we have many different philosophies about transfer payments, but we have the greatest difficulty in accepting those that are designed to provide for the daily necessities of life for the aged, the disabled, and the young, the unskilled, and the uneducated. This problem of ideologies has been summed up as follows:

> In relation to poverty, the myths of a classless society and a melting pot mobility pattern still have wide acceptance. * * * Holdovers from the Puritan ethic still dominate the patterns of welfare subsidies and dehumanize the poor with what almost amounts to vengeance. The possibilities of technology clearly demand a radical revision of these old attitudes and a commitment to provide, without charity, adequate food, clothing, and shelter to all men.[1]

The Poorest of the Poor

While poverty itself is relative, the extent of destitution within the total poverty population is also relative. There is no denying the fact, however, that the poorest of the poor are the recipients of public assistance, including general assistance, and those equally eligible in terms of need but not currently receiving aid. These are the men, women, and children whose sheer survival depends upon the public assistance check, a population at the bottom of the economic ladder which includes both a hard core of recipients and large numbers in marginal circumstances who come on and go off assistance as their conditions change.

The statistical picture, including both people and costs, has already been well documented.[2] However, in programs of such magnitudes, averages tend to conceal the acuteness of the problem. Recent data on aid to families with dependent children may be cited to illustrate the variations. In December 1966, the number of children aided by AFDC per 1,000 population under age 18 ranged from 111 in West Virginia to 19 in Texas. The average monthly payment per individual on this program ranged from $55.10 in New York to $9.25 in Mississippi.

States establish minimum standards of need for their public assistance programs, albeit generally too low to meet costs of a health and decency level of living. Then in varying degree they fail to meet their own standards through arbitrary ceilings on grants or percentage reductions of payments, restrictive eligibility requirements, inequity in providing for specially defined needs, and/or failure to price their standards at current costs. Such variations have meant that even within an inadequate total program, gross inequities exist from State to State and even within States in meeting the needs of the poverty stricken.

---

[1] Hall, Cameron P., *Human Values and Advancing Technology*, Friendship Press, Inc., New York, 1967, pp. 141–142.

[2] "Federal Programs for the Development of Human Resources," Joint Committee Print of the Joint Economic Committee, 89th Cong., 2d sess., especially vol. I, pp. 6–19, 76–80; vol. II, pp. 887–925. For further data, see Hearings Before the Committee on Finance, U.S. Senate, 90th Cong., 1st sess., on H.R. 12080, pt. 1, especially pp. 285 and 288.

The situation is compounded by the Federal matching formulas which discriminate against children—our productive human resources of the near future—and their parents as compared with the aged and disabled. Thus, the Federal Government will match, on a variable formula, up to an average of $75 per month in the adult categories, but only up to an average of $32 per month per child.

In support of this basic discrepancy, emphasis is generally placed on the rising cost in the Federal budget of public assistance. Actually this is a shared cost with the States and localities. Thus, through action by State legislatures and local governing bodies, non-Federal appropriations have also maintained a strong upward trend, roughly doubling since 1950.[3]

While not all the data one might wish for are available, due in large part to the lack of adequate research funds, there are sufficient indications from the information available and from empirical evaluation to provide support for attacking the problem through major changes in the present public assistance program. It is time for a large-scale overhaul as contrasted with efforts, however admirable and useful, to strengthen by amendments a basically unchanged approach to income maintenance.[3]

Strategy for Major Change in Public Assistance

As the concept of social responsibility for individual and family problems in our modern affluent society has become much more widely accepted, there has emerged a more general recognition of the need for certain basic guarantees—guarantees that include not only income maintenance but also a broad spectrum of social services. While they in essence represent two sides of the same coin, only public assistance payments are discussed here for two main reasons. First, public assistance is one of a variety of income-maintenance programs currently receiving careful study and hence needs to be viewed separately from social services for purposes of comparison. Second, no other income-maintenance proposal would do away with the residual need for a reoriented public assistance program of some undetermined size. Only a program that takes into account varying needs, as contrasted with uniform payments, can provide variable benefits that will protect individuals and families against special crises and emergencies. Furthermore, the public assistance administrative machinery, already in existence in every local jurisdiction, can practically and feasibly be reoriented to provide a guaranteed minimum income floor nationwide for all who require public support. Many of the required changes were first blueprinted in the report by the Advisory Council on Public Welfare.[4] They have since been elaborated on in a number of speeches and professional papers.

Among the criteria for a reoriented public assistance program are the following:

1. There would be only one criterion for financial aid; namely, need. The present categories, including old-age assistance, aid to the blind,

[3] See annual reports of Division of Research, Bureau of Family Services, Welfare Administration, U.S. Department of Health, Education, and Welfare, including reports for fiscal year ended June 30, 1966.

[4] See "Having the Power, We Have the Duty," U.S. Department of Health, Education, and Welfare, June 1966.

aid to the permanently and totally disabled, aid to families with dependent children, and general assistance, would be eliminated. In addition to persons presently eligible for financial assistance, those persons in need, but excluded under present Federal law, would be included in one comprehensive public assistance program.

2. There would be minimum standards, federally established, which each State would be required to meet as conditions for receiving Federal matching funds. These would include standards (1) for public assistance payments geared to an acceptable level of living and (2) for objective, impersonal determination of need.[5]

3. There would be full protection of the long-established principle of unrestricted money payments, crucial to the development of independence and self-respect.

4. States would continue to participate in the financing of the program but on a new, simplified plan for Federal-State sharing in costs. After objective determination of each State's share, taking into account relative fiscal ability, the Federal Government would assume responsibility for the difference between the State's share and the total cost of meeting the national standards.[6] This would support the objectives but would at the same time serve to maintain substantial State fiscal responsibility for the program.

There are many exponents of some new completely federally financed program of income maintenance, federally administered. In contrast this proposal for a reoriented public assistance program should lead to the strengthening of State administrative responsibility and accountability and at the same time preserve the principle of Federal-State sharing of the costs of the program.

Also important is the fact that this proposal would be sufficiently flexible in its operation to provide support at the specified level for those with no income or to supplement up to the standard those with inadequate resources or earnings to meet the established level of living. Necessary provisions would have to be incorporated so that people who can work or should work would do so, including a work-incentive formula.

The deficit between what people have in the way of income and what would be required to meet an established standard, such as the "poverty level," is basically the same regardless of method for filling the gap.[7] The issue is whether the focus should be on zeroing in so that every dollar counts in reducing destitution, as in an improved public assistance program. By taking a broader approach, either the payments must be so limited that the people in the poverty group will not be brought up to a decent level of living or far greater amounts of money will be called for.

Why the Proposed Plan

"There is growing recognition that every family or individual should have access at all times to purchasing power that will assure a socially acceptable minimum level of support; and that, to this end, steps should be taken to establish an effective income floor for all those

[5] A declaration or affidavit as the basis for eligibility determination has been acceptable to the Federal agency since early in 1967. It is, however, permissive rather than mandatory.

[6] Of course, any State could set higher standards, financing the excess entirely from non-Federal funds.

[7] While estimates vary according to type of income-maintenance program advocated, $12 billion is widely cited. Also see Orshansky, Mollie, "Recounting the Poor—A Five-Year Review." *Social Security Bulletin,* April 1966, pp. 10–12.

who are incapable of self-support, whether by reason of age, ill health, physical or mental impairment, family burdens, or other handicaps." [8]

By this definition there is acceptance of the fact that there will always be individuals incapable of self-support. Recognizing public assistance as a right, the basic problem is how to meet need at a decent level taking into account differing circumstances. As pointed out earlier, some of the methods for doing this were detailed in the report of the Advisory Council on Public Welfare. Others were issued in policy statements of the Welfare Administration of the Department of Health, Education, and Welfare, during 1966 and the early months of 1967. Still others are dependent upon legislative action.

The crucial strategies relate to children and youth, almost 4 million of whom are recipients of aid to families with dependent children in any given month, or somewhat less than one-third of the children in poverty.[9] More strikingly, one out of every six young people reaching the age of 18 has at some time received AFDC. While the period on assistance may have been short, the program's potential for helping children toward a better and more productive life is tremendous. So far the assistance in most States has been far too little, but were it geared to a national standard the beneficiaries would receive not only the essentials of decent living but also be encouraged to strive to attain higher goals.

The United States has no generally accepted public social policy with respect to the centrality of the family. It is true that public assistance laws refer to strengthening family life, even though both law and policy often negate this goal. However that may be, it is crucial that family life be improved for the millions of children, on and off assistance, in families with substandard levels of living. The integrity of family life is essential to the soundness of the Nation. In a money oriented society, cash is an essential to support the integrity of the family.

On every hand, there are complaints about inadequate manpower. Children born into extreme poverty are already gravely handicapped, and they quickly become retarded in many areas of functioning as a result of acute deprivation.[10] This cultural retardation becomes more marked as the child continues in school. Recognition of the problem is inherent in the efforts to improve and enrich the educational program for the child of the ghetto and other slum areas. It is not enough, however, to offer special programs away from home. The child spends not only the first years of life under the influence of home and family but also the major portion of his time even when he is attending school. It is essential, therefore, both to protect and to enhance human resources through making possible decent family living for all children and youth. In fact much of the money made available for education and health results in small dividends where home life is grossly substandard. The schools are greatly handicapped when children are

---

[8] "Federal Programs for the Development of Human Resources," *op. cit.*, vol. I, p. 79.
[9] Nicol, Helen D., "Guaranteed Income Maintenance—Another Look at the Debate," *Welfare in Review*, vol. V, June–July 1967, p. 3.
[10] Among many studies, the work of the National Laboratory on Early Childhood Education in examining intellectual development during the first 3 years of life is pertinent here.

malnourished, have little or no intellectual stimulation at home, are sent to class irregularly because of family malfunctioning. Health programs can accomplish far less, for example, when the child who entered the hospital as a result of problems associated with malnutrition returns to the same destitute home. Public assistance must indeed provide a floor below which no one will need to exist and make possible full return from other service programs.

In summary, too often the attention in public assistance has been upon the separate categories and upon the increasing cost. Rather the focus needs to be upon one comprehensive program, adequate in terms of both coverage and amount. So far as children, our human resources, are concerned public assistance should be viewed as an investment. As a properly administered investment at sufficiently high levels, it should produce returns in productive adults commensurate with those from the more fortunate children and youth of this Nation.

INCOMES OF THE ELDERLY

BY I. M. LABOVITZ*

CONTENTS

*Senior specialist in social welfare, Legislative Reference Service, The Library of Congress. Views expressed in this paper are those of the author and are not necessarily those of the Legislative Reference Service or the Library of Congress.

Introduction

The development of social security and other public programs in the United States during the last 35 years has pretty well demonstrated nationwide concensus upon this broad conclusion: That in our highly organized, interdependent economy, the competitive forces of the marketplace and the initiatives and resources of individuals cannot assure to the elderly population, considered as a group, an adequate or even a barely essential share of command over the goods and services currently available to the Nation as a whole.

Earnings and the yield on accumulated assets are major sources of current income for many elderly individuals. They are important sources for the elderly as a group, but they are not the leading income sources for the great majority of them. For many of these people, the possibilities of gainful employment are diminished by the concomitants of old age—the chronic afflictions, the dwindling of energies, and the obsolescence of skills or absence of special skills required for contemporary technology. At the same time, some of these developments in the lives of individuals create new and compelling consumption requirements, often more costly than the items of middle-aged expenditure which have been dropped from the individual or family budget.

A principal public policy expression of the developing social consensus has been the creation and expansion of public retirement-income systems and encouragement of private group pension plans. By far the dominant element in this still evolving structure is the national old-age, survivors, disability, and health insurance system. Within OASDHI, old-age, survivors, and disability cash benefits assure at least some monthly cash income to nearly all elderly persons and to many younger dependents of workers who are retired, severely disabled, or dead. Health insurance provides financing in an area of especially acute need for this age group—hospital, medical, and related services.

Efforts to diminish or overcome poverty have drawn special—though wavering—public support in the last several years. One byproduct of these efforts is a series of studies that count the numbers of the poor, measure the degrees of their poverty, and summarize other characteristics. As far as the elderly are concerned, these reports have emphasized that disproportionate numbers of the poor are elderly—and disproportionate numbers of the elderly are poor.

Economic Claims of the Elderly

As these and other relevant data draw public attention, basic long-term issues and propositions take on new significance.

In a comprehensive study published a dozen years ago, John J. Corson and John W. McConnell commented as follows:

> * * * Many of the problems that arise in relation to the aged reflect our inability as a people to adjust time-honored and persistent customs of our youth, when older people typically belonged to a large family group, to the more limited functions the family is able to perform in the civilization of the 1950's.[1]

[1] John J. Corson and John W. McConnell, "Economic Needs of Older People" (Twentieth Century Fund, 1956), p. 15.

Concerning the political implications of the problems, these authors wrote:

* * * In the face of the demographic trends of the past half century, the American people are confronted with the social necessity of inventing social institutions which will make the aged an asset, not a burden; which will view the aged as participants in our society rather than as cold statistics.[2]

At about the same time, Eveline M. Burns discussed attitudes toward the claims of the aged on the economy:

* * * The very fact which, in the long run, will increase the numbers in the productive age groups, namely an increase in the birth rate, involves in the short run a heavy burden on current earners of support and education until children enter the labor market. People may feel less generous toward the aged if they have to carry heavy costs of child rearing and be more likely to insist that public action be limited to provision for persons who are presumptively in need, i.e., who have retired from the labor market, and to frown upon proposals to pay automatic pensions at some specified age, especially if this is set relatively low.

Even more broadly, the sum total of all other demands (such as defense and other tax-financed expenditures) which cut into the freely disposable income of current earners will have an influence on the community's willingness to shoulder an additional burden in the form of a lowering of the pensionable age or complete abolition of a retirement test.

Attitudes to the claims of the aged as compared with other groups will also affect the community's response to the undeniable loss of output occasioned by retirement of the elderly prior to the time when their economic effectiveness falls to zero. For even if there is a general labor shortage, a favorable sentiment toward the aged as such may foster efforts to utilize other available labor reserves (such as employment of married women or recourse to overtime) rather than to raise the retirement age of public social security systems.[3]

In this connection it may be noted that Congress, since 1960, has lowered eligibility ages for monthly cash retirement benefits under old-age and survivors insurance to 62 for both men and women and to age 60 for widows (with reduced benefits to balance the longer duration of payments).

Retirement Benefits, Prices, and Wages

Individual benefit amounts have been increased from time to time by congressional action since 1950. The benefit increases during 1950–54 raised payments generally by somewhat more than was required to overtake consumer price advances that followed the previous adjustments of benefits. Benefit increases in 1959 and 1965 did not in either instance overtake consumer price advances since the preceding benefit increase.

Nine-tenths of those drawing benefits in 1967 retired after 1954, so that monthly payments to most beneficiaries provided less purchasing power in 1967 than when they retired. For the person who retired in 1954, the average monthly benefit in December 1966 was more than 7 percent below the amount needed to maintain parity with consumer prices. For the worker who retired in 1959, the average benefit fell short of its original purchasing power by about 4.5 percent.[4]

---

[2] Ibid, p. 16.

[3] Eveline M. Burns, Social Security and Public Policy (1956), p. 106.

[4] Saul Waldman, "OASDI Benefits, Prices, and Wages: 1966 Experience," Social Security Bulletin, June 1967, pp. 8–12, 36.

In adjusting for price changes, the Social Security Administration uses the Bureau of Labor Statistics Consumer Price Index. Because this index measures changes in prices paid by an urban worker's family, it is generally considered to be at best a crude general deflator for all family incomes and even less suited for adjustments related to elderly

In the year following December 1966 there was a further advance of 3 percent in the Consumer Price Index.

The 1967 Social Security Amendments provide benefit increases of 13 percent or more for all beneficiaries, effective in checks for February 1968, that will be distributed in March. For most beneficiaries, the result in 1968 should be some net gain in purchasing power over the amount awarded at the time of retirement. The analysis for 1966 suggests, however, that the positive margin will generally be small.

Comparisons with spendable wages and disposable personal income indicate a more substantial lag. In most years, spendable wages have risen faster than prices.

*** * * Even for the older beneficiaries, who have some relative advantage in terms of purchasing power, benefits have fallen far behind the increases in spendable wages over the relatively long period they have been on the rolls. Retired workers have therefore not shared to any appreciable extent in the increasing standard of living enjoyed by the working population.[5]**

The worker who retired in 1950 and was still on the rolls in December 1966 received in that month an average benefit payment that was 22 percent less than the amount needed to maintain the same relationship to spendable wages as the benefit he received when he retired. The worker who retired in 1954 was 27 percent behind on this basis, and the one who retired in 1959 was 16 percent behind.[6]

The comparison with disposable personal income was made in an earlier social security report, also following the 1965 benefit increases:

*** * * A beneficiary who came on the rolls in 1940 with a benefit equal to the average for that year and who is still drawing benefits now receives a benefit with purchasing power about 16 percent greater than that of his first benefit check. Over this same span of years, real per capita disposable income—that is, the average purchasing power of the entire population after payment of personal taxes—increased about 69 percent.[7]**

The gap between these measurements might be somewhat less in 1968, when further benefit increases take effect, but the data still would support a broad conclusion that the cash benefits of retired workers and their dependents under OASI have risen much more slowly than the incomes of the working population.

Disposable income comprises income from all sources. Monthly OASI benefits are only one source—though a major one—of income for the elderly. But other income sources of the elderly have lagged even more than OASI.[8]

---

families, whose consumption patterns differ substantially from those of younger families. Medical care prices have risen more rapidly than other prices, and their advance has a greater relative impact on expenditures of elderly families than of other families—or would have, to the extent that these expenditures are not financed through medicare. Waldman (pp. 10–11) notes these problems and evaluates the applicability of the Consumer Price Index to retirement benefits of elderly persons. He concludes that appropriate adjustments in the price index would not modify it significantly and that it is, therefore, a useful deflator. See also footnote 21, below.

[5] Waldman, loc. cit., p. 11. The wage index is based on Bureau of Labor Statistics data for annual average spendable weekly wages for production workers (no dependents) in manufacturing industries.

[6] Derived from Waldman, loc. cit., p. 11. Percentages used here measure the extent to which actual average benefits paid to these workers in December 1966 fell short of amounts required to maintain wage parity, whereas Waldman's higher percentages measure the relative increase in actual benefits that would be necessary to provide wage parity.

[7] Department of Health, Education, and Welfare, Social Security Administration, "Social Security Programs in the United States" (January 1966), p. 9.

[8] Because there are no comprehensive reports showing comparative amounts from the major sources of income of the elderly population over a period of years (or even for scattered years), the statement in the text is necessarily a conjecture, supported by data such as the following:

From 1947 to 1964, mean income from all sources (in current dollars) increased by 85 percent for families with a head 65 or older and for unrelated individuals 65 or

Benefits as an Inducement To Retire

Even though they fall behind advances in earnings, benefit increases under OASI and other retirement systems enhance the attractiveness of withdrawal from the labor market. The inducement thus given is reinforced by the earnings test for benefits (described in a later section). Prof. Lowell E. Gallaway, in an analysis contributed to another compendium issued by the Joint Economic Committee, suggests that—

> * * * changes in elderly labor force activity have been primarily voluntary in character and, consequently, * * * the experiments in income maintenance for the aged which were inaugurated in the 1930's have led to significant improvements in the economic position of the elderly.[9]

But the degree of voluntary reduction in labor force activity is conditioned by the institutional conditions surrounding it—and for nearly all but the highest income earners, the terms of eligibility for benefit payments are overwhelmingly persuasive in the decisionmaking process of each elderly individual. Gallaway recognizes this, for he observes:

> * * * If we are interested in developing an optimal old-age benefit system, it would seem that it is possible to increase the social welfare generated by the present system by offering elderly individuals a range of options beyond age 65 in place of the flat-rate-benefit system employed for all beneficiaries who elect retirement at age 65 or older.[10]

In the Social Security Amendments of 1967, Congress directed the Secretary of Health, Education, and Welfare to study the existing retirement test and proposals for its modification. This study is to include proposals for increases in old-age insurance benefit amounts on account of delayed retirement. The report and recommendations are to be submitted by January 1969, without waiting for the more comprehensive report to be made by a quadrennial advisory council, to be appointed in 1969 to survey the whole social security system.

Central Role of Social Security Program

For the Nation as a whole, the central problem of economic policy toward the aged is—as has often been observed—"the basic economic problem of the distribution of national product among productive and dependent groups." The allocation is made by a succession of decisions—the decisions of individuals, employers, unions, and government:

---

older. The average OASI benefit to a retired worker in current payment status increased over the same period by 212 percent (in current dollars). Mean income from sources other than OASI rose 61 percent for families and 40 percent for unrelated individuals.

When the comparison is based on dollars of constant purchasing power (reported in terms of the median rather than the mean), the increase in income from all sources during 1947–64 was 33 percent for families with a head 65 or older and 48 percent for unrelated individuals of this age. The average OASI benefit to a retired worker in current payment status advanced in this period by 134 percent. Measured in constant dollars, income from sources other than OASI appears to have declined—by 1 percent for the median family and by 31 percent for the median unrelated individual.

Yearly data on mean incomes in current dollars and median incomes in constant dollars, 1947–64, appear in Department of Commerce, Bureau of the Census. "Trends in the Income of Families and Persons in the United States," 1947–64, Technical Paper No. 17, by Mary F. Henson (August 1967), pp. 57, 62, 182, and 187. OASI benefits are from Waldman, loc. cit., table 2.

[9] Gallaway, "The Economic Impact of OASDHI on the Aged," in 90th Cong., first sess., Joint Economic Committee, Subcommittee on Fiscal Policy, Old Age Income Assurance: A compendium of papers on problems and policy issues in the public and private pension system (committee print, December 1967), pt. II (in press).

[10] Ibid.

* * * Employers' decisions as to whether they will hire or retire older workers, and workers' decisions as to whether they prefer to work or retire are determining whether, as a people, we are going to use the productive power of the aged and whether the aged as individuals will have the opportunity to contribute their economic worth as long as they can. Employers' and unions' decisions on provision of private pensions are affecting the distribution of income between worker and nonworker. Governmental decisions on provision of public benefits for the retired have raised tax levels materially and altered the distribution of income.[11]

Each of these categories of decision and action remains important. But developments of the last decade, resulting in extension and increase of OASDHI, now accord the dominant role to congressional decisions on this program and related public policies.

The war against poverty has aroused public interest in other types of measures, particularly minimum income guarantees, negative taxes, and Government guarantees of employment for all who wish jobs and cannot get them. President Johnson has appointed a commission on income maintenance to examine proposed programs and reforms. Whatever policy choices may emerge from these studies and discussions, it is predictable that OASDHI will remain for many years a cornerstone in the structure of programs for meeting needs of the elderly population.

Scope of This Paper

The remainder of this paper, for the most part, summarizes data from several recent compilations that bear on economic conditions and prospects of the elderly in the United States. These deal with (1) the numbers and age distributions, recent and prospective, of the elderly in the whole population; (2) the incomes of the aged in relation to consumer incomes generally; and (3) a few recent developments in OASDHI that have enhanced its importance to the elderly. Space and time limits prevent a more extensive and systematic presentation.

Data summarized here, and the more extensive detail in the studies from which they are drawn, are relevant for evaluations of present policies and programs and formulation of proposals for further action. Yet these materials are admittedly insufficient, by themselves, as a basis for broad conclusions and comparative judgments, and none are attempted in this paper. The objective, rather, is to provide some possibly helpful facts that are not yet widely disseminated and to call attention to important reference materials and studies pertinent to income maintenance and economic support for the elderly.

The Increasing Numbers of the Elderly

Lifespans and Life Expectancies

During the first two-thirds of the 20th century, the average lifespan of persons born in the United States has been substantially extended. Boys born in 1900 could be expected to live, on an average, 48.2 years if white and 32.5 years if nonwhite. Girls born in 1900 could be expected to live, on an average, 51.5 years if white and 35 years if nonwhite. By 1965, the average life expectancies for males were 67.6 years for whites and 61.1 years for nonwhites; and for females, 74.7 years for whites and 67.4 years for nonwhites.[12]

[11] Corson and McConnell, op cit., pp. 14–15.

[12] Data in this subsection are from the Statistical Abstract table cited as the source of table 1.

Although the number of years added to life was far greater for nonwhites than for whites, nonwhites still had appreciably shorter life expectancies than whites. In fact, nonwhite children had about the same average expectation of life at birth in 1965 as white children had at birth a quarter-century earlier, on the eve of World War II.

Similarly, substantial improvements appear in estimated life expectancies for young and middle-aged people, still with a lag for nonwhites behind their white contemporaries. At more advanced years, and especially for males, gains in longevity have been decidedly narrower, as examples in table 1 demonstrate. Recent dramatic advances in medical science and technology have indeed added many years to many lives, but most of these gains are for early and middle age—for persons who, in earlier decades, would not have lived to celebrate their 65th birthdays.

Among 100 white males born in 1835, only 39 were alive in 1900. Of each 100 of their sons born in 1865, 53 still lived in 1930. And of each 100 of *their* sons born in 1900, 66 were living in 1965.

Among white women, of each 100 born in 1835, 44 remained alive in 1900. Among their granddaughters, from each 100 born in 1900, 81 were living in 1965.

Nonwhites born in the United States in 1835 were, in a large proportion of cases, born in slavery. Only about one-fifth of those born in that year lived to celebrate a 65th birthday in 1900—19 of each 100 males and 22 of each 100 females. Of each 100 nonwhite males born in 1900, 50 were alive in 1965; and of each 100 females, 63 were alive in 1965.

Among persons born in 1900, the following percentage of each group lived 65 years or more:

	Percent
White female	81.1
White male	65.9
Nonwhite female	62.6
Nonwhite male	50.4

In short, the gains have been large—and their effect has been quite uneven. The odds favoring a long life are best for white females and worst for nonwhite males.

CURRENT AND PROJECTED NUMBERS

At present, 9.4 percent of the population of the United States—about one in every 11 persons—is 65 years old or older. The composition of the elderly population, shown in table 2, reflects the differences in life expectancies reported earlier. Elderly women are proportionately more numerous than elderly men, and elderly whites proportionately more numerous than elderly nonwhites. This appears clearly in percentages for 1960 and 1965 in table 3.

Population projections issued recently by the Bureau of the Census indicate that over the next 25 years the population aged 65 and over will remain at least as large a proportion of total population, and may rise somewhat more rapidly than total population. Separate projections by sex and race indicate that, within the total, differences of trend may be expected. In all the projections, women aged 65 and older are expected to become a larger proportion of the whole population—with a substantial relative increase if the rise in total population decelerates.

The proportion of elderly women will rise more rapidly for nonwhite than for white women. For men aged 65 and older, both white and nonwhite, the projections indicate a substantial percentage decrease if total population grows rapidly, and a slight percentage increase if it grows slowly.

For the quarter century from 1990 to 2015, the projections suggest a change in direction. They show the numbers aged 65 or older continuing to rise but at a rate less rapid than total population or, at most, maintaining their proportions at the relative levels of 1990. If total population grows rapidly over the next 50 years, the elderly population may fall to as little as 7.1 percent of the total in 2015 (2.9 percent males and 4.2 percent females). If total population grows slowly, the elderly population in 2015 may hold the level of 10.6 percent projected for 1990, with only a slight change in its composition (4.4 percent males and 6.2 percent females in 2015, against 4.3 percent males and 6.3 percent females in 1990).

The projections indicate for 1990 a ratio of 148 women aged 65 or older to every 100 men of this age. For 2015, they indicate a ratio of 142 women to every 100 men 65 or older. These ratios are substantially above the current numbers: In 1966 there were 131 women to every 100 men aged 65 or older.

The Census Bureau projections are presented in four series, all based on assumptions designed to be consistent with a continuing high level of economic activity over the next five decades. The differences are in assumptions about fertility rates, with series A assuming the most rapid growth of future population and series D the least rapid growth.

ASSUMPTIONS ABOUT LONGEVITY

The same mortality rates are used in all four series of projections. The assumption is that mortality rates will be "slightly declining"—an assumption "intended to reflect a relatively pessimistic view" of the future course of mortality. An underlying consideration is that "the momentum of the pace of improvement in death rates has slowed down considerably in the last decade or so," and that "recent mortality trends for chronic diseases * * * retard any reduction in the overall death rate."

> * * * To achieve a significant increase in life expectancy, it would be necessary to accomplish a major "breakthrough" in the prevention and treatment of the major diseases characteristic of later life * * *.

That kind of achievement, the Bureau observes, "is not yet in sight." [13]

All persons who will attain age 65 in the next half century are already alive. With uniform assumptions as to mortality and migration rates, the four series of population projections carry identical numbers for all the older age groups. But the ratios of these age groups to total population differ among the series because the estimates of total population differ. (Table 2 gives series A and series D projections for total population of all ages, subdivided by sex and race.) These percentages are highest for series D (the projections with slowest population growth) and lowest for series A. Accordingly, a range of projected percentages is shown for each specified age group

[13] Department of Commerce, Bureau of the Census, Population Estimates, series P–25, No. 381 (Dec. 18, 1967), pp. 36–37.

in tables 3 and 4. The base in every case is the total for the class to which the figures refer. For example, male population aged 65 or over in 1990 is projected at 7.4 to 8.7 percent of the total male population in that year. The number of white men 65 and over in 1990 is projected at 7.9 to 9.2 percent of the total number of white males in that year. The number of nonwhite males 65 and over in 1990 is projected at 4.2 to 5.4 percent of the total population of nonwhite males.

Since the mortality assumptions are deliberately "pessimistic," numbers of aged shown for future years in table 2 and the percentage ranges in table 3 and 4 might appear to be low estimates. That inference as to the percentages is not necessarily valid, however, for it would be unrealistic to assume significant mortality reductions for the elderly without some accompanying improvements for younger people. Higher survival rates at earlier ages might indirectly affect fertility rates. Consequently, the application of more optimistic mortality rates might not alter materially the percentages of the population in older age groups. Still, the experience of recent decades has shown that the increase in the number of persons surviving to a given age is likely to be larger at age 65 than at ages 40 or 20. The relatively greater effect on numbers in older age groups appears, also, in Census Bureau illustrations of the implications of different mortality assumptions.[14]

In any event, the mortality assumptions are crucial to estimates of the future size of the elderly population and consequently the relative magnitude of income-maintenance needs in coming decades. If more persons survive beyond their 65th birthdays into their seventies, eighties, or beyond, the aggregate income-maintenance needs for older people will be significantly altered.

The slightly declining mortality rates assumed in the Census Bureau projections represent (with small modifications) the higher of "high" and "low" mortality projections prepared in 1957 by the actuary of the Social Security Administration. In these projections, average life expectancy of the whole population at age 60 is assumed to be longer by one-fifth of a year in 1975 and 0.9 of a year in 2000 than it was in 1965. The greatest increase would be for nonwhites—1.2 years added to the life expectancy of males at age 60 and 1.4 years for females.[15]

If population projections were based on an assumption of rapidly declining mortality rates, that is, on the "low" mortality rates in

---

[14] In assessing the effects of alternative mortality assumptions on population projections, the Census Bureau explained that the assumption of slightly declining mortality results in about 2.1 million more persons in 1990, or a population 0.7 percent larger than if mortality had been assumed to remain constant at 1963 levels. About half of the difference (more than 1 million persons) is in the age group 65 years and over, so that the estimated 1990 population in this group, 27 million, would be 3.8 percent lower if constant mortality rates had been projected. If "rapidly declining" mortality had been assumed, the whole population projected for 1990 would have been 1.3 percent higher than it was. Three-fifths of the difference would have been in the older age group. This difference, 2.2 million persons aged 65 and over, would have made the 1990 projection 29.2 million persons 65 and over. The proportionate changes would be largest for elderly males. For them, the constant mortality assumption gives a 1990 projection 5.1 percent lower, and the rapidly declining mortality assumption gives a projection 11.8 percent higher, than the assumption that was used for the basic projections. Ibid., pp. 39–41. See also Department of Commerce, Bureau of the Census, Statistical Abstract of the United States, 1967, p. 55, tabulating effects of past changes upon several age groups.

[15] Bureau of the Census, Population Estimates (series P–25, No. 381), p. 39, table Y. Life expectancies at age 60, projected for the year 2000, are 16.9 years for all males, 20.6 for all females. The Social Security Administration in its later projections shows a revised "low-cost" (high-mortality) estimate of life expectancy in the year 2000 at age 60 at 17 years for all males and 20.6 years (20.55 instead of 20.58) for all females. Department of Health, Education, and Welfare, Social Security Administration, Office of the Actuary, "United States Population Projections for OASDHI Cost Estimates" (Actuarial Study No. 62, December 1966), p. 14, table 6. The table compares estimates in the 1957 and 1966 studies.

the 1957 Social Security Administration study, average life expectancy of the whole population at age 60 would be about 3⅓ years longer in the year 2000 than it was in 1965. A newer actuarial study reduces the assumed improvement to about 2 years. On this basis, life expectancy for males at age 60 in the year 2000 would be 2.3 years greater than in 1965, and for females, 1.7 years. For the whole population, the assumed improvement over 1965 would be about 1.6 years at age 65 and about 1.1 years at age 70.[16]

This degree of improvement in life expectancies, though substantial, is not necessarily unattainable or fanciful. In a recent 35-year period, from 1930 through 1965, life expectancies at ages 60 and 65 actually gained nearly 2½ years.

The trend toward increased longevity may indeed be "approaching a limit," as the Census Bureau report suggests, but recent projections by both the Bureau and the Social Security Administration are strongly influenced by mortality experience of the period 1954–66. This is described as a period in which there was a deceleration of the earlier gains in mortality rates, especially at the older ages, "where there are now indications of possible increases in future mortality." [17] But 1954–66 was a period also in which an unprecedented volume of health-related research was undertaken, while only beginnings were made in such areas as the systematic application of new health knowledge to daily life, increase in the numbers of trained health personnel, and the expansion of health-supporting facilities and services oriented to the particular needs of the elderly. This was, at the same time, a period in which the Nation recognized and undertook to meet special housing needs of the elderly. Universal health insurance as a means of financing health care for the elderly population did not come into operation until 1966. Measures inaugurated to improve the lot of low-income people of all ages—better housing, nutrition, education, disease and accident prevention, and opportunities to earn income—also are relevant to long-term prospects for mortality rates in the upper age groups.

Implicit in the assumption of slightly declining mortality rates is a conclusion that all the recent advances in medical sciences and arts and all the other recent improvements in living conditions will contribute little to prolonging the lifespan of people now alive—or else that detrimental developments, such as deterioration of the physical environment, will cancel nearly all the gains.

Each of these recent developments taken separately may have only a small impact on longevity. Nor should rapid or spectacular results be expected from all of the favorable together. Yet the list itself suggests that prospects for some further mortality gains in the older ages may have been enhanced in a degree that is not reflected in vital statistics of 1954–66 and in the official population projections based on those statistics. Whether these possibilities should be heavily discounted in assessing the future probable size and income needs of the elderly population is at least debatable.

---

[16] Social Security Administration, Office of the Actuary, op. cit., p. 14, table 6; Bureau of the Census, Population Estimates (series P–25, No. 381), p. 39, table Y; and Department of Health, Education, and Welfare, Public Health Service, Vital Statistics of the United States, 1965, vol. II (1967), pp. 5–10. Combined rates for males and females in the year 2000 are my interpolations.

[17] Social Security Administration, op. cit., p. 12; Bureau of the Census, Population Estimates (series P–25, No. 381), p. 38.

THE OLDEST OF THE ELDERLY

The population projections, even though they employ "pessimistic" mortality assumptions, show rising numbers at ages 75 and over. An "optimistic" assumption about mortality rates—that is, an assumption of "rapidly declining" mortality—would have its greatest relative effect on this age group. Such an assumption is not reflected in any of the projections summarized here.

Women of these ages would be a larger proportion of the whole population in 1990 than in 1966, even if the total grows rapidly. The projected rise is relatively larger for nonwhite than for white elderly women, though the nonwhite numbers remain small.

Men aged 75 and older—white and nonwhite—would be a smaller percentage of the whole population in 1990 if the total rises rapidly. They would barely hold their proportion even if total population increases more slowly.

In 1966, the population aged 75 or older consisted of 142 women to every 100 men. The projections indicate that the ratio will climb above 170 to 100 in the years 1990 and 2015.[18]

INCOMES OF THE ELDERLY IN 1966 AND EARLIER

INTRODUCTION

Incomes of elderly families and unrelated individuals are substantially below those of the younger population and derived in much smaller part from current gainful employment. (An "unrelated individual" is one who lives alone or in a household where he has no relatives.)[19]

The latest collection of comprehensive income data for the aged is a Bureau of the Census report on income in 1966 of all families and persons in the United States. It is based on inquiries in the current population survey conducted in March 1967—one of a long series of annual reports derived from field surveys of a representative sample of households. These indicate the size and distribution of money incomes of the whole population in the calendar year 1966 in relation to age and other individual characteristics in March 1967. Some tabulations are for units consisting of families and unrelated individuals; those that differentiate income-recipients by sex use "persons" as the reporting units.

Special supplementary tabulations provide an analysis of incomes of the poor in 1966, using indexes of poverty and near poverty developed by the Social Security Administration. These tabulations, prepared by the Bureau of the Census and reported by the Social Security Administration, include selected data for persons aged 65 and older, additional to information for this age group in the Census Bureau report of the current population survey.[20]

---

[18] See tables 2, 3, and 4.

[19] Except as other sources are specified, data in this section are from Department of Commerce, Bureau of the Census, Current Population Reports, Consumer Income, series P–60, No. 53, Dec. 28, 1967, "Income in 1966 of Families and Persons in the United States."

[20] For data from the special tabulations and comparisons with earlier years, cf. Department of Health, Education, and Welfare. Social Security Administration, Office of Research and Statistics, Research and Statistics Note No. 23–1967, "Who Was Poor in 1966?" by Mollie Orshansky (Dec. 6, 1967). This study is cited hereafter as Orshansky, op. cit.

MEDIAN INCOMES

One-half of all U.S. families reported money incomes of $7,436 or more for 1966. But for families headed by a person aged 65 or more, the median money income was $3,645—less than half as much as for all families. Money income was less than $2,500 for one of every eight families of all ages—but it was less than $2,500 for one of every three elderly families.

For all unrelated individuals, median money income was $2,270. For these aged 65 and older, the median was $1,443, or less than two-thirds as much.

Table 5 gives distributions by amount of total money income. It shows data not only for families headed by persons 65 or older and for unrelated individuals in this age group, but also for families and unrelated individuals in the next younger group, 55 to 64. Whereas median incomes of the elderly in 1966 were significantly below those for the whole population, median incomes for ages 55 to 64 were above the medians for the whole population—and especially so for unrelated individuals. The comparative amounts and percentages dramatize the degree to which old age is a period when incomes are sharply below those of younger contemporary families and individuals.

In reporting on family and individual incomes in 1966, the Census Bureau pointed out that median family income, at $7,436, was at a new peak. In current dollars, the median amount in 1966 was 124 percent greater than in 1950. Adjusted for price increases, the gain in real income was about 66 percent from 1950 to 1966.

The new high level for the median of all family incomes was accompanied by a new peak at a much lower level for older families. The median amount in current dollars was almost double that of 1950. Adjusted for price advances, the gain in real income for the median elderly family was 41 percent from 1950 to 1966.[21]

THE POVERTY INDEX

Poverty, as defined by the Social Security Administration for its periodic reports on the numbers and progress of the poor, is measured by an income standard that takes account of family size, composition, and farm or nonfarm place of residence. The Department of Agriculture "economy food plan" is used as a basic element in calculating family needs for food (estimated for 1966 at 75 cents a day for each person in an average four-person family). For all family living items other than food, the SSA index adds twice the minimum amount required for food. The index postulates cash income needs for a farm family at 30 percent less than for a similar nonfarm family. For living arrangements that typify the elderly—individuals living alone and two-person families—some allowance is made for the fact that consumption needs are not strictly proportionate to household size. Except to allow for rising prices, the poverty index has been applied without change to income data for each year since 1959.

---

[21] Data for elderly families, 1950, are derived from Bureau of the Census, Technical Paper No. 17, p. 62 (for a full citation, see footnote 8 above). Data for 1966 are from Bureau of the Census, series P–60, No. 53, cited earlier, especially pp. 2 and 24.

In adjusting for price changes, the Bureau of the Census used the Consumer Price Index and noted its shortcomings as a deflator for family income data. See Technical Paper 17, noted above, p. 33. See, also, footnote 4 above.

A related measure of near poverty (sometimes designated a "low-income level") uses a "low-cost food plan" instead of the "economy food plan." This index assumes living costs about one-third higher than at the poverty line.[22]

In 1966, among nearly 30 million persons living in households with incomes below the poverty level, 18 percent were 65 or older. This was a disproportionate ratio, in light of the fact that persons 65 or older were only 9.4 percent of the whole population.[23]

More than one-third of the one- or two-person households of the elderly were poor in 1966, compared with less than one-seventh of all other households. Almost half of all elderly households were among the poor and the near poor.[24]

In nonfarm households of persons 65 or older, poverty in 1966 was much more frequent in households headed by women than in those headed by men. The numbers of these poor households and the incidence of poverty are shown in table 6, drawn from a compilation for the 1968 Annual Report of the Council of Economic Advisers. In nonfarm households of persons 65 or older, income in 1966 was below the poverty line in one-fourth of all households headed by white men, in one-half of all households headed by white women or by nonwhite men, and in more than two-thirds of all households headed by nonwhite women.

Moreover, although the incidence of poverty diminished in all these categories between 1959 and 1966, the greatest relative decline was in nonfarm households headed by elderly white males. This was the category with the lowest incidence in 1959, as well as 1966.

The summary comparison indicates that there was an increase in the number of poor households among the nonfarm elderly—from 3.9 million in 1959 to 4 million in 1966—during a period when the total number of nonfarm households in poverty declined substantially. As a result, households of the elderly comprised an increased proportion of all poor households.

Somewhat more detailed comparisons for 1959 and 1965, covering both farm and nonfarm households, confirm that households of the elderly were a higher proportion of all poor households in 1965 than in 1959—nearly 37 percent, compared with 33 percent.

Increased incomes are derived primarily from increased earnings, and these come from more jobs and higher pay rates—gains that are not fully shared by the elderly population. As Miss Mollie Orshansky has observed:

> **The fact that aged men and women are less likely to work regularly than younger persons and that they earn less when they do work is the main reason why poverty is so much more prevalent among the aged.[25]**

---

[22] For the composition of the Social Security Administration indexes of poverty and low income, and their application to income data from current population reports, see Social Security Bulletin, January and July 1965, and April, May, and December 1966.

[23] Derived from Orshansky, op. cit., table 3; for population percentage, see table 3 in this paper.

[24] Orshansky, op. cit., p. 8, states that "two out of five households consisting of one aged person or an elderly couple fell below the poverty line, compared with but one in seven of all other households." The numbers and percentages are not explicitly specified on this basis in Miss Orshansky's tables, but approximations from her table 4 indicate that in 1966 some 35 percent of all households of the elderly had incomes below the SSA poverty level and 12 percent had incomes above that level but below the low-income level. For all other households, the proportions were 14 and 6 percent.

Wilbur J. Cohen, "Improving the Status of the Aged," in Social Security Bulletin, December 1966, p. 5, using earlier data, observed that "A significant proportion of this Nation's poor are aged (16 percent), and a large proportion of the aged are poor (31 percent)."

[25] Orshansky, op. cit., pp. 6 and 9 and table 12.

As the incidence of poverty is narrowed by advancing prosperity, a rising proportion consists of groups who are especially vulnerable. The elderly constitute one of these susceptible groups. The women among them are especially apt to have low incomes—and the older they are, the smaller are their incomes.[26]

The low-income "gap"—the dollar-deficit of actual incomes below amounts needed to reach the threshold of poverty—was estimated for all the poor at $13.7 billion in 1959 and $11 billion for 1965. For households of persons aged 65 and older, the amount was $3.3 billion in 1959 and $2.6 billion in 1965, slightly less than one-fourth of the total in each year. Measured in dollars, the gap declined for both elderly families and unrelated elderly individuals—but the percentage for unrelated elderly individuals was higher in 1965 than 1959.[27]

INCOME GAINS, 1947–66

Real incomes of elderly families improved significantly over the last two decades—but not so dramatically as for the population as a whole. Adjusted for price changes,[28] upward shifts in family incomes (expressed in 1964 dollars) may be indicated by the summary comparisons that follow. It should be emphasized that the first four blocks give cumulative percentages. Percentages in the last three blocks comprise 100 percent for each year in each column.

Family income for year (1964 dollars)	Percentage of families[29]	
	Head aged 65 or older	All ages
Under $1,000:		
1947	21	7
1964	5	3
Under $2,000:		
1947	42	17
1964	23	9
Under $3,000:		
1947	56	31
1964	44	18
Under $4,000:		
1947	67	47
1964	58	26
$4,000 to $10,000:		
1947	26	46
1964	31	52
$10,000 and over:		
1947	7	7
1964	11	22

[29] Percentages in the last 3 blocks comprise 100 percent for each year in each column. The first 4 blocks show cumulative percentages for each year in each column, based on the total for the column as 100 percent. The data are drawn from Department of Commerce, Bureau of the Census, "Trends in the Income of Families and Persons in the United States, 1947–64," Technical Paper No. 17, by Mary F. Henson (August 1967), table 3.

[26] Cf. Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, "The Aged Population of the United States: The 1963 Social Security Survey of the Aged" (Research Report No. 19, by Lenore A. Epstein and Janet H. Murray, 1967), p. 53, chart 3.4. This study is discussed in a later section. It is cited hereafter as Epstein and Murray, "The Aged Population."

[27] Orshansky, op. cit., pp. 11–12 and table 12. Aggregate deficits, as calculated, represent a needs-resources gap remaining after payments of public assistance, OASDHI, and other public programs, as well as all other money income. (Ibid., p. 11.)

[28] See footnotes 4 and 21, above, on use of the Consumer Price Index for these adjustments.

Price-adjusted distributions for elderly families are not available for 1965 and 1966, but comparisons without adjustment are as follows:

Family income for year (current dollars)	Percentage of families [30]	
	Head aged 65 or older	All ages
Under $1,000:		
1964	5.1	3.2
1966	3.7	2.3
Under $4,000:		
1964	57.6	26.0
1966	54.3	21.1
$4,000 to $10,000:		
1964	31.0	51.6
1966	34.6	49.3
$10,000 and over:		
1964	11.4	22.5
1966	11.1	29.6

[30] Derived from Bureau of the Census Current Population Reports, series P–60, Nos. 46 (1965) and 53 (1967), p. 24 in each report.

From these data it appears that incomes of elderly families as a group continued to rise, but far less rapidly than for the population as a whole—and an increasing proportion of the elderly was left behind in the general advance. Among the elderly, the proportion of families with incomes under $4,000 declined by 3.3 percent of the total during 1964–66, and those with incomes of $4,000 to $10,000 rose by 3.6 percent. The percentage with incomes above $10,000 actually declined. Fewer than one-half of all elderly families had incomes of as much as $4,000 in 1966 (the median for all elderly families was $3,376), whereas nearly four-fifths of all U.S. families had incomes above $4,000 and nearly one-third had more than $10,000.

Elderly families with less than $4,000 income in 1966 comprised 36 in every 100 families at that income level. In 1947, when low incomes were more common, elderly families with less than $4,000 comprised fewer than 17 in every 100 families at that income level.

COMPARISONS OF WHITE AND NONWHITE FAMILIES

Median income for all white families, without adjustment for price changes, had risen from $3,445 to $7,722. On the same basis, median income for nonwhite families had risen from $1,869 to $4,628. The gap between the white and nonwhite median amounts was somewhat narrower in 1966 than during the two preceeding decades, but the median amount for nonwhite families still was only three-fifths as high as for white families.[31]

The Census Bureau tabulations for 1966 do not show separate income distributions and medians for white and nonwhite families headed by elderly persons. In all likelihood, the gap is proportionately less than for younger families. One reason is the major importance of social security benefits as a source of income for all the elderly. Another (and closely related) reason is that nonwhite families above the

[31] The ratio for Negro to white median family income in 1966 was 58 percent, compared with 60 percent for all nonwhites, but data for Negro families separately are available only from 1964. Department of Commerce, Bureau of the Census, and Department of Labor, Bureau of Labor Statistics, "Social and Economic Conditions of Negroes in the United States," October 1967, p. 15.

poverty line are more likely than white families to be older families without children.[32]

FULL-TIME WORKERS

Generally, a family headed by a full-time worker has more income than other families. As table 5 indicates, the median 1966 income for all families headed by full-time workers (excluding the Armed Forces) was $8,693. This was 17 percent above the median for all families. Where the family head was a full-time worker aged 65 or older, the median (from all sources—not earnings alone) was $4,956, and this was 37 percent above the median for all families headed by older persons. But, as the table also indicates, only 15 percent of the older family heads were full-time workers, whereas two-thirds of all other family heads were full-time workers.

In the case of unrelated individuals, median income for a full-time worker 65 or older was more than 2½ times the median for all unrelated individuals in this age group and, in fact, was substantially above the median income of unrelated individuals of all ages. But only 7 percent of unrelated individuals aged 65 and over were full-time workers, compared to 35 percent for all ages (excluding the Armed Forces).

Tables 7 and 8 present data for persons, rather than families or households. They indicate that most persons aged 65 and older had some money income. Less than 1 percent of all males and about one-sixth of all females in this age group reported no money income in 1966. Among elderly men, fewer than 15 in each 100 were year-round full-time workers, and among elderly women, fewer than 5 in 100. Those few women who worked full-time after age 65 had median money income (from all sources—not earnings alone) well above the median for full-time women workers generally; but the elderly male full-time workers had median incomes (from all sources) considerably below the median for all full-time male workers. Money incomes of these full-time workers were substantially lower, as a rule, than the incomes of full-time workers of the same sex who were in the next lower age bracket, 55–64. Moreover, the incomes of older persons may be somewhat overstated in these comparisons, since some persons who were 65 years old in March 1967, had not reached that age in 1966, the year for which incomes were reported.

INCOME SOURCES

Elderly persons derived their money incomes not primarily from current earnings but from these and a variety of other sources. Among other sources are old-age, survivors, and disability insurance benefits; railroad retirement benefits; other public retirement or benefit systems; interest on savings or bonds; dividends; distributions from estates or trust funds; property income, such as rentals; public assistance or other welfare payment; and contributions from relatives and others outside their households. The major sources, particularly Federal Government programs, are considered separately in a later section.

Even including full-time earners, persons 65 and older had much

[32] Among nonwhite families in poverty in 1966, about 8 percent were two-person families with a head aged 65 or older. Among white families in poverty, about 25 percent were in these circumstances. Ibid, p. 23.

lower money incomes than the population generally. Elderly income recipients among men had median incomes in 1966 of $2,162, barely two-fifths as much as the median for all male income recipients who were 14 or older in March 1967; and among women, $1,085, or two-thirds of the median for all female recipients 14 or older. For persons living on farms the discrepancy was less, but median money incomes on farms were barely half of the comparable nonfarm medians.[33] Among men, the number of income recipients aged 65 and older was only slightly less than the number aged 55–64. but the proportionate difference in numbers was greater among farm than nonfarm residents. Among women, the number of income recipients aged 65 and older was considerably greater than the number aged 55–64, and the difference was substantial for both farm and nonfarm residents. (See table 7.)

The distribution of men and women income recipients among various income levels is shown in table 8. The comparative concentration of elderly persons in lower income brackets is evident in this table, as it is for families and unrelated individuals in table 5.

With the median amount of money income at $2,162 for men 65 and older, nearly one-third of these men had 1966 incomes of less than $1,500, and more than two-thirds had less than $3,000. For women aged 65 and older, the median amount was half as much—$1,085. More than two-thirds of these women had less than $1,500 income for the year, and nine-tenths had less than $3,000.

Only 4 percent of the older men received incomes of $10,000 or more, compared with 14 percent of all men. High incomes were much more rare for women of all ages. Among elderly women, 0.8 of 1 percent had incomes of $10,000 or more, compared with 1 percent of all women income recipients.

EDUCATION AND INCOME

Nowadays, persons of 65 and over typically have less formal schooling than do their juniors. Table 9 illustrates for several population groups that there is a positive association between years of schooling and total income from all sources.

For elderly men, total money incomes in 1966 ranged from a median of $1,738 for those with less than 8 years of elementary school to $6,433 for those who had 5 or more years of college. For elderly women, although the absolute amounts of income were about half as much as for men, the range was proportionately even wider—from a median of $870 for those with less than 8 years of elementary school to $3,500 for those with 5 or more years of college. Quite likely, the amount of money incomes of widows is directly associated with the number of years of schooling completed by their husbands, as well as with their own education.

Persons who were aged 55–64 in March 1967 (when the 1966 income data were collected) typically had considerably more formal education than those 65 and older. The median period of school attendance was 1.4 years longer for men (halfway through high school) and 2.1 years longer for women (nearly through the junior year of high school).

[33] Among income recipients, the proportions aged 65 and older were about the same for farm and nonfarm residents—14 percent on farms and 15 percent of the much larger number of nonfarm income recipients.

Whether this additional schooling will be associated with higher money incomes after age 65 for this next generation of the elderly is necessarily problematical. Available data suggest that a positive correlation will persist. Several factors and relationships are involved. Not least of these are some of the lasting effects of higher levels of income experienced in earlier years. These higher incomes may have been derived in more instances from occupations that permit a longer working life. They have financed better living conditions than were available to the current elderly population, including better nutrition and health care, thus contributing to individual abilities to continue longer in gainful employment. Also, the higher earlier income may serve as a foundation for higher retirement benefits and property income in old age.

Social Security Survey of the Aged

INTRODUCTION

In the Census Bureau compilations, data on the incomes and other economic characteristics of the elderly are incidental to surveys of the whole United States population. The first major nationwide survey focused specifically on the elderly population was conducted by the Office of Research and Statistics of the Social Security Administration. Information was collected early in 1963 and most of the data—particularly on incomes—are for the calendar year 1962. Findings from the study were used in formulating legislative proposals that led to the 1965 and 1967 amendments of the Social Security Act, including provisions for medicare and medicaid.

A comprehensive report, published in 1967, presents information about the size and sources of income and the assets of all Americans aged 62 and older, their living arrangements, medical expenditures, and other major aspects of their lives.[34] Like the Current Population Reports, the Social Security study was based on a field survey of a representative sample of all the aged. To the extent feasible, data were verified by comparison with old-age, survivors, and disability insurance operating records. Throughout the report, data are given separately for beneficiaries under the old-age, survivors, and disability insurance program and for nonbeneficiaries.

The 1963 survey is regarded by the Social Security Administration as an important benchmark from which to measure the economic progress of the elderly, particularly changes in their status following introduction in 1965–66 of the Federal Government program of health insurance for the elderly. In conjunction with later surveys, this compilation should be useful, also, in assessing the effects of increases in social security benefits under legislation enacted in 1965 and 1967.

It is not feasible to summarize here even the main findings of the Social Security survey, but it may be useful to indicate the principal sources of income and their relative importance for specified groups of elderly people. This is a point on which comparable details are not available in the Census Bureau survey reviewed in earlier sections or from other studies.

[34] Epstein and Murray, "The Aged Population." For a complete citation, see footnote 26 above.

Couples were considered 65 or older if the husband had attained that age.[35] Nonmarried persons comprised those who were widowed, divorced, separated, or never married.

MAJOR SOURCES OF INCOME

Information on sources of income is summarized in table 10 and on amount of income in table 11 (pp. 370–371).

Among elderly couples who reported on income sources, 55 percent had earnings from employment during 1962. Among nonmarried persons, 28 percent of the men and 23 percent of the women had earnings. The proportions were higher among couples and nonmarried men who were not beneficiaries of old-age, survivors, and disability insurance than among beneficiaries. In fact, earnings were the primary source of income for these nonbeneficiaries, yielding 69 percent of all income for nonbeneficiary couples and 48 percent for nonbeneficiary nonmarried men. The median amount for these men was $3,470.[36]

Among nonmarried women, however, the distribution differed. Income from earnings was reported by one-third of the retired beneficiaries, compared with only one-sixth of the widowed beneficiaries and the nonbeneficiaries. Moreover, for those retired nonmarried women beneficiaries who had earnings, the median amount was nearly twice as much as for widowed beneficiaries, $885 versus $485. For nonmarried nonbeneficiary women with earnings, the median amount was $1,025. All the earnings of all these nonbeneficiary women provided 23 percent of their aggregate income. Public assistance provided a larger sum—27 percent of all the income of nonmarried nonbeneficiary women—and this was the leading source of income for this category of elderly women.

Retirement benefits were the most prevalent source of income for all categories of the elderly except those who were not drawing OASDI benefits. Among couples, 84 percent received retirement payments during 1962. The median amount was $1,605. Proportions were smaller and amounts appreciably less for nonmarried persons. Among nonmarried men, 72 percent reported retirement benefits, with a median of $980 for the year. Among nonmarried women, 64 percent received similar benefits, and their median amount was $770.

OASDI benefits dominated. Other public retirement systems (such as railroad, military, and Federal, State, and local civilian government employees' retirement systems) yielded larger median amounts to those couples and individuals who drew on them but were available to relatively small percentages of the elderly population. For example, among married couples, 3,743,000 drew OASDI benefits throughout the year, with a median amount of $1,405. Another 1,120,000 elderly couples were nonbeneficiaries; only 25 percent of them drew retirement benefits of any kind, nearly all from other public retirement systems. The median benefit of some 270,000 couples who received such payments was $2,460. Only 3 percent of nonbeneficiary couples had private

[35] A few couples in which the husband was under 62 and the wife was older were classified according to the age of the wife.

[36] Data on the percentage of income derived from particular sources are from Social Security Administration, op. cit., p. 291, table 3.6. Distinctions between beneficiaries and nonbeneficiaries are from the source tables from which table 10 is derived—that is, ibid., pp. 286–287, tables 3.1 and 3.2.

group pensions (whereas one-fifth of all beneficiary couples had private group pensions to supplement their OASDI benefits).

Almost as many couples received veterans benefits as private pensions, and the median amount was about the same. For the nonmarried elderly, veterans benefits were more numerous and the median amounts were larger than private pensions.

The number of couples and nonmarried persons who had income from interest, dividends, and rents was larger, in each group, than the number with earnings from employment, but the median amounts were comparatively small—one-fifth or less of median incomes from earnings. Only 14 percent of the income of married couples and 17 percent of the income of the nonmarried was from this source. Interest, dividends, and rents were equal to earnings in relative importance for nonmarried women who were nonbeneficiaries.

Public assistance (old-age assistance) was an important type of income for the elderly in 1962. One couple in every 12 drew public assistance. The median amount, $850, though less than OASDI or earnings, was nearly one-third of the median income of all elderly couples. Nevertheless, public assistance provided only 2 percent of all the income of all the elderly couples. Larger proportions of nonmarried men and women drew public assistance, and this was the source of 9 percent of all the income of these elderly individuals. Among nonbeneficiaries classified as nonmarried, 30 percent of the women and 35 percent of the men depended at least in part on old-age assistance. As already noted, these payments were the largest single source of income for nonmarried nonbeneficiary women, supplying 27 percent of all their income. For nonmarried nonbeneficiary men, old-age assistance provided one-sixth of their aggregate income.[37]

Income from private annuities, unemployment insurance, and contributions by relatives also was identified in the 1962 survey. These sources were found to be minor for all categories of the elderly.

DISTRIBUTION BY AMOUNT OF INCOME

Table 11 reports the distribution of elderly couples and individuals in 1962 according to their total incomes. Nonbeneficiary couples generally had the higher incomes. Half of the couples receiving OASI cash benefits had total incomes of less than $2,710. Half of the nonbeneficiary couples had total incomes of less than $3,580. The distribution about these medians differed substantially for the two groups. Beneficiary couples were mostly in the low- and middle-income brackets; nonbeneficiary couples, in the low and the high brackets. Thus, one-third of all nonbeneficiary couples and 28 percent of all beneficiary

[37] Percentages of overlap in receipt of old-age assistance and old-age, survivors, and disability insurance cash benefits have moved in generally opposite directions over the last two decades. In yearly reports, usually for February, the number of persons receiving both types of payments first passed 1 million in 1966. In February, 1962 (the year of the Social Security survey), 744,000 persons 65 or older drew both types of payments. These were 33.7 percent of all old-age assistance money recipients and 6.4 percent of all OASDHI cash beneficiaries 65 or older. The 6.4 percent was the lowest ratio during 1948–67. In February 1967, 1,096,000 persons 65 or older drew both types of payments. They were 53.1 percent of all old-age assistance money payment recipients and 7.0 percent of all OASDHI cash beneficiaries 65 or older. This was the first year in which more than half of all OAA cash payment recipients were also cash beneficiaries under OASDHI. Department of Health, Education, and Welfare, Social and Rehabilitation Service, Assistance Payments Administration, "Concurrent Receipt of Public Assistance Money Payments and Old-Age Survivors and Disability Insurance Cash Benefits by Persons Aged 65 or over, 1948–67 and February 1967" (multilithed release, December 1967), table 1.

couples had total income under $2,000. Nearly one-half of the beneficiary couples (46 percent) were in the income range from $2,000 to $4,000, but less than one-fourth (23 percent) of the nonbeneficiary couples were between these limits. At the upper end of the scale, 45 percent of all nonbeneficiary couples and only 26 percent of all beneficiary couples had total incomes of $4,000 or more.

Income levels were generally much lower for nonmarried elderly individual nonbeneficiaries than for beneficiaries. Half of the men had incomes below $1,365 for the year—$1,375 for beneficiaries and $1,145 for nonbeneficiaries. Half of the women had incomes below $1,015—$1,300 for retired beneficiaries, $1,105 for widowed beneficiaries, and $755 for nonbeneficiaries.

Among nonmarried beneficiaries, 26 percent of the men, 36 percent of the retired women, and 44 percent of the widows had total incomes under $1,000 for the year. At the upper end of the scale, with incomes of $3,000 or more, were 9 percent of the nonmarried men beneficiaries, 10 percent of the retired nonmarried women beneficiaries, and only 6 percent of the widowed beneficiaries.

Nonmarried beneficiaries reported incomes under $1,000 for 46 percent of the men and 65 percent of the women. In fact, less than one-third of the men and one-seventh of the women had as much as $2,000. In contrast, however, 21 percent of the men in this category had incomes above $3,000—reflecting their earnings from continued employment beyond age 65 (and in some instances, perhaps, the fact that they turned 65 during or after the survey period). Of the women who were nonmarried nonbeneficiaries, only about 6 percent had incomes above $3,000.

Federal Programs

If retirement benefits were a major source of money income for most elderly persons in 1962, as table 9 indicates, it may be assumed that they have mounted to greater importance in 1968. OASDI benefits, which dominated in 1962, have been extended since then to a larger proportion of the elderly population and increased in individual amounts.

This section notes summarily the retirement and other programs of the Federal Government that provide current money income to the elderly or supplement money incomes by making available goods or services (such as food or health care) of value to the elderly. The following section describes selected aspects of the OASDHI program.

Major Federal Government programs that provide cash benefits to individuals, including the elderly, are identified for the period 1960–68 in table V of the paper on "Federal Programs for Human Resource Development," by Michael S. March, elsewhere in this compendium. Mr. March's table reports total expenditures in each program, whether for money income payments directly to individuals or families or, in the case of public assistance, through State and local government agencies. Subdividing the payments between youth and the aged, Mr. March assigns to the aged (in his table VIII) 56 to 57 percent of the cash benefit payments in each of the fiscal years 1960, 1963, 1966, 1967, and 1968. When expenditures for cash benefits are combined with health and medical care programs and other service

programs, principally for housing, Federal expenditures for programs assisting the aged are estimated as follows:

Fiscal year:	*Amount (billions)*
1950	$2.6
1960	13.3
1963	17.4
1966	21.3
1967	25.9
1968	29.4

Not included in these totals are special Federal tax benefits for the elderly, estimated for the calendar year 1967 at $2.3 billion.[38]

Selected Aspects of OASDHI

COVERAGE

Among the 19 million Americans aged 65 or older, more than 16.2 million were drawing old-age and survivors monthly cash benefits at the beginning of 1968. The sum of their benefits was more than $1.5 billion a month. The average benefit in November 1967, was $85 ($1,020 a year) for retired workers and $75 ($900 a year) for aged widows and widowers. These averages were rising gradually month by month, and all benefits were scheduled to advance 13 percent or more at the end of February 1968, under the Social Security Amendments of 1967.

Many persons 65 or older who do not draw benefits would be eligible if they applied. Several million have qualified for health insurance protection under the medicare program. Most of them probably worked in employment that was not covered for monthly cash benefits (such as the Federal civil service); many others presumably are still at work, with earned incomes in excess of the amount specified in the earnings test for cash benefits.[39]

Among those reaching age 65, some 92 percent or more are eligible for old-age or survivors benefits. The principal groups of elderly outside OASI, about 5 percent of all those 65 or older, are protected under the railroad retirement or Federal civil service retirement systems or, in some instances, State or local government employees' retirement systems. Fewer than 3 percent of those who turned 65 in 1965 lacked protection under either OASI or some other public retirement program. About 15 percent of the elderly drew private pensions, and some of these may have been outside the public programs.[40]

Nearly 7.5 million persons under age 65 also drew monthly benefits at the beginning of 1968. More than 2 million of these were in their early sixties (the numbers here are for those receiving cash payments in October 1967):

---

[38] March, op. cit., table VIII.

[39] In the 50 States and District of Columbia (the areas covered in the population statistics in tables 2 and 3), 18,755,000 persons were enrolled in the hospital insurance program Jan. 1, 1967. Another 144,000 persons were enrolled in the outlying areas. (Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, Health Insurance Statistics, "Current Data from Medicare Program," by Dorothy P. Rice, HI–1, Nov. 20, 1967, table 1.)

[40] Cohen, loc. cit., p. 3. See also Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, Research and Statistics Note 2–1968, "OASDHI Beneficiaries Receiving Government Pensions," by Lawrence D. Haber and Rena Kling, Jan. 26, 1968.

Category	Thousands
Retired workers, 62 to 64 (actuarially reduced benefits)	1,030
Disabled workers, 62 to 64	252
Survivors and dependents, 62 to 64	706
Widows and divorced wives, 60 to 61 (actuarially reduced benefits) [41]	120

[41] Eligible beginning in September 1965.

Another small group aged 60 and 61 will be added under the 1967 amendments. Reduced benefits are payable at ages 50 to 62 for certain disabled widows and widowers of deceased workers. Under earlier law, these surviving dependents were not eligible, excepting as mothers with dependent young children in their care.

Modification of the retirement or earnings test by the 1967 amendments, described in the next subsection, also will permit payment of benefits to thousands of persons, aged 60 to 72, who earlier were disqualified by the earnings test. In addition, it will permit increases in benefits for larger numbers whose payments were reduced because of their earnings.

RETIREMENT TEST

Ever since monthly retirement benefits became payable, the Social Security Act has required recipients to show that they were substantially retired from gainful employment. Retirement status is measured by monthly and yearly earnings.

The test affects persons who work after applying for monthly cash benefits. If a beneficiary under age 72—whether a retired worker or dependent—earns income of more than $1,680 a year (lesser amounts prior to 1968), he may be denied part or all of the benefits otherwise payable—$1 of benefits for each $2 of earnings from $1,680 to $2,880 a year, and $1 for each $1 of earnings above $2,880. The disallowance does not apply, however, in any month in which earnings are less than $140, regardless of the yearly sum.

Prior to 1950, benefits were denied whenever the recipient earned more than $14.99 a month in covered employment, a limit that required almost complete withdrawal from covered employment. The earnings test was raised to $50 a month in 1950. Subsequently it was advanced in several steps to higher levels and modified in various details (1952, 1954, 1958, 1960). The present two-step form of the earnings limit was introduced in 1960, with benefits reduced $1 for each $2 of annual earnings between $1,200 and $1,500 and $1 for each $1 above $1,500. These limits were raised in 1965 to $1,500 and $2,700, and to their current levels in the 1967 amendments, effective for 1968.

Since 1954, earnings from work not covered under OASDI have been counted equally with those from covered employment. As coverage broadened, the effect of this change was narrowed.

Congress finds the retirement test a difficult issue to deal with. The test is an important mechanism for holding down benefit outlays. Its relaxation—as proposed in every congressional session by many members—would entail a larger increase in benefits than most other changes that are urged. Yet this would not augment the benefits paid to more than 90 percent of all beneficiaries aged 65 or older—those who earn less than the lower limit specified in the earnings test and those who are exempt because they are over age 72. Most of the gain would accrue to persons who, because of continuing employment, have the highest incomes in the OASDI beneficiary groups. At the same time,

the retirement test is widely criticized as a compelling inducement to most elderly persons to leave or limit their productive employment, whether or not they are able to work. Their continued employment would benefit them directly and would help broaden the labor force.

Retention of the test is based—apart from cost—upon the argument that the principal purpose of the OASDI system is to insure against a loss of earnings, and that payments therefore should not be made to persons who continue to earn at substantially the same levels as in earlier years. Critics look upon the test as a heritage from the great depression, when there appeared to be an excess of available labor, and the retirement system was viewed as a means of encouraging and enabling elderly people to withdraw from the labor market. In the high employment economy of the 1960's, that kind of objective seems much less important.

SPECIAL PROVISIONS FOR PERSONS 72 OR OLDER

Since 1950, the Social Security Act has provided distinctive treatment for some elderly persons in the form of exemption from the usual earnings or retirement test that applies to younger beneficiaries.

The retirement test in its original form applied to beneficiaries of all ages. The 1950 amendments exempted those aged 75 or over; they could draw benefits regardless of their earnings. In 1956 the age exemption was changed to 72 and older.

The age exemption, it was explained, "recognizes that some people go on working and paying taxes to the end of their lives and might otherwise never get any monthly benefits." [42]

Further special treatment was accorded in 1965, when a "transitional" provision bestowed insured status upon older persons who had some covered employment but not enough to qualify for benefits. The same legislation established the national health insurance program for the elderly (medicare). Everyone eligible for transitional insured status was brought into the hospital insurance plan and permitted to enroll in the voluntary supplementary medical insurance plan. The law provided also that persons who were then 65 or would become 65 in the next few years would likewise be covered by hospital insurance and eligible to enroll for supplementary insurance even though they were not insured under OASDI or the railroad retirement program. Those who reached age 65 after 1967 would have to have six or more quarters of covered employment, but many would be able to qualify for health insurance who were ineligible for monthly cash benefits.

In 1966, eligibility for cash benefits was extended—without any requirement for previous covered employment—to all persons who were already 72 or would reach that age before 1968.

On the basis of the 1965 and 1966 provisions for special payments, more than three-quarters of a million elderly individuals drew monthly cash benefits without meeting the usual requirements for in-

---

[42] Department of Health, Education, and Welfare, Social Security Administration, Social Security Programs in the United States (January 1966), p. 27; and cf. Social Security Bulletin, August 1960, "The Social Security Act; Its First 25 Years," p. 25.

On the retirement test generally and expected effects of the 1967 amendments, see "The Retirement Test Under the Social Security Program," 90th Cong., first sess., House Committee on Ways and Means, President's Proposals for Revision in the Social Security System, hearings, March 1–3, 1967, pt. 1, pp. 316–19.

sured status. By the end of 1966, benefit awards had been made to 131,452 persons entitled to insured status under the 1965 transitional provision. Of these, 127,510 were receiving current monthly payments at the end of the year. Under the 1966 amendment, 748,632 awards were approved by the end of 1966. Cash benefits were, however, denied to many of these new claimants because of offset provisions noted below. At the year's end, 634,313 were in current payment status.[43]

Men who were 72 before 1964 and women who reached that age before 1967 could qualify under the 1965 "transitional" provision with as little as one quarter of coverage for each year that elapsed after 1950 and before retirement age. The minimum requirement was three quarters, and the maximum five. Benefits under the special provision are not related to the applicant's record of earnings. They were payable at a flat rate of $35 a month for the worker or his widow, and $17.50 for the wife of a retired worker. A widow was eligible if her husband was living in September 1965, and met the reduced work requirements.

The 127,510 persons who, at the end of 1966, were receiving special benefits under the 1965 amendments comprised 89,950 retired workers, 8,350 wives, and 29,210 widows. Their benefits were financed from the OASI trust fund.[44]

The 1966 amendment, part of the Tax Adjustment Act of 1966, extended eligibility for the special $35 and $17.50 monthly benefits to persons 72 and older who could not qualify for the "transitional insured" benefits. The new "transitional noninsured" benefit was made available to any U.S. citizen resident in a State or the District of Columbia and any alien lawfully admitted for permanent residence who had maintained continuous residence in the United States for 5 years immediately preceding application.

Persons attaining age 72 after 1967 are eligible for these special benefits only if they have at least three quarters of covered employment for each calendar year after 1966 and before they reach age 72. After a few years, this requirement will equal or outweigh the regular eligibility requirements for fully insured benefits—in 1970 for women and in 1972 for men attaining age 72 in those years.

For transitional noninsured hospital benefits, the special coverage requirements will continue for a longer period to be less than requirements for fully insured cash benefits. Men attaining age 72 before 1981 and women attaining this age before 1979 will be eligible until then for hospital insurance benefits with fewer quarters of coverage than for fully insured cash benefits.[45]

Although the monthly cash benefits were set at the same level as for transitional insured beneficiaries, benefits under the 1966 amendment are subject to reduction by the amount of any government pension that the individual or his spouse is receiving or is eligible to receive. (Gov-

---

[43] The number of awards made and in current payment status are reported in Social Security Administration, Office of the Actuary, "OASDI: Summary benefit data: Calendar year 1965" (release B–1, Feb. 10, 1966), and ibid., 1966, tables 1 and 2 (release B–1, Feb. 16, 1967). The number of recipients under the 1966 provision was 731,000 in July 1967. Despite changes resulting from deaths and the eligibility of new applicants, the number of monthly benefit payments to this group continued at 731,000 during the next several months. Cf. Social Security Administration, Office of Research and Statistics, "Current Social Security Program Operations," monthly release.

[44] Office of the Actuary, "OASDI: Summary benefit data: Calendar year 1966," table 2.

[45] See Wilbur J. Cohen, Robert M. Ball, and Robert J. Myers, "Social Security Payments to Noninsured Persons," Social Security Bulletin, September 1966, p. 6, for a table comparing coverage requirements under the several provisions for fully insured and transitional benefits.

ernment pensions include State or local government employee retirement system payments, as well as various Federal Government payments.) Also, payments are not made to recipients of public assistance unless they leave the assistance rolls. The public assistance condition was included in the legislation in recognition of State practices in administering public assistance. Almost invariably, they reduce assistance payments to individuals or families by the amount of income from other sources. Without this provision in the Federal law, the payment of special OASI benefits would be tantamount to a substantial increase in Federal financing of old-age assistance.[46]

The 1967 amendments raised benefit rates for the uninsured aged 72 and older from $35 to $40 a month for a single person and from $52.50 to $60 a month for a couple. The increased rates are first payable for February 1968.

Benefits to the new noninsured are disbursed initially from the OASI trust fund, but they are almost all financed by reimbursements from general revenues of the Federal Government. Only those benefits paid to persons with three or more quarters of covered employment are charged to the trust fund. Of the 749,000 benefits awarded during the first 3 months when they were available (October–December 1966), 743,000 were chargeable to general revenues and only 5,400—less than 1 percent—to the trust fund.

The 749,000 eligible claimants included 106,000 for whom benefits were completely offset and no payments made under these statutory provisions—71,000 offset by government pensions and 35,000 by public assistance. In another 4,800 cases, government pensions offset part of the benefit.[47]

The provision for noninsured benefits, as adopted in 1966, was a conference committee modification of a Senate-approved amendment that would have qualified many more persons for OASI cash benefits. As first approved in the Senate, this amendment would have awarded the minimum monthly benefit, then $44, to nearly all U.S. citizens (and certain alien residents) aged 70 and older who were not already drawing benefits. A married woman aged 70 would receive half of her husband's minimum benefit or, in some instances, have the benefit increased to $44. Persons drawing less than the minimum benefit (for example, some who retired early on reduced benefits) would be raised to this amount at age 70. In this form, the new benefit provision would have eliminated prospectively the transitional insured status created in 1965, since it would assure benefits above the special primary rate of $35 a month and the wife's rate of $17.50.[48]

Conclusion

This descriptive survey has indicated that the Federal Government plays a central role in assuring incomes to the elderly, primarily through the old-age and survivors insurance program. The paper does

[46] Ibid., pp. 5–6.

[47] "Special Awards to Persons 72 and Over, October–December 1966," in Social Security Bulletin, August 1967, pp. 18–20. See also, "OASDI: Summary benefit data: Calendar year 1966," cited above, table 1.

In reviewing characteristics of special beneficiaries, the article in the Social Security Bulletin notes that relatively few nonwhites were able to claim noninsured awards, possibly because of the public assistance offset provision.

[48] Cohen, Ball, and Myers, loc. cit., pp. 4–5, report the legislative history.

not examine several aspects of the social security system that are of special importance to the elderly, such as the national health insurance system and the Federal-State old-age assistance program. Nor does it take up numerous matters of public policy and private practice that affect the incomes of the elderly. The reader will find many issues identified and examined in studies cited in this paper.[49]

The central issue, basic to national policy, relates to the distribution of the Nation's goods, services, and wealth among producing and dependent members of the population—and within these groups.

The United States no longer relies on the forces of marketplace and the resources and initiatives of individuals to assure its elderly people an allotted share in the current output of the Nation.

But what is the share to be allotted to the elderly, and by what mechanisms shall it be delivered?

Shall this share be determined for each individual by his earnings history in earlier years? Or by his earnings history adjusted for subsequent price changes? Or by his earnings history adjusted not only for price changes, but also for productivity advances and other economic, social, and cultural developments?

Is there a level of support for individuals and families above which public policy does not attempt to reach—a level at which further provision for their needs and wants is left to private decisions and private responsibility? If so, how is that level defined?

These difficult questions are in fact inescapable, though their solutions for any given period may not be explicitly formulated. Whether the responses are prescribed deliberately in public policies, with which private actions may be concerted, or whether the responses are fortuitous and uncoordinated, some kind of economic distribution will nonetheless emerge. It will, inevitably, determine the share of national output made available to the elderly as a group and to each of them.

In his paper on Federal programs for human resource development, elsewhere in this compendium, Michael S. March compares Federal expenditures for the elderly with those for the young. He finds that within the totals of Federal expenditures for education and training, health, and cash benefits during the fiscal years 1960–68, approximately 45 to 47 percent were for the elderly (65 and older) and 10 to 18 percent for the young (under 21).[50]

These differences led Mr. March to pose a series of interesting analytical questions, including these:

> What are the comparative social and economic returns to the society for expenditures to increase the level of consumption of these two groups?
>
> What are the comparable returns for human "investment" expenditures for the two groups?
>
> What, in sum, is the relative importance to the Nation of expenditures for children and youth as compared to expeditures for assistance for the aged?

---

[49] Particular attention is directed to the 6-part compendium issued by the Subcommittee on Fiscal Policy of the Joint Economic Committee: Old Age Income Assurance: A Compendium of Papers on Problems and Policy Issues in the Public and Private Pension System. Part I, General Policy Guidelines; Part II. The Aged Population and Retirement-Income Programs; and Part VI, Summaries and Bibliography, are currently in press. The following parts were issued in December 1967: III, Public Programs; IV, Employment Aspects of Pension Plans; and V, Financial Aspects of Pension Plans.

[50] March, loc. cit., table IX.

Merely to present such questions seems to suggest a conclusion that there is disproportion in the present allocation of Federal expenditures. Mr. March's paper provides only fragmentary evidence that would point to this conclusion. If there is a disproportion, it will not appear from Federal Government program expenditures alone. It will be expressed as a judgment derived from examination of the support that our society as a whole accords to measures that will develop the capacities of children and youth and promote the good life for all of them, and an evaluation of that support against explicit choices in the realm of social goals and social values. Ultimate answers may be achieved not through arithmetic but only through a reconciliation of individual judgments and subjective values.

Governmental decisions on the allocation of resources among public programs necessarily respond to many influences and considerations that are not especially or primarily rational or objective, let alone quantifiable, but nevertheless are important to individuals and the community. In our society the elderly may vote and the very young do not. The elderly, though less numerous than their juniors, often occupy the role of respected and beloved parents and grandparents. Our legal system maintains support laws under which parents and children are assigned reciprocal obligations. Most young and middle-aged persons look forward to an old age in which their children will accord to them, in their turn, a measure of economic security and an assured share in the national product. These and myriad other considerations are bound to affect personal attitudes and therefore public policy toward the elderly population.

A strict analysis of costs and benefits might indicate that the most efficient policy toward the elderly as a group would be one of abandonment. Or it might suport the widespread application of a common contemporary arrangement—consignment of the very elderly to mental hospitals and other custodial institutions. Behind institutional walls society may sequester and leave to their own introspective solitude those elderly persons whose marketable skills have dwindled or disappeared, and may give a minimum of attention to those others whose minds and bodies have grown feeble.

Again, analysis might suggest that resources would be more productively applied if they were used to invigorate young life rather than to promote longevity in those already old.

But in the United States at this time that kind of conclusion would not be widely aceptable as a guide to policy. It would, indeed, reflect inadequacies of the appraisal more than inefficiencies of current policy.

Something similar may be said about analyses that assume, for their basic premise, an opposition of interest or competition for preference between the young and the old. There are, of course, persistent rivalries and oppositions of this kind, and they are thoroughly visible in contemporary life. Yet it may be suggested that an equally valid premise for analysis of public policy, and perhaps more acceptable in our culture, is one that assumes a community of interest, an identity of social objectives and ultimate values, for young and old and everyone in between—and therefore counts all persons in all age groups as full members of the society that is being served.

TABLE 1.—AVERAGE LIFE EXPECTANCY AT AGE 65 IN SELECTED YEARS, 1900–65

Year when age 65 was attained	Average life expectancy, in years				
	Total population	White		Nonwhite	
		Male	Female	Male	Female
1900	11.9	11.5	12.2	[1] 10.4	[1] 11.4
1930	12.2	11.8	12.8	[1] 10.9	[1] 12.2
1950	13.8	12.8	15.0	12.8	14.5
1960	14.4	13.0	15.9	12.8	15.1
1965	14.6	12.9	16.3	12.6	15.5

[1] Negro population only. Negro population comprised 95 percent or more of corresponding nonwhite total.

Source: Department of Commerce, Bureau of the Census, Statistical Abstract of the United States, 1967 (88th edition; Washington, 1967), p. 55; except that data for total population, 1900–60, are from Department of Health, Education, and Welfare, "Health, Education, and Welfare Trends," 1968 edition, pt. 1, pp. 3–13, and 1965 from Public Health Service, Vital Statistics of the United States, 1965, vol. II, pp. 5–10.

TABLE 2.—ESTIMATES AND PROJECTIONS OF THE POPULATION OF THE UNITED STATES—TOTAL AND SELECTED GROUPS: SELECTED YEARS, 1960–90

[In thousands]

Coverage and year	Total population			White population			Nonwhite population		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Series A projections:									
All ages:									
1960	180,684	89,332	91,352	160,033	79,266	80,767	20,651	10,066	10,585
1966	196,842	96,900	99,942	173,253	85,432	87,820	23,589	11,468	12,121
1970	208,615	102,541	106,075	182,815	90,005	92,811	25,800	12,536	13,264
1975	227,929	111,994	115,935	198,453	97,660	100,793	29,476	14,334	15,142
1980	250,489	123,185	127,304	216,588	106,669	109,919	33,901	16,516	17,385
1985	274,748	135,305	139,443	235,841	116,310	119,532	38,907	18,995	19,911
1990	300,131	148,056	152,075	255,610	126,271	129,340	44,521	21,785	22,735
Series D projections:									
All ages:									
1960	180,684	89,332	91,352	160,033	79,266	80,767	20,651	10,066	10,585
1966	196,842	96,900	99,942	173,253	85,432	87,820	23,589	11,468	12,121
1970	204,923	100,656	104,268	179,794	88,457	91,339	25,129	12,199	12,929
1975	215,367	105,581	109,787	188,211	92,414	95,798	27,156	13,167	13,989
1980	227,665	111,533	116,133	198,117	97,208	100,910	29,548	14,325	15,223
1985	241,731	118,450	123,280	209,427	102,778	106,648	32,304	15,672	16,632
1990	255,967	125,518	130,449	220,805	108,442	112,363	35,162	17,076	18,086
All series:									
60 to 64 years:									
1960	7,162	3,418	3,744	6,566	3,128	3,438	596	290	306
1966	7,931	3,757	4,174	7,215	3,410	3,805	716	347	369
1970	8,451	3,957	4,494	7,717	3,610	4,108	734	347	386
1975	9,279	4,293	4,986	8,445	3,909	4,536	834	384	450
1980	9,770	4,467	5,303	8,895	4,075	4,820	875	392	483
1985	10,442	4,747	5,695	9,460	4,313	5,147	982	434	548
1990	10,022	4,577	5,444	9,024	4,148	4,875	998	429	569
65 years and over:									
1960	16,659	7,537	9,121	15,393	6,938	8,455	1,266	599	667
1966	18,457	8,003	10,452	17,023	7,353	9,671	1,434	651	783
1970	19,585	8,336	11,249	18,031	7,641	10,390	1,554	694	859
1975	21,159	8,836	12,323	19,472	8,102	11,370	1,687	735	953
1980	23,063	9,507	13,556	21,193	8,712	12,480	1,870	795	1,076
1985	24,978	10,175	14,802	22,925	9,326	13,599	2,052	849	1,203
1990	27,005	10,904	16,102	24,722	9,978	14,743	2,284	925	1,359
65 to 69 years:									
1960	6,264	2,929	3,335	5,746	2,682	3,063	518	247	272
1966	6,378	2,901	3,476	5,882	2,668	3,214	496	233	263
1970	6,883	3,123	3,760	6,304	2,851	3,453	579	272	307
1975	7,470	3,341	4,129	6,865	3,065	3,800	605	276	329
1980	8,223	3,635	4,588	7,529	3,328	4,202	694	307	386
1985	8,681	3,794	4,887	7,947	3,478	4,469	734	316	418
1990	9,299	4,043	5,256	8,469	3,690	4,779	830	353	477
70 to 74 years:									
1960	4,769	2,195	2,574	4,417	2,027	2,391	352	168	183
1966	5,190	2,261	2,929	4,791	2,083	2,708	399	178	221
1970	5,214	2,230	2,984	4,841	2,065	2,776	373	165	208
1975	5,721	2,439	3,281	5,270	2,239	3,030	451	200	251
1980	6,234	2,624	3,611	5,758	2,419	3,340	476	205	271
1985	6,889	2,869	4,020	6,338	2,638	3,699	551	231	321
1990	7,302	3,011	4,292	6,713	2,771	3,943	589	240	349
75 years and over:									
1960	5,625	2,413	3,212	5,229	2,229	3,002	396	184	210
1966	6,888	2,841	4,048	6,349	2,602	3,749	540	241	299
1970	7,488	2,982	4,505	6,886	2,725	4,214	602	257	345
1975	7,969	3,055	4,913	7,327	2,796	4,540	632	259	373
1980	8,606	3,248	5,358	7,905	2,966	4,758	701	282	418
1985	9,407	3,511	5,896	8,640	3,208	5,431	767	303	465
1990	10,405	3,849	6,554	9,540	3,516	6,022	865	333	532

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P–25, No. 381, "Projections of the Population of the United States, by Age, Sex, and Color to 1990, With Extensions of Population by Age and Sex to 2015" (Dec. 18, 1967).

TABLE 3.—PERCENTAGE OF U.S. POPULATION IN AGE GROUPS 60 AND OLDER, IN TOTAL AND BY RACE, 1960, 1966, AND CENSUS BUREAU PROJECTIONS FOR 1970-90

Age group and year	Percent of total population						Percent of white population						Percent of nonwhite population					
	Total		Male		Female		Total		Male		Female		Total		Male		Female	
	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D	Series A	Series D
60 to 64 years:																		
1960	4.0	4.0	3.8	3.8	4.1	4.1	4.1	4.1	3.9	3.9	4.3	4.3	2.9	2.9	2.9	2.9	2.9	2.9
1966	4.0	4.0	3.9	3.9	4.2	4.2	4.2	4.2	4.0	4.0	4.3	4.3	3.0	3.0	3.0	3.0	3.0	3.0
1970	4.1	4.1	3.9	3.9	4.2	4.3	4.2	4.3	4.0	4.1	4.4	4.5	2.8	2.9	2.8	2.8	2.9	3.0
1975	4.1	4.3	3.8	4.1	4.3	4.5	4.3	4.5	4.0	4.2	4.5	4.7	2.8	3.1	2.7	2.9	3.0	3.2
1980	3.9	4.3	3.6	4.0	4.2	4.6	4.1	4.5	3.8	4.2	4.4	4.8	2.6	3.0	2.4	2.7	2.8	3.2
1985	3.8	4.3	3.5	4.0	4.1	4.6	4.0	4.5	3.7	4.2	4.3	4.8	2.5	3.0	2.3	2.8	2.8	3.3
1990	3.3	3.9	3.1	3.6	3.6	4.2	3.5	4.1	3.3	3.8	3.8	4.3	2.2	2.8	2.0	2.5	2.5	3.1
65 years and over:																		
1960	9.2	9.2	8.4	8.4	10.0	10.0	9.6	9.6	8.8	8.8	10.5	10.5	6.1	6.1	6.0	6.0	6.3	6.3
1966	9.4	9.4	8.3	8.3	10.5	10.5	9.8	9.8	8.6	8.6	11.0	11.0	6.1	6.1	5.7	5.7	6.5	6.5
1970	9.4	9.6	8.1	8.3	10.6	10.8	9.9	10.0	8.5	8.6	11.2	11.4	6.0	6.2	5.5	5.7	6.5	6.6
1975	9.3	9.8	7.9	8.4	10.6	11.2	9.8	10.3	8.3	8.8	11.3	11.9	5.7	6.2	5.1	5.6	6.3	6.8
1980	9.2	10.1	7.7	8.5	10.6	11.7	9.8	10.7	8.2	9.0	11.4	12.4	5.5	6.3	4.8	5.5	6.2	7.1
1985	9.1	10.3	7.5	8.6	10.6	12.0	9.7	10.9	8.0	9.1	11.4	12.8	5.3	6.4	4.5	5.4	6.0	7.2
1990	9.0	10.6	7.4	8.7	10.6	12.3	9.7	11.2	7.9	9.2	11.4	13.1	5.1	6.5	4.2	5.4	6.0	7.5
65 to 69 years:																		
1960	3.5	3.5	3.3	3.3	3.7	3.7	3.6	3.6	3.4	3.4	3.8	3.8	2.5	2.5	2.5	2.5	2.6	2.6
1966	3.2	3.2	3.0	3.0	3.5	3.5	3.4	3.4	3.1	3.1	3.7	3.7	2.1	2.1	2.0	2.0	2.2	2.2
1970	3.3	3.4	3.0	3.1	3.5	3.6	3.4	3.5	3.2	3.2	3.7	3.8	2.2	2.3	2.2	2.2	2.3	2.4
1975	3.3	3.5	3.0	3.2	3.6	3.8	3.5	3.6	3.1	3.3	3.8	4.0	2.1	2.2	1.9	2.1	2.2	2.4
1980	3.3	3.6	3.0	3.3	3.6	4.0	3.5	3.8	3.1	3.4	3.8	4.2	2.0	2.3	1.9	2.1	2.2	2.5
1985	3.2	3.6	2.8	3.2	3.5	4.0	3.4	3.8	3.0	3.4	3.7	4.2	1.9	2.3	1.7	2.0	2.1	2.5
1990	3.1	3.6	2.7	3.2	3.5	4.0	3.3	3.8	2.9	3.4	3.7	4.3	1.9	2.4	1.6	2.1	2.1	2.6
70 to 74 years:																		
1960	2.6	2.6	2.5	2.5	2.8	2.8	2.8	2.8	2.6	2.6	3.0	3.0	1.7	1.7	1.7	1.7	1.7	1.7
1966	2.6	2.6	2.3	2.3	2.9	2.9	2.8	2.8	2.4	2.4	3.1	3.1	1.7	1.7	1.6	1.6	1.8	1.8
1970	2.5	2.5	2.2	2.2	2.8	2.9	2.6	2.7	2.3	2.3	3.0	3.0	1.4	1.5	1.3	1.4	1.6	1.6
1975	2.5	2.7	2.2	2.3	2.8	3.0	2.7	2.8	2.3	2.4	3.0	3.2	1.5	1.7	1.4	1.5	1.7	1.8
1980	2.5	2.7	2.1	2.4	2.8	3.1	2.7	2.9	2.3	2.5	3.0	3.3	1.4	1.6	1.2	1.4	1.6	1.8
1985	2.5	2.8	2.1	2.4	2.9	3.3	2.7	3.0	2.3	2.6	3.1	3.5	1.4	1.7	1.2	1.5	1.6	1.9
1990	2.4	2.9	2.0	2.4	2.8	3.3	2.6	3.0	2.2	2.6	3.0	3.5	1.3	1.7	1.1	1.4	1.5	1.9
75 years and over:																		
1960	3.1	3.1	2.7	2.7	3.5	3.5	3.3	3.3	2.8	2.8	3.7	3.7	1.9	1.9	1.8	1.8	2.0	2.0
1966	3.5	3.5	2.9	2.9	4.0	4.0	3.7	3.7	3.0	3.0	4.3	4.3	2.3	2.3	2.1	2.1	2.5	2.5
1970	3.6	3.7	2.9	3.0	4.2	4.3	3.8	3.8	3.0	3.1	4.5	4.6	2.3	2.4	2.1	2.1	2.6	2.7
1975	3.5	3.7	2.7	2.9	4.2	4.5	3.7	3.9	2.9	3.0	4.5	4.7	2.1	2.3	1.8	2.0	2.5	2.7
1980	3.4	3.8	2.6	2.9	4.2	4.6	3.7	4.0	2.8	3.1	4.5	4.9	2.1	2.4	1.7	2.0	2.4	2.7
1985	3.4	3.9	2.6	3.0	4.2	4.8	3.7	4.1	2.8	3.1	4.5	5.1	2.0	2.4	1.6	1.9	2.3	2.8
1990	3.5	4.1	2.6	3.1	4.3	5.0	3.7	4.3	2.8	3.2	4.7	5.4	1.9	2.5	1.5	1.9	2.3	2.9

Note: Series A and D identify the highest and lowest of the Census Bureau projections.

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P-25, No. 381, "Projections of the Population of the United States, by Age, Sex, and Color to 1990, With Extensions of Population by Age and Sex to 2015" (Dec. 18, 1967).

TABLE 4.—AGED POPULATION RELATED TO TOTAL POPULATION, 1990, 2000, 2015 (SERIES A AND D, BY SEX)

[In thousands]

Category	Number			Percent of total [1]		
	1990	2000	2015	1990	2000	2015
Total population:						
Total:						
Series A	300,131	361,424	483,371			
Series D	255,967	282,642	325,785			
Male:						
Series A	148,056	179,070	240,626			
Series D	125,518	138,901	160,382			
Female:						
Series A	152,075	182,354	242,747			
Series D	130,449	143,741	165,404			
65 years and over:						
Total:						
Series A	27,005	28,184	34,451	9.0	7.8	7.1
Series D				10.6	10.0	10.6
Male:						
Series A	10,904	11,284	14,225	7.4	6.3	5.9
Series D				8.7	8.1	8.9
Female:						
Series A	16,102	16,900	20,227	10.6	9.3	8.3
Series D				12.3	11.8	12.2
75 years and over:						
Total:						
Series A	10,405	12,291	12,321	3.5	3.4	2.5
Series D				4.1	4.3	3.8
Male:						
Series A	3,849	4,488	4,540	2.6	2.5	1.9
Series D				3.1	3.2	2.8
Female:						
Series A	6,554	7,802	7,782	4.3	4.3	3.2
Series D				5.0	5.4	4.7

[1] Base is the total for the class to which the figures refer.

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P-25, No. 381, "Projections of the Population of the United States, by Age, Sex, and Color to 1990, with Extensions of Population by Age and Sex to 2015" (Dec. 18, 1967), tables 9, 16, and 17.

TABLE 5.—TOTAL MONEY INCOME OF FAMILIES AND UNRELATED INDIVIDUALS, BY AGE OF FAMILY HEAD OR INDIVIDUAL, 1966

[Incomes in calendar year 1966 of persons of specified ages in March 1967]

Total money income or other measure	Families			Unrelated individuals		
	Total	Age of family head		Total	Age of individual	
		55 to 64	65 or older		55 to 64	65 or older
Number (in thousands)	48,922	7,689	6,929	12,368	2,502	4,879
Median income	$7,436	$7,586	$3,645	$2,270	$2,788	$1,443
Percent distribution by amount of money income:						
Under $1,000	2.3	2.7	3.7	20.3	19.9	25.1
$1,000 to $1,499	2.3	2.5	7.0	16.4	12.0	28.2
$1,500 to $1,999	3.1	2.4	10.9	9.3	7.5	14.8
$2,000 to $2,499	3.4	3.1	10.1	7.5	7.4	8.9
$2,500 to $2,999	3.2	2.7	8.9	5.4	5.6	5.4
$3,000 to $3,499	3.5	3.3	7.6	6.1	8.1	4.2
$3,500 to $3,999	3.3	3.5	6.1	4.0	4.1	2.1
$4,000 to $4,999	7.1	7.5	8.6	8.1	11.1	3.2
$5,000 to $5,999	8.4	8.7	7.8	6.4	7.3	2.3
$6,000 to $6,999	9.4	8.7	6.1	5.1	5.4	1.8
$7,000 to $7,999	9.3	8.5	5.5	3.3	3.5	.9
$8,000 to $9,999	15.1	13.9	6.6	3.6	3.5	1.3
$10,000 to $11,999	11.2	11.0	3.9	1.9	2.0	.7
$12,000 to $14,999	9.2	9.1	3.1	1.1	1.2	.3
$15,000 to $24,999	7.5	9.8	3.0	1.0	1.0	.6
$25,000 and over	1.7	2.7	1.1	.4	.6	.2
Families with full-time worker as head: Percent of total in this column, excluding Armed Forces	67.1	67.3	15.0	34.8	43.8	7.2
Median income	$8,693	$8,691	$4,956	$4,263	$4,615	$3,847
Nonfarm families:						
Number (in thousands)	46,225	7,074	6,395	12,068	2,441	4,730
Median income	$7,582	$7,825	$3,709	$2,300	$2,834	$1,444
Farm families:						
Number (in thousands)	2,697	615	534	300	61	149
Median income	$4,841	$4,563	$2,989	$1,295	(1)	$1,354

[1] Not ascertained (base less than 75,000).

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P-60, No. 53, "Income in 1966 of Families and Persons in the United States" (Dec. 28, 1967), p. 24, table 3.

TABLE 6.—NUMBER OF POOR HOUSEHOLDS AND INCIDENCE OF POVERTY FOR ALL NONFARM HOUSEHOLDS AND THOSE HEADED BY PERSONS AGED 65 AND OVER, 1959 AND 1966

Characteristics of household head	Number of poor households (millions)[1]		Incidence of poverty (percent of category total)	
	1959	1966	1959	1966
Total nonfarm (heads of all ages)	11.6	10.3	22.5	17.6
Total nonfarm, heads 65 years or older	3.9	4.0	46.0	37.0
White male	1.7	1.5	34.0	24.7
White female	1.8	2.0	59.3	48.9
Nonwhite male	.2	.3	64.4	51.4
Nonwhite female	.2	.2	76.3	69.9

[1] Households are defined here as the total of families and unrelated individuals.

Source: Economic Report of the President, together with annual report of the Council of Economic Advisers (February 1968), p. 143, table 21.

TABLE 7.—TOTAL MONEY INCOMES OF PERSONS, BY EMPLOYMENT, RESIDENCE, AGE, AND SEX, 1966

[Incomes in calendar year 1966 of persons 14 years of age and older in March 1967]

Category	Male			Female		
	Total	55 to 64	65 or older	Total	55 to 64	65 or older
Number of persons (thousands)	65,335	8,178	7,785	72,224	9,039	10,152
Number with income (thousands)	60,088	8,066	7,729	44,067	5,771	8,453
Median income of income recipients	$5,306	$5,750	$2,162	$1,638	$2,214	$1,085
Year-round full-time workers:						
Percent of civilian income recipients in this column	60.2	68.4	14.8	30.0	39.1	4.8
Median income	$6,955	$6,680	$5,382	$2,827	$4,067	$3,379
Nonfarm:						
Number of persons (thousands)	61,352	7,511	7,214	68,407	8,457	9,619
Number with income (thousands)	56,474	7,402	7,168	42,139	5,497	8,066
Median income	$5,457	$5,936	$2,179	$1,696	$2,287	$1,103
Farm:						
Number of persons (thousands)	3,983	667	571	3,817	582	533
Number with income (thousands)	3,614	664	561	1,928	274	387
Median income	$2,846	$3,053	$1,926	$838	$936	$806

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P-60, No. 53, "Income in 1966 of Families and Persons in the United States" (Dec. 28, 1967), p. 38, table 20.

TABLE 8.—AMOUNT OF TOTAL MONEY INCOMES OF PERSONS, BY AGE AND SEX, 1966

[Incomes in calendar year 1966 of persons 14 years of age and older in March 1967]

Total money income	Male			Female		
	Total	55 to 64	65 or older	Total	55 to 64	65 or older
Percent of income recipients in this column receiving—						
$1 to $499 or less	6.9	2.5	2.5	19.7	14.2	11.9
$500 to $999	6.0	4.3	12.2	16.2	14.5	33.9
$1,000 to $1,499	5.8	4.9	17.5	12.1	11.3	24.3
$1,500 to $1,999	4.3	4.3	14.1	7.3	6.8	9.8
$2,000 to $2,499	4.6	4.2	11.8	7.2	7.4	5.5
$2,500 to $2,999	3.6	3.4	8.9	5.4	5.9	3.5
$3,000 to $3,499	4.2	4.4	6.1	7.1	7.8	2.6
$3,500 to $3,999	3.7	4.0	4.5	4.5	5.1	1.4
$4,000 to $4,999	7.8	9.4	6.1	8.2	9.3	2.2
$5,000 to $5,999	9.7	11.3	4.3	5.6	6.7	1.6
$6,000 to $6,999	10.0	11.2	3.1	3.0	4.5	1.1
$7,000 to $7,999	8.6	8.8	2.4	1.5	2.2	.5
$8,000 to $9,999	10.8	10.3	2.4	1.3	2.4	.7
$10,000 to $14,999	9.8	10.7	2.3	.7	1.3	.5
$15,000 to $24,999	3.1	4.4	1.2	.2	.3	.3
$25,000 and over	1.1	1.7	.7	.1	.1	.1
Median income of income recipients	$5,306	$5,750	$2,162	$1,638	$2,214	$1,085

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P-60, No. 53, "Income in 1966 of Families and Persons in the United States" (Dec. 28, 1967), p. 38, table 20.

TABLE 9.—MEDIAN MONEY INCOMES IN 1966 IN RELATION TO YEARS OF SCHOOL COMPLETED

[Income in 1966 for persons of specified ages in March 1967]

Years of school completed	65 and older	Ages 55 to 64	Nonfarm, 25 and older		Farm, 25 and older	
			White	Nonwhite	White	Nonwhite
MALE						
Total	$2,162	$5,750	$6,526	$3,800	$3,862	$1,295
Elementary school:						
Total	1,885	4,339	3,865	2,765	2,903	1,233
Less than 8 years	1,738	3,335	3,032	2,522	2,329	1,176
8 years	2,185	5,331	4,781	3,723	3,451	(1)
High school:						
Total	2,681	6,412	6,811	4,801	5,089	(1)
1 to 3 years	2,541	6,117	6,258	4,385	4,311	(1)
4 years	2,841	6,755	7,178	5,214	5,376	(1)
College:						
Total	4,270	9,177	9,095	5,953	5,383	(1)
1 to 3 years	3,685	7,602	(2)	(2)	(2)	(2)
4 or more years:						
Total	4,856	10,933	(2)	(2)	(2)	(2)
4 years	4,207	10,840	(2)	(2)	(2)	(2)
5 or more years	6,433	11,081	(2)	(2)	(2)	(2)
Median number of years completed	8.5	9.9	12.2	9.1	9.0	[3]8.0
FEMALE						
Total	$1,085	$2,214	$2,030	$1,629	$1,199	$613
Elementary school:						
Total	926	1,404	1,268	1,067	826	590
Less than 8 years	870	1,070	1,093	978	741	599
8 years	1,010	1,788	1,441	1,353	926	(1)
High school:						
Total	1,360	2,583	2,462	2,097	1,518	(1)
1 to 3 years	1,221	2,148	1,988	1,736	1,132	(1)
4 years	1,480	2,894	2,742	2,505	1,722	(1)
College:						
Total	1,846	4,513	3,554	3,993	2,707	(1)
1 to 3 years	1,660	3,547	(2)	(2)	(2)	(2)
4 or more years:						
Total	2,237	5,890	(2)	(2)	(2)	(2)
4 years	1,912	5,214	(2)	(2)	(2)	(2)
5 years or more	3,500	7,452	(2)	(2)	(2)	(2)
Median number of years completed	8.7	10.8	12.1	9.9	11.2	[3]8.0

[1] Median not ascertained (base is under 75,000).
[2] Not available.
[3] Less than number shown.

Source: Department of Commerce, Bureau of the Census, Current Population Reports, series P–60, No. 53, "Incom. in 1966 of Families and Persons in the United States" (Dec. 28, 1967), p. 40, table 22.

TABLE 10.—INCOME SOURCES AND MEDIAN INCOME OF COUPLES AND NONMARRIED PERSONS AGED 65 AND OLDER, 1962

Source of income	Percent reporting income from specified source			Median income of recipients from specified source		
	Married couples	Nonmarried persons		Married couples	Nonmarried persons	
	(percent)	Men (percent)	Women (percent)		Men	Women
Earnings from employment	55	28	23	$1,485	$1,065	$840
Reirtement benefits	84	72	64	1,605	980	770
Old-age, survivors, and disability insurance [1]	79	68	60	1,405	905	740
Other public systems	12	8	7	2,030	1,380	895
Private group pensions	16	10	3	775	630	645
Veterans' benefits	14	11	6	780	770	765
Interest, dividends, and rents	63	45	50	280	180	175
Private individual annuities	4	1	3	(2)	(2)	(2)
Unemployment insurance	3	1	1	(2)	(2)	(2)
Public assistance	8	18	17	850	650	735
Contributions by relatives [3]	3	1	6	75	(2)	285
Payments under any public program	89	87	78	(2)	(2)	(2)
Public and private retirement benefits other than OASDI	(2)	(2)	(2)	980	910	810
Number of units (thousands)	5,445	2,402	6,329			
Number reporting on source (thousands)	5,443	2,345	6,267			

[1] Excludes part-year benefits and parent beneficiaries.
[2] Not shown in survey report.
[3] Relatives or friends not in same household.

Source: Department of Health, Education, and Welfare, Social Security Administration, Division of Research and Statistics, The Aged Population of the United States: The 1963 Social Security Survey of the Aged (Research Report No. 19, by Lenore A. Epstein and Janet H. Murray, 1967), pp. 286–287, tables 3.1 and 3.2.

TABLE 11.—SIZE OF INCOME OF COUPLES AND NONMARRIED PERSONS AGED 65 AND OLDER, 1962

Amount of income in 1962	Married couples			Nonmarried persons							
					Men			Women			
	Total	Beneficiaries [1]	Non-beneficiaries	Total	Total	Beneficiaries [1]	Non-beneficiaries	Total	Beneficiaries [1]		Non-beneficiaries
									Retired	Widowed	
Number (thousands):											
Total	5,445	3,743	1,120	8,731	2,402	1,490	803	6,329	1,912	1,502	2,543
Reporting on income	4,719	3,289	932	7,709	2,173	1,384	685	5,536	1,690	1,325	2,192
Total (percent)	100	100	100	100	100	100	100	100	100	100	100
Less than $1,000	5	4	10	44	32	26	46	49	36	44	65
$1,000 to $1,499	10	9	12	22	25	32	13	21	23	27	14
$1,500 to $1,999	14	15	11	13	12	14	10	13	17	16	7
$2,000 to $2,499	13	16	5	8	11	13	6	7	9	6	4
$2,500 to $2,999	12	14	6	4	5	6	3	3	5	2	2
$3,000 to $3,999	16	16	12	4	6	5	4	3	4	2	3
$4,000 to $4,999	11	11	10	2	3	2	4	1	2	1	1
$5,000 to $9,999	15	12	24	4	6	2	12	3	4	2	2
$10,000 or more	5	3	11	(2)	1	(2)	1	(2)	(2)	1	(2)
Median income	$2,875	$2,710	$3,580	$1,130	$1,365	$1,375	$1,145	$1,015	$1,300	$1,105	$755

[1] Excludes part-year and parent beneficiaries. Retired women receive benefits based on their own wage record, regardless of eligibility as widows; widows receive benefits based on husband's wage record.

[2] Not shown where 0.5 percent or less.

Source: Survey report cited for table 10, above, p. 288, table 3.3.

HARNESSING SOCIAL PROGRAM DESIGNS TO PUBLIC POLICY OBJECTIVES

BY GEORGE F. ROHRLICH*

I. THE PRESENT SOCIAL SCENE: A PROBING WAR ON POVERTY AND GROWING MALAISE

Despite the Government's solemn proclamation, in 1964, of a war on proverty, legitimate doubts can be raised as to whether the abolition of poverty in our time is truly an accepted social policy objective and whether, in fact, we have an agreed-to social policy agenda at all to which this Nation is committed at the present time.

In a country which has fought some tough wars and has never lost one, the proclamation of a war is bound to be fraught with high expectations of success, even, if need be, at the price of heavy sacrifices. Those most concerned with the outcome of this war, the poor and the underprivileged, expect no less. If successes in the war on poverty to date are a far cry from its victorious conclusion and there is little evidence of our preparedness to incur heavy sacrifices to achieve such victory, no one ought to be surprised if disappointment is widespread and intense and a feeling of letdown prevails that is more acute now than if the vision had never been raised.

Yet, at the time this is being written, a serious question exists as to whether the antipoverty program will continue and, if it does, whether it will receive even as much as its previous annual appropriations of around $2 billion that were geared to preliminary explorations and pilot programs on a relatively small scale. Of course, it must be remembered that the array of measures lumped together under the economic opportunity program constitute merely a part of the war on poverty. In fact, one might dub the OEO program the "little" war on poverty. For a multiple of the amount invested in it is being spent on related programs of education and manpower development, and amounts equivalent to many times the annual OEO appropriation have been spent for many years on long-established social programs, including social insurance, public health and welfare programs. However, and this is the important point, that which is new in the war on poverty is being treated niggardly, not to say starved, in light of the magnitude of the tasks that confront us and that are left substantially untouched by the traditional programs. Moreover, we have come to recognize, at least in bare outline form, not only the primary problem areas, but some of their reverberations—secondary and even more serious social dangers born of problems that have gone unmet far too long.

As yet, we have not dared to call these phenomena by their proper names. An unappealing, yet truthful, identification of some of the principal ones must include: (1) The development of a homegrown

*Professor of Political Economy and Social Insurance, Temple University School of Business Administration, Philadelphia, Pa.

proletariat: a distinct minority living under conditions which the large majority would consider utterly intolerable for itself and whose alienation from society is fast becoming a way of life, (2) the large-scale existence in our urban centers and elsewhere of *de facto* non-family, non-home, and non-parent situations which leave masses of young children and teenagers to their own devices, untutored, unprotected, and unguided, many of them in settings in which all they can acquire by learning from observation and through imitation is either asocial or antisocial attitudes and practices. Partly as a consequence, and partly as a cause of the foregoing, (3) the existence of large numbers of practically untaught and, with traditional methods unteachable, children who lose further ground in their education as they pursue it over the years; and, to make matters still worse, (4) a widely prevalent neglect of their even more basic needs, i.e., for proper nutrition, for their general, dental, and mental health. (It is by no means implausible that (4) in the earliest years of life may account for much of (3) in the years following.)

Many of these perturbing phenomena are no longer hidden from public view as they once were when we could talk about our "invisible poor." By now, they must be known far and wide. We read about them in our daily papers, numerous magazine articles, and books; we view and review them in any number of TV programs depicting the festering slums of our cities. In increasing numbers we come to know at first hand some of the fallout in the form of the acute insecurity of our cities and the vanishing opportunities for the carefree enjoyment of one's personal gains in the form of such simple and supposedly free pleasures as the use of public facilities for recreation, be they public parks, public waters, or simply fresh air.

Instead of improvements in the slums we see slum-like conditions brought into being by a widespread spillover of disaffection with society among youngsters from families of middle-income and upper-income levels ostentatiously displaying and proclaiming their lack of socialization, i.e., their resistance to espousing the values of, and to taking their rightful places within the established social organizations of the Nation. We bemoan the transformation of cleancut teenagers into beatniks and hippies, and of gifted boys and girls into aimless seekers after opiates of all kinds. But we seem to be spellbound by this observation and at a loss for a remedy.

Presumably, we continue to expect government to attend to its oldest function, that of maintaining law and order. Yet we make it increasingly more difficult, if not impossible, for government to do this by allowing a social situation to perpetuate itself which is simply not conducive to the maintenance of law and order. We are hesitant to appeal to authority for the enforcement of compliance with traditional (not just "middle class") values of respect for one's neighbor and the proprieties of civilized living. But neither are there in evidence on the social horizon any other ideals that would radiate authority of a different kind, namely, a self-enforcing moral code.

Curiously, we seem to be eager—or, at any rate, content—to delegate to "private initiative" the most gigantic of tasks, such as slum clearance and urban renewal, while expecting government to operate some rather marginal ameliorative programs to cope with some of the most difficult and involved problem complexes of our time at cut-rate prices, cheaply, so as to ease our tax burden.

Meanwhile, presumably pending improvement in the situation by "letting things work themselves out," there is a growing flight from the cities into safer and more attractive suburban areas. The out-migration of dwellers is being followed by the exodus of shopkeepers and churches. In consequence of this movement, the 1960's, which have stood so prominently under the sign of a concerted effort toward greater integration of the races, have seen a pronounced rise rather than a fall in the rate of segregated living in all the urban centers of the United States.

History is replete with precedents that tell the outcome of letting a social climate deteriorate for too long. The decomposition and polarization of nations has traditionally led either to their demise or to violent efforts at national regeneration and reintegration. Frequently, these have taken on either radical or reactionary demagogic and not seldom dictatorial forms and contents, avowedly in an effort to revive national or other collective virtues and ideals, but in fact accompanied by bestial antihumanism and disregard for the individual and individual values. Not infrequently, new social cohesion was brought about at the expense of minority groups singled out as scapegoats and whipping boys, so that the accumulated frustration of the masses could translate itself into steered aggression and thus serve as a unifying bond for the majority.

To know this much, we hardly need more social research. What we do seem to need is the courage of our professed convictions and a resolve to act on the basis of that which we already know. If we want to cut the ground from under the demagogs that are sure to make their appearance and to gain a mass following if things continue as at present, then we simply cannot permit conditions to persist indefinitely that we readily recognize as pernicious for civic coherence and incapable of engendering any positive feeling of commitment to the Nation. Rather, we must be able to point to some very substantial improvement, or at least a visible all-out effort at such improvement. Most certainly, the conditions outlined cannot be left to piddling approaches and spontaneous small-scale displays of goodwill. In particular, the government cannot wish off on private civic interests, no matter how eager they may be to make their contribution, the quest for a solution of problems that go far beyond the radius of effectiveness of mere "community action."

Even the most selective enumeration of some of the major social ills that beset us suggests the need for a comprehensive and coherent national approach to meeting them. It must consist of an overall plan for action and a tentative time schedule. At the present time, nothing of this sort exists. Among the reasons that may account for this lack, two principal ones suggest themselves. One is the apparent lack of avowed national domestic goals other than the purely macroeconomic ones of stability and growth in terms of national output, measured in dollars of final sales. The second is the unavailability at the present time of suitable measures of the diseconomies and losses outside of the purely economic sector that society sustains as a result of the existence of the social problem conditions referred to, and many more like them.

II. National Goals and Policy Choices

In the last resort, any attempt to evolve a set of national goals hinges upon what one considers to be the proper relationship between

individual wealth and commonweal, between individual and collective rights.

Any clarification of the relationship between additions to individual wealth or personal gains generally and collective gains must take cognizance of a fact of paramount importance, to wit: that it is both *de facto* impossible, and also quite out of keeping with the basically egalitarian character of American society, to maximize *and enjoy* individual enrichment in close proximity to very much less favorably situated fellow citizens. Such enjoyment would be possible today only if it were feasible to hide the more favored mode of living from the eyes of those who cannot partake of it. This, clearly, involves the geographic and social segregation between those who have the means to enjoy the "better things in life" and the rest, who do not. Needless to say, this goes directly counter to any and all moves for greater integration.

The enjoyment of wealth, or well-being in general, thus becomes much more nearly than ever before comparable to the enjoyment of good health; the ever broader access to the respective type of enjoyment is a prerequisite for both. What we have long known to be a necessary preventive of the spreading of communicable diseases is their eradication from an entire neighborhood, community, or country. What we have been slower to understand is that protection against the dangerous fallout from acute poverty and underprivileged conditions of other kinds, similarly, depends upon their eradication from an entire neighborhood, region, or country. Short of complete ignorance by the poor of how the folks on the other side of the tracks live, the only alternatives are either the willing acquiescence on the part of the have-nots in the given state of affairs, or the imposition of such acquiescence by sheer force or threat of force. None of these conditions would appear to be present in the United States today, nor do we hope or can we wish that this be different.

If it thus turns out to be impossible to confine the fallout of poverty by quarantining the poor, and if such fallout threatens the lasting achievement and unconditional enjoyment of whatever social betterment is sought after by the rest of the population, then it would appear self-defeating for the more affluent majority to resist whatever tax burdens are necessary to establish conditions whereby the incidence of extreme poverty and suffering due to it can be reduced significantly, and in due course abolished altogether. About this point, more later.

Just what are the components of greater well-being, anyway? Surprisingly, there is little that is explicit about the components of our contemporary quest for individual or collective maximization of well-being. Nothing even approaching a systematic or exhaustive analysis can be attempted in the present limited context. However, as a first approximation, our most general answer might be simply: "the realization of our potential as individuals and as a commonwealth." Leaving aside the principal foreign policy aims, whether political or economic, viz., to assure the survival and security of the Nation, this twin objective would appear to include the following more specific goals: the largest possible production of goods and services needed or desired of which our Nation is capable; full employment, i.e., jobs for all those who wish to work or who need to work, with Government as employer of last resort. But, most certainly, it must mean, too, the avoidance of mass alienation and of proletarianization of society, i.e., the existence

of a substantial group of dispossessed and disaffected individuals that feel neither loyalty nor commitment for the Nation. To that end, it must involve the abolition of avoidable conditions of need and dependence, and adequate provisions for those unavoidably dependent on the help of others. This in turn suggests the necessity of maintaining in dignity the underprivileged and dependent of all ages and caring especially for the children, most particularly the neglected ones. Safety and peaceful pursuit and enjoyment of the fruits of one's labor, surely, need to be part of this enumeration, as must be a healthful and appealing environment in which to do so. This suggests, last, but probably of topmost importance, that we must aim at a reasonable ratio of population to the area and resources of the country, i.e., a growth that is controlled so as to reduce the rate of our population increase and, eventually, lead to stabilization of the population. This, more than any other single condition of our well-being, constitutes, at the same time, an important safeguard for future generations of a degree of well-being at least equal to our own.

Virtually each of these broad objectives—if adopted as a national policy resolve—raises important and difficult moral choices. To name just a few: Does the objective of the largest possible production of goods and services take precedence over the preservation of local conditions that make for cohesion, i.e., take precedence over social stability? Quite frequently, the maximization of economic efficiency involves rapid and radical changes of all kinds: outmigration from areas with depleting resources; the abandonment of obsolescent trades and production methods; the fastest possible introduction of the newest labor-saving machinery and industrial organization in general, etc. But at the same time that each of these radical innovations may cause us to take a giant step forward in the direction of making our economy more efficient and, hence, toward greater abundance of the things people need or want, it may also yield us severe social dislocation and decay in the form of rundown and rapidly deteriorating neighborhoods or even entire ghost towns, human obsolescence, premature withdrawals from the labor force, and other forms of social disintegration and waste of human resources. Which, on balance, is the greater: the net economic gain or the net social loss?

In connection with the full-employment objective, many such questions pose themselves. E.g., in referring to someone's need for a job, what about the conditions that cause that need? The participation of women in the labor force, for instance, raises the thorny issue concerning working mothers as against mothers who stay home to rear and educate their children. Is the need of a job on the part of either woman to be judged merely in terms of the possible good that the addition of her income as a secondary earner in the family can make to the family's well-being? If it is, then full employment must mean the availability of jobs for all those mothers who consider that they need, or who simply desire, paid employment.

Yet, it seems quite conceivable, and perfectly reasonable, to posit an alternative, viz., that the first *social* need is for a mother to take care of her children. It could very well be argued, however, that either answer ought to carry with it a corollary obligation on the part of society. The first, the provision, if necessary, of suitable jobs for mothers (of young children) seeking work *and* suitable day care centers for their children; the second, some compensatory payment or

allowance for those families in which mothers stay home to take care of their children and thereby forfeit the opportunity to supplement the family income.[1]

The problem becomes more intricate, and an unequivocal answer even more difficult, when the mother is the only parent in the family. It assumes further complications in both two-parent and one-parent families in which the parents are either unwilling or unable to take proper care of their children, i.e., provide for them a suitable home.

Further social-policy questions related to the problem of working mothers arise in connection with the necessary—or, at any rate, highly desirable—assurance of proper prepartum and postpartum care.

One aspect of this whole problem area takes on special significance in connection with the objective of avoiding mass alienation. It concerns the pros and cons of relatively large-scale institutionalization of certain types of persons, especially young persons. This problem has many policy angles. It presents itself in connection with the unsuitable-home predicament already touched on; in connection with wayward and defiant youngsters, notably criminal or precriminal problem youths; in respect of youthful dope users and vagrants. Is "individual freedom" of those proximately concerned the only or supreme value at stake here—as is alleged or implied most of the time—or are there countervailing values, perhaps equally worthy of society's concern, viz., the protection of both individual and collective health and integrity, i.e., self-respect and responsibility?

Could it be argued that in between custodial provisions and facilities for neglected infants and youngsters on the one hand, and the needs to train some of the more mature youths for purposes of national defense on the other hand, there is a need for some broader socialization and acculturation programs for teenagers?

Both relative to the mass alienation problem complex and the kindred problem area of proletarianization, merely two out of a large number of the most vexing moral and social problem quandaries may be mentioned here. One refers to the opportunity for work and work training, the other to the entitlement of individuals and families to at least minimum-adequate support by society and the corresponding rights, if any, of society vis-a-vis such individuals and families.

If paid employment and training for such employment are two of the main levers of massive emancipation from the slums—as, undoubtedly, they are—then the role of labor unions in their capacity as regulators of effective labor force participation and apprenticeship training assumes new aspects. In addition to being a one-sided interest representation of organized labor, a union thus becomes vested with the public interest. Given this broadening of responsibility, is it possible any longer to reconcile exclusive rights of representation in the bargaining about wages and working conditions and, most particularly, the admission to apprenticeship training without simultaneous guarantees and enforcement of open membership and equal opportunity policies and practices in each of the unions involved?

Once society assumes and implements the twin obligation of supplying subsidiary employment opportunities for those unable to obtain other gainful employment and of furnishing the essentials of living

[1] In France, a social benefit in the nature of an allowance for "single-earner" families exists, based on this reasoning.

at levels of at least minimum adequacy to those in need for whatever reason, should it not exact in return any productive contributions to the extent of their respective abilities from those aided, notably the acceptance by them of suitable opportunities for gainful employment (unless socially contraindicated, e.g., because of a greater need for a mother of young children to remain at home) or training for such employment? If not, why not? What other guarantees can and should society adopt to protect the integrity of an up-to-date, truly adequate and dignified mode of aiding the needy—if it is not to fall back on miserly and demeaning forms of relief giving? If the answer is "yes," what shape are such subsidiary employment and training programs to take? How are they to be administered so as to make them both constructive and free from any punitive or vindictive characteristics?

Probably the greatest issues of morals and of public policy are those stirred up in squarely facing the need to limit our population growth. The questions to be faced in this connection are far broader than those relating to public welfare. Yet it is within the problem area of public welfare that they take on a specific form and urgency. Given the commanding necessity for population growth control, there no longer can be any question of the right of individuals and families to the fullest knowledge of contraceptive methods and of the general access to contraceptive devices medically certified as safe for human use. The additional and more far-reaching, yet plausible, inferences pertain to the advisibility of granting equally or nearly as ready access to legal abortion and voluntary sterilization. This issue takes on added weight in cases where offspring is foreseeably going to become a public charge, let alone predictable medical or social problem cases. To put the problem succinctly: Does or doesn't society, in return for an unconditional guarantee to maintain all its members in freedom from want, have a right to expect each member to consider himself or herself under some obligation to limit avoidable dependence on this guaranty? Weighty as is the question, it is by no means novel or revolutionary. It is really in the nature of a "*certiorari*" of Justice Holmes' famed dicta to the effect that taxes buy us civilization, and that "three generations of imbeciles are enough." [2]

These are moral quandaries that are bound to raise severe disagreements concerning the corresponding policy goals at stake, and in respect of the values of individual and collective self-realization. We must face them squarely and decide them in common council as intelligently as we can, in light of all relevant information that we possess. In a democracy, obviously, such matters must be discussed and thrashed out on the widest possible basis, and the decisions must be made through the duly elected representatives of the people. But, before this can take place, the issues must be clarified and put for public discussion. This is now the awesome responsibility of government and of academic circles—not to the exclusion of others, of course, but certainly of these.

Beyond this clarification of issues, goals, and policy choices to reach them, there lies the laborious exploration of alternative approaches by which the desired aims can be achieved: Their comparative advantages and drawbacks, the costs and other sacrifices involved in each, and the weighing of their respective merits in light of them.

---

[2] *Buck* v. *Bell*, 247 U.S. 200, 47 S. Ct. 584, 71 L. Ed. 1000 (1927).

Again, it is the particular responsibility of government and academia to come forward with thought-through program proposals that can be discussed publicly and can be submitted to policymakers for their evaluation and support and, ultimately, to the public for adoption or rejection. To date, very little—if any—such comprehensive planning (in the sense of fostering informed expression and intelligent judgment on vital issues to be faced) appears to have taken place.

Clearly, then, a comprehensive social policy agenda needs to be spelled out and in principle agreed upon. It should contain a listing of the problem areas that confront the Nation and that will call for an answer during the next several decades. Obviously, any projections and quantifications involved cannot be expected to be precise or unequivocal. What counts, however, is not just the exact quantitive aspect but is equally, and in some respects primarily, the nature of the problem and of its likely evolution (a) in light of possible efforts to cope with it, and (b) in the absence of such efforts. Only when this type of agenda exists can we begin to program effectively and meaningfully, and to budget realistically for adequate programs. Only then can we evaluate progress by examining whether and to what extent the problems thus identified have been alleviated or met—although this task in itself will require much conceptual and developmental work on reliable gauges of social progress that we are currently lacking.

III. Goal-Specific Social Programing and Costing

Obviously, all programs avow a purpose. What they lack is a built-in evaluation mechanism and feedback devices to point up their own shortcomings and to bring about the corresponding amendments. Nor do existing programs necessarily possess the facility of readjusting themselves in light of changing needs. Nor are they susceptible to easy readjustment in light of such needs, unless these needs can be spelled out very concretely and couched in operational terms that readily fall in place. Unless some overall frame of reference exists, it is difficult if not impossible clearly to comprehend the role that given programs are to play, to evaluate the extent to which they serve their proper function, and to appraise both gaps and overlapping between programs. In addition, the effective coordination of programs and further program development planning depend upon the clarification of the approach that needs to be chosen to achieve certain ends. Both the mechanical cumulation of programs of similar type when a given type of program could well serve related needs and, conversely, the inherent limitations of given approaches as against possible new approaches will become manifest only within such a need-oriented comprehensive frame of reference.

For the longest time, for example, we have used the vehicle of social insurance almost exclusively for the purpose of income maintenance. Thus, we have chosen to reduce a variety of needs to a common cash level. Only with considerable delay was the recognition forced upon us that certain needs cannot be met by cash benefits alone, for example rehabilitation and training needs, the need for hospitalization, and related services. Admittedly, most needs could be translated into cash needs. But it is highly questionable whether this procedure would be the most effective in terms of costs and in terms of being conducive to

achieving the desired objectives. Among the unmet needs—aside from the primary one for gainful employment for all those able and willing to work—the need for proper housing and basic health care, certainly, command priority attention. Subsidization and reimbursement measures, undoubtedly, could go a long way to help improve the situation. But a look at our housing problem, e.g., would make clear that the mere availability of purchasing power in the hands of those seeking decent housing would not of itself provide the supply of such housing, or would—if at all—supply it only at unduly high costs or with such unconscionable delays as to defeat the purpose. Health care for those not now included in our public programs, or for certain sectors of those not now included, notably children, presents a similar problem. It is precisely because we have been stuck in grooves long established that we have failed to recognize that our ever-growing problem of dependent children, e.g., cannot be solved by merely supporting their mothers at minimum rates and under conditions that are not conducive to changing their predicament.

These examples should suffice to illustrate the fact that not just any available program can or should be adapted to changing needs. Rather, in the interest of getting the most mileage out of program development, any extensions of existing programs or their transformation, with or without the addition of new ones, should be made in the form in which and at the strategic points at which their effect can be maximized in terms of doing the most for the attainment of any given purpose. By the same token, to derive the greatest beneficial impact from programs that do effectively serve given purposes, as in the case of our basic social insurance programs serving the principal income maintenance needs in the common contingencies, the benefits available ought to be such as to meet these needs adequately. There is little logic in establishing a program properly focused on the needs to be met by it and then shortchanging it, i.e., undercutting its effectiveness, by denying it the proper dimensions.

No one familiar with the exigencies of the political process and the limitations which it imposes upon the realization of purely rational proposals, would maintain that the mere existence of a well-reasoned proposal is all it takes, or will assure its adoption in precisely the fashion proposed. However, the important thing is to spell out, first of all, each of the problems that beset us in all its aspects and ramifications, then to research all possible ways of alleviating or meeting each of them, to put price tags on these various possibilities. The nearest that we have come in recent years to performing this task was the appointment in 1960 of a Presidential Commission on National Goals. Its report, *Goals for Americans*, set forth a list of policy targets in rather broad terms. These targets were later subjected to cost estimates by Mr. Leonard A. Lecht of the National Planning Association (*The Costs of Our National Goals**). Not only in terms of specifics, but in one other important respect, would the type of estimate here advocated have to go beyond such traditional concepts of costing. It seems no longer sufficient to estimate the probable costs of a program that is being proposed or even the possible costs of a variety of programs that may be advocated to meet given needs. What is called for, in addition,

*National Planning Association, Washington, D.C. 1965. 66 pp.

within a framework of meaningful social accounting, are estimates of the costs to society that are likely to arise if no program is established and no other measures are taken to meet an identifiable need and if therefore, as a consequence, that need goes unmet. The latter costs must take account not only of proximate results that can be anticipated but also of any induced conditions that will predictably follow if action is not taken, or is not taken in adequate or timely ways.

Several examples may serve to illustrate what is meant here. To arrive at a realistic valuation of the task of urban renewal, e.g., it would not be enough merely to cost possible rebuilding programs, or even to include in one's cost estimate the financing of the accompanying positive socio-economic policies that are conditions of success. To be sure, this is an impotrant part of the job of costing. However, the price tag is put in proper perspective and assumes meaningful proportions only by taking account of the alternative costs to society of letting present developments take their course. What are these?

Increasingly, the percentage of uniracial occupancy of our urban areas has risen over the past several years. Negro occupancy has attained the two-thirds mark in Washington, D.C., and is approaching the halfway mark in other cities. Conversely, the suburbs have remained or have increasingly become white uniracial areas of occupancy. There has been some exodus from the slum areas. However, of 1½ million slum-dwelling nonwhite families at least half a million are not poor and could leave the slum if there were places to move to. Remaining in the slums increases for them the danger that they or their children will backslide into poverty. How much will it cost us to court that danger rather than to free the path for the exodus? Among those who have to remain because they have neither the opportunity nor the means to move out, the concentration of festering social sores is bound to be all the greater. How much will it cost us to conserve in the midst of our big cities cesspools of asocial and antisocial manifestations and ways of life?

Similarly, the immediate cost of "savings" involved in paying or refusing to pay the maintenance cost (and of doing or refusing to do a lot more) for dependent children of needy families (or mothers) gives but a very incomplete and, therefore, highly deceptive idea of the full cost to society that either of these alternative courses of action entails. Even with Head Start and other compensatory trial programs we find that the return of the disadvantaged children to their disadvantaged homes undoes whatever gain the compensatory program bestows upon them. The ramifications of this problem are vast and staggering. The total impact on our society is bound to exceed by many times over both the present (or presently proposed) program costs and the savings that can be obtained by any further cuts in them.

While we have succeeded in increasing the average schooling of Negroes and have brought it within a small margin of that for whites (near 12th grade), their actual educational attainment still hovers 3 years below the nominal one (around ninth grade)—barely enough for functional literacy in terms of modern-day living. Again, it stands to reason that neither true costs nor savings are meaningfully portrayed in the dollar total of educational expenditures incurred.

It is to be expected that the way in which taxpayers will view the costs of goal-specific programs—once these have been fully explained

and widely discussed—and the burdens of new taxes needed to finance them will be greatly influenced by a realistic appreciation of the real (even though to-date still largely hidden) costs of stable or lightened tax burdens which can be achieved only at the price of letting things take their course. Put differently, it is to be expected that the real value of our private dollars (after taxes) may fall very greatly, even if their purchasing power remains intact, in consideration of that which these private dollars can—or cannot—buy in the way of enjoyable living. By the same token the undoubtedly heavy burden of taxation required to improve in a significant way the conditions under which private life will be carried on—even if staggering compared with the past—would look much more acceptable as the even heavier costs of "cheaper" alternatives are made explicit. "Value received" for taxes paid may come to be acknowledged at last, and the Holmes view of taxes as the price of civilization might come to prevail.

IV. Social Welfare Programs and the General Welfare

As may be apparent from the foregoing examples, the view here taken of the "welfare state" conceives of it as a state of mind. In this view, Sweden and Great Britain, e.g., are "welfare states" not because they have achieved extraordinary levels of general welfare, but because their citizens have accepted the essential causal connection between individual sacrifice and reward attendant to living in the closely interconnected society of our complex and technically advanced age, and because they willingly pay the necessarily high price of orderly living in this kind of society.

Such a civic state of mind cannot come about, nor could it be long maintained, without the understanding and acceptance by the average citizen who must support it and provide the wherewithal. Without the proper education of the younger generation the sustained and sympathetic endorsement and continued evolution of this social philosophy would be unthinkable.

The awareness of this prerequisite is bound to arouse concern among informed observers of the American social scene over the ever-recurrent attempts to discredit the American social security system. These attempts are particularly worrisome when addressed openly or in thinly veiled fashion to the young who are about to enter the labor force, and they are most objectionable when they fall little short of charging that a "fraud" is being committed (e.g., with regard to the social security reserves), and that the young are being "exploited." The exact opposite message to the young would seem more in keeping both with the truth and with the public interest. With regard to old-age and survivors' insurance, for example, it would seem to be important to make youngsters understand the great social significance and importance of programs of this kind and to make them more, rather than less, willing to shoulder the cost thereof—in their own interest as well as society's—rather than to instill suspicion and ill will against "costly" old-age schemes that allegedly "will never really benefit" the young.

Of course, the question of individual freedom, notably the freedom of choice, and the ways and whys of certain limitations that society puts on it through the establishment of social programs and through other necessary measures must forever be discussed and debated anew.

But the common good for whose protection these limitations presumably exist is no less deserving of being examined and reexamined. There are, indeed, good reasons for asking whether a free society that expects its citizens to protect themselves and their families through their own efforts and by means of a variety of private measures against poverty and dependence may not at the same time demand protection in its turn against the very real possibility that such foresight may be lacking in some or may not effectively be exercised—whether by choice or by limited ability—and may thus, ultimately, cause to be passed on to society the task of providing the essentials in these instances.

Or, more broadly stated, why not raise and thoroughly discuss throughout the Nation this fundamental question: Does society, in assuming, as it must, at least a residual responsibility to prevent want, avoidable disease, ignorance, squalor and idleness (Beveridge's "five giants"), have the right or does it not have the right to establish safeguards against wanton action on the part of individual citizens that manifestly add to the dimensions of these problems?

Stated in an even more comprehensive way, the big and important problem that needs to be spelled out bluntly and discussed widely consists of the two sides of the question (as it presents itself from the individual's and from society's point of view) about the right against unwarranted impositions (*on or by* an individual) that have neither a rational nor a charitable justification. For example, does an individual's rights of freedom of movement necessarily preclude any system of registration whereby a society threatened by crime, both organized and unorganized, could more easily keep track of the offenders? Or, to cite yet another example, does the constitutional right of citizens to bear arms preclude registration of such arms and other, substantive, qualifications to its exercise in a society cursed with violence and among whose most effective bearers of arms are gangsters and a motley crew of wayward individuals threatening the rest of citizenry in its peaceful pursuits?

Clearly it is a country's social philosophy which determines to what extent and by what means society through its agents is to act as the safekeeper of the public interest. The same can be said in respect of a society's approach to social welfare in the narrower sense, e.g., its social security fabric. Yet, in the latter connection the economic attainment of a nation appears to have an important bearing. To be sure, affluence is not a prerequisite of the "welfare state" as here understood, witness the examples of Sweden and the United Kingdom cited above, both countries in which the foundation of the welfare state predates the attainment of affluence. However, there seems to be some connection between the achievement of a certain degree of comfort for the majority and a shift toward more sophisticated welfare programs including, in particular, service programs and special bring-up-the-rear schemes.

At the start of capitalist development, when the majority lives in poverty, the approach is a highly selective poor law. As the majority attains a measure of comfort, the approach to welfare programing appears to take on a risk-specific nature. At that point, there is still selectivity, but the important transition is made from needs programs toward programs awarding benefits as of right. Most recently, as in our society and in Western society in general, a greater measure of

affluence is gradually being attained by substantial numbers of citizens; their desire for greater comfort increasingly results in a demand for amenities and for a change in the quality of living toward a more cultured and gracious kind of life. Concomitantly, the approach to welfare also is broadened. The concepts of "poverty" and "dependency" tend to be redefined in a more liberal light and to include larger numbers. Service programs take on greater emphasis as well as special programs referred to above as "bring-up-the-rear schemes." These gradually emerge as the necessary conditions—part of the infrastructure, as it were—of a more affluent sort of life which would otherwise constantly be in jeopardy due to envy, hate, and active threats from the underprivileged sector.

This transition is difficult if not impossible to perform, however, unless a pronouncedly individualistic philosophy makes some room for an admixture of another ideology, that of social solidarity. It is this admixture of social concern which makes for the acceptance of the broader, more inclusive, and more generous programs. The combination of a broader and deeper, more enlightened understanding of self-interest and the greater generosity born from greater affluence (which is less concerned with minor seepages through abuse and other enforcement gaps) makes possible the institution and operation of such newer and more broadly conceived social schemes.

If, and when, this needed breakthrough toward a more widely aware and perceptive social conscience is achieved, a plausible agenda for social action can readily be derived from our poverty-research findings of the past several years. While the approaches and mechanics are wide open for discussion, the human targets are clearly earmarked in terms of the types of persons and families to be benefited. They are, in order of magnitude: (1) Minor children and families with minor children, notably those headed by females and those headed by males whose employment and/or earnings are insufficient to raise them above the poverty line; (2) the aged; and (3) the disabled.

A variety of approaches readily offer themselves: provision of cash benefits, where necessary on the basis of a humanized means test, and selective subsidies to stretch incomes of those now working and earning or prepared to work but insufficiently equipped to achieve earnings that will support them. One particular subsidy that would go a long way may take the form of children's allowances. In one fell swoop they would pull out of poverty a substantial number of the poor who have nothing wrong with them, except that they have more children than they can support. Other subsidies which could accomplish much at the present juncture are employment and housing subsidies for marginal wage earners, and aged persons respectively, and for other persons with limited resources.

For the benefit of those who have clearly something wrong with them, that cash alone cannot heal, the infusion of cash, obviously, will not be the answer, at least not the whole answer. What they require will be services of a specialized kind. The sundry cash programs can, however, serve as screening devices to help identify those cases. Apart from large-scale day-care centers and other institutions to take care of children who cannot be left in unsuitable homes or roaming the streets, and possibly a large-scale service corps for teenage youth who

are not in school and not employed, it will become necessary to engage in a planful rebuilding of the slum areas and the systematic rehabilitation of many of the inhabitants of these areas.[3]

Along with these huge efforts there may well have to come about a reorientation of our rule-enforcement systems. It seems of little promise to go further along the road whereby effective enforcement power is withheld from persons in authority, because authority without power is helpless. This applies especially to our teachers and to our law-enforcing agents. Short of respect for some set of groundrules, and respect in general, not even the eradication of the sources of discontent can be expected to restore a vision, let alone a realization of some latter-day version of "the good life."

[3] In far greater detail than is here possible, specific suggestions along the above lines have been made in an article by this writer entitled, "Social Assistance, Social Subsidies, and Social Services to Underwrite the Essentials," which will appear in ch. 1 of a forthcoming volume titled, "Toward Freedom From Want," edited by S. A. Levitan, *et al.*, to be published in March 1968, under the auspices of the Industrial Relations Research Association, by Harper & Rowe, New York.

UNEMPLOYMENT INSURANCE

BY JOSEPH M. BECKER, S.J.*

In terms of the numbers benefited, the program of unemployment insurance (UI) represents the major effort of our society in aid of the unemployed. The poverty programs and the many related activities of the Employment Service bring help to those among the unemployed who most need assistance. These other important programs have been relatively neglected in the past, and they now deserve all the attention currently being accorded them. However, in our preoccupation with the new, we should not overlook the central importance of the UI system. Many more of the unemployed receive income from the unemployment insurance system than are placed in jobs by the Employment Service or are aided by a poverty program.[1] Any improvement effected in the UI system improves the lot of more of the unemployed more immediately and more certainly than does a change in any other program.

The UI program was subjected to an unusually thorough review in 1965, when the House Ways and Means Committee held hearings on the administration's bill, H.R. 8282. Nearly everybody in the country who had anything to say about unemployment insurance was given an opportunity to be heard. The hearings lasted through nearly 4 weeks and produced five volumes of evidence. The import of this evidence was then considered by the Ways and Means Committee during 7 weeks of executive sessions.

Out of these extended deliberations came H.R. 15119, which represented the committee's judgment on what revisions in the UI program were most needed and would have the best chance of acceptance by the Congress. H.R. 15119 passed the House by an overwhelming majority (374 to 10), but it was rejected by the Senate at the request of organized labor. Labor considered H.R. 15119 to be "half a loaf"—primarily because it lacked certain Federal standards—and believed that by rejecting the half loaf it could obtain the whole loaf later. As a result, the 89th Congress enacted no substantive UI legislation, and at this writing (November 1967) the stalemate persists.

Following upon the massive consideration of UI issues which preceded H.R. 15119, this essay would serve no useful purpose by offering a series of detailed recommendations. If there is any function to be performed by still another discussion of UI—a debatable proposition—it might be to review general principles. Accordingly, this essay is mainly taken up with a consideration of two fundamental issues—

* Cambridge Center for Social Studies, Cambridge, Mass.

[1] The insured unemployed and the poor (technically defined) overlap only slightly. A relatively small number of the "poor" are among the insured unemployed and a relatively small number of the insured unemployed are "poor." Unemployment insurance is related to the poverty programs chiefly through its preventive function; that is, it saves some unemployed persons from slipping into poverty.

the proper order of UI objectives and the relative control to be exercised over the UI program by the Federal and State Governments.

From these two considerations grows a possible solution of the present legislative stalemate. A viable solution might be found in trading Federal standards. A standard regulating the maximum benefit amount might be added to H.R. 15119 and one or more standards currently in the bill might be dropped. Reasons for considering such a trade are suggested below.

Order of UI Objectives

There is wide agreement that unemployment insurance has two general objectives: to aid the unemployed individual and to promote the efficiency of the economy. There is almost equally wide agreement that the first of these is primary—that the most important objective of UI is to provide adequate income insurance for unemployed workers when suitable jobs are not available.[2] All other objectives are secondary.

Among the secondary objectives, which relate to the efficiency of the general economy, four predominate. The program is designed to prevent: (*a*) The dispersal of an employer's trained labor force during lulls in production and (*b*) the breakdown of labor standards during temporary unemployment. The program is also designed to stabilize employment by (*c*) maintaining purchasing power and (*d*) providing employers with an incentive to avoid unemployment. These objectives reflect a concern about the effect of unemployment benefits on the supply of labor (*a*) and (*b*), and on the demand for labor, (*c*) and (*d*).

The objectives have been listed in order of decreasing importance, according to the weight which the factor of unemployment benefits has relative to other factors in the attainment of a given objective. Thus, in the attainment of the primary objective, the maintenance of an individual's income at the adequate level while he is unemployed, unemployment benefits are the major factor because they are nearly always the principal source, often the only source, of income for the unemployed. Moreover, in whatever way "adequate level" is defined, the contribution of unemployment benefits to the attainment of that level can be measured fairly accurately.

In the attainment of the other objectives, unemployment benefits are a minor factor and their impact is difficult to measure. Consider, for example, the objective of stabilizing unemployment by bolstering private spending during recessions. In the most recent recession year, 1961, the UI programs (Federal and State) put into the hands of consumers over $4 billion while taking from employers about $2.5 billion in taxes. The $4 billion in benefits amounted to 0.8 percent of the gross national product—a miniscule determinant of total spending. (The net addition to private spending, benefits minus taxes, amounted to less than half this amount.) Although unemployment benefits have a greater impact than this small proportion of the GNP would seem to indicate, since the payments enter the economic system at a strategic time and place, the facts remain that such payments are a very small part of total spending and that it is total spending, not just a strategic

[2] U.S., Congress, Joint Economic Committee, *Federal Programs for the Development of Human Resources.* A compilation of replies from departments and agencies of the U.S. Government to a questionnaire formulated by the Subcommittee on Economic Progress, 89th Congress, 2d sess., 1966, pt. 1, 416.

part of it, that determines the overall level of employment.

Similar observations could be made about the relationship between unemployment benefits and the other secondary objectives. The program's importance to the secondary objectives is neither so great nor so certain as its importance to the primary objective.

Those who have the responsibility for UI legislation may array the above objectives in a different order, but every decisionmaker must determine explicitly for himself how he does order these objectives, since his decisions will depend greatly on the relative importance he assigns to each of them. Let us suppose, for example, that a decision must be reached on the amount and duration of UI benefits. One who accepts the above order of obpectives will not proceed as the original Committee on Economic Security felt forced to proceed when, operating in the depth of the great depression, it decided first how much of a tax burden could be imposed on employers without producing an unfavorable effect on the business cycle, and then how much in the way of benefiits could be paid out of this tax. He will rather first ask what proportion of wages the typical covered worker needs when unemployed and whether the program is currently meeting this need.

This crucial question is usually answerable because, as remarked before, it is readily quantifiable. Only after considering this question will the decisionmaker turn to the other objectives for whatever additional guidance they may offer. Although the discussion here has considered only the order of the objectives, it goes without saying that the decisionmaker will also have to consider the costs of attaining each objective.

Federal-State Relationships

A second general consideration revolves about the relationship between the Federal and State Governments. Differing concepts of the power roles of the Federal and State Governments in the UI program can lead to very different decisions regarding UI provisions. It was the Federal-State issue that, more than any other, determined the fate of the 1965–66 legislative efforts. H.R. 15119 failed to become law primarily because it lacked provision for certain Federal standards which organized labor considered essential. This issue is certain to remain a paramount one.

In the resolution of this issue, a useful starting point is provided by the social principle of subsidiarity, which states that a higher unit of society should not undertake to perform functions which can be handled as well by a lower unit but rather should offer help (subsidium) where necessary to enable the lower unit to function adequately. The principle carries the crucial implication that decisionmaking should be shared as widely as possible.[3]

Some general applications of the principle of subsidiarity to unemployment insurance are contained in a paper I recently prepared on the role of the States in unemployment insurance. The conclusions of this paper are given in the accompanying appendix, the gist of which is that the existing Federal-State partnership is working out reasonably well and had best not be burdened with additional Federal standards, with one possible exception.

[3] For a fuller discussion of this principle, see Joseph M. Becker, S.J., *Shared Government in Employment Security* (New York: Columbia University Press, 1959), pp. 5–9.

The one exception is a standard governing the maximum benefit amount. If a valid case for a Federal standard can be made anywhere, it is here; for in this area there is enough social consensus and enough tested experience to support sound conclusions.

First, there is a general consensus that the UI system should pay at least 50 percent of lost wages to most beneficiaries. If we do not specify too minutely the meaning of "wages" and "most," we have here a consensus broad and deep enough to constitute a socially significant force. Second, there is an equally general consensus that if any class of worker should be protected adequately by the system it is the worker with dependents. Third, it is certain that workers with dependents, who constitute a little more than half of all beneficiaries, get relatively less protection than workers without dependents. Fourth, it is certain that workers with dependents get relatively less protection largely because their wages average higher than those of other workers and the maximum benefit is not sufficiently high to allow these primary workers to receive 50 percent of their lost wages. There is an extensive body of reliable data available to establish the last two points.[4]

It is rare in unemployment insurance that the issues and the relevant facts stand out so clearly as they do here. It is, therefore, easy to reach the conclusion that the maximum benefit in most States should be raised in order to allow the majority of primary workers to receive a weekly benefit equal to at least 50 percent of their lost wages.

It is not entirely clear what maximum benefit would accomplish this. Still less is it clear what sort of Federal standard in this area, if one were desired, would be most workable. Difficult technical problems remain to be solved. However, given the decision to establish such a standard, technical difficulties could be overcome.

A Suggested Solution

There were many provisions in H.R. 15119 which, it was generally agreed, the unemployment insurance program sorely needed. If they could somehow be rescued, the program would become a more effective instrument of social policy.

Any hope of a successful rescue operation depends on finding a viable compromise that will resolve the current stalemate. If such a compromise is to be achieved in the near future, it will probably have to be based on H.R. 15119—rather than on the original bill, H.R. 8282, or the more recent bill, S. 2377.

As indicated above, H.R. 15119 was the product of almost a month of hearings and over a month of executive sessions on the part of the influential Ways and Means Committee. In the history of unemployment insurance, this was an unprecedented amount of preparation, and it was fittingly capped by the unprecedented mode of introducing the proposed legislation: At the same time that Chairman Wilbur D. Mills introduced H.R. 15119, the minority leader, John W. Byrnes, introduced H.R. 15120, an identical bill. The House of Representatives then proceeded to pass this legislation by an overwhelming majority,

---

[4] For a fuller development of these propositions, see Joseph M. Becker, S.J., (ed), *In Aid of the Unemployed*, (Baltimore: The Johns Hopkins Press, 1965), pp. 79–111.

374 to 10. H.R. 15119 comes very close to reflecting the dominant mood of the country with respect to unemployment insurance and defines, at least roughly, the boundaries within which final agreement must be reached.

If H.R. 15119 could be amended so as to present a further compromise on the issue of Federal standards, it might become acceptable to all parties. One possible compromise would be to eliminate several Federal standards now in the bill and substitute the key standard regulating the maximum benefit amount.

H.R. 15119 proposed four new Federal standards, which would: (1) Require that a benefit claimant have some intervening employment before continuing to draw checks in a second benefit year; (2) restrict total cancellation or reduction of benefit rights to certain specified situations; (3) generally provide that benefits not be interrupted while otherwise eligible claimants were undergoing work training; (4) provide that benefit payments not be denied or reduced by a State merely because a claimant had filed his claim in another State.

Compared with a standard regulating the maximum benefit amount, these proposed standards are less important and are supported by a less broad consensus. Their sacrifice would seem to be a small price to pay for a standard regulating the maximum benefit amount and the passage of the entire bill.

If this compromise should prove acceptable to the proponents of the original H.R. 8282, and if a solution can be found for the technical difficulties relating to the exact nature of the standard, there is reason to think that it would also be acceptable to the opposing side. As pointed out above, this standard rests on a body of experience that has been measured with some exactness and on principles that enjoy a wide consensus. This is the one Federal standard approved by the administrators of the State employment security programs when they met in Phoenix, Ariz., in January 1966 to record the attitudes of the States toward H.R. 8282. Furthermore, "usually reliable sources" report that the chairman of the Ways and Means Committee came closer to approving this Federal standard than any other of those finally eliminated. Altogether, there seems to be a solid probability that H.R. 15119 thus modified might become acceptable to all parties and serve as the vehicle for modernizing the Nation's principal program of aid for the unemployed.

Although this essay has been primarily directed to the legislative problem that is of immediate concern to the unemployment insurance program, it ought not close without at least mentioning a long-range problem that will eventually require action. The two parts of the employment security program—the Employment Service and the Unemployment Insurance Service—have been growing apart from each other year by year. The manpower revolution is catching up the Employment Service and linking it with a large number of activities only remotely connected with unemployment insurance. It is likely that, in a somewhat similar fashion, the Unemployment Insurance Service will become involved with the growing movement toward a universal program of income maintenance.

It is not too early to consider seriously the eventual results of these lines of development. It may be necessary, for example, to arrange an administrative divorce that would leave the Employment Service in

the Department of Labor, while giving it its own source of financing, and would return the Unemployment Insurance Service to the Social Security Administration in the Department of Health, Education, and Welfare where it could breathe the same air as the other social insurances. A study of this problem could begin with the experience acquired by the New England States, all of which have, for varying periods of time, been operating their employment security programs with an almost complete separation between the employment service and unemployment insurance divisions.

APPENDIX*

General Economic and Political Considerations

If it were quite clear that a notable improvement in the unemployment insurance program would result from changing it into a completely State or completely Federal program, I would favor the indicated change. The absence of evidence for such a result leaves me free to follow my political preferences and retain the present Federal-State partnership.

My political preference is for a society in which power and responsibility are widely dispersed and yet in which there is a significant amount of central planning and control. How much dispersion and how much centralization? Who can say precisely? All I can say is that I like the mixture I have found in the unemployment insurance program.

Why do I like this political arrangement? Out of all the reasons I might give, let me select two—one of which leads me to welcome a significant amount of State responsibility and the other of which leads me to welcome a significant amount of Federal influence. Both are very general considerations. But general considerations are important: They establish the framework within which we make particular decisions. We would probably find, if we were fully aware of our motivations, that general considerations explain many of our particular decisions.

The Economic Function of Unemployment

The first consideration, which is primarily economic in character, constitutes a reason for a significant amount of State responsibility in the UI system.

Unemployment—more precisely, the possibility of unemployment—performs an essential economic function. The threat of unemployment is an indispensable part of the machinery that allocates resources, capital or labor, in a free economy. The allocation of resources in a free economy is unavoidably dependent on the possibility that a resource will be unemployed if its owner sets too high a price on it or otherwise makes bad economic use of it. If the threat of unemployment did not operate to allocate resources, this function would have to be assumed by the coercive power of the Government, and there would no longer be a free economy.

---

*Joseph M. Becker, S.J., "The Role of the States in Unemployment Insurance," reprinted by Cambridge Center for Social Studies (Cambridge, Mass.) from *State Labor and Social Legislation*, a symposium in honor of Elizabeth Brandeis Raushenbush. Industrial Relations Research Institute, the University of Wisconsin (Madison, May 20, 1966).

What is involved here is not merely the need to maintain an incentive to work; the issue involved is much broader. Aid to the unemployed can affect all labor costs, including wage rates and conditions of work, and labor costs are the larger part of total costs of production. Unemployment and the cost of unemployment benefits affect the competition between industries and between firms within an industry. Unemployment and unemployment benefits also affect the bargaining relationship between labor and management. The Director of the Federal Mediation and Conciliation Service, William E. Simkin, was adverting to this broader economic function of unemployment when he said, in explanation of increased strike activity in 1966:

> "With the decline in unemployment, unions have more muscle than they have had in years."

The need to preserve the regulatory function of the possibility of unemployment—while always striving to reduce unemployment to a minimum—makes it more difficult to devise appropriate programs of aid for the unemployed than for those groups in society who are not competitive members of the labor force—the aged, the blind, the ill, the children. Such persons are behind the battlelines, withdrawn from the fray. But the unemployed are still in the economic struggle, still part of the forces that determine the size of the national product and its distribution among the factors of production. Aid for the unemployed inevitably becomes a force that directly influences the competitive struggle.

Unemployment insurance is thus closely tied to the economy of each State. Since the States are to a significant extent independent economic units in competition with each other, it is desirable that each State have considerable freedom to determine the provisions of its own unemployment insurance law. A remark made by Justice Louis Brandeis back in 1912 is still pertinent:

> They urge that the days of competition are past; that to insist upon competition is to go backward, not forward; that the path of progress lies in regulation. * * * But to abandon competition and rely upon regulation as a safeguard against the evils of monopoly would be like surrendering liberty and regulating despotism.

If economic competition between the States were notably holding back the unemployment insurance program from its proper development, I should be in favor of ending the competition in this area and of setting up a complete set of Federal regulations. But the evidence is lacking for such a notable deterrent effect. The effect is certainly not evident in the duration provisions, and it is almost equally lacking in the other provisions, such as the benefit amount.

Competition could affect the benefit amount, for example, by causing the States with the highest relative costs (taxes as a percentage of total wages) to have the lowest relative benefits (average weekly benefit as a percentage of average total wage). But in 1964, only five States (Michigan, Mississippi, Nevada, Pennsylvania, and Washington) had costs significantly above the average and benefits significantly below the average. Of these, only three (Mississippi, Pennsylvania, and Washington) provided a benefit which was lower in dollar amount than the average benefit ($35.96).

I am more impressed by the operation of two contrary effects of competition and State independence. One is the opportunity it provides for the more industrialized States, where most of the affected workers live, to move out ahead. A single national program, with benefits determined by the balance of forces in Congress, might well provide benefits less liberal than those currently provided by, say, California, New York, or Wisconsin.

The other is the stimulus that competition provides for the laggard States. On more than one occasion I have been present at a State legislative planning session and heard the argument made that liberalization of the UI program was in order because the State in question was falling behind other States that were in its own economic bracket. And this argument carried weight with legislative committees. The stimulating effect of competition is at least as real an influence as the deterrent effect of competition. This might be called the Commons effect, after John R. Commons, who relied on such competition for much social progress. In the Commons formula, you induced a few leaders (firms or States) to step out in front in the hope that the rest would eventually follow.

I readily grant that the Commons effect can be too slow in its operation. Then recourse must be had to direct Federal regulatory action. The lag in the maximum benefit amount may be a case in point. A good argument can be made for a modest Federal standard in this area after this length of time.

Possibly, also, the extension of benefits beyond 26 weeks in normal, nonrecessionary periods, may prove to be a task best handled by the Federal Government. However, two recent developments give some ground for thinking that even here the States may prove to be competent to manage for themselves. In the first place, a few States, of which Wisconsin is one, have begun to explore the duration region lying beyond the border of 26 weeks. Perhaps the Commons effect will begin to operate here as it has operated in the region below the border; if experience proves favorable, other States may begin to follow the leader. The other development is the serious proposal, generally associated with Walter Heller, to return a portion of the Federal tax money to the States to be used as the States wish. The States could use general revenues from this source to establish their own special programs for the long-term unemployed. The reapportionment of State legislatures in recent years makes them more representative of their total citizenry and removes much of the former criticism directed against them on this score.

The Lesson of Evolution

When one takes the longest view of all, the evolution of the universe or at least of the planet earth, a reason emerges for expecting the central government to retain and even enlarge its role in political and economic society.

As far as we can decipher the story of man's past, the solar system of which he is a part has been in evolution for about 5–10 billion years, and the spacecraft in which he rides, called earth, has been in evolution for about 5 billion years.

This long process of evolution has shown a consistent pattern, marked by two characteristics which are really just opposite sides of

the same coin. The world has evolved from the simple to the complex, and in so doing has developed forms of unity which are internal rather than external.

At the beginning, our solar system was a gaseous mass of undifferentiated atoms. One part of the mass was much like any other part, and all parts were held together chiefly by the external force of gravitation. The atoms gradually cooled and coalesced into differentiated molecules. Each molecule had an internal principle of unity that set it off from other molecules. The earliest molecules were simple crystals that grew by accretion; that is, by the simple addition of similar crystals. They could grow to any size, and they could be split up into any number of identical parts, for their degree of differentiation and internal unity were still relatively slight.

Much later in the evolutionary process came the giant, immensely more complex, protein molecules, the building blocks of living things. Their principle of internal unity was so strong that they could not grow beyond a certain size nor could they be subdivided beyond a certain point without losing their specific nature.

From these molecules, living things evolved simultaneously in many directions, but the pattern remained of evolving from the less to the more complex and from the less to the more unified. Fish evolved into reptiles, and reptiles into mammals, and among the mammals there finally appeared one which could do the strange, strange thing that the reader is now doing: it could think. And because it could think, it could also exercise free choice. Man had appeared, rational and moral.

If a moving picture had been taken of this total evolutionary process, and if the machine were now speeded up so that the 5 billion years of film could be run off in 2 hours, then living things would make their appearance in the last 20 minutes of the show, and human beings would appear in the last one-half second.

Man is the climax of the evolutionary pattern. He is the most complex of all existing creatures, and yet he also has the strongest principle of internal unity. It is relatively easy to graft the branch of one tree onto another tree, but extremely difficult—usually impossible—to graft the organ of one man into another man. The reason is that man's principle of internal unity is stronger than that of a tree: hence it rejects the transplant as something foreign to itself. Man's principle of internal unity is so strong that he is a complete something in and of himself. This is what we mean when we speak of the uniqueness and the dignity of the human person.

This is the picture of man's physical evolution. The picture of his social evolution—man in society—is sufficiently similar to raise the fascinating possibility that there is at work here some fundamental law of nature that governs the social as well as the physical development of mankind. Without at all subscribing to the extreme Spencerian view that the patterns of physical and social evolutions are necessarily identical, I hold that the general process of cultural evolution resembles that of physical evolution in moving from the less to the more unified and at the same time from the more to the less coerced. That is, we are moving toward greater cultural unity of a kind that is compatible with freedom because it is based on the free consent of the united parts.

Man's social world, like his physical world, has grown increasingly complex and increasingly unified. An institution like the United Nations would not have been possible a century ago; but it exists today. Eventually there may be a world government in the full sense of the term. Perhaps—who knows?—we are evolving toward a society in which all mankind will be as closely unified and yet as greatly governed as a religious order or an individual family.

Applied to Federal-State relations, the general pattern of evolution would incline us to expect a growing area of central planning and hence a growing function for the Federal Government. It would also incline us to welcome such a development as in accord with the main thrust of evolution—which, we assume, is a movement from the less perfect to the more perfect.

But while it would be shortsighted to overlook the lessons to be learned from the general pattern of evolution, it would be dangerous to apply that pattern blindly and rigidly to Federal-State relations. We must recognize that the evolutionary proces of the future will have to work under a different set of conditions. With the advent of man into the world, a new force has come into action, the force of free choice.

Evolution can no longer be the automatic product of necessarily acting forces. The evolutionary process is now subject to unpredictable modification by the free choices of men. To cite the most obvious example, the free choice of mankind, exercised through war, could lead not to the further development but to the complete destruction of life in this world.

This new condition of freedom has two implications for the problem we are discussing today. First of all, if men are free, then men—all men, including the rulers of governments—can be evil. History establishes only too cogently the danger of oppression from a strong central government, and the story of growing human freedom is partly the story of increased restrictions placed on governmental power. There is much discussion currently of the limitations that hamper law officials in their pursuit of criminals. Each of these limitations is an eloquent testimonial of the fear the citizens once had of their own government. A basis for that fear still remains, and despotic government is still a possibility.

Moreover the danger of too much government is not merely that of despotism; it is also the more subtle danger of paternalism. Because paternalism turns adults into children, it is to be avoided. Because each of us is a person, each should have the opportunity to make his own decisions—even, sometimes, when big brother could make them more efficiently. As we grow wealthier, we may be able to afford this sacrifice of efficiency more often.

The danger of oppressive or paternalistic government is one implication to be drawn from the very recent appearance of this new creature, called man, possessed of this new power, called freedom. But there is a second and more optimistic implication to be drawn from the same fact. The grand sweep of evolution shows a development toward a more interior type of unity, that is, toward a unity which is the product of internal rather than of external forces. It would seem therefore that we may look forward to a time when human society will become not only more unified but will be held together more by internal consent and less by external coercion.

Self-determination does not necessarily mean anarchy. It is possible for free individuals to choose freely to work together for agreed-upon ends and thus achieve unity through consensus. Unity through consensus—this is the center through which the axis of evolution must now pass if the whole process is not to fail finally.

In the process of moving toward this goal of unity with freedom, I suggest that we make use of four practical guidelines:

First, let us move actually and willingly toward more unity. This is the thrust of evolution, and it is a good one. It is as though we were now in the upper regions of mountains, cold and bare, with all the roads leading downward, under the inexorable pull of gravity. The roads lead toward a greener and richer country—out of conflict toward cooperation.

Second, let us move slowly. Mankind improves slowly, and it is difficult to know when men have changed enough to be ready for the next step forward. The frontispiece of Alfred Marshall's famous *Principles of Economics* carries the motto: *Natura non facit saltum*—Nature does not advance by leaps. Marginalism is the technique to be used. Make changes experimentally at the margins of new situations. As we descend the mountain road let us keep our brakes in working order. We know we have an excellent unemployment insurance system now, perhaps the best in the world. Let us change it cautiously.

Third, as we move toward the goal of greater unity, let us be aware of possible distinctions. Reality is complex, and different situations call for different solutions. Precisely because the central government must exercise a greater influence in modern society than ever before, the central government must not be overloaded with tasks that can be performed as well, or nearly as well, by smaller units of society. The society of the future will be particularly prone to the disease described by Lamennais as consisting of apoplexy at the center and anemia at the extremities.

To avoid this dread disease, we will need to distinguish carefully between what the central government must do and what it can but need not do. For example, central planning would seem to be needed to a greater extent in the employment service than in unemployment insurance—or, to put the proposition in reverse form, there is less need for Federal direction in unemployment insurance than in the employment service. Again, within unemployment insurance itself there is room for distinctions. Currently, there is more need for a Federal norm to regulate the maximum benefit amount than for one to regulate the maximum benefit duration; and within duration, there is more room for Federal concern about the very long-term unemployed than about those who are employed for less than 6 months.

Fourth, let us move toward greater unity in such a way as to require the upper levels of authority to elicit the consent and participation of the lower levels. Otherwise the unity achieved will be the unity not of persons but of things. This is why we must achieve some form of judicial review in employment security. This is why, also, it is desirable to make more effective use of advisory councils. And this is why I rejoiced to see H.R. 8282 discussed so thoroughly and so openly. I consider it a healthy development that the Committee on Ways and Means invited the Interstate Conference of Employment Security Agencies to prepare alternative provisions to those in H.R. 8282 and allowed representatives of the conference to participate

in the executive sessions of the committee along with representatives of the Federal Bureau of Employment Security. Procedures of this sort increase the probabilities of achieving unity through consensus. In fact, I consider the steps by which H.R. 8282 has become H.R. 15119 to constitute an almost—if not quite—perfect example of the creative federalism described by Gov. Nelson Rockefeller in his 1962 Godkin lectures at Harvard and endorsed by President Lyndon Johnson in his 1964 speech at the University of Michigan.

To sum up the four guidelines: Let us continue to move toward greater unity, but let us move slowly, with an eye to differences, and in such a way as to achieve unity less by coercion and more by consensus.

I close with an optimistic observation. Unity through consensus is not merely possible; it is probable. The people of the United States are more of a unity, and their Government has more effective control over their lives than was true in the most despotic empires of antiquity. Yet the Government of the United States is also more democratic than the ancient governments. Up to now, at least, the evolutionary process has been able to produce both more unity and more freedom. I am optimistic that this process can continue.

IMPROVING INCOME MAINTENANCE THROUGH NEGATIVE TAXATION

BY CHRISTOPHER GREEN*

The revival of interest in the problems of low-income people has stimulated a search for programs which will reduce and ultimately eliminate poverty in the United States. Even if some people believe that the final elimination of poverty is too grandiose a goal, there is at least general optimism that the incidence of poverty can be reduced well below present levels. The chief tool for eradicating poverty is improved education and vocational training designed to meet the demands of our rapidly changing economy. Education and training programs are designed to provide relevant skills and thus raise the economic productivity of potentially poor people. Greater productivity may be translated into increased income. Increased income means a less binding budget constraint and the ability, therefore, to satisfy wider combinations of expenditure preferences.

I think it is now widely recognized that the education and training route to the reduction and elimination of poverty is a long and time-consuming one, although it undoubtedly yields the greatest individual and social benefits in the long run. Persons who lack education and training are often not quickly and easily educable; moreover, the chief educating institution, the system of public primary and secondary education, does not seem to adapt easily to changes required by a rapidly changing technology. New institutions for providing vocational training and education have been created, but the number of persons graduating from these programs is necessarily small relative to the number of persons who are unskilled or who have increasingly unusable skills. For example, it is estimated that from 1964–67, 940,000 persons were approved for training under the MDTA and Job Corps programs compared to 10 million poor persons including 4 million heads of poor families who had some work experience in 1965.[1] Even, then, many of the 866,000 MDTA trainees were not poor but were replacing old, technologically obsolete skills with new ones.

INCOME MAINTENANCE IN THE UNITED STATES

Recently, interest has developed in new programs which may have an important impact on the income of poor families in the short as well as the long run. These programs include those that would guaran-

*The author is assistant professor of economics at North Carolina State University, at Raleigh.

[1] Data on the MDTA and Job Corps program is found in U.S. Congress, Joint Economic Committee, *Federal Programs for the Development of Human Resources,* vol. I, 89th Cong., second sess., December 1966 (Washington: U.S. Government Printing Office), table 1, p. 374 and table 1, p. 117. Data on the working poor is from the Survey of Economic Opportunity taken by the Census for the Office of Economic Opportunity. Relevant data were made available to me by the Institute for Research on Poverty at the University of Wisconsin.

tee a minimum income and supplement the low earnings, property, and transfer income of our low-income population. The logic of guaranteed minimum income programs stems in part from an evaluation of the poverty problems in America. The logic also stems from an examination or our present income transfer programs. At present we are transferring over $40 billion of income through one or another of our public income maintenance programs.[2] These programs include Old-Age, Survivors, Disability, and Health Insurance (OASDHI), certainly the most important of our income maintenance programs; unemployment compensation, the public assistance programs, and veterans' pensions and compensation. Interestingly, it is estimated that only about one-half of these transfer payments go to persons who are poor before having received transfers and a little less than a quarter of total transfers is received by persons who are defined as poor after receiving transfer payments, i.e., on a total money income basis.[3] While few persons would suggest that our income maintenance programs should help only poor people, the statistics perhaps suggest that not enough is now being done for the persons who really need aid.

One of the problems with our income-maintenance system is that none of the programs makes need a sufficient condition for receiving public transfer income. The OASDHI and unemployment compensation programs protect against income loss due to age or involuntary unemployment. Thus they provide varying measures of income security to only some of our poor population while many of the benficiaries of these programs are in no sense poor. The "welfare" or public assistance programs make need a necessary condition for receipt of public aid, but they do not make need a sufficient condition for receiving assistance payments. Only certain categories of the poor have any claim to public assistance. Assistance is usually available to families or individuals who fall into certain categories, such as the blind, the permanently and totally disabled, dependent children in families lacking a breadwinner, and the indigent aged. However, the statistics on poverty indicate that the poor, who now number nearly 30 million, are a much more diverse group than the public assistance categories and the 8 million recipients of public assistance (including general assistance) would suggest. In fact, the "able-bodied" poor and their dependents make up a substantial portion of the poor population and yet our federally assisted public assistance programs simply bypass this particular group.

Further examination of our present income-maintenance programs suggests that they alone cannot achieve the objectives of effectively improving income maintenance for all of the poor. The reasons for this can be fairly simply stated. The OASDHI program is for the aged but four-fifths of the poor are not aged. The unemployment compensation program is for the insured unemployed and yet many of the poor are living in families with a head who works full year full time. For example, it is estimated that in 1965, 2.4 million poor families were headed by a person who worked full year full time. Another 350,000 heads of poor families worked full year at part-time jobs and they

---

[2] For fiscal 1966 the total was $40.8 billion. *Social Security Bulletin,* October 1966, p. 2.
[3] Christopher Green, *Negative Taxes and the Poverty Problem* (Washington: The Brookings Institution, 1967), p. 18. Also see Robert J. Lampman "How Much Does the American System of Transfers Benefit the Poor," *Economic Progress and Social Welfare,* ed. Leonard H. Goodman (New York: Columbia University Press, 1966).

too would not be eligible for unemployment compensation.[4] Moreover, unemployment compensation programs cannot easily be made to meet the needs of many persons with very spotty employment records. The veterans' programs are for veterans with particular disabilities.

The public assistance programs presently aid only certain categories of the poor. While they could be modified to cover all of the poor, two factors suggest that this route would not be a recommended one. One factor is that under the present public assistance program benefits are usually reduced dollar for dollar of any increase in the other income of the recipient. Therefore, earners of income would be "taxed" at a 100 percent rate on their increased earnings. This means that as long as an individual's or family's earnings are below the maximum level of public assistance for which it is eligible there is no monetary gain from earning. Thus, making public assistance available to all of the poor could severely undermine the incentives to work of the poor population, reduce their already meager earnings, and considerably raise the cost of any program aimed at raising the income of poor families. A further objection to the use of public assistance to meet the needs of the able-bodies poor is that the present public assistance programs are already unpopular with the taxpayer and the recipient alike. Therefore, receipt of public assistance by the able-bodied poor may continue to carry with it the stigmas that are attached to the present relief recipient.

In recent years interest has grown in developing new systems of income transfers. One of these systems is called negative income taxation. Negative income taxation is related to the idea of a guaranteed minimum income which has roots that go back at least several decades. The suggestion for a negative income tax can be traced to plans devised during World War II by the Englishwoman Lady Juliet Rhys-Williams. Another system is family allowances which already exist, in practice, in almost every other Western industrialized nation. Family allowances are payments to families with children for the benefit of the children and they are usually paid irrespective of family income. Perhaps the chief difference between negative income taxation and family allowances is that payments under the former are conditioned on family income while those under the latter represent a demogrant to all children irrespective of family income. From the standpoint of public policy toward low-income people this difference is a very important one.

Negative Income Taxation

Negative income tax proposals would use the income tax system to transfer income to low-income families. One version of negative income taxation would use the value of exemptions and deductions in the income tax system as a criterion for determining eligibility to receive negative income tax payments.[5] Families with income amounting to less than the value of exemptions (EX) and minimum standard deduc-

---

[4] From Survey of Economic Opportunity. See footnote 1.

[5] Milton Friedman, *Capitalism and Freedom* (Chicago, University of Chicago Press, 1962), pp. 191–94. For a suggestion that a family's poverty income gap be used instead of its unused exemptions and minimum standard deductions in calculating negative tax payments, see Christopher Green and Robert J. Lampman, "Schemes for Transferring Income to the Poor," *Industrial Relations* (February 1967), pp. 121–37.

tions (MSD) allowed them under the present tax system (i.e., families with "unused EX & MSD") would be eligible to receive negative income tax payments equal to some percentage of the shortfall. The percentage would be considerably less than 100 percent in order to preserve incentives to work. To take an example, assume the percentage (i.e., the "negative tax rate") is 50 percent. Under present income tax law a four-person family is allowed $2,400 in personal exemptions plus $600 in minimum standard deductions, or a total of $3,000 in nontaxable income. If the family's income is $1,000 in a given year it would have $2,000 in unused EX & MSD against which is applied the 50-percent negative tax rate. The family would receive a $1,000 payment from the Government.

Another version of negative income taxation is illustrated in table 1. This version specifies a basic allowance, say, $400 for each member of the family and a tax rate(s) applied to total income which indicates what portion of the basic allowance actually comprises a net addition to family disposable income. For example, a five-person family has a total of $2,000 in basic allowances. If the family's income (not including the allowance) is $2,400 and the tax rate is 33⅓ percent, then the family's disposable income is increased by $1,200 ($2,000−0.33⅓ ($2,400)=$1,200). This plan, proposed by James Tobin,[6] would "tie in" to the present income tax schedule if taxpayers are allowed to choose either the basic allowance plus the 33⅓-percent schedule *or* the present personal exemptions and standard deductions plus the present tax schedule. The levels of income at which the two tax schedules meet is indicated at the bottom of table 2. For a four-person family the tax schedules meet at $6,306.

TABLE 1.—*A negative income tax plan of Tobin type* [1]

Family income before income supplement	Size of family					
	1	2	3	4	5	6
	Income supplement family is entitled to					
0	$400	$800	$1,200	$1,600	$2,000	$2,400
$600	200	600	1,000	1,400	1,800	2,200
$1,200	0	400	800	1,200	1,600	2,000
$1,800	----------	200	600	1,000	1,400	1,800
$2,400	----------	0	400	800	1,200	1,600
$3,000	----------	----------	200	600	1,000	1,400
Breakeven income [2]	1,200	2,400	3,600	4,800	6,000	7,200
Tax breakeven income [3]	1,422	3,000	4,636	6,306	8,050	9,928

[1] Income guarantee or basic allowance of $400 per person and a tax rate of 33⅓ percent.
[2] Level of income at which net allowances are reduced to zero by the 33⅓-percent tax rate.
[3] Level of income at which the tax liability under the Tobin 33⅓-percent schedule equals the tax liability under the present individual income tax schedule.

It is useful to ask what are the basic similarities and differences between the two types of negative income tax plans described. In their essentials the proposals are similar. Each has three basic variables. These three variables are: (1) A basic income guarantee—that is, the basic allowance or the level of negative tax payments a family may

[6] James Tobin, "Improving the Economic Status of the Negro," *Daedalus*, vol. 94 (fall, 1965), pp. 878–898.
Also see James Tobin, Joseph A. Pechman and Peter Mieszkowski, "Is a Negative Income Tax Practical?" *Yale Law Journal*, vol. 77, No. 1 (November 1967), pp. 1–27.

receive when it has no other income; (2) a tax rate or set of tax rates which reduces the basic guarantee to zero at some level of income; (3) a break-even level of income—that is, the level of income where negative income tax payments are reduced to zero. For example, in the case of the EX & MSD version, the break-even level of income is the value of the tax unit's exemptions and deductions.

Any two of these three basic variables determines the third. For example, in the EX & MSD plan the two variables that are chosen are the break-even level of income and the tax rate or rates. These determine that the guaranteed minimum income for a family of four is $1,500 if the negative tax rate is 50 percent [(3000–0)·50]. In the other version of negative income taxation the two variables chosen are the level of basic allowances and the tax rate(s). These two combined determine, as table 1 shows, that the break-even level of income is $4,800 for a family of four.

The fact that these two plans are basically similar suggests that any plan to guarantee a minimum income will be similar to the negative income tax plans described here. That is, any plan to guarantee a minimum income will have three variables: a guarantee, a rate at which that guarantee is reduced as family income rises, and ultimately a break-even level of income at which persons no longer receive *net* benefits under the plan.

The discussion of the three basic variables also suggests something about adoption of family allowances. In most countries, family allowances are paid to all children no matter what the income or wealth of their parents. Many families receive children's allowances even though there is very little need for them. This means that it is necessary to raise income taxes or other taxes in order to pay for what will ultimately be a very costly plan. To some American social planners family allowances seem to make sense, especially when one realizes how unpopular the present aid to dependent children program is. But in a very real sense, family allowances would be an inefficient way of helping the poor, although they may be justified on other grounds such as protecting the security of children. An example of the "inefficiency of family allowances is indicated by a proposal to pay monthly allowances of $8 to each child under 6 and $12 to each child between 6 and 17 years of age.[7] Since less than a quarter of all children live in poor families only about $2 to $2½ billion of the $9 billion estimated cost of the proposal would be received by poor families, although perhaps an additional $½ billion to $1 billion would be received by near-poor families. The majority of the payments would be received by families who are reasonably well off.

In a system of family allowances only one basic variable is operative: the level of family allowances. There is no explicit tax rate which reduces the family allowances as family income rises and thus there is no break-even level of income. (There is, however, a tax on income or payrolls which raises revenues to pay for the plan.) Thus the essential difference between family allowances and negative income taxes, aside from the fact that the former applies only to families with children, is that negative income taxation makes *need* a condition for

[7] Daniel Moynihan, *New York Times Magazine*, Sunday, February 5, 1967, p. 13. Also see Alvin L. Schorr, "Against a Negative Income Tax," *The Public Interest*, (fall, 1966), pp. 110–117 and James Tobin's comment. *Ibid.*, pp. 117–119.

eligibility to receive income transfers. To repeat an earlier point: negative income taxation is an income-condition plan and family allowances are demogrants.

The two negative income tax plans described above differ in an important respect. Aside from differences in cost, the basic difference between these plans lies in the extent to which the income tax system that exists today would need to be reformed in order to accommodate the proposals described above. In the case of the unused EX & MSD plan, only minor reforms are needed. Reforms could be largely if not wholly confined to the negative side of the system. For example, it is important to broaden the definition of income for negative income tax purposes to include all income, not simply that income which is defined as taxable under the present income tax system. An important source of income that should be included is public transfer income which is presently nontaxable. One exception might be the continued exclusion of public assistance payments on the grounds that public assistance is payable only after calculating all family resources including the level of negative income tax payments for which the family is eligible. It is also important to redefine the tax unit to more closely approximate a family or consumer unit. Individuals with no income may not be poor if they are living in a family with a productive breadwinner.

In contrast, the basic allowance plan would necessitate undertaking more general tax reform. This is so, because a much larger portion of the population would be affected by the plan. By effectively substituting a refundable $400 per capita tax credit in place of the present $600 per capita personal exemptions for taxpayers using the 33⅓-percent tax schedule, the Tobin basic allowance plan would alter the present positive tax system and make net allowance recipients of some present taxpayers. Moreover, the cost of the plan, which is estimated at $14 billion on the basis of 1962 data [8] bluntly raises the question of financing negative income taxation. The estimated (1964) cost of the EX & MSD plan is approximately $6 billion—about equal to the annual increase in Federal income tax revenues produced by economic growth.

The question of financing negative income taxation is often overlooked by its proponents. Perhaps there is a good reason: proponents perhaps think the public needs to be convinced that negative income taxation is desirable and workable. Once this is accomplished the question of financing negative taxes is similar to the question of financing any other type of Government expenditure. Yet, negative taxation need not be treated as a welfare program necessitating an "initial" authorization and annual expenditure appropriations. Earl Rolph has suggested simply converting the $600 personal exemption into a refundable tax credit of, say, $400.[9] Taxpayers would calculate their gross tax [tax rate(s) times income] and deduct the tax credits for which they are eligible. Refunds would be made when total credits exceed gross tax liability. The elimination of personal exemptions would broaden the tax base; further tax "reforms" which would more carefully limit personal deductions and would broaden the definition of adjusted gross income to include most public transfer income and

[8] Tobin, Peehman, & Mieszkowski, *op. cit.* p. 24.
[9] Earl R. Rolph "The Case for a Negative Income Tax Device," *Industrial Relations* (February 1967), pp. 155–165.

the portion of capital gains presently excluded would both be desirable and necessary if income tax rates are to be kept low.

Rolph proposes financing his plan, the cost of which includes converting personal exemptions into refundable tax credits plus financing other expenditures normally financed by income taxes, with a proportional tax schedule. Rolph roughly estimates that a rate of 25 to 30 percent applied against a broadened tax base is required. Assuming, perhaps too optimistically, that a flat rate of 25 percent is sufficient, table 2 shows how a four-person family with varying levels of taxable income would fare under a revised tax system utilizing $400 per capita, refundable tax credits.

TABLE 2.—*Negative tax device using refundable tax credits* [1]

(1) Taxable income	(2) Net tax [2]	(3) Disposable income (col. 1 minus col. 2)	(4) Marginal tax rate (percent)	(5) Average tax rate [3] (percent)
0	−$1,600	$1,600	+25	−∞
$500	−1,475	1,975	+25	−295.0
$1,000	−1,350	2,350	+25	−135.0
$2,000	−1,100	3,100	+25	−55.0
$3,000	−850	3,850	+25	−28.3
$4,000	−600	4,600	+25	−15.0
$5,000	−350	5,350	+25	−7.0
$6,000	−100	6,100	+25	−1.7
$8,000	+400	7,600	+25	+5.0
$10,000	+900	9,100	+25	+9.0
$12,000	+1,400	10,600	+25	+11.7
$15,000	+2,150	12,850	+25	+14.3
$25,000	+4,650	20,350	+25	+18.6
$50,000	+10,900	39,100	+25	+21.8
$100,000	+23,400	76,600	+25	+23.4

[1] Assumes a family of 4.

[2] Net tax is $t \cdot I - c \cdot n$ where t=tax rate; I=taxable income; c=tax credit; per capita=$400; n=number of persons in the taxpaying unit.

[3] The average tax rate is defined as: $\frac{t \cdot I - c \cdot n}{I}$; i.e. (gross tax−total tax credits)÷taxable income.

Interestingly, a reform along the lines just described (including the adoption of a proportional marginal tax rate schedule) would produce greater horizontal equity and might increase vertical equity among taxpayers. A broadened tax base would increase the likelihood that taxpayers in equal circumstances will pay equal taxes. The refundable tax credits would provide progressivity in the schedule of average tax rates even though there is a proportional (marginal) tax rate schedule. (See cols 4 and 5 of table 2.) The rise in average tax rates from minus 8 to 25 percent reflects the fact that income is being generally redistributed in a manner that is usually accomplished with a graduated set of tax rates. The negative average tax rates at the lower end of the income distribution reflect the direct redistribution of income and purchasing power to lower income groups. If it is felt that the rates are not progressive enough at the upper end of the income distribution, a surtax amounting to perhaps another 15 percentage points of tax (raising the total tax rate to 40 percent) could be added to the tax on income in excess of, say, $50,000 or $75,000.

CONCLUSION

This decade's concern with poverty amidst affluence has inevitably attracted interest in guaranteeing a minimum income. The idea of a negative income tax is a relatively new and significant addition to the

arsenal of methods to improve income maintenance in the United States. The preceding comments have indicated how the tax system can be used as a vehicle to guarantee a minimum income and supplement the income of low-income taxpayers.

Negative income tax plans range from those which represent an additional income transfer program to those which are part and parcel of income tax reform. Whatever the approach, constructing a workable and effective negative income tax plan is not easy. Aside from the problem of precisely defining the income tax base and tax unit there are problems of administration including the thorny one of timing payments.[10] (To low-income people the payments will mean the most if received at regular intervals during the current year—not in April of the succeeding year.) In addition, the level of the income guarantees in conjunction with the schedule of tax rates should not be allowed to interfere with incentives to work. The planned negative income tax experiment to be carried out in New Jersey should provide clues about the work incentive effects and administrative feasibility of alternative negative income tax plans. An experiment is a logical tool for determining how well negative income taxation will work in practice. In the meantime, the logic of using negative taxes to improve income maintenance commands our attention.

[10] See Tobin, Pechman, and Mieszkowski, *op. cit.*, pp. 20–23.

HEAD START: IT IS NEVER TOO EARLY TO FIGHT POVERTY

BY SAR A. LEVITAN*

CONTENTS

"PREP SCHOOL" FOR POOR KIDS

Without a doubt, Head Start is the most popular component in the antipoverty Community Action Program. By focusing upon early physical and educational development, the program is designed to prepare impoverished preschool children for primary school experience. Head Start was sold to the American public as an expanded kindergarten program for the poor with health and nutrition components. In turn, the child oriented and educationally conscious public has not seriously questioned the need or value of such efforts.

Public preschool education as a concept is the product of natural historical evolution. The battle for free primary and secondary public education was fought and won in the late 19th century. With public education achieved for 6- to 18-year-olds, attention in recent years has shifted to other categories of students. One recent thrust of the public education movement has been directed at the older age groups and has manifested itself in the growth of higher education in colleges and universities and an increasing Federal commitment. Almost a quarter of the costs of higher education are now financed with Federal funds.

A second thrust of recent years has been the increasing realization of the need for publicly supported preschool efforts. The preschool

*Center for Manpower Policy Studies. The George Washington University. The paper is part of a chapter from the author's forthcoming volume, *The Great Society's Poor Law: A New Approach to Poverty*, a study devoted to the Economic Opportunity Act. The project is financed by a grant from the Ford Foundation. The counsel and assistance of Mr. Jeffrey Burt in preparing this paper is gratefully acknowledged.

movement received great stimulus during World War II, when daytime care for preschool children of working mothers became a necessity. Twenty-eight States now provide support for kindergartens. However, these and other facilities for younger children are still unavailable in many areas of the country: only three of every 10 preschoolers were attending nursery school or kindergarten in 1966. Very few of the 3- and 4-year-olds were attending any school at all (table 1). Participation in these programs is related in part to income level; and children of minority groups face limited opportunities for attending preschool programs (table 2). Fifty-three percent of 5-year-old nonwhites in the Nation are enrolled in kindergarten or nursery school, compared to 64 percent for the white population.

Lack of preschool educational experience affects the disadvantaged child more adversely than a middle-class child. In recent years it has become increasingly clear that many impoverished children experience serious difficulties throughout their school years because they lack intellectual stimulation and emotional adjustment in their early family life.[1] For these children, therefore, preschool programs might compensate for their background deficiencies and bring them closer to the achievement level of their middle-class peers. This reasoning seemed to have particular relevance to the war on poverty. By extending free public education to the poor during a formative period of growth—ages 3 to 6—the advocates of Head Start hoped to give children a better chance to succeed in school. The program was thus conceived as a "prep school" for poor kids.

TABLE 1.—*Children aged 3 to 5 enrolled in kindergarten or nursery school, October 1966*

	Population (thousands)	Number enrolled in school [1] (thousands)	Percent enrolled in school [1]
3- to 5-year-olds	12,486	3,674	29.4
White	10,514	3,142	29.9
Nonwhite	1,970	532	27.0
3-year-olds	4,087	248	6.1
White	3,431	193	5.6
Nonwhite	655	54	8.2
4-year-olds	4,155	785	18.9
White	3,499	659	18.8
Nonwhite	656	126	19.2
5-year-olds	4,244	2,642	62.2
White	3,584	2,290	63.9
Nonwhite	659	352	53.4

[1] Excluded from this table are the following enrollment figures: (*a*) 505,000 5-year-olds in elementary schools (above the kindergarten level); (*b*) 127,000 6-year-olds in kindergarten; and (*c*) 2,000 6-year-olds in nursery school.

NOTE.—Details may not add to totals because of rounding.

Source: U.S. Department of Health, Education, and Welfare, Office of Education, *Nursery-Kindergarten Enrollment of Children Under Six*, October 1966.

[1] Julius Richmond, "For the Child of Poverty," *American Child*, Spring 1966, pp. 5–10.

TABLE 2.—*Percent of children in kindergarten or nursery school by family income level, October 1966* [1]

	Annual family income			
	Under $3,000	$3,000 to $4,999	$5,000 to $7,499	$7,500 and over
3-year-olds	5.4	3.5	3.9	8.8
4-year-olds	13.2	11.1	15.1	29.1
5-year-olds	40.1	49.6	66.5	72.2

[1] Footnote, table 1.

Source: Same as table 1.

MAGNITUDE OF PROGRAM AND NEEDS

The Economic Opportunity Act of 1964 did not specifically address itself to the educational problems of poor children. In fact, Dr. Urie Bronfenbrenner, a leading educator, argued during the congressional hearings that the act placed too much emphasis on 16- to 22-year-olds. A program for preschoolers, he said, could accomplish much more with the same funds.[2] But Congress deferred the educational component of the war on poverty until passage of the Elementary and Secondary Education Act of 1965. Advocates of a preschool program for poor children persisted in pushing for their cause and they found Shriver sympathetic to the idea. Even though the Economic Opportunity Act emphasized exit from poverty through training and work, it was never too early to start building the bridges leading from poverty. Deprived of cultural advantages at home, the poor child is disadvantaged when he enters school and falls even further behind as he progresses through the elementary grades. And the failure of the school system to help many poor children is well documented.

The fact that CAP made no explicit provisions to aid the poor children did not preclude allocation of funds for this purpose. Indeed, initially Congress had given OEO administrators maximum flexibility in allocating CAP funds. Communities were expected to initiate and design their own projects, the funding of which would be provided by the Federal Government. But no more than a handful of communities were in a position to take this sort of initiative in 1964. Except for a few communities, mostly large metropolitan areas, which had experience in the earlier prototypes of community action—the Ford-funded "gray area" projects or the federally sponsored juvenile delinquency programs—most areas were slow in organizing community action agencies and filing acceptable applications. During the initial months of OEO, R. Sargent Shriver and his staff were therefore faced with the dual tasks of determining the components that would go into CAP and actually committing the $240 million which Congress allocated to the program. Fearing that inaction might be interpreted as an absence of need, they sensed the urgency of committing the available funds before the next budgetary authorization request was to be presented to Congress. Packaged programs were desperately needed.

The idea of a preschool program—it was eventually given the name "Head Start"—was developed in a series of conferences that Sargent

[2] U.S. Congress. House. Committee on Education and Labor. *Hearings on the Economic Opportunity Act of 1964* (Washington: U.S. Government Printing Office, 1964), pt. 3, pp. 1336–1346.

Shriver held with key staff members in the fall of 1964 and was enthusiastically supported by a panel of medical and educational experts appointed by Shriver "to consider the kinds of programs which might be most effective in increasing achievement and opportunities for the children of the poor." [3] The decision to Fund Head Start under the antipoverty program was announced in President Johnson's education message on January 12, 1965. Even then, some advisers counseled delay, noting that a large-scale summer preschool program would require considerable planning. Shriver rejected the cautious approach, and the Head Start program was officially announced on February 19, 1965, at a White House tea party. The rapid public acceptance of the program went beyond the fondest expectations of its most ardent advocates. "It stands," as one observer commented, "as a rare example of the speed with which a Government program can be organized—given money, adroit press agentry, and cooperation by volunteer groups." [4]

The program was to be largely experimental, and few knew what to expect. The need for a Head Start program seemed obvious, but how would communities around the Nation respond? OEO intended to commit only $17 million for the summer of 1965, enabling 100,000 children to participate. Although more CAP funds were available, OEO did not anticipate a greater demand. Sargent Shriver followed up the tea party announcement with personal letters to 35,000 school administrators, welfare officials, mayors, and other community officials. Within a few weeks, localities had besieged OEO headquarters with an unforeseen volume of requests. OEO officials in Washington suddenly were faced with the decision of whether to meet the demands and support a program many times larger than originally conceived

OEO decided to meet the demand, and Head Start grew like Topsy. Program costs rose to $103 million for the summer of 1965, providing places for 560,000 children as well as some staff training. This expansion easily swallowed uncommitted CAP funds, equaling more than two-fifths of the total CAP budget. OEO rapidly expanded Head Start's skeleton national staff, and many temporary hands were recruited to help with the tremendous workload.

Aside from its intrinsic value in the education of the poor, Head Start presented several attractive features that motivated OEO to launch the large crash program for the summer of 1965. Perhaps most importantly, with most CAP funds uncommitted in the spring of 1965, OEO officials were searching for innovative and attractive programs to catch the public eye and help the agency to secure greater appropriations for fiscal 1966. Head Start was tailor made for the situation: it used large amounts of funds quickly and was capable of gaining wide support.

Moreover, Head Start was a prepackaged project that could be adopted immediately by newly formed community action agencies struggling to design appropriate and acceptable programs for their community. Where no CAA (community action agency) existed, other public and nonprofit private agencies could use the funds. School boards were interested in preschool programs and were eligible re-

[3] "Improving the Opportunities and Achievements of the Children of the Poor," memorandum by Dr. Robert Cooke to Sargent Shriver (mimeographed). The memo is undated, but was prepared in January 1965 and is a blueprint of Head Start.

[4] Mark R. Arnold, "Opportunity Knocks for Pre-Schoolers in Slums," *The National Observer*, May 24, 1965.

cipients; moreover, they possessed the expertise and administrative capacity to deploy the funds during fiscal year 1965.

Head Start also had the potential for forming the core of a community action program in areas that had failed to set up CAA's. Often the established political powers were neither concerned nor sufficiently skilled to set up CAA's. OEO thus conceived that Head Start might be part of a "building block" approach to a CAA. Because Head Start required the establishment of advisory boards by the fall of 1965, it was hoped that these boards would serve as nuclei for CAA's around which other CAP projects could be added. In communities hostile to the idea of a CAA, Head Start was the brightest hope for their creation.

Underlying all the factors that induced OEO to fund Head Start on such a large scale for the summer of 1965 was the program's enormous popularity. No doubt the secret of Head Start's success was that it directed attention to preschool poor children, who readily aroused the public sympathy. Whatever its source, the unexpected and overwhelming response contrasted sharply with the apathy and antagonism with which many other OEO programs were greeted. OEO needed no public relations experts to discern that Head Start was a best seller. Head Start continued to be the antipoverty program's most popular effort. In contrast to other poverty programs, which were plagued by charges of waste and inefficiency, Head Start was widely acclaimed. The major criticism seemed directed at delays in grants and inadequacy of funds.

At first Head Start was viewed as a temporary summer effort, and it was not until late summer of 1965 that it was expanded to a year-round program and also achieved a permanent status. President Johnson, anxious to bolster the poverty program, announced on August 31, 1965, that Head Start would be expanded to a full-year effort.

During its first 3 years of operations, Head Start was by far the largest single component of CAP. The $653 million allocated to Head Start accounted for two-fifths of total CAP budget during the period (table 3). But, despite the relatively large expenditures, the demand for projects exceeded available funds. By 1967 Head Start was funding 1,500 summer projects and 500 full year projects. Sharing the general public's enthusiasm for the program, Congress displayed unusual generosity: for fiscal 1967, for example, Congress allocated $42 million more than the administration requested.

Head Start's popularity was not universal, however. Paradoxically, opposition to expansion of the program came most frequently from the most ardent supporters of community action programs, including CAP officials at the national and local level. Their resistance stemmed from a conviction that Head Start and other prepackaged projects contributed little to the community action concept, only diverting limited funds and attention from the more important task. In many communities, moreover, Head Start became an adjunct to the school system, thus reducing the resources available to CAA's as leverage to encourage their communities to mount a coordinated attack on poverty. Finally, Head Start was a national program, and some CAP personnel feared that it and other "national emphasis" programs would destroy local initiative.

TABLE 3.—*Head Start expenditures and enrollment*

[Dollar amounts in millions]

Fiscal year	Total	Summer programs		Full year		Training	Evaluation and research
		Children	Amount	Children	Amount		
1965	$103	560,000	$84	----------	$8	$11	---------
1966	198	573,000	97	160,000	83	16	$2
1967	352	473,000	[1] 119	193,000	211	16	6
1968 (requested by President)	352	520,000	104	213,000	224	18	6

[1] An additional $15,000,000 obligated from fiscal year 1967 funds to supplement summer 1966 program in 9 large cities. Total costs for summer 1966 programs was $111,000,000. Summer 1967 program costs estimated at $104,000,000.

Source: Office of Economic Opportunity, community action program.

The summer efforts have remained at the size established during the first summer, and the major program expansion has involved full-year projects. The median length of the full-year program is 9 months. Projects average 5 hours a day, but 28 percent of the full-year programs run 7 hours or more—reflecting the fact that many such projects double as day-care centers for children of working parents.

Full-year programs tend to serve younger children than the summer projects. And in the fall of 1967 OEO announced the launching of projects in 36 cities—at an annual cost of $6.6 million—for families with at least one child below 3 years of age. The funding of nursery and even prenursery projects reflects the belief of Head Start's officials that programs should be made available to children at the earliest age possible. Also, in some communities where kindergarten facilities are universally available, Head Start has been able to fund nursery projects. As table 4 shows, a majority of children enrolled in full-year Head Start projects were less than 5 years old; many of these no doubt enter kindergarten upon completion of the Head Start experience. The objective, however, has been to provide year-round Head Start facilities during the year prior to entry into public schools. Some have questioned this policy and have urged that Head Start concentrate its funds in areas where public schools do not offer kindergarten. The issue is the most effective way of distributing limited resources. The need for giving poor children a head start as early as possible is acknowledged.

TABLE 4.—*Age distribution of Head Start participants, 1965–66*

[In percent]

Age	Summer, 1965	Full year, 1965–66	Summer, 1966
Total	100.0	100.0	100.0
Below 4	.7	10.9	2.3
4 to 5	14.0	47.9	18.2
5 to 6	44.2	31.6	44.5
6 and over	41.0	9.6	35.0

Source: Office of Economic Opportunity, community action program.

Despite its growth and increasing expenditures, Head Start has been able to serve only a portion of the poor children in the poverty category. To reach all impoverished children between the ages of 3 and 6, facilities would have to be provided for more than 2 million chil-

dren—more than three times the number who attended Head Start programs during fiscal 1967 (table 5). OEO officials estimate that the initial annual cost of establishing a full-year Head Start effort for all eligible children would amount to $6.5 billion, nearly four times the total OEO budget for 1967.[5] Lack of funds is therefore the major barrier to expansion of Head Start to serve all eligible children. Inadequate facilities and shortage of teachers are other constraints upon expansion. More than $3.6 billion would be needed to construct, renovate and equip adequate facilities to provide full-year Head Start for the 2.2 million poor children needing such a program.

Program Components and Operation

Head Start differs from the usual nursery school or kindergarten. These latter presumably enrich the preschool experience of children and allow mothers a respite from their offspring. As envisioned by educators, the nursery school function is to "influence only the child's emotional and social development—not his mental growth. . . . Deliberate teaching of mathematical or language skills is frowned upon in these schools."[6]

Table 5.—*Head Start universe, 1966*

	3-year-olds	4-year-olds	5-year-olds
Total poor children	1,000,000	1,000,000	1,000,000
Poor in other programs (excluding title I preschools)	----------	34,000	467,000
Total universe for Head Start	1,000,000	966,000	533,000
Universe which program may reach (80 percent)	800,000	773,000	427,000
Net Head Start universe (including 10 percent nonpoor)	880,000	850,000	470,000

Source: Office of Economic Opportunity, community action program.

The objectives of Head Start are much broader (though the social-emotional element is still very strong in Head Start) and stress the educational development of the preschool child. Because the average child from a poor home is already intellectually "retarded" by the time he enters Head Start, the brief summer program is presumably designed to offer a "cram" course to prepare him for school. Although this would appear to be the formal rationale for the summer program as presented by OEO, spokesmen for Head Start have refrained from making such broad claims. Rather, they indicate that it is basically a diagnostic program. That is, Head Start is concerned with improving the nutrition and health of the children in the hope of increasing their readiness for school. In more affluent homes these aspects of the child's development are normally assumed by the family as a matter of course. Finally, as a component of the community action program, Head Start provides a further avenue for involving the poor in the planning and administration of projects.

Child Development

The children with whom the program deals come from deprived backgrounds and broken homes (table 6). More than a fourth came

[5] U.S. Congress. House. Committee on Education and Labor. *Economic Opportunity Act Amendments of 1967* (Washington: U.S. Government Printing Office, 1967), pt. 2, pp. 1413–14.

[6] Maya Pines, "Slum Children Must Make Up for Lost Time," *The New York Times Magazine*, October 15, 1967.

from families who were on welfare or whose father was absent from home. Given the deprivation of these children, the program planners felt that only a "comprehensive program" would be adequate, preferring the name "child development centers" to emphasize the broad scope of the program's concern.

TABLE 6.—*Family profile of Head Start children*

[In percent]

	1965–66 full-year program	Summer 1966
Past medical treatment:		
No visit to doctor in past 2 years	14	15
No visit to dentist in past 2 years	41	43
Their families:		
Father not living with child	26	31
Parent or guardian with 8 grades or less of education:		
Father or male guardian	31	37
Mother or female guardian	29	31
Parent or guardian presently employed:		
Father or male guardian	65	69
Mother or female guardian	26	25
On welfare	24	20
Their homes:		
No running water inside	9	17
Telephone not available	34	32

Source: U.S. Department of Commerce, Bureau of the Census.

Because the program concentrates on meeting the individual needs of each child, it demands a high teacher-student ratio. One professional teacher for every 15 children is considered optimal, although the median has actually been one to 16. Head Start also favors a ratio of one adult (including volunteers) for every five children.

In line with other community action operations, Head Start has emphasized the employment of subprofessionals and volunteers to relieve the workload of the teacher and provide additional attention to the child. Each teacher is expected to have one paid subprofessional and one volunteer assistant. Over 50,000 subprofessionals and nearly twice that number of volunteers have worked on Head Start projects during each of the summer efforts. About 19,500 subprofessionals and 39,000 volunteers participated in the full-year projects during fiscal year 1967.

Many of the subprofessional workers and volunteers have been parents, primarily mothers, of children participating in the program. Parental involvement is a major goal of the Head Start effort, for it is recognized that the child's needs cannot usually be met without parental cooperation and even changes in the home environment. Bringing parents into the day-to-day operation of the centers has proved an effective way to enlighten parents concerning child-rearing practices and increase their interest in the schooling of their children. According to Bureau of the Census surveys of the 1965 and 1966 summer programs and the full-year 1965–1966 programs, the Head Start professional staff observed that the parents themselves benefited from the children's preschool experience—they were more involved in community activities and the child's education.[7]

[7] U.S. Department of Commerce, Bureau of the Census, "Project Head Start: One Percent Sample Summarizations of Summer 1965 and Summer 1966 Head Start" and "Five Percent Sample of 1965–66 Program."

Parents of Head Start children frequently suffer from their own physical, psychological, or educational handicaps. OEO has estimated that a third of Head Start parents need adult literacy training. Some centers provide special education courses in homemaking and literacy and, when called upon, refer parents to other community action or welfare services.

Perhaps the most conspicuous difference between Head Start and traditional kindergarten programs is the attempt to provide health care. Parents in rural and urban poverty often lack the knowledge or resources to provide their children with adequate medical attention. To meet this need, Head Start projects were required to plan for identifying and attempting to correct or alleviate the children's medical or dental problems.

The medical examinations of Head Start children revealed that many poor children were in serious need of medical help. In Los Angeles, for example, more than 60 percent of the Head Start participants in the 1966 summer program needed medical, dental, or psychological services.[8] While Head Start has been successful in diagnosing the medical problems of the children, it has experienced great difficulty in providing adequate treatment following examination. There was very little followup medical assistance for the summer 1965 program. Improvements were sought in subsequent programs but, according to Dr. Frederick North, senior pediatrician for Head Start, treatment remained "rather like the care of the poor child—fragmented, discontinuous, and often not too well recorded." [9] Dr. North blamed the inadequacy of treatment in part to "institutional and bureaucratic inertia, especially in large cities." He charged that in some cases public health departments responsible for Head Start diagnostic examinations refused to allow their doctors to treat even the minor defects they discovered, thus creating the need for cumbersome and often unsuccessful referral systems. Data from the 1966 summer health program indicate that the medical needs of the children were very great, with almost 40 percent requiring dental care. More than half the disorders discovered, according to OEO figures, received some treatment.[10]

Health services in the child development centers highlight the problem of coordinating Federal and local resources. Funds are available for health services designed for Head Start from a multitude of Federal, State, and local sources. The diversity of fund sources poses a serious challenge to local Head Start officials. The national office recommends that plans and budgets for examination, treatment, and preventative services be prepared only after extensive consultation with physicians, dentists, and public health officials. Head Start funds are to be used only to fill gaps in the community's existing health programs. Knowing all the sources that may be tapped, much less integrating them in a medical program, has proved an enormous task. To improve the health services offered to Head Start children, OEO has contracted with the American Academy of Pediatrics to select

[8] *Los Angeles Times*, October 21, 1966.
[9] Gerald Grant, "Head Start: Not Enough," *Progressive Magazine*, vol. 31, No. 3, Mar. 1967, p. 31.
[10] Office of Economic Opportunity, community action program, "National Summary—Health Report, Summer 1966 Head Start," April 24, 1967.

300 physician consultants to help plan and evaluate project health programs.[11]

Head Start also attempts to provide a mental health program for the children. Many children of illiterate and unemployed parents suffer from special psychological problems that require treatment if they are to progress through school. National guidelines specify that Head Start centers provide psychologists and social workers to help disturbed children. In actuality, however, the delivery of such services has not been realized. Census Bureau surveys of summer and full-year programs revealed that professional, subprofessional, and volunteer workers believed the psychological and psychiatric services were among the least adequate features of Head Start programs. Budgetary limitations and the shortages of qualified personnel are the major obstacles.

The Head Start centers also offer one and often two meals and snacks per day to the children, who frequently lack proper diets. Head Start centers are also expected to educate parents concerning proper nutritional requirements.

Head Start is thus conceived as a multifaceted program to meet the needs of the impoverished child. It understandably has excited the imagination of many communities throughout the country. At the same time, the program has been very expensive, the full-year program more than doubling the costs normally spent per child in kindergarten programs (table 7).

COMMUNITY ROLE

As a component of the community action program, Head Start has other objectives which extend beyond providing for the children. These added goals have occasionally produced politically explosive issues.

TABLE 7.—*Head Start costs per child by program component, 1965–67*

Service	Summer			Year round			
				Part day		Full day	
	1965	1966	1967	1965–66	1966–67	1965–66	1966–67
Total program cost per child	$150	$192	$200	$900	$1,050	$1,260	$1,380
Costs per child per month	75	96	100	90	105	105	115
Percent distribution	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Parent and family (recruitment, social services, parent involvement)	1.7	6.3	6.3	4.6	5.9	5.3	5.4
Health, (medical, dental, psychological)	6.0	18.2	18.2	4.6	6.3	3.3	4.0
Nutrition	8.0	13.5	13.5	12.7	13.7	13.8	14.0
Daily activities (personnel, equipment, child transportation)	77.3	54.2	54.2	70.4	66.9	69.6	68.6
Research, evaluation, and training	2.0	4.7	4.7	3.4	3.2	4.0	4.0
Administration	5.0	3.1	3.1	3.4	4.0	4.0	4.0

Source: Office of Economic Opportunity, community action program.

One controversial issue has been parent participation on Head Start advisory committees. Each Head Start project was required to have an advisory committee, in order to involve the parents and the local community in the administration of the Head Start program.

[11] American Academy of Pediatrics, "AAP Launches Head Start Medical Consultation Program," *Newsletter*, Oct. 10, 1967, p. 1.

Half of the committee's members were required to be parents of the children; the other members would be representatives of community groups, professional organizations, and teachers and school administrators. Through participation, parents would, it was hoped, develop more active concern for the education of their children. Moreover, it was thought that involvement in the planning and administration of Head Start projects would give parents the experience and self-confidence to pressure local school boards to undertake increased efforts to educate the poor.

These advisory committees were officially assigned broad responsibilities, including concurrence in the selection of the Head Start director and review of Head Start applications. Many school boards running the Head Start projects did not take kindly to such "outside" interference with their traditional authority. In the Washington, D.C., suburban area, for example, superintendents from Maryland and Virginia joined in objecting to Head Start criteria which gave the advisory committee power to reject the school board's choice and pick its own staff director.[12] The superintendents charged that this was contrary to Maryland and Virginia laws which gave them exclusive power to hire school personnel. In the face of this opposition, the OEO's mid-Atlantic CAP director "clarified" the role of the advisory committee, stressing that it was a consultative body whose views must be sought concerning the Head Start director.[13] Head Start officials are apparently leaving the power struggles for the local level. But they continue to press for advisory committee involvement and that it be consulted in the selection of project directors, even though the official appointment is left to the agency selected to run the program. Jule M. Sugarman, Director of Head Start, has stressed the need of involving low-income parents in Head Start projects.

Inasmuch as Head Start programs could be administered by private nonprofit and public agencies (including civil rights organizations) as well as by school boards, Head Start projects could serve to challenge backward school systems which cling to traditional education methods. In some areas, in fact, the availability of other agencies has forced reluctant school boards to sponsor projects.

Although Head Start has been relatively free of the political controversy which has plagued community action agencies generally, the case of the Child Development Group of Mississippi (CDGM) was a conspicuous exception. In conducting Head Start centers for more than 6,000 children in 28 counties, CDGM won two OEO grants (in summer 1965 and February 1966) and unusual praise from OEO officials and outside observers. However, the group's association with civil rights activism proved irritating to local politicians, including the State's congressional delegation; and in September 1966, OEO suddenly announced it was withdrawing support in favor of a newly formed biracial group, Mississippi Action for Progress (MAP), which had the Governor's backing. The ostensible reason for OEO's action was a charge of "fiscal irresponsibility" against CDGM, though the group's difficulties were apparently no more serious than those experienced by many CAA's—inexperienced administrators who were

[12] Ellen Hoffman, "Head Start Rules Scored by Educators," *The Washington Post*, Dec. 11, 1966.
[13] Ellen Hoffman, "Head Start Revises Its Guidelines After Protest by Schools," *The Washington Post*, Jan. 5, 1967.

unaccustomed to General Accounting Office procedures. Vocal outside pressures resulted 2 months later in OEO's renewal of the CDGM grant, though with several stipulations designed to circumscribe the group's operations. OEO did not escape unscathed from this incident: civil rights advocates were embittered at what they saw as a compromise with the segregationists, while the latter continued to resent OEO's financing a project associated with Negro activists.[14]

TRAINING AND RESEARCH

As with any education program, an appraisal of Head Start must begin with the quality of the teachers and other personnel associated with the programs. The studies that have been undertaken have indicated that the quality of the Head Start program and the caliber of the teaching personnel are vitally important factors that affect the gains of the children.

Teachers for full-year programs are, where possible, given 8-week child development and education programs in qualified colleges and universities. Some subprofessional aides also participate in these courses. Summer teachers, teachers' aides, and occasional volunteers are provided a week of orientation sessions.

The training program has apparently been inadequate. Only 2,700 of the 18,000 teachers for the full-year 1966–67 program participated in the 8-week orientation course held in the summer of 1966. About 32,000 persons participated in the summer 1966 week-long orientation session, although there were 46,000 teachers, 57,000 nonprofessionals, and 93,000 volunteers involved in Head Start that summer. It may also be questioned whether the week-long summer orientation courses are long enough to import the necessary level of training.

Most of the training for Head Start staff workers is presumably accomplished by inservice training programs. However, Bureau of the Census surveys of the summer and full-year Head Start staff have revealed that many staff members of the Head Start programs receive no training at all, despite the fact that more than a third of the staff workers had no prior experience with preschoolers and almost a half had never before worked with poor children.

Costs, facilities, and trained personnel are the major constraints upon expanding the training program. Reports on the 1967 summer training (week-long orientation courses for summer personnel) indicate that 43,000 people were trained at a cost of $7 million, or $165 per trainee. To provide a similar course for the entire summer teaching staff—including volunteers—would require a program almost five times as large.

Head Start has devoted considerable funds to research and evaluation. Two million dollars was spent in fiscal year 1966 and more than $6 million in fiscal year 1967; a similar amount was planned for the subsequent year. The budget has been distributed about equally among independent research projects; research, demonstration and development projects; and 13 evaluation and research centers located at various universities that specialize in child development. In addition to conducting independent research, the centers are assigned a major re-

[14] William C. Selover, "Shriver Turnabout on Poverty Project Criticized," *The Christian Science Monitor*, Oct. 19, 1966, and Patrick Young, "Why There's a Big Fuss in a Pre-School Program," *The National Observer*, Oct. 10, 1966.

sponsibility for collecting data on the operation of various Head Start centers in their region and evaluating the success of each program. More than a million dollars was devoted to this effort in fiscal year 1967, the objective being to pinpoint the factors that make Head Start programs successful.

The research, demonstration, and development projects consist of half a dozen experimental Head Start programs where new techniques of training are developed, curricula are enriched, and controlled research projects are undertaken. Although these demonstration projects are expensive to administer, Head Start officials justify the extra costs on the basis that the projects provide an opportunity to test hypotheses and use control groups—conditions impossible to obtain in regular Head Start projects run by school systems or CAA's.

The research effort is directed to testing a number of hypotheses and discovering the factors that determine the success of Head Start programs. The effect on the child of the quality of teaching and nature of the school have been measured. Then, research was directed to the question of whether family involvement in the Head Start program is related to the child's development; tentative results support the relationship. Much of the research is inconclusive, however, since measuring and testing the factors associated with the complex educational and psychological development of preschoolers is exceedingly difficult. And there is little to build upon, because our educational system has neglected the evaluation of motives, values, goals, and attitudes related to school achievement. A great deal of research is thus devoted exclusively to developing better instruments of measurement.

Although the Head Start effort cannot as yet claim any major breakthroughs, it is one of the most comprehensive Federal research designs ever undertaken in a single area of educational development. The research has attracted outstanding authorities in the field. Nevertheless, because of the experimental nature of the effort, Head Start officials have emphasized that satisfactory answers to some of the pressing questions cannot be expected for several years.

Administration of the Program

The Head Start program is administered through OEO's seven regional offices. The grantee at the local level is the community action agency for the locality; where no CAA has been formed or special conditions prevail, a public or nonprofit organization may administer a Head Start project as the grantee, provided the agency organizes a policy board including representatives of the poor before submitting an application. In the summer of 1965, a number of Head Start projects were administered by public and nonprofit organizations as the direct grantees. By 1967 most participating localities had formed functioning CAA's and the number of other Head Start grantees had dwindled. About nine of every 10 grantees of full-year Head Start projects are community action agencies.

The CAA as the direct grantee may delegate the operation of Head Start programs to public or nonprofit agencies, including religious organizations. The public school systems have played a large role in administering the program (table 8). In the 1966 summer program, 67 percent of the agencies administering the program were school systems (the direct grantee in most cases being the local CAA). Fifty-

five percent of the children participating in Head Start that summer were in school-administered programs. In the full-year Head Start projects, the school systems play a smaller role; and the CAA's themselves, with the help of private nonprofit institutions, run a majority of these projects.

TABLE 8.—*Delegate agency affiliation* [1]

[In percent]

	Full year, 1965–66	Summer, 1966	Full-year, 1966–67
School	35	71	35
Church	10	2	10
Private nonprofit	26	18	26
CAA's (actually running the program)	29	9	29

[1] A delegate agency is a grantee actually administering a Head Start program or an organization that has contracted with a direct grantee. Most of the schools, churches, and private nonprofit agencies administering the program were delegate agencies of the local CAA's.

Source: Office of Economic Opportunity, community action program.

Among the various States, the distribution of delegate agencies differs greatly. In the Southeast region (Alabama, Florida, Georgia, Mississippi, South Carolina, and Tennessee), 24 percent of the agencies administering the full-year 1965–66 programs were public schools, compared with the national figure of 35 percent. Fifty-two percent of the delegate agencies in the Deep South for full-year programs are community action agencies. Difficulties in complying with civil rights criteria account for the diminished role of the schools in this region.

The eligibility of religious institutions as recipients of Head Start funds has been a unique feature of the program. While religious institutions are not eligible for direct Federal grants under the Elementary and Secondary Education Act of 1965, Head Start has supported projects sponsored by church groups providing that the funds would not be used for religious instruction, proselytization or worship. Admission to church-sponsored projects cannot be based on religious affiliation or atendance in a church-related school. In addition, facilities used for Head Start classes must be devoid of sectarian or religious symbols.[15]

National Head Start guidelines regulate the proportion of the nonpoor that can participate. The rationale for nonpoor participation was that children from more advantaged families would enrich the classroom activities, providing better programs for the poor children. During the 1965 summer, 15 percent of the children in a Head Start project were allowed to come from families which exceeded the standard poverty criteria. The figure was later decreased to 10 percent in response to public and congressional criticism. However, Head Start continues to permit a greater number of more advantaged children to participate in projects, if the cost of the additional children is met by other than Federal funds or the required local contribution.

Although there were a number of reports that the percentage of children from above the poverty line exceeded the national guidelines (one Congressman reported that his son had been invited to join a Head Start program), the requirement was by and large met.

[15] Office of Economic Opportunity, community action program, *How to Apply for Head Start Child Development Programs*, September 1966.

The median family income of the children was $3,400 and the median family size was six. This was well within the national poverty income criteria.

Some communities, on the other hand, have criticized the exclusion of middle-class children, charging that the program was guilty of "economic discrimination." [16] The Educational Policy Commission of the National Education Association advocated expansion of Head Start programs to 4- and 5-year-olds who fall in that "large middle group—between the well-to-do and the disadvantaged." [17] The Commission was understandably concerned that more than half of the Nation's 4- and 5-year-olds are not enrolled in any school. Head Start maintains an ambivalent position on enrolling nonpoor children. For political reasons, Head Start officials prefer to limit enrollment on Head Start projects to the poor, in order to avoid charges that the program serves children from affluent homes. In addition, broadening the enrollment base creates policing problems—checking that local and State funds cover the costs of enrolling children ineligible on the basis of income criteria. On substantive grounds, however, Head Start spokesmen acknowledge the advantages of adding children from more affluent homes in the program, as long as the extra costs are covered by non-Head Start funds.

FUND ALLOCATION

Head Start projects are included as part of CAP appropriations, a fact which has led to wide controversy. Opponents of the community action program no doubt have resented the bolstering of CAP appropriations with the popularity of Head Start; friends of Head Start feel the program has been damaged by its inclusion in CAP requests.

Allocation of Head Start funds among the States and counties is based on the same criteria governing distribution of other CAP funds.[18] In January 1966, OEO itself added the requirement that no CAA could spend more than one-third of its funds on Head Start projects. This limitation was prompted by concern among CAP advocates that CAA's would concentrate disproportionately on the popular Head Start program. This restriction triggered widespread opposition, and it soon became apparent that CAA's as well as Head Start programs would be damaged.[19] The issue became moot when Congress earmarked more than a third of the total CAP authorization for Head Start.

LOCAL SCHOOL BOARDS, TITLE I AND HEADSTART

The administrative relationship of the public school systems has proved one of the most controversial aspects of the Head Start program. Public schools administering Head Start projects operate by and large as "delegate agencies" responsible to the local CAA. Each CAA submits its application to the regional office and is held responsible for obtaining public school compliance with regional and national

---

[16] "Next Year's Head Start," *The Washington Post*, Jan. 31, 1967.
[17] *Philadelphia Inquirer*, June 5, 1966.
[18] Office of Economic Opportunity, community action program, Community Action Program Memorandum No. 21, Feb. 28, 1966.
[19] U.S. Congress. House. Committee on Education and Labor. *1966 Amendments to the Economic Opportunity Act of 1964* (Washington: U.S. Government Printing Office, 1966), pt. 1, pp. 195–208.

Head Start standards. National and regional headquarters have urged CAA's to keep a watchful eye on public-school-administered Head Start projects, in order to insure quality programs and persuade the school systems to devote more effort to the poor in the primary grades.

Too frequently, however, CAA's act as little more than a funnel for school board requests. Some school boards deal directly with the OEO regional representative on financial matters without even going through the CAA. The CAA's and school-administered Head Start programs often operate autonomously. Many school boards continue to show no special concern for the poor, and are jealous of Head Start interference in educational matters.

The administrative arrangement favored by OEO contrasts sharply with the structure of traditional Federal-State programs in education and welfare. The Elementary and Secondary Education Act of 1965, for example, provides for the distribution of Federal money to State education authorities allocated according to congressional guidelines with little Federal monitoring. State education authorities distribute the funds to local school boards and are the only agencies in a position to exert control over the use of the money. The Federal role is strictly circumscribed. Head Start, in contrast, follows no rigid allocation formula, bypasses State education authorities, and assigns local CAA's the role of Federal watchdog over Head Start funds.

The controversy over the administration of Head Start is heightened by the availability, under title I of the Elementary and Secondary Education Act, of funds for preschool programs that might be devoted to the same efforts as Head Start. In fiscal 1966, more than 100,000 children participated in prekindergarten programs under title I assistance, and 380,000 children attended kindergarten financed in whole or in part from title I.[20] The provision of funds for similar purposes from two different Federal sources through different administrative channels has increased the pressure for a unified program.

Head Start officials have resisted proposals to merge the two programs in the Office of Education. More than narrow bureaucratic interests are involved. Head Start's monitoring of projects to assure that grantees follow national goals, albeit not always successful, would be precluded by the Elementary and Secondary Education Act structure, in which the Office of Education serves primarily as a funnel in allocating Federal funds to State educational authorities. Moreover, Head Start is the most popular CAP component, and in many areas the only justification for the existence of the local CAA. Removal of Head Start would unquestionably damage, if not kill, some CAA's.

The more sanguine Head Start advocates have also hoped that Head Start could give Federal authorities and local CAA's leverage in persuading local school boards to devote more effort in the primary grades to the education of the poor. Channeling Head Start funds through State educational authorities would completely shield school boards from Federal and CAA pressures for change.

Another advantage of the present Head Start administrative structure, as seen by OEO officials, is that Head Start projects can be operated by nonpublic school agencies. As table 8 indicates, a large number of Head Start projects are administered by agencies other than school

[20] U.S. Department of Health, Education, and Welfare, Office of Education. *The States Report: The First Year of Title I* (Washington: U.S. Government Printing Office. 1967).

boards. This makes it possible to establish Head Start in areas where school boards are unable or reluctant to embark on the venture, and may increase the pressure that can be brought to bear on local school boards to establish Head Start projects. Restricting Head Start grants to State education authorities and local school boards would destroy this leverage. And where local authorities are willing to undertake projects, the Federal strings are often necessary to induce them to run the programs on other than traditional kindergarten lines. If Head Start were administered by the Office of Education under a title I structure, State and local authorities would be free to run their own kind of program.

The experience of title I operations to date supports OEO's argument. The Council on the Education of Disadvantaged Children, an advisory group of distinguished educators established under the Elementary and Secondary Education Act of 1965, reported to President Johnson that most summer programs under title I were "piecemeal, fragmented efforts of remediation or vaguely directed enrichment." [21] The content of title I preschool programs is left almost entirely to local school boards, many of which seem to place little value in parent participation or medical, dental, and nutritional services for participants. Rarely did the Council find a "strategically planned, comprehensive program for change." Many Head Start projects may fall short of such a high standard, but at least here Federal pressures operate in that direction.

Critics of OEO respond that transfer of Head Start to the Office of Education would still be desirable in eliminating unnecessary duplication and returning educational matters to local and State school authorities. The Federal "interference" that accompanies Head Start funds is rarely welcomed and usually resented by the educational establishment. Many Congressmen, finding the arguments of school administrators persuasive, have favored the transfer of Head Start to the Office of Education. However, an amendment to that effect, introduced during consideration of the extension of the Economic Opportunity Act (EOA) in 1967, was defeated—not because the majority loved OEO more than the Office of Education, but because southern segregationists objected to transferring Head Start to an agency headed by Harold Howe II, who happened at that time to be more unacceptable to them than Sargent Shriver. As one southern Congressman put it, somewhat inelegantly, "The best thing * * * to do is to keep all the trash in one pile, do not scatter it." [22]

COORDINATION

The availability of funds for preschool programs from both OEO and the Office of Education has prompted demands for coordination efforts at the local and national level. Though title I funds can be used to supplement Head Start projects, in most cases the two programs have operated independently. An exception is cooperation between the Office of Education and OEO to prevent school systems from violating civil rights standards. Also, the two national offices have agreed to

[21] U.S. Department of Health, Education, and Welfare, Office of Education, "Summer Education for Children of Poverty," Report of the National Advisory Council on the Education of Disadvantaged Children (Washington: U.S. Government Printing Office, 1966), pp. 39–40.

[22] Joe D. Waggonner, *Congressional Record* (daily edition), Nov. 13, 1967, p. H15091.

open Head Start summer orientation programs to personnel participating in the Elementary and Secondary Education Act (ESEA) programs.[23]

Despite such efforts at coordination, OEO and OE do not integrate major policy determinations. Each agency goes its separate way. OEO, for example, asked the Office of Education to earmark its title I funds for followthrough programs in the primary grades for former Head Start participants. The Office of Education refused, reducing title I preference provision drafted by OEO into a hortatory message. The lack of coordination between the two agencies reflects in part the fact that the Office of Education makes no major policy decisions: Under title I, decisions on how money is spent are left to local school systems and State educational authorities. Hence, the coordination of preschool programs that Head Start officials might have preferred could not be made with the Office of Education, because major decisions on the use of title I funds are lodged elsewhere. The difficulty of coordinating title I and Head Start funding is illustrative of the relationships between other CAP-funded projects and older grant-in-aid programs which distribute funds through State agencies.

Local coordination between school boards and CAA's concerning title I and Head Start funds is also weak. A formal checkpoint procedure which is prescribed, forms the basis of cooperation. In the case of Head Start, the applying agency must certify in its application that it has consulted with the local school board about obtaining title I funds. And under title I, the Office of Education must require that local boards of education consult with CAA's in the formulation of their plans and programs. However, these procedures for coordination are usually little more than paper efforts.

Information on the use of title I money in Head Start projects is scarce. The national offices of Head Start and the Office of Education have no data on the utilization of funds, and only impressions can be gleaned from disparate sources. In Washington, D.C., the summer 1966 programs used Head Start funds for 5 year olds and title I funds for younger preschoolers. In Dade County (Miami), Fla., title I funds were used to provide facilities for Head Start programs, with Head Start financing the operating costs. Los Angeles combined three sources of funds for its Head Start program, using $4 million from Head Start, $2.6 million from California's Compensatory Education Act, and $100,000 from ESEA. In New York City, the Board of Education paid teachers' salaries for a program involving 28,000 Head Start children, while OEO provided the health, social services and food components.[24]

RACIAL SEGREGATION AND HEAD START

The large number of Head Start programs in the South, combined with OEO's insistence on racially integrated efforts, has been a continuing source of conflict. For Head Start officials, segregated pro-

---

[23] U.S. Congress. Senate. Committee on Labor and Public Welfare. *Amendments to the Economic Opportunity Act of 1964*, "Memorandum from Office of Education and OEO on Cooperation" (Washington: U.S. Government Printing Office, 1966), pp. 71–72.

[24] "District of Columbia Summer Head Start," *The Washington Star*, June 16, 1967; Office of Economic Opportunity, Community Action Program, "Miami Receives $2 Million Head Start Grant," release 66–6, Jan. 21, 1966 and "Los Angeles Receives $4 Million Head Start Grant," release 66–99, Jan. 19, 1966.

grams pose a severe dilemma. On the one hand, violations of civil rights guidelines warrant a termination of Federal funds. Yet such action may result in denying the poorest children a preschool program.

Three methods have been used frequently to evade the civil rights requirement.[25] First, local officers may refrain from using any white schools, even where there are substantial numbers of eligible white children. Second, virtually all Negro staffs may be selected. And third, recruitment efforts for Negroes and white children may differ (Negro families are actively solicited, whereas white families get only a general notice). The nature of these violations indicates that such communities run their Head Start programs primarily for Negro children.

Southern Head Start projects, while not exclusively serving Negro children, have indeed concentrated on them. Although there are twice as many white families in the South in the poverty category as there are nonwhite poverty families, Negro children have constituted the majority of Head Start participants in the South. The South has not been alone in mounting racially imbalanced Head Start programs: nationally, Negro preschoolers have participated in the program in greater proportions than their percentage of the poverty population (table 9). Nonwhite children have accounted for about two of every five Head Start enrollees, though they constitute a third of the poor children in the Nation.

TABLE 9.—*Estimated percentage of 1st grade students participating in Project Head Start, 1965–66*

	Negro	White
Metropolitan (total)	13.8	2.0
Northeast	14.2	1.9
Midwest	2.7	1.3
South	25.1	7.4
Southwest	19.3	1.2
West	1.9	.5
Nonmetropolitan (total)	32.6	7.0
North and West	35.8	4.1
South	35.2	12.0
Southwest	12.8	4.8

Source: James S. Coleman et al., *Equality of Educational Opportunity*, U.S. Department of Health, Education, and Welfare, Office of Education (Washington: U.S. Government Printing Office, 1966), p. 492.

Some communities have utilized the availability of funds for preschools under title I to avoid integrated programs. Under this tactic, title I money is used to finance all-white projects while Negroes are provided Head Start facilities.

Another favorite technique of segregated southern communities was allowing parents and children "freedom of choice" between two or more centers.[26] Negro families gravitated toward Negro centers while white children attended all-white centers. Subsequently, OEO had forbidden "freedom of choice" and has required children to attend the Head Start center in their geographical district.[27] But while this action

[25] Office of Economic Opportunity, "OEO Expresses Concern Over Segregation in Head Start," release 66–340, May 20, 1966.
[26] Office of Economic Opportunity, "Freedom of Choice Disallowed for Head Start," release 65–1594, Dec. 8, 1965.
[27] Footnote 15.

prevents segregation in mixed housing communities, it does nothing to assist integration where the housing pattern is itself segregated.

Although segregation is most flagrant in the South, it is a national malady. More than one-third of all Head Start projects visited during the 1966 summer by the Head Start inspection staff lacked an integrated program. The administrators were not necessarily guilty of civil rights violations: often the projects only reflected the racially segregated nature of the communities.

Though OEO could cut off funding to projects where civil rights were violated, it has used this authority sparingly. Faced with the choice between a segregated preschool program and no preschool program at all, Head Start officials have usually limited themselves to threats short of actual project termination. OEO announced that out of the 2,400 Head Start projects during the summer of 1965, about 60 communities ultimately were denied funding for failing to meet OEO's nondiscrimination requirements. In the fall of 1966, 25 summer Head Start programs were refused funds for civil rights violations. In all cases the communities were offered an opportunity to defend their action in public hearings.

An Assessment

The Head Start program has sustained its original popularity, managing to avoid the disenchantment that has beset other poverty projects. But popularity does not necessarily connote program effectiveness; and despite congressional and public support, many basic questions regarding Head Start's value, scope, and direction remain unanswered. For example, should the limited resources be concentrated upon summer programs preparing poor children for school entry, maximizing the number of participants? Or should the funds be devoted to year-round facilities? Would cost effectiveness be maximized by restricting participation to even fewer children and providing them preschool facilities beginning with age 3 or even lower? And, assuming a determination to concentrate resources on a limited number of children, should the resources be restricted to preschool facilities or should they be devoted to enriching the primary grade education of poor children? Finally, what are the merits of expending funds on services—health, nutrition, and family stability—compared with concentrating on classroom education?

Head Start research efforts have not yet produced conclusive answers to these basic questions. For that matter, it is not even clear whether the resources devoted to the child development program produce lasting improvement in the poverty child's school performance. This issue is fundamental to an assessment of Head Start, for tentative findings suggest that a 2- or 3-month summer program for preschoolers may yield only temporary benefits. The children's needs are simply too overwhelming to be met by a summer program.

During its early phases Head Start was able to capitalize on research that indicated significant educational advancements. Dr. Leon Eisenberg of Johns Hopkins University found improvements of eight to ten points in the IQ's of 480 children who took part in Baltimore's 1965 Head Start project.[28] These findings were supported by additional studies during the following year. Followup studies of the 1965 sum-

[28] Footnote 19, pt. 2, pp. 1138–41.

mer program participants indicated, however, that benefits of the Head Start program soon faded. A widely publicized study by Max Wolff and Annie Stein of Yeshiva University among kindergartens in four public New York City elementary schools showed that, 6 months after the Head Start program ended, participants scored no higher on achievement tests than nonparticipants from similar socioeconomic backgrounds.[29] A later study by the Office of Education also indicated that several months after completion of the 1965 summer program, discernible gains could be detected only among those participants from the most disadvantaged backgrounds—especially among the rural Negroes of the South.[30] For most Head Start children achievement level gains were not appreciably different than those of non-Head Start children. These studies do not indicate that Head Start in any way harms the children, but only that their classmates from similar backgrounds soon catch up to the Head Start children.

Most studies of the effect of Head Start have focused on the cognitive achievements of participants. Because Head Start aspires to broader goals, however, consideration must also be given to other factors in appraising the total impact of the program. Most significant, perhaps, is the fact that Head Start children are frequently observed to be more motivated and responsive than their non-Head Start peers. More than 90 percent of a sample of Head Start parents and workers testified that the children gained self-confidence and developed new interests, according to the Census Bureau survey. The Wolff and Stein study, as well as the Office of Education survey, also indicated that Head Start participants showed much more noticeable gains in educational motivation than in achievement levels several months after the 1965 summer program.

The rapid "fadeout" effect in achievement levels, together with the realization that a short summer course is often inadequate to meet the comprehensive needs of the preschool children, have provided the major impetus for the creation of full-year Head Start programs. These projects, with a median length of 9 months, are expected to have greater impact on the achievement levels of disadvantaged children. The study conducted by the Office of Education showed that first-grade children from almost all socioeconomic backgrounds who attended kindergarten had significantly higher achievement levels than comparable first graders who had not attended kindergarten. The Head Start program, which is generally conceded to be of higher quality than kindergarten, might be anticipated to produce even better results. But the benefits of the longer Head Start program may not be sustained unless the primary school programs the children enter are similarly enriched.

In terms of their impact on the children, then, full-year Head Start projects would seem preferable to the shorter summer programs. Though a summer's exposure is beneficial in some respects, it is not long enough to leave an indelible effect and its benefits (at least in terms of achievement levels) soon fade. If these facts are accepted, why has OEO not been quicker to shift its resources to the full-year concept? The answer is as much political as programmatic. Because

---

[29] Max Wolff and Annie Stein, Six Months Later: *A Comparison of Children Who Had Head Start, Summer 1965, With Their Classmates in Kindergarten,* Aug. 18, 1966 (mimeographed).

[30] James S. Coleman *et al.*, Office of Education, U.S. Department of Health, Education, and Welfare. *Equality in Educational Opportunity* (Washington: U.S. Government Printing Office, 1966), pp. 491–522.

they are much more expensive, full-year programs cannot serve nearly as many children as the summer programs—given current levels of spending. The administration has shown a distinct penchant for the "numbers racket," and the huge numbers of children who have passed through summer Head Start projects unquestionably provide impressive publicity. In this case, however, it has probably amounted to a sacrifice of quality for quantity—a result no doubt regretted by Head Start officials themselves. Aside from these political considerations, it is also more difficult to mount full-year projects than summer programs, when school facilities are unutilized and regular teachers are available.

Head Start's inability to turn impoverished preschoolers into middle-class high achievers should not be surprising. For a realistic evaluation of Head Start must begin with a realistic appraisal of its objectives. Prominent educators and social observers, like Martin Deutsch and S. M. Miller, have warned that it would be naive to expect the children to emerge from the program transformed.[31] The fadeout effect that researchers have observed, then, only confirms what Head Start officials early realized:

> The initial gains are not likely to hold up in the absence of a continuity in educational experience. Preschool education, then, to be meaningful, must be followed by appropriate educational experiences of good quality in the subsequent years.[32]

In view of these considerations, it is unfortunate that OEO attempted to sell Head Start as an educational panacea. While preschool programs unquestionably have an important place in the total school strategy, it would be a mistake to feed the public the notion that preschool is the central focus of such a strategy.[33] And although the politics of the moment dictated emphasis on a popular program, claims that the program was capable of giving the children a head start must be considered part of OEO's penchant for rhetoric. Regrettably, one byproduct of OEO's sloganeering has been skepticism when the program proved incapable of achieving the lofty goals established by OEO spokesmen.

A more realistic role for Head Start—as acknowledged by educators as well as Head Start planners—is as a preschool program in the context of a strategy for change in the primary schools offering poor children an enriched and more comprehensive education. On this count, Head Start has made a contribution.

Another significant innovation—first initiated on a large scale by Head Start and now being adopted by educational systems throughout the country—is the use of subprofessional aids to provide individual attention to students, to relieve teachers of routine tasks and to provide employment to the poor.[34] In fiscal year 1966, title I money was used to employ 73,000 teachers' aids and 44,000 other subprofessionals, equaling more than one-quarter of all new staff positions created by title I funds.[35] It should be noted, however, that the maximum potential of many subprofessionals has not been utilized; they are often assigned routine duties—as lunchroom monitors or keepers of simple

---

[31] S. M. Miller, "Strategy for Change," *American Child*, Spring 1966, pp. 22–24.
[32] Edmund Gordon, "What Did We Learn," *ibid.*, pp. 11–13.
[33] Frank Riessman, "The New Pre-School Mythology: Child-Centered Radicalism," *ibid.*, pp. 19–21.
[34] National Advisory Council on the Education of Disadvantaged Children, Report of Jan. 31, 1967 (mimeographed).
[35] Footnote 20, p. 105.

records. Rarely have they been allowed to contribute to the educational process as Head Start planners would prefer.

Head Start's broad approach to services for children in another concept which has been widely and profitably copied. The Council on the Education of Disadvantaged Children reports that provision of breakfasts and lunches, basic medical care and eyeglasses, shoes and overcoats—the most elemental bodily requisites for learning—is one of the most rapidly spreading practices under title I. Only a few years ago one major Southern city refused to provide free lunches to poor children on the ground that food had no relation to the child's education. Today, that southern city provides free lunches.

There is also increasing concern among local educational authorities for improving teachers' attitudes toward poor children and increasing the flexibility of the curriculum. Although these changes cannot be attributed entirely to Head Start, appropriate credit must be given to its emphasis on innovative techniques dealing with each child as a developing individual. Several colleges and universities have established new departments or programs in early childhood education. One-third of the States have commissioned studies on establishing certification standards for early childhood teaching. Iowa, Vermont, and Oklahoma established such certification early in 1967. This preoccupation with certification is a well established trait of educators, but Head Start may, at least, claim credit for generating the added interest in preschool education.

Head Start programs have increased demands for State support of public kindergartens. Virginia has established free public kindergartens, using Head Start money to finance the costs allocated to special services for poor children. Several States are now considering legislation for statewide kindergartens.

One of the most controversial concepts in the Head Start package has been the involvement of parents in an advisory capacity. School boards and educators have been ambivalent, if not antagonistic, at the prospect of integrating this practice into the regular school program. Head Start has succeeded, however, in successfully raising the issue of parent involvement in some communities. In Washington, D.C., for example, the Superintendent of Schools proposed in April 1967 the establishment of parent advisory councils that were to have a major voice in determining policies of each neighborhood school.[36]

The Head Start preschool program has also stimulated inquiry and research into the educational needs of 2- and 3-year-olds. A project sponsored by the National Institutes of Health found that after a year of tutoring for an hour a day, the average IQ of a group of 30 slum children was substantially raised while the average IQ of a control group of untutored poor children actually dropped.[37] The psychologist in charge of the project has concluded that slum children will make more educational progress if they are properly stimulated by words when they are learning to talk.

Findings of this nature prompted a White House task force on preschool children to recommend to the President that treatment centers be established for very young children and their parents. The parent and children centers would be designed to offer Head Start-type pro-

[36] Susan Jacoby, "Hansen Shifts on Parent Council Plan," *The Washington Post*, June 24, 1967.
[37] *The Washington Post*, June 24, 1967.

grams for the pre-preschoolers and to provide trained workers to go into homes to help improve home life and education.

Potential Catalyst

Whether Head Start is a successful strategy for change ultimately depends on the evolution of the American public school system. More than a change of heart is necessary, for substantial funds must be forthcoming. Commissioner Harold Howe of the Office of Education has estimated that an annual expenditure of $1,200 per child would be a useful benchmark in estimating the cost of providing adequate education for poor children in the primary grades.[38] This is more than twice the amount currently spent per pupil for primary education. Indeed, the financial commitment necessary to make Head Start's goal of quality education for the poor a reality may conceivably lie beyond realistic expectations for the immediate years ahead.

The Federal Government will undoubtedly be the major source of the financial support necessary to provide educational programs adequate to meet the needs of poor children. Congress approved in December 1967 President Johnson's request for a followthrough program to extend Head Start services for Head Start and title I (Elementary and Secondary Education Act) preschool graduates into the first year of regular school. Under an agreement between HEW and OEO the program would be delegated to the Office of Education, with funds allocated directly to local public educational agencies.[39] However, OEO hopes to be able to set program requirements for followthrough, leaving local school boards less discretion in the use of the money than they enjoy under title I funds. Indeed, OEO has urged Congress that followthrough "would be operated by local education agencies, with a view to changing their approaches to primary grade education of disadvantaged children."[40]

The followthrough proposal would enable 190,000 children to participate during the 1968–69 academic school year. Projects would be directed to improving language skills and analytical abilities and providing medical and dental attention and other noncurricular assistance. Special training programs would be provided for teachers and other personnel dealing with followthrough children. Parents would participate in activities aimed at increasing their ability to assist their children, and would help in planning and executing the program. The program package for followthrough thus includes essentially the same components as Head Start. Half of the followthrough participants will be Head Start graduates. If the program is continued over the next 3 years, solid evidence might be developed about the value of the program.

The $120 million in Federal funds requested by the President for followthrough can only be considered a downpayment. The magnitude of the task confronting the American educational system has long been recognized, and competition for scarce educational resources is not new. From its very beginnings Head Start has been a demonstration project to mobilize the resources and thinking of educators and administrators to the needs of poor children. By this standard Head

[38] "Address to school administrators in San Diego," *The Washington Post*, Dec. 7, 1966.
[39] Memorandum of understanding between OEO and HEW relative to the administration of the followthrough program, June 26, 1967.
[40] OEO congressional presentation, April 1967, C–19 (mimeographed).

Start has been successful in dramatizing the educational needs of the poor and in selling to the Nation a program package whose components provide guidelines for dealing with the poor child's needs. Head Start has challenged local school boards which lacked the understanding, concern, or commitment for the education of the poor. Its impact must ultimately be judged by the changes it induces in the American educational system.

GOVERNMENT AS EMPLOYER OF LAST RESORT

BY GARTH L. MANGUM*

CONTENTS

In February 1966, unemployment in the United States fell to 3.7 percent. In the same month, the National Commission on Technology, Automation, and Economic Progress recommended that the Federal Government commit itself as the "employer of last resort," guaranteeing useful employment opportunities to those not absorbed by more orthodox private and public employment.[1] In January 1967, after a year in which unemployment fluctuated between 3.6 and 3.9 percent, the Council of Economic Advisers announced the return of "full employment." [2] In spite of this proclamation, "Government as employer of last resort" has become a familiar and frequently endorsed phrase in public policy discussions. Bills to make the concept reality have been sponsored by Senate and House committees. Their $3 and $4 billion price tags indicate serious political consideration, even though passage is highly unlikely.

The upsurge of support for a program to guarantee employment opportunities coincidental with low unemployment can be understood if one accepts the conceptual device used by the Automation Commission to justify its proposals. Imagine, the Commission said, a gigantic "shapeup" of the kind once common on the east coast docks with workers and potential workers queued up in order of their relative attractiveness to employers.[3] Relative standing in the line would be

*Garth L. Mangum, Research Professor of Economics, The George Washington University, is directing an evaluation of Federal manpower policies under a Ford Foundation grant.

NOTE: This paper was prepared for inclusion in the annual volume of the 1968 Industrial Relations Research Association, *Toward Freedom From Want* (forthcoming).

[1] *Technology and the American Economy*, Report of the National Commission on Technology, Automation, and Economic Progress (Washington, D.C., U.S. Government Printing Office) February 1966, pp. 35–37, 110.
[2] *Economic Report of the President*, January 1967, p. 37.
[3] Technology and the American Economy, p. 23.

determined by such rational criteria as education, skill, experience, and wage expectations, and by employers' prejudices against race, age, or sex. The lineup would not be a static one and new entrants would not necessarily go to the end of the line but would be slotted in according to their own relative attractiveness.

Using this expository device, the Commission pointed out that employers could be encouraged to reach more deeply into the queue by increasing the demand for their products. In addition, those further back in the line might, through education, training, relocation, and other labor market services, be enabled to compete more effectively with others toward the front of the line. However, employers, rather than hire more of the increasingly less attractive workers, could choose to mechanize, extend overtime, and compete with each other for the already employed. Therefore, as unemployement fell, progressively greater increments of demand might be required for additional reductions.

In addition to the shortages and bottlenecks typical of such periods, increased demand could be exploited by business and labor organizations to increase prices and wages. In a country characterized by geographical differences, wide range of educational attainment, heterogeneity of race and national origin and low tolerance for inflation, price increases would become unacceptable long before a job was available for everyone "able, willing, and seeking to work." If the implied promise of the 1946 Employment Act was to be kept, even in a period of low general unemployment, the Commission maintained, the Government would have to guarantee public service jobs to the hard to employ.

The "shapeup" concept has been supported by the experience of the past several years. Stimulated by tax cuts and increased Vietnam spending, unemployment plunged from 5.6 to 3.7 percent between February 1964 and February 1966, with those who had borne disproportionately the burdens of slow economic growth receiving a more than proportionate share of the new jobs. Since the more attractive were already employed, employment increased more than proportionately for the young, the old, the nonwhite and the blue-collar worker. However, with the rapid influx of young workers, inexperienced but better educated than their predecessors, the undereducated were shoved further back in the queue.

The price of the 3 million new jobs during 1965 and 1966 was a 5-percent increase in the Consumer Price Index. As policies shifted to "cool" the economy, the relative progress of the disadvantaged was again halted. It had been hoped that the manpower programs introduced during the previous 4 years (i.e., the Manpower Development and Training Act, the Neighborhood Youth Corps, Job Corps, and related programs) would have reduced the trade-off rates between full employment and stable prices. Instead, the price level experience was similar to that of 1957 when unemployment fell below 4 percent. All the manpower programs together were involving no more than 300,000 of a labor force of 76 million. Any improvements these programs may have made were more than offset by the speed with which unemployment fell during the winter of 1965–66 and by the inherent inflation proneness of war. When unemployment stabilized, even though at a low level, the price pressures continued heavy for the following 6

months, then eased. While it is still possible to believe that, given other conditions, unemployment could be brought below 3.5 percent without politically unacceptable price increases, there is no experience to support that faith.

Inevitably those whose employment opportunities depend upon the final increments of demand bear the major costs of efforts to stabilize prices. That concern for their welfare continues despite low general unemployment is a hopeful sign of increasing sensitivity to economic distress. It is in part, at least, because efforts to increase demand pressure have been stymied by inflation and because the slow processes of labor force and labor market improvement have not yet had a measurable impact that attention is swinging to possibilities of direct attack on the problems of those at the back of the line.

Experience with Job Creation Programs

The frankly attention-seeking phrase, "government as employer of last resort," may have been original with the Automation Commision, but the use of public employment as a weapon against joblessness and inadequate incomes had been tested by considerable historical experience. During the 4 years preceding the Commission report several federally supported programs had experimented with direct employment but the goal was never explicitly stated and the programs were carried out under the guise of other objectives. However, the concept had much longer history, though with national significance only in the 1930's.

New Deal Job Creation Programs

The New Deal marked the belated recognition that the United States had become an interdependent, national, and industrial economy. The emotional opposition it fomented marked the transfer from voluntary private charity to public responsibility and from local courthouses to the Nation's Capital. The permanent reforms have proven unimpeachable, insomuch that all of the accumulated antipathy has concentrated upon the temporary employment programs which disappeared with the war. The Civilian Conservation Corps (CCC) and National Youth Administration (NYA) were well-accepted youth programs and became models for the Job Corps and the Neighborhood Youth Corps. The Public Works Administration (PWA) was an expansion of the normal contracting out of the construction of public facilities. However, it was the adult work relief program, the Works Progress Administration (WPA) which more than any other seemed to symbolize the New Deal.

The "shovel leaning" WPA worker became a universal image in the minds of people still driving the roads, crossing the bridges, walking the sidewalks, and enjoying the economically costless facilities created from otherwise idle resources. The brief list in table 1 is only a minor indication of the accomplishments of that program, apart from its primary employment-generating purpose. It does not include the many conservation, art, and research projects which also had important long-term payoffs and the minor construction, service, and cultural

projects which contributed significantly to the quality of life at the time and whose impact still persists.[4]

TABLE 1.—*What $10 billion bought: Major construction projects of the WPA*

	Number		
	New construction	Additions	Reconstruction or improvement
Highways, roads, streets and related facilities [1]	651,087	----------	----------
Bridges and viaducts	77,965	----------	46,046
Sidewalks and paths [1]	23,607	----------	6,972
Curbs [1]	25,073	----------	3,441
Public buildings	35,064	4,792	85,254
Stadiums, grandstands, and bleachers	2,302	129	797
Parks	1,668	189	6,335
Playgrounds	3,085	107	9,581
Athletic fields	3,026	68	2,457
Swimming pools	805	----------	339
Utility plants	2,877	123	1,172
Water mains and distribution lines [1]	16,117	----------	3,658
Landing fields	353	131	469

[1] Miles.

Source: U.S. Federal Works Agency, "Final Report on the WPA Program 1935-43," (Washington, D.C., U.S. Government Printing Office, 1947), pp. 131-132.

The Public Works Administration provided an average of about 225,000 jobs each year for 7 years from a total budget of $1.7 billion but followed normal contracting procedures with private contractors paying prevailing wages and competing for the best available labor. The WPA was also under pressure to pay prevailing rates but spread employment to over 2 million per year by reducing the number of days an individual worked per month. It also obtained a "bigger bang" for its 7 year, $10 billion Federal budget by furnishing free labor to State and local governments which provided supervision, equipment, and materials.

The CCC provided income and work experience for nearly 2.5 million youth in its residential camps at a cost of $3 billion while contributing to conservation and constructing roads and recreational facilities. However, like the Job Corps conservation centers which are direct descendants, it offered not only employment and experience but basic education as well as some formal occupational training.[5] The NYA like its Neighborhood Youth Corps descendant spent its $500 million, 6-year budget on both in-school and out-of-school components with 1.5 million passing through the former and over 1 million through the latter. Many of them also received occupational training in job shops operated under the program.

Poor planning and hasty undertakings during the early months doubtless gave basis to the "leaf-raking" reputation which has followed WPA and other programs to the current day (though it is difficult to find fault with leaf raking if there are leaves to be raked). Since the primary purpose was employment and production was secondary, labor intensive methods were deliberately used. Aggressive fiscal policies could have minimized the need for work relief but the necessary

[4] The "Roosevelt Monuments" (sanitary privies) became a joke to city dwellers but they were a rare touch of luxury in the otherwise Spartan lives of rural people. For a more complete enumeration of both the employment and physical accomplishments of the WPA see U.S. Federal Works Agency, *Final Report on the WPA Program,* 1935–43 (U.S. Government Printing Office, 1947).

[5] W. S. Woytinsky & Associates, *Employment and Wages in the United States* (New York, the Twentieth Century Funds, 1953), p. 176.

economic sophistication was not available. There was criticism of the programs but little documented abuse. Cost-benefit analysis rather than uninformed emotion would undoubtedly sanction the programs of the period.

JOB CREATION PROGRAMS OF THE 1960'S

The 6.7 percent unemployment which faced the "New Frontier" in 1961 hardly compared to the unemploment of the thirties but it displaced many workers with substantial skills and long labor force attachment. Yet some of its earliest efforts were patterned after the less criticized New Deal programs. The accelerated public works program of 1962 was a PWA for depressed areas. The 1962 Social Security Act Amendments contained an experimental community work and training program offering the States matching funds to encourage unemployed parents of dependent children to gain work experience and training by public assistance payments.

The initial concern of the New Frontier, reflected in the Manpower Development and Training Act (MDTA), was the unemployment of adult family heads with long labor force attachment. However, by 1963, unemployment of white adult men was falling and concern shifted to youth and minority groups and others at the fringes of the job market. The Senate passed a Youth Employment Opportunities Act, using the CCC and NYA as models, but it was bottled up by the House Rules Committee which opposed racially intergrated camps. More successful that year were major improvements in Federal assistance to vocational education, including a work-study program to provide earning opportunities to needy students and reorientation of MDTA to enlarge the youth compenent. Unsuccessful at the time, but a harbinger of things to come was the proposal by the Senate Employment and Manpower Subcommittee in the winter of 1963–64 for a special program of public service employment to allow residents of especially distressed areas and neighborhoods to work at rehabilitating their deteriorating environments.[6]

By mid-1964, the general unemployment level was receding following a sizable tax cut. Yet, rather than relax in satisfaction, the Nation chose that moment to launch its "war on poverty." [7] The Economic Opportunity Act, a grab bag of existing and new programs focused on "breaking the poverty cycle," contained five programs which had job creation components and later amendments added three more with job creation as their explicit goal.

The NYA portion of the youth employment opportunities bill became the Neighborhood Youth Corps which offered publicly sponsored full- and part-time employment to poor youth ranging from park maintenance to subprofessional positions. The Job Corps which enlisted poor, undereducated youths, had rural conservation centers modeled on the CCC and urban centers which were essentially residential vocational schools. The community work and training concept was adopted for welfare recipients and other adult poor under title V of EOA as the work experience and training program. Labor Department participants working on the design of EOA pushed for a

---

[6] *Toward Full Employment: Proposals for a Comprehensive Employment and Manpower Policy in the United States,* U.S. Senate, Committee on Labor and Public Welfare, Subcommittee on Employment and Manpower, 88th Cong. 2d sess., pp. 58–60.

[7] Sar A. Levitan, *The Design of Federal Antipoverty Strategy,* Institute of Labor and Industrial Relations, University of Michigan, Ann Arbor, March 1967.

general adult work relief program but were assured that title V provided sufficient discretion to fill that function. A college work-study program was added to assist needy youth attending college. In addition, the community action program financed neighborhood service centers which employed some poor persons in "new careers" as subprofessional aides.[8]

All these programs were ambivalent between the goal of job creation and that of providing experience and skill leading to more orthodox jobs. Then in 1965, Senator Gaylord Nelson, picking up the 1964 proposal of the Senate Subcommittee on Employment and Manpower, successfully introduced an amendment to the Economic Opportunity Act which had the frank objective of employing the adult, mostly elderly, poor on conservation and beautification projects. This was supplemented in 1966 by Congressman James H. Scheuer's amendment to add employment as subprofessional aides and another cosponsored by Senators Jacob K. Javits and Robert F. Kennedy to provide money for public employment of the poor in areas burdened by heavy concentrations of poverty.

In 1966–67, this combination of acts and programs provided authorization for about 475,000 public service jobs (nearly one-half of them part time, see table 2) for those who would not have been absorbed by more orthodox employment. Considering that training programs had been authorized for only 280,000, the Government had made a substantial start toward becoming "employer of last resort." However, the commitment was only to particular groups under specific conditions and within the limits of budget constraints. There was no open guarantee of employment to any group in any locality. The authors and administrators of the Neighborhood Youth Corps, the Job Corps conservation centers, and the work experience and training program were unclear as to the objectives of their programs and were, therefore, uncertain as to direction and unable to measure success or failure. The remaining programs were new and had no experience to evaluate. Those underway identified numerous problems and provided lessons for advocates of a more generalized job guarantee.

TABLE 2.—*Employment and funding of the job creation programs*

[Dollar amounts in millions]

	Fiscal year 1966		Fiscal year 1967	
	Jobs	Funding	Jobs	Funding
Total	474,000	$557	469,000	$682
Neighborhood Youth Corps:		272		325
In school	106,000	----------	125,000	----------
Out of school	55,000	----------	60,000	----------
Summer	209,000	----------	165,000	----------
Work experience and training	64,000	112	46,000	100
Community action program	40,000	160	40,000	160
Adult work program (Nelson-Scheuer)	(1)	13	25,000	72
Special impact program (Kennedy-Javits)	----------	----------	−8,000	25

[1] Not available.

[8] Arthur Pearl and Frank Riessman, *New Careers for the Poor* (New York, the Free Press, 1965).

The Neighborhood Youth Corps [9]

It is paradoxical that the first work program of significant size to become politically acceptable was one for youth who have the least to gain from such a program. The hope is that work experience will lead to better things. It is possible that a mere holding action, offering employment until age increases employability, may be worthwhile. However, with few family responsibilities and a full working life ahead, NYC must be compared in expenditures and time to the potential gains from training programs offering direct skill improvement.

The NYC has been one of the most popular of the Economic Opportunity Act programs, providing, as it does, almost free labor to public and private nonprofit agencies for in-school, summer, and out-of-school projects. Project sponsors match Federal funds with 10 percent "in kind" facilities and services. The participants, with some exceptions, receive $1.25 an hour for an average of about 12 hours a week in-school, 28 hours out-of-school, and 24 hours in summer programs.

The primary objective of the in-school program is to encourage and enable poor youth to remain in school and in this it appears to be successful. The out-of-school and summer programs are the ones with ambivalent objectives. Extending employment opportunities to the greatest number dictated a "lean" program with a minimum of counseling remedial education, or skill training. This leanness could be expected to reduce the impact on the trainees' future employability. Though administrators resent the term and prefer to think of it, too, as an incentive for school attendance, the summer program appears to be primarily "antiriot insurance" in response to unrest among minority group youth in central city slums.

No useful data are available on the nature of the "work stations" to which the youth are assigned. The limited evidence suggests that it is much easier to provide meaningful work for girls who can be interspersed with other employees in clerical activities. Attractive jobs are more difficult to find for boys and they are more likely to be assigned in work gangs to maintenance, custodial, and cleanup activities.

As a result of high turnover, a large number of youth are served by NYC projects, perhaps as many as 500,000 per year. However, the short stay probably limits any potential training contribution. The one major followup study of NYC graduates suggests that the long-term effects of NYC participation are limited.[10] Duration of stay in NYC appears to have made little difference in the proportion employed, in school, in training, in the Armed Forces, unemployed, or out of the labor force after leaving NYC. The program provides employment and income, though some sponsors do not appear to use the "free labor" as efficiently as they might. The contribution is a positive one. The relevant question is whether or not better alternatives were or can be made available.

---

[9] This section draws heavily on Sar A. Levitan, *Antipoverty Work and Training Efforts: Goals and Reality*, Policy Paper No. 3, the Institute of Labor and Industrial Relations, the University of Michigan, Ann Arbor, Mich.

[10] Dunlap and Associates, "Survey of Terminees From Out-of-School Neighborhood Youth Corps Projects" (Darien, Conn.: Dunlap and Associates), May 1967.

Work experience and training [11]

The work experience and training program established under title V of the Economic Opportunity Act has also suffered from an ambivalence of purpose. The administering agency, the Bureau of Family Services, saw its objective, not as providing skills and giving or finding employment, but as a total family approach to a "better life" including home management, nutrition, and hygiene as well. Participants in the program receive, not "wages" or "training allowances," but only public assistance benefits. The administrators play down the work relief aspects of the program work, preferring that it be viewed as a source of education, training, and work experience leading from dependency into the mainstream of employment. The expenditure patterns give a different impression. Available expenditure data for half of fiscal 1966 are an example: About 72 percent spent for public assistance payments, 13 percent for administration and social services, 1 percent for medical care, 4 percent for child care, 1 percent for adult basic education, and only 2.2 percent for vocational education.

Judging from this expenditure breakdown, title V has been first a source of income maintenance and secondarily of work relief with remedial education and training trailing behind. This pattern is not necessarily undesirable. Public assistance is notoriously capricious in its eligibility rules and niggardly in its payments. Eligibility requirements favor some and discriminate against others for reasons unrelated to need. Seventeen States still refuse help to families where there is an employable adult regardless of the availability of work, and only 22 States have adopted the federally supported program of aid to families where an employable father is present. States estimate the "basic needs" of welfare families, then two out of three provide less than that level of assistance, some as low as 22 percent. Therefore, the major portion of title V funds is used as a supplement to bring public assistance payments up to full "need" and to provide assistance to needy but ineligible persons.

As a rehabilitation program, title V can point to little success. Three out of four have failed to complete the assigned course of training or work experience. For all persons leaving title V, only one-third have found or been placed in jobs. With the average family with an employable parent remaining on relief 9 months, a substantial number of them might have been expected to find employment without the program. Half have continued on public relief after leaving the program.

Title V has been unable to deliver on its promises of rehabilitation. However, viewed as a straightforward work relief program supplemented by basic education and vocational training as funds permitted, many of the criticisms which have brought title V into congressional disrepute lose relevance. There have been other funds and programs for remedial education and training. In fact, more skill training has probably been provided to public assistance recipients under the Manpower Development and Training Act than under title V. Until 1965, the only other federally supported source of public employment for disadvantaged adults was title V's own diminutive parent, community work and training.

---

[11] Sar A. Levitan, *op. cit.*, p. 66.

The role the title V program has been most reluctant to accept has probably been the one of highest potential contribution. The Bureau of Family Services claims an estimated 31.5 million hours of useful work performed in a little over 1 year by title V trainees. The single most criticized title V project has been the so-called "Happy Pappy" project covering a 19-county eastern Kentucky area which has practically no alternative employment opportunities. The project is well designed for the local situation, employing in 1966 a highly immobile group of about 6,000 men while providing them with some basic education. Similar projects have been undertaken in West Virginia and Puerto Rico. Since employment and income are the objectives, these programs have only to assure that some useful work is done to be more desirable than the public assistance alternative.

New careers for the poor

Probably the most imaginative development in the creation of job opportunities for the disadvantaged has been the concept of employing the poor to serve the poor in the community action program. Interpreting the Economic Opportunity Act's directive for "maximum feasible participation" of the poor to mean in staffing, as well as planning, some 40,000 low income persons have been employed full- or part-time as so-called "nonprofessionals" in various community aspects of the poverty program. They are involved as teacher aides in education and day care activities, clerical aides, neighborhood aides who visit homes, bars, and other places to recruit persons eligible for available services, counseling aides and employment aides who assist professionals in counseling, job development, and placement.

No dependable data are available on numbers employed, their performance or the results of efforts to build meaningful careers. Two limited evaluation studies found over two-thirds of those employed were women, about one-fourth worked part-time, most were family heads, nearly three-fourths lived in or near the neighborhoods in which they were employed. The pay for most was only slightly above the poverty line.[12] However, the nonprofessionals were not drawn from "hard core" groups. Only 25 percent had not graduated from high school and 20 percent had some college or were college graduates.

The nonprofessionals were found to have performed well but there was almost universal disappointment with the "career" concept. While a few efforts were made to devise career ladders, the steps were few with apparently impenetrable barriers closing off professional ranks. In most cases the jobs were completely dependent upon the vagaries of community action budgets with no analogous positions available in private or other public employment to allow lateral movement.

Despite these difficulties, the concept of "new careers" has risen to prominence as one of a long line of "solutions" to poverty and the employment problems of disadvantaged groups. The subprofessional career amendment introduced by Congressman James Scheuer, of New York, provided $365 million to create public service jobs for poor adults in fiscal 1967. The guidelines require that, to be funded, projects must open career opportunities leading to permanent jobs.

---

[12] National Committee on the Employment of Youth, *The CAP Aide Study*, New York 1966 and Daniel Yankelovich, Inc., *A Study of the Nonprofessional in the CAP*, Office of Economic Opportunity, Washington, D.C., 1966. See also Frank Riesman, Edith Lynton, Mitchell Ginsberg, Sherman Burr, and Mark Battle, "The New Nonprofessional," *American Child*, vol. 49, No. 1, winter 1967.

Individual projects sponsored by cities, police departments, universities, hospitals, and private nonprofit organizations, are experimenting with nonprofessional aides in teaching, police work, corrections, recreation, mental health, and similar activities. The Elementary and Secondary Education Act and medicare offer promise for a substantial increase in the number of jobs available.

However, the fact that almost all are financed in whole or part by Federal funds is indicative of a number of barriers which have yet to be hurdled. Many remain reluctant to see their activities subdivided to involve subprofessionals despite shortages of personnel. This may change as secure professionals see the advantages of relief from some of their less challenging activities. The more difficult question is, given acceptance of the subprofessional concept, who will get the jobs? It is one thing for a poverty program to employ the poor—even though the best prepared of them appear to get the jobs—it is quite another to expect administrators whose goal is the provision of service to allocate the new jobs to the disadvantaged. Unless special incentives are provided, new employment opportunities will go to the most attractive personnel available.

Other CAP programs

Like the Scheuer amendment, the Nelson amendment to EOA in 1965 and the Kennedy-Javits amendment in 1966 did not share the NYC and title V reluctance to be considered work programs. Despite its small $13 million budget, the Nelson amendment effectively supplied useful jobs and income to 5,700 aged poor and Indians in fiscal 1966. Its fiscal 1967 budget was $36.5 million. The Kennedy-Javits special impact program had not been actually initiated by the end of fiscal 1967, but its $25 million can be used to pay "labor costs for physical rehabilitation of neighborhood facilities and other necessary specific services and costs attendant to improving the employability of the unemployed and effecting their placement in jobs."

REALLOCATING AVAILABLE JOBS

Experimentation has not been limited to creating public employment for those whose skill, education, or social handicaps make competition for available jobs difficult. It is conceptually possible to offset employer reluctance to hire particular categories of workers by subsidizing their employment, by reducing their undesirable qualities, by eliminating alternative sources of labor, or by appeals to social conscience. Experiments with all these tools provide significant lessons.

MDTA on-the-job training in effect subsidizes training by reimbursing employer's training costs. It played a limited role until fiscal 1967 when it grew to nearly one-half of new MDTA authorizations. Unfortunately, the willingness of employers to employ the disadvantaged in return for subsidization to offset additional labor costs was not tested. Under pressure to expand their program, administrators made little effort to assure that those accepted for training were not those the employer would have hired and trained at his own expense. High proportions of current employees were enrolled for upgrading on the assumption that entry level jobs would be opened to the disadvantaged but there were no means of assuring this. Enroll-

ments of the disadvantaged dropped proportionately but once enrolled 90 percent remained employed.

The Chicago JOBS NOW project is one example of many innovative approaches to reducing undesirable personal characteristics that act as a bar to employment.[13] Its trainees are drawn from gang youth and given 2 weeks of orientation in work attitudes before being placed in high support jobs with cooperating employers. A coach is provided to work with each youth and his supervisors on the job and in the street until the employment relationship is stabilized. The success rate in the first 6 months of the project was only 35 percent but, considering the backgrounds and personal limitations of the youthful clients this was considered satisfactory. However, when offered funds to substantially expand the project, its staff demurred believing 125 trainees every 2 weeks was all that Chicago employers were willing to provide jobs for, despite the city's tight labor market.

Cutting off the flow of Mexican and other foreign nationals as farm laborers in 1965 and 1966 did increase the number of jobs available to domestic laborers, but only at a 1-to-4 ratio. The difference was made up by increased mechanization with wages rising substantially in the process.

Presidential appeals to employers to provide summer jobs for youth were answered by letters reporting they had done so. Youth employment was up 920,000 in July 1966 over July 1965 but economic conditions were better and adult employment was up over 2 million. All but 33,000 of the additional youth jobs went to white youth. The results of the 1967 campaign appear similar. Only those jobs provided by NYC funds and Federal hiring were clearly a net addition over what would have occurred in absence of the appeals.

These experiences seem to offer the same conclusion: jobs can be reallocated to the benefit of disadvantaged only when public funds offset the inherent economic obstacles.

The Need for an Employment Guarantee

As experience with manpower programs for the disadvantaged grows, so does the recognition of the primacy of job opportunities. Followup surveys of MDTA completers find one in seven unemployed when last contacted following training. NYC and Job Corps graduates face the same job markets after leaving the programs that they faced before; one-half of the public assistance recipients who enter work experience and training programs return to public assistance when they leave it. Special Labor Department surveys discover ghetto unemployment three times the national average along with extraordinarily high rates of nonparticipation in the labor force and low wage employment, resulting in a "subemployment" rate of 34 percent.[14] A 1966 job order for 3,000 automobile workers attracted 16,000 applicants to the employment service in Detroit. More than twice as many Washington, D.C., youths responded to 1967 publicity than there were jobs available. A "jobmobile" touring Philadelphia slums

[13] Arnold Nemore, "Transferability of Manpower Programs," in U.S. Congress. Senate. Subcommittee on Employment, Manpower, and Poverty, *Examination of the War on Poverty* (Washington: Government Printing Office, August 1967), pp. 197–232.

[14] Manpower Report of the President, 1967, p. 75.

in July 1967 to advertise 500 low paid city jobs resulted in 2,700 applicants, some of whom had stood in line for hours.

Someday, absent the inflationary potential of war, general levels of employment may absorb a higher proportion of the available labor force. Education and training efforts may make a significant contribution to the number who can be employed without unacceptable price level impacts. The political balance may even change to increase the tolerance to inflation. Ways may be found to attract employers to depressed areas and neighborhoods, or to get the unemployed and underemployed out of them. Until then—and the day appears far off—reasonably adequate solutions to the social and personal problems of the employable but competitively disadvantaged will require the Government to act as "employer of last resort." A significant start has been made in this direction but the ambivalence with which work programs have approached their objectives has limited the value of the experience. We now know more about what doesn't work well than what does. We can only guess how many require an employment guarantee but clearly the number exceeds the willingness of the public to foot the bill. Advocacy requires an estimate of the universe of need and an exploration of administrative problems.

The Potential Clientele

Even with the relatively low levels of unemployment which prevailed during 1966, 2.4 million persons were unemployed at least 15 weeks during the year and 840,000 were unemployed more than half the time during the year. An additional 1.2 million looked but did not find work at any time. An average of 2 million persons was working part time but sought full-time jobs. There were 1.8 million men between the ages of 25 and 64 who, though able to work, were neither working nor seeking jobs; 500,000 were between 25 and 49. Nearly three-quarters of a million households were headed by men who were not workers.[15] At least 5 million persons were working at wages below the Federal minimum.

It would be from these persons, and those women and youth now outside the labor force who would prefer to work, that a program to guarantee employment opportunities would draw its clientele. The number taking advantage of such a program would depend upon the pay, location, working conditions, and eligibility rules. Experience upon which to make a judgment is totally lacking, but speculation can provide a budgetary estimate.

Many of those who experienced a total of between 15 and 26 weeks of unemployment at some time during the year may have been reasonably satisfied with their situation. However, a larger proportion of those unemployed continually for between 15 and 26 weeks and most of the 840,000 who were unemployed at least half the year might consider themselves better off in a guaranteed employment program. The one-half million prime working age males, able to work but out of the labor force, may be a reasonable estimate of all of those out of the labor force who would re-enter. The one-half of the involuntary part-time employed who customarily work full time are a reasonable estimate of the potential applicants among that group.

[15] Manpower Report of the President, 1967, p. 132.

Based on 1966 labor force data, the potential clients of an employment guarantee would appear to be the following:

Unemployed more than 26 weeks	840,000
Unemployed continuously, 15–26 weeks	295,000
Unemployed continuously, 15 to 26 weeks	295,000
Males, 25 to 49 years of age, out of labor force but able to work	500,000
Involuntary part time who usually work full time (nonagriculture)	1,183,000
Total	2,818,000

An estimate of nearly 3 million potential applicants is probably conservative, particularly in view of the 5 million low wage workers who have demonstrated their commitment to the larbor force and might apply if wages were significantly higher than their current ones. Some of the part year employed might prefer the public program for the remainder. Thus a larger estimate might be defended but, considering the odds against a program of that magnitude being undertaken in the near future, it is a reasonable outside boundary for program and budgetary consideration. Adding likely overhead costs to annual employment at, for instance, $1.50 an hour would bring the total budgetary cost to $12 billion a year.

LIMITING ELIGIBILITY

Regardless of the merits, it is inconceivable that a nation which is not now willing to appropriate more than $1 billion a year for remedial work and training programs would provide public service jobs to 3 million persons at an annual cost of $12 billion. While the number who would be attracted by such employment opportunities is probably far less, it far exceeds the number for whom appropriations could be wrung by a reluctant administration from a reluctant Congress representing an even more reluctant citizenry. A guarantee of employment opportunities, if it is to avoid the frustrating partialities and unfulfilled promises which have produced cynicism among potential clientele of current programs, must be clearly restricted in its coverage either by demographic characteristics or by location, yet must offer universal opportunity to those eligible.

Eligibility could be limited to the long-term unemployed. However, to use unemployment as a criterion would penalize those in low-wage employment, it would do nothing to alleviate the poverty of the nearly 2 million families with fully employed family heads (95 percent of them male), and might encourage part-time workers and low-wage employees to seek unemployment as the first step to a better paid public service job. Those currently out of the labor market could become eligible simply by registering at the employment service and awaiting the passage of time.

A program for heads of families

Experience with the Neighborhood Youth Corps suggests the possibility of restricting an employment guarantee to family heads. The assumption that the out-of-school, out-of-work youth are eager for steady, low-wage, low-prestige jobs clashes with experience. Adventure and experimentation are more likely to motivate this age group. For some, college is a socially acceptable "aging vat" where they can attend a few classes a week and wander among varied experiences. The middle-

class but noncollege youth has family support to rely upon during his periods of escape from routine. The plight of rural youth is hidden by isolation and underemployment. There is no socially acceptable "aging vat" for the youth of the urban slums, but the motivations are the same—and the latter often have more lucrative sources of income than unattractive work at a minimum wage. The Job Corps has experienced recruitment difficulties and the popularity of the Neighborhood Youth Corps is clearly dependent upon local wage levels and alternative employment opportunities. This youthful reticence about work programs may not be irrational. With limited income needs and a full working life ahead, basic education and skill training or a job with a built-in ladder to the competitive world of work is the need, and this has not been offered by current programs.

About 130,000 of those experiencing over 6 months of unemployment during 1966 were 20 years of age and under as were similar numbers of the involuntarily part-time employed. Opportunities for youth deserve high priority among manpower programs. A public job creation program of current NYC size will undoubtedly continue to be needed for those who cannot or will not be fitted into better alternatives. However, expansion of youth programs should emphasize remedial education and training and wage subsidization or other means of opening access to private employment for disadvantaged youth.

A little over one-half of the long-term unemployed men and one-quarter of the long-term unemployed women in 1966 were married family heads. So were 60 percent of the out-of-the-labor-force males and half of the involuntary part-time workers. Limiting an employment guarantee to family heads would, therefore, be a defensible initial approach. Although many current out-of-the-labor-force female family heads might apply, particularly if child care were provided, any expenditures for them would be largely offset by reductions in current public assistance expenditures. Many of the nearly 2 million poor families headed by a full-time employed person are probably poor as a result of family size rather than wages below the Federal minimum. Thus, the cost of a guarantee to family heads, though expensive, might become politically feasible.

Area saturation

A preferable approach would be an open guarantee to all applicants within specified labor market areas of greatest need. This would allow the program to live within whatever budget was made available, would eliminate the need for any eligibility criteria, and would provide an estimate of the total national need. The main obstacle would be political: the desire of every Congressman and Senator to assure a "fair share" for his constituents. The Department of Labor estimated in 1966 that close to 400,000 unemployed workers lived in the high unemployment and poverty areas of 37 cities. The Michigan State Employment Service at the time of the 1967 Detroit riots estimated that 56,000 were unemployed, 11,000 of them long-term unemployed, while 70,000 represented the "hidden unemployed" who might apply for an employment guarantee. At the same time, all the manpower and antipoverty programs in the city were employing or training a total of only 7,000 persons. Similar concentrations of unemployment and underemployment could be identified in rural areas.

The concentrated employment program which the Labor Department introduced in 1967 in specially selected target slums in 19 cities and two depressed rural areas is a limited start toward area saturation. However, the work and training offerings are limited to the objectives and eligibility rules of the programs from which the funds have been drawn. The average of $5 million allocated to selected slums in hopes of employing or training 2,500 to 4,000 people in each is far less than that required for a significant impact. The Detroit data suggest a $200 million need for a single large city. Frustration will not be reduced by a program which limits its availability to small sections of a city with widespread employment problems nor to only a few of those with equal needs.

The job potential

Conceiving of useful work to be done by participants in a guaranteed employment program is not difficult, but overcoming opposition of interest groups and administering projects might be. There are few, if any, alternatives to publicly sponsored employment for immobile rural adults with inadequate education. Hard physical work still has dignity in these areas and there is plenty of useful work in conservation and related activities to be done. Since the incidence of broken families is less, the clientele would be primarily male. The always deficient education and health systems could, with minimal training, offer outlets for women as well. A guarantee of rural employment would slow outmigration to urban areas, which might ease their current problems.

The situation in the urban slums is much more complex. Housing discrimination and transportation deficiencies limit the access of slum residents. Personal limitations such as inadequate education and training, police records, low motivation, or family burdens tend to block them from the most rapidly growing urban jobs. Vested interests control many of the jobs for which slum residents might qualify. Self-esteem appears to be more threatened by low wages and distasteful tasks than by idleness and dependency.

Remedial education and training, transportation improvements, industrial development efforts, and open housing in the suburbs could reduce the need for publicly supported jobs. Absent those, subprofessionalization in the poverty program, in education and in health offer a demonstrated potential for the most able and motivated. Most of the opportunities are for women but theirs is a high percent of the need, particularly if child care is available. Since the employers are hard-pressed public agencies, professional reluctance can be overcome by availability of funds, political leverage, and assistance in restructuring jobs. Job needs for women will far exceed the supply of potential subprofessional openings. The better prepared males can also work as subprofessional aides of various kinds but the opportunities are limited. Vast opportunities exist in slum rehabilitation but tapping them will require funds, the overcoming of resistance from institutions with a vested interest in such work and the development of methods to utilize low-skilled labor. It is the lower skilled public service jobs which must provide the major source of employment to unemployed adult male residents of urban slums. Local and State governments and more particularly the Federal civil service have been guilty of insisting on unrealistically high eligibility requirements for

low-skilled, nonsensitive jobs. These governments are more likely than private employers to demand high school education and clean police records even when both are irrelevant to the job.

The desirable supplemental approach to guaranteeing job opportunities for urban slum residents, but one which can work anywhere else as well, is the subsidization of private employment. Few, if any, new jobs would be created, but job vacancies might be filled where previously the costs of employing the quality of labor available exceeded the market value of the product or service. MDTA–OJT provides a subsidy, ostensibly equal to training costs. It is a short step from there to a subsidy equal to the added costs of employing the disadvantaged or even equal to the employer's reluctance to hire a particular individual.

A final employment source of almost unlimited potential is the expansion of public employment to those activities which would pay if labor were free, assuming the employment of the idle, the dependent and the lowly paid to be an objective equal to the value of the wages paid. Two mails a day in residential areas has been suggested. The number of unskilled but useful tasks in cleaning, repairing, and refurbishing public buildings, streets, parks, and neighborhoods is without limit, though it may be difficult to avoid the stigma of make work.

Lessons From Experience

That willing people should not be idle while useful tasks remain undone is a truism. However, the experience of the various programs related to job creation for the competitively disadvantaged suggests there are problems in moving from concept to practice.

Some of the difficulties are inherent in the nature of the programs themselves and are separate from their objectives. Current programs reflect crisis-by-crisis legislative reaction. Each has been separately administered and, though consolidation is underway, they have yet to be rationalized in common terms. There is no justification but history for the continuance of a variety of job creation programs with differing rules, funding, and clientele.

Other problems arise from uncertainty and ambivalence among objectives. Job creation is itself an adequate goal for those whose competitive disadvantages make orthodox employment unlikely. For others it is justified only insofar as it provides a step on the road to the competitive job market. Work experience can have training value but only when it is carefully structured and supervised. On the other hand, failure to accept job creation as an end in itself has caused some programs to suffer undeserved criticism.

It is equally necessary to be clear about the objectives of a guaranteed employment program. Income for the poor is not the primary goal though it would be a byproduct. For an estimated $10 billion every one of the 11 million family units with incomes currently below the accepted poverty definitions could be brought to the threshold of those norms, either by a guaranteed income plan or by beefing up the current public assistance system. An economy which can absorb with so little dislocation a war costing over twice that amount could have absorbed an abolition of poverty. Even given the war, a tax increase sufficient to finance such a program would not entail unreasonable sacrifices.

There is doubtless some proportion of the population for which the costs involved in employing them exceed those of income maintenance. Our society has shown itself reluctant, though willing, to support dependent children, the aged and the seriously handicapped, but never adequately and with dignity. Mores are subject to change and there are signs society may eventually come to deal more generously with those dependent groups for whom it presently provides support. There is no indication that it is becoming more willing to accept idleness among able-bodied adults. As long as the laws of scarcity prevail, despite increasing abundance, one man's idleness is another's burden, though an increasingly lighter one. That is not the issue, however. A characteristic of an achieving society is disrespect for the nonachiever, not only in society's eyes but his own and his family's.

Useful jobs can be created commensurate with the abilities of almost anyone simply by the expenditure of public funds to purchase their services. However, there is no assurance that supervisors will use free labor wisely. Making the jobs lead to the mainstream of the job market is a difficult and as yet unsolved task. The redistribution of jobs in favor of the disadvantaged is more difficult. Employers, whether public or private, have a natural preference for the best qualified worker available and no current program has a sufficiently attractive "carrot" to substantially effect recruiting decisions. While alternatives merit continued experimentation, current experience would suggest that, for those for whom remedial education, training and better labor market services are inadequate to make successful job market competition possible, direct public expenditure, either for public employment or subsidized private employment is the only dependable road to jobs. However, maintenance of dignity requires both that the jobs be productive and that all stigma of depency be absent.

The foray into social philosophy suggests three principles which make the design and implementation of programs more difficult: (1) The Government as employer of last resort must be the last resort for the Government as well as the individual; (2) every possible bridge into regular competitive employment must be supplied and its crossing encouraged; and (3) if the guaranteed employment is identified and stigmatized it will have failed its purpose. Since the goal is income with dignity, the criteria should not be family income but employability and the need for a productive role. Willingness to expend funds is only the first step to guaranteeing employment opportunities. More ingenuity and imagination are needed than that required for the relatively simple task of providing a competitive job in the private economy.

Part IV

Health Care and Improvement

88-744—68—vol. 2——9

THE CONTRIBUTION OF HEALTH SERVICES TO ECONOMIC GROWTH AND WELL-BEING

BY HERBERT E. KLARMAN*

This paper aims at a conceptual analysis of the contribution of health services to the economy. The focus is not on the size of the health services-medical care industry, as defined in the national income accounts[21], but on the contribution of health services to economic growth, on the one hand, and to consumer satisfaction, on the other hand.

Economic growth in this context has two meanings. One meaning is the usual one: how much does improved health contribute to an increase in national output in the future? The other meaning is more special: to what extent do expenditures for one set of health services lead to a reduction in expeditures for another set, so that ultimately resources may be transferred from the health services industry to other uses? The contribution of health services to consumer satisfaction is straight-forward in meaning. In the absence of any contribution to growth, as here defined, what are the reductions in pain, discomfort, and bereavment today, in the initial round?

In outlook the paper is limited to the United States and to a relatively short period of time—approximately 30 years backward (to 1935) and 15 years forward (to 1980). In other, developing countries it would be necessary to take account of the effects of improved health both on the size of population[4] and on aspirations or life attitudes ([12], p. 280).

The paper is divided into four sections. In turn, they pose these questions:

1. What estimates exist of the contribution of improved health to economic growth and how valid are they?
2. What relationship is there between more and better health services and improved health status?
3. What are the problems of economic valuation for purposes of economic growth?
4. Looking ahead, are increased expenditures for health services in the United States likely to contribute more to economic growth and to consumer satisfaction?

I. CONTRIBUTION OF REDUCED MORTALITY TO THE GNP

In recent years a number of estimates have appeared that purport to measure the contribution of improved health to national output. Health is usually represented by reduced mortality, and only occa-

*The Johns Hopkins University. Revision of paper prepared for 36th Annual Conference of the Southern Economic Association, November 1966. Financial support from the U.S. Public Health Service is acknowledged under CH002S2, Research Program in the Economics of Health.

sionally by morbidity as well. In effect, these estimates derive from two steps: (1) a population is identified, which has avoided premature death or disability; and (2) its potential economic output is valued.

NUMERICAL ESTIMATES

Mushkin has estimated the contribution to GNP in the year 1960 attributable to the reduction in the death rate before 1960. If the starting point for the decline in mortality is taken as the year 1900, the output in 1960 of survivors (persons who would not be alive in 1960 in the absence of the specified decline in mortality) is $61 billion on a total GNP of $500 billion. If the base line is taken as 1930, the gain attributable to survivors is $16 billion ([42], p. 146). Mushkin does not connect the decline in mortality with any particular set of factors.

Elsewhere I have employed the Mushkin findings as the basis for estimating the contribution to the GNP in 1960 of the decline in the death rate for heart disease. The figgure was $2.6 billion if the base year was 1900 and $1.4 billion if the base year was 1930—4 to 9 percent of the total contributions given above ([26], p. 297).

Arthur Little & Co., at the request of the American Pharmaceutical Association, have prepared similar estimates for four diseases—tuberculosis, syphilis, influenza, and pneumonia. New drugs for treating these diseases became prominent in the 1930's, and the base line chosen is the year 1935, marking the advent of a new pharmaceutical era. The estimated gains in output in the year 1960 due to the cumulative reduction in mortality are $4.9 billion for the four diseases and $10.4 billion for all causes of death. The latter, global figure is reasonably consistent with Mushkin's for base year 1930. The same report also contains a 1-year estimate of the economic value of the reduction between 1943 and 1959 in workdays lost from acute conditions—$2.5 billion ([34], pp. 17, 20).

The American Medical Association's Commission on the Cost of Medical Care published cost-benefit analyses of changes in the care of four diagnostic conditions—polio, cancer of the uterus, surgery for congenital heart defects, and tuberculosis [1]. The data for tuberculosis are taken from the Arthur Little study and are the same. The study of polio is elaborate, with the numbers of survivors and their characteristics calculated with care. The present value of benefits—comprising lifetime earnings of survivors and averted medical care costs—is estimated at $6.7 billion, without discounting for the time of occurrence. Benefits are more than 10 times the total costs of polio programs, which are estimated at $650 million ([1], p. 45). For the other two diagnostic conditions—cancer of the uterus and surgery for congential heart defects—the studies are sketchy and the figures incomplete.

Fuchs has prepared estimates of the present value of the earnings of males whose lives were saved in the single year 1960, owing to the fact that the higher 1929 death rate no longer prevailed ([17], p. 35). However, his figure of $14 billion is not comparable with the others. It is much lower, on the one hand, because survivors are not accumulated between 1929 and 1960. It is higher, on the other hand, because he employs present value estimates, rather than lesser for a single year. For certain policy purposes, however, his type of figure, which repre-

sents the present value of a specified reduction in mortality, may be the most appropriate one.

CRITIQUE

A large, and growing, literature exists today on the economic costs of diseases ([26], [21]; [29]). Increasingly the calculations are performed with refinement of technique and with more reliable and more detailed data. However, most of these estimates do not bear directly on the question of the contribution of health services to the economy.

In addition, there is a considerable body of cost-benefit literature in the health field, cultivated mostly by persons without formal training in economics, that is not likely to withstand scrutiny, though honorable exceptions exist. Many of these estimates display inconsistencies and double counting ([26], [20] and [27]).

With respect to the studies cited in this section, it is evident that those dealing with total mortality must not be interpreted to attribute the entire reduction in mortality, and gain in economic output, to the provision of more and better health services. Other factors may play a part.

The studies of individual diseases warrant further comment.

One, they lack uniformity in technique and, more serious, in the basic economic data.

Two, it is not so difficult a task today as formerly to rework each of these studies with the assistance of Rice's data and tables [48, 49].

Three, such a reworking on a large scale would be justified only if the underlying data were also reviewed. It seems to me mistaken to employ deviations in mortality from a fixed base year, rather than from points where marked discontinuities in trend are displayed. It may be misleading to concentrate exclusively on mortality (see below).

Four, this approach may be capable of extension to more diseases. The resulting data would be useful in themselves, and together the findings might serve as an approximate check on the more general approach, that of total mortality. The sum of the parts is bound to exceed the total, however, owing to the simultaneous presence in some patients of multiple, complicating diseases.

II. Relationship Between Health Services and Health Status

To begin with, there is the well-known, and still unsolved, difficulty of identifying or defining health status. Definitions abound, and there is no agreement. They range from the highly positive, almost utopian, view of the World Health Organization that health means optimum physical, mental, and social efficiency and well-being to the essentially negative view of Roberts that health is the individual's capacity to resist disease and death ([50], p. 21).

Increasing attention is being given to developing indexes of health status. The National Health Survey collects a good deal of information on morbidity and disability ([33], [58]). Yet intensive efforts at quantification, health levels are still frequently represented by mortality rates. The reasons are practical—the measure is objective and data are readily available, comparable over time and place. Moreover, as Fuchs emphasizes, considerable variation in death rates still exist in this country ([18], p. 75).

In broad terms it is possible to identify several schools of thought on the determinants of health status.

Some hold that environment is all. Encompassed by environment may be the genetic inheritance and the ability to adapt, as well as the level of living ([8], pp. 59, 167–69). Ginzberg frequently points to the high level of civilian health in wartime when income was high, despite the diversion of large numbers of physicians to the military forces ([19], p. 737).

McKeown has adopted a historical position, as a result of his studies of population growth ([37], [38]). He concludes that the chief factors in reduced mortality were diet in the 18th and early 19th centuries, the sanitary movement—pure water, sewage disposal, and personal hygiene—in the late 19th and early 20th centuries, and health services beginning about 1925 ([36], p. 39). No specific evidence is presented for the last phase, however.

Another position that appears to be emerging is that the attitudes and motivations of the individual are central in dealing with the chronic diseases.

FINDINGS BY FUCHS AND ASSOCIATES

Among economists Fuchs has addressed himself systematically to this question in recent years ([17], [18]). Specifically, he asks: what are the relationships between mortality (reflecting health status) and the level of living (represented by income) and medical services and technology?

It is not at all necessary to dwell on the fact that urban life would be impossible without clean water and waste disposal. This tells us that the sanitary services have a high average value but says nothing about the contribution to health of additional quantities of the same services.

Fuchs arrives at the position that the contribution of health services to health status may be appreciable. He appeals to two sets of data, one from the United States and one international.

The domestic data are as follows. The age-adjusted death rate in the United States has remained stable for a decade. Yet cross sectional multiple regression analyses across 48 States in the years 1940, 1950, and 1960 show a positive relation between mortality and income for the white population. Only infant mortality shows the expected, inverse relationship ([17], pp. 14–18). His interpretation of these findings is that environmental forces exert a negative effect on adult mortality. The death-rate has been stable because health services have provided an offset.

The international comparison employs data from six countries. Age-adjusted death rates in five western European countries with a per capita income in 1960 equal to that of the United States in 1925 are close to the death rate in the United States in 1960. The point here is that the medical technology of these countries is similar to that of the United States in 1960. From this comparison Fuchs concludes that changes in medical technology have been the principal cause of the decline in the United States death rate between 1925 and 1960 ([18], pp. 82–83).

The widespread presumption that an increase in per capita income has favorable implications for health ([1], p. IX; [56], p. 5) is, there-

fore, questionable (in part, the association goes the other way; see below), except for infant mortality. It may be that the United States has passed the peak with respect to the favorable effect of a rising level of living on health status ([18], p. 81). A rise in the unfavorable elements may outweigh the favorable ones—such as pollution of the air, nuclear fallout, obesity, sedentary occupations, tension, alcohol, narcotics, smoking, and automobile accidents ([19], p. 735; [23], p. 68).

Coronary heart disease is generally considered a concomitant of modern living. Lerner and Anderson summarize existing knowledge, "Coronary heart disease * * * flourishes where general levels of living are highest and diets richest * * *. Mortality rates * * * are highest where urbanization has progressed the furthest, where the largest proportions of the population are engaged in sedentary, white-collar business and professional occupations, and where * * * psychological stress may be the most intense * * * ([32], p. 65)."

Fuchs falls back on a residual explanation. Although education is a factor favorable to health, the largest share of the credit for improvement in health levels in the past 25 years should go to improvements in medical services and technology—new drugs, knowledge, diagnostic tools, etc. ([18], p. 81).

Colleagues of Fuchs at the National Bureau of Economic Research followed up on his work by developing and estimating a clearly specified multiple equation model that takes into account the interrelationship among the supply and demand factors producing health services, the process by which health services are produced, and the relationship of health status both to health services and to those environmental factors that affect health [3]. Following Fuchs, health is measured negatively by the death rate.

The analysis employs statewide data. The principal numerical findings are: health services are associated with lower death rates and environmental factors, except for education, with higher death rates. Each set of factors has a fairly low elasticity (relative responsiveness) with respect to a change in the death rate—0.2. To insure that the effect of income as an environmental factor has been truly captured, separate calculations are made for each race and with the omission of education. Again, the adverse effect of income on the death rate is confirmed.

Finally one test of the model and the findings is highly successful in explaining the stability of the age adjusted death rate in the United States between 1955 and 1963. Another test tries to account for the difference between white and nonwhite death rates, and two-thirds of the difference in death rates falls on the difference in the use of medical specialists. This explanation is implausible.

CRITIQUE AND APPRAISAL

As for Fuchs' international comparison, it seems to me that if the factual situation in the year 1960 had also prevailed in 1925 or earlier, the power of the medical technology explanation would be undermined. A test is, therefore, indicated.

Fuchs' findings and interpretations have been questioned on several grounds.

It has been observed that although the correlation between male mortality and occupational status is positive for a few diseases, such

as coronary heart disease and diabetes in England ([32], pp. 61–62, 89), the correlation is negative for most disease entities [20]. Overall the higher occupational levels have the lower mortality rates ([41], p. 66).

The methodology of areal correlations has a basic weakness. This method receives widespread application in the health field on a local level, where the basic units are generally census tracts. The implicit assumption is that the average characteristics of an area apply to its individual residents. This general assumption is dubious, and becomes increasingly so as the size of the unit increases to that of the State, thereby permitting greater heterogeneity among the area's population. In discussing Fuchs' paper, Stockwell observed, "The fact that a State with a low median income also has a low death rate is not a sufficient basis for concluding that individuals in that State with low incomes are exposed to a smaller risk of mortality than individuals with high incomes [57]." It is important to investigate health relationships on an individual level.

In general, it is recognized in the statistical literature that individual relationships inferred from areal correlations may be seriously biased as to magnitude and even erroneous as to sign. The reason is that the component of variation within the area is lost. Under specified favorable circumstances it is, however, possible to estimate individual relationships from areal data ([9], p. 9).

Another technical argument pertains to possible colinearity among the independent variables in any multiple regression linear model, whether the observations be areal or individual. The following example may serve as an analogy. Several years ago an economist for the first time applied multiple correlation to data on the use of hospital services, and found that age was not a significant factor ([51], p. 36). In this case it is likely that age was confounded with income ([25], p. 138). It took principal components analysis before the author was able to retrieve the well-known positive relationship between hospital use and old age ([52], pp. 89–90).

It should be noted in qualification of this particular criticism of Fuchs' findings that its factual basis is perhaps not so strong as it might be. Since the data on deaths and those on occupation derive from different sources—the former from death certificates and the latter from census returns—they do not always jibe, and the death rates may be distorted ([41], pp. 61–63). A study to develop such data from a single sources has been undertaken, but the findings have not yet been published ([2], p. 60).

It is recognized that poor health and disability also lead to a decline in income ([18], p. 83; [56], p. 10). In studies of health status or health services expenditures, economic status should not be represented by actual income, which is affected by the illness of the breadwinner, but by family expenditures or some notion of normal or permanent income ([15], pp. 97–98; [25], pp. 27–28). For long-term studies, occupational status may be the superior indicator of economic status. In fairness to the critics of Fuchs, however, their mortality data relate to occupational class, not to income.

Following a method different from Fuchs', Lerner and Anderson have found that the relationship between mortality and income varies with age. (They do not segregate by sex.) The age class studied most

intensively by Fuchs, 45 to 54 years, is close to the borderline of a positive correlation in the Anderson-Lerner study, with the younger age classes showing an inverse correlation between mortality and income and the aged showing a positive one ([32], pp. 112, 126). Moriyama and Guralnick report some overlapping in the death rate among four of the five occupational levels at ages 45 to 64 ([41], p. 67).

Moreover, it appears that the reported differentials in death rates by occupation (and by income, in the Lerner-Anderson data) may have narrowed over time ([2] p. 65; [32], p. 125; [56], p. 4). If so, the years studied would make a difference. Fuchs' findings would gain credence if his method were applied to an earlier era when medical technology was not yet so developed, say 1900 or 1910, and found to give opposite results.

Another criticism of the Fuchs thesis is that for many diseases, such as rheumatic fever, scarlet fever, diptheria, and tuberculosis, the long-run trends in the incidence of the disease, severity, and mortality have been continuous, steadily downward. They appear to display no abrupt change when new technology and practices were introduced [18a].

This point of view runs counter to the assumptions and findings of the studies of individual diseases listed in section I. Nor is it consistent with the data presented by Fuchs and some of the time series plotted by Lerner and Anderson ([18], p. 78; [32], pp. 152, 184). It may well be, however, that for chronic diseases it is unreasonable to expect a discontinuity in mortality rates. Thus, in tuberculosis the new drugs introduced in the early 1950's made a more direct impression on hospital use than on the death rate. Similarly rheumatic fever is reflected in the death rate through rheumatic heart disease, which occurs years later. For these reasons, as well as the fact that changes in medical practice tend toward uniformity in local areas and would display sharper discontinuities, if any, the latter may furnish better sites for study than the Nation as a whole.

A major obstacle to studying local areas is the lack of data on the size and characteristics of the population for intracensal years. However, for short intervals, changes in numbers are just as useful as changes in rates, thereby obviating the problem of getting corresponding denominators.

The above discussion does suggest one conclusion: there is need for a good deal of research in this area.

In addition, it is evident that frequently the underlying technological data required by the economist are lacking or suspect ([18], pp. 77–80). It is erroneous to make economic calculations on the assumption that a disease will be totally eradicated ([29], p. 1949) and only somewhat less misleading to proceed as if specified percentage reductions in the size of a disease problem can be achieved ([24], p. 29). Rather, it is necessary to try to ascertain with some specificity what can be accomplished through the provision of health services and what opportunities confront medical research and development.

By intent, if not in practice, the economist bases his calculations on specific events, such as death, disability, employment, episodes of illness, and use of health services. Each event has a place on the calendar and some have additional dimensions, such as duration and intensity. Ideally such data should extend as close as possible to the observed person's full lifetime. Although a person cannot die twice, he can be ill or disabled from the same disease more than once.

Such information is frequently lacking, however. Elsewhere I have listed some reasons for this. Briefly, physicians may disagree on both diagnosis and evaluation, because medicine is not an exact science; the best data would come from longitudinal studies, which are not common in this country; the presence of asymptomatic disease may impair the conclusions drawn from control groups; the possibility of disease induced by medical care itself leads to a requirement for field studies of normal populations, which are far more costly than clinical studies of hospitalized patients; and populations vary in the simultaneous presence and detection of multiple diseases ([26], pp. 698–99).

Heretofore, economists have not really specified the types of data they require. Epidemologists have enough interesting work of their own in ascertaining the distribution of diseases in populations, and are not likely to assume the initiative in developing data useful for the economist in the health field.

The best solution is for the two professions to work together in formulating this type of problem, to join in getting their hands dirty, and to revise their questions and approaches as they carry on the work.

A brief comment is in order on medical research. From the standpoint of health—neglecting the economic value of the pursuit of curiosity for its own sake, of the cultivation of sheer knowledge—the value of research, as well as of development, derives from the value of the health services rendered. Thus the expected value of a research program is the expected value of the health problem toward which it is directed times the probability of a discovery capable of reducing the problem by x percent times the probability that the discovery will be applied, that is, the services will be offered and used ([28], pp. 152–53). It is important to take account of the timing of the discovery and the timing of its application. Presumably the probabilities can incorporate allowances for the uncertainty of success that pervades all research ([39], p. 138). It is not evident how to handle the related problem of uncertainty concerning the field in which the discovery, if achieved, will be applied.

III. Economic Valuation of Output

It is generally accepted today that data on the number of deaths do not convey a great deal about economic value. Even life-years gained, which take account of age at death, may fail to do so. An extreme example of disparity between life-years gained and economic value is the estimate that the instantaneous elimination of heart disease as a cause of death would raise life expectancy for males of age 30 by 12.4 years but would extend work life expectancy by less than 1 year ([26], p. 702; [64], p. 588).

Better health can contribute to the national output by improving the quality of labor and by increasing the quantity of labor.

QUALITY OF LABOR

What the literature says about the quality of labor does not appear to be immediately useful. Three diverse sets of comment may be gleaned concerning absenteeism, debility, and life attitudes.

Absenteeism.—In his basic work Weisbrod has a brief discussion of absenteeism from work associated with illness. He concludes that short-

term and fluctuating types of absenteeism have the most adverse effects on labor productivity ([61], p. 41).

The Jewkes have a more extended discussion in which they compare absenteeism rates due to sickness between the United States and England and between civilians and the military. Absenteeism from military duty is presumably most influenced by health considerations, and has shown a decline over the years. This is not true of the civilian rates, which have been constant in both countries but consistently higher in England. They note an inverse correlation between absenteeism rates and unemployment and emphasize the importance of social insurance provisions ([24], pp. 10–15).

Debility.—Only Mushkin has said anything concrete about the possible gains in productivity due to a decline in debility. After observing that the distinction from disability may not be clear cut, she proposes several ways to measure debility. These may be broadly classified as involving comparisons of observed performances in production, the conduct of controlled experiments, and laboratory tests of work capacity ([42], pp. 139, 142).

Fuchs reports few findings from a thorough search of the literature. In one study of occupational health programs only 12 percent of the executives associated with the responding companies thought that improved productivity on the job was a benefit from such programs ([18], p. 90).

Life attitudes.—The literature is full of references to possible, favorable shifts in attitudes toward most things in life as a result of health programs ([11], p. 180; [12], p. 280; [18], p. 92). These would constitute appreciable external, or spillover, benefits from such programs. So far there is not a great deal of evidence in support of this hypothesis, which nevertheless may be valid.

CONTRIBUTION OF INCREMENT IN SUPPLY OF LABOR

With respect to the quantity of labor, a large literature already exists on how the economist measures gains in output due to reductions in mortality or morbidity. A good deal of the literature on cost-benefit analysis in the health field appears to be readily adaptable to the measurement of the contribution of health services to growth.

The procedure is straightforward. One begins with the number of persons whose death or disability will have been averted. How such numbers are, or may be, obtained was discussed in sections I and II above. The total numbers are distribtued by as many demographic and socioeconomic characteristics as is practicable and are likely to be useful, from the standpoint of the availability of economic data. In recent years it has become practicable to move from breakdowns by sex and perhaps also by age to breakdowns by race and educational level, as well ([47],[48],[49]).

The economist then proceeds according to the following steps.

He applies a labor force participation rate. This is usually the prevailing set of rates by sex and by age or the set that has prevailed under conditions of full employment. It may be desirable, looking ahead, to posit probable changes in these rates, particularly by sex ([16], p. 19).

He assumes full employment (or 3 to 4 percent unemployment). This is believed to be necessary, in order to isolate the contribution to

output of health; unemployment is then attributed to the failure of fiscal and monetary policies. The appropriateness of the assumption of full employment has been questioned, however, for areas with persistent underemployment ([45], p. 441). Nor is is always a realistic assumption for many disabled persons who have been rehabilitated and cannot find jobs ([19], p. 736). It may be necessary to spend effort and money to validate this assumption in reality. Perhaps it would be useful to distinguish between cure and prevention programs, as exemplified by the measurement of "stigma" associated with having had venereal disease ([27], p. 395.).

The economist employs earnings plus wage supplements as the measure of labor's ouput. He abjures per capita income, except in the event of wholesale catastrophe, on the ground that the total product, including income from property, must not be attributed to a single factor of production, labor. He excludes transfer payments whether income taxes or public assistance grants, in order to avoid double counting. It may be necessary, however, to consider transfer payments when dealing with accepted social goals other than maximum national output, such as a belief in the desirability of a universal opportunity to work, or with the realities of intergovernmental fiscal relations ([27], pp. 377–378).

The proper measures of average earnings is the arithmetic mean, not the median of the distribution, which is lower. In the past it was necessary to employ the median because it was the only figure published. Today mean earnings are available ([48], 49).

Increasingly, the custom is to allow for probable changes in productivity when estimating future streams of earnings ([27], p. 374). A possible refinement is to adjust expected changes in productivity by age class ([42], p. 148).

Nevertheless, there are some differences between measurement procedures for cost-benefit analysis and for economic growth.

Measurement for Growth

It has been argued that the above procedure measures the contribution of improved health to the national income and not to total economic well-being, by neglecting leisure. By choosing additional leisure, the people have shown their true preferences. Moreover, leisure-time activity in nonprofit and voluntary organizations enhance the social welfare ([63], pp. 50–51).

This point is apart from the question of how to treat the intangible components of better health, such as avoidance of pain, discomfort, and bereavement. Such health benefits are still largely intractable to measurement; quite properly they must be pursued in cost-benefit analysis and are disregarded for purposes of economic growth. The literature on measuring these benefits is growing ([14], pp. 5–6; [26], pp. 700–701; [29], p. 1951; [42], pp. 155–56; [53]; [59]; [63], p. 54).

The services of housewives are excluded from the national income and would be disregarded in measuring economic growth. In cost-benefit analysis the services of housewives must be included because the distribution of disease by sex is not uniform. To exclude them would be to bias expenditure decisions for health services in favor of men ([25], p. 170).

For cost-benefit analysis the treatment of consumption by beneficiaries still remans a point of controversy, although more and more economists tend to disregard it. In developing nations, where essential consumption is of consequence, a different treatment may be warranted. It is clear that in measuring economic growth consumption must be disregarded; to deduct it would be to understate total output ([25], pp. 168–69).

Pervading all public policy decisions that are implemented over time is the rate of discount to be employed. Benefits and costs accrue over time, and this is especially true of benefits that represent economic growth. Discounting makes disparate streams commensurate by converting them to a single measure, the present value. The level of the discount rate has a strong effect on the size of the present value if the time period considered is long, with a high rate reducing the present value and conversely.

Although economists agree on the need for discounting, they do not agree on the proper level of the discount rate. Some economists would employ the market rate of interest, after adjustment for taxes and degrees of uncertainty. Others doubt that this is the appropriate rate. In general, the discount rate is meant to balance the opportunity cost of capital in production and time preference (preference for consumption now). Private and collective estimates of time preferences are likely to diverge, because the individual has a shorter time horizon than society. Moreover, time preference may vary by social class, with the upper classes more willing to postpone gratification, and also by country, depending on life expectancy.

Accordingly, it is customary to prepare and present alternative estimates of the present value of costs and benefits, specifying two two or more discount rates. Another reason advanced for the dual rate procedure is that the discount rate represents a judgment on the relative importance of successive generations, concerning which differences of opinion are likely ([10], pp. 453–463; [22], pp. 115–123; [46], pp. 686, 697–700).

The rate that appears in the literature most frequently is 6 percent, which is also intermediate in the range of rates that have been employed.

For purposes of economic growth, as well as for cost-benefit analysis, the proper criterion for choosing among programs is the maximum *difference* between the present values of benefits and costs, not the highest ratio ([14], p. 2; [22], pp. 137–138; [35], pp. 107–113). The largest difference in amount will make the greatest percentage contribution to the rate of growth, to the extent that tangible economic benefits are embodied in the the numbers.

IV. The Probable Direction of Health Service Expenditures

It is worth distinguishing between questions 2 and 4. Question 2 dealt with the contribution of health services to health status. It was concluded on the basis of the existing evidence that their relative importance may be rising. Clearly a great deal more work remains to improve models and to obtain requisite numbers.

Question 4 aims to assess the probable direction of change in the distribution of the health services dollar between economic growth and consumer satisfaction.

It should be emphasized that a reduction in expenditures for health services is by itself tantamount to an increase in the national output, because in the long run resources freed from the health services industry would be available for other uses. However, a reduction from expenditures of $50 billion in calendar year 1967 is not in sight. Health services expenditures in the United States amount to 6 percent of GNP, having risen from 3.6 percent in 1929 and 4 percent in 1940 [40]. The lowest projection reported for 1975 is 7 to 7.5 percent ([44], p. 30). Projections of 9 and 10 percent have appeared ([31], pp. 20, 27; [55], p. 203).

Let us assume that the future level of health services expenditures in this country is known or given. Is the predominant impact of the increase likely to be on consumption or on economic growth? (Frequently the latter type of expenditure is termed investment. This is confusing usage because strictly speaking investment applies to any economic good that renders services beyond one accounting year.)

TWO SCHOOLS OF THOUGHT

Most programs of health services yield bundles of services that have both consumption and growth aspects which are difficult to disentangle. Viewed broadly, there are two schools of thought concerning the probable effects of additional health services in the future.

Among others, Weisbrod has but recently urged recognition of the fact that a substantial—though unquantified—fraction of all health expenditures are "investment" in longevity and labor productivity. He envisions health expenditures for prevention as doubly valuable, on the ground that certain health resources are freed for other uses ([62], p. 7).

Denison has dealt with this thesis explicitly. He emphasizes that no evidence has been presented in behalf of health services as a potential source of economic growth. He proceeds to present some data and to analyze them.

First Denison calculates the effects of a one-tenth reduction in the death rate of all persons under 65, and finds that it would add 0.20 percentage point to the annual growth rate in the period 1960–80. He concludes, "* * * The calculation indicates that efforts to reduce the death rate do not offer a promising means of stimulating economic growth very much over the next 20 years ([7], p. 48)."

Next he estimates the value of a reduction by one-fourth in time lost on the job due to sickness and accidents. The effect would be to raise the annual growth rate by 0.05 percentage point. If this gain were costless, another 0.06 percentage point of the national output might be freed for other uses. However, if expenditures for additional resources were required in order to realize the gains, these would constitute an offset to the increase in the growth rate, which should be deducted ([7], p. 51).

To avoid misunderstanding, it might be emphasized that the above argument is oriented toward the future; that is, at the prospect of further gains in output owing to improvement in health status. It is not meant to deny that a good deal of medical care rendered today is productive, by reducing disability and restoring young and middle-aged persons to the labor force. The implicit assumption of this argument is that much of the potential gain has already been realized.

Denison adds that some costless ways to improve health are well known. To eat, drink, and smoke less and to drive more carefully would only require shifts in individual preferences and attitudes ([7], p. 52).

Conley concludes that vocational rehabilitation programs are not likely to enhance the economic growth rate. One reason is that clients who have completed the program successfully find jobs and earnings hard to come by. Another is that this group has low incomes with a below average propensity to save ([6], pp. 137, 97).

Fuchs calculates the economic value or reducing the 1960 death rate of American males to the Swedish level at $7.5 billion, but does not indicate whether such a reduction is attainable or how it can be attained ([17], p. 11).

ASSESSMENT

My own view is that there does exist a tendency to overstate the tangible economic benefits of health services programs. The sources of overstatement are several. One has been noted, the assumption of full employment for all beneficiaries.

Another source is the failure—or inability—to take account of the simultaneous presence of multiple diseases. If every person with a disease had only one, it would be appropriate to add the benefits of reducing the effects of disease A and of disease B. When a person has two or more diseases at the same time, reducing or even eliminating one still leaves him with one or more diseases, which may also contribute to the continuance of economic losses. If the treatment for two diseases is the same, the problem posed here is removed. When treatment for the diseases simultaneously present differs and prognoses vary, failure to consider the second disease results in an overstatement of potential economic benefits. The risk of overstatement is greater when diseases are interdependent in origin, as they are in families with multiple problems ([27], pp. 375–76).

A third source of overstatement is the counting of all institutional expenditures, including hotel-type costs, in the cost of medical care. In the case of long-term illness the patient no longer maintains a home and the potential benefit of a cure is overstated to that extent ([25], p. 170).

In sum, too little attention is paid to the characteristics of the probable beneficiaries from the services and to a realistic appraisal of the opportunities for realizing reduction in mortality or morbidity.

I believe that, at least for the United States, the above sources of overstatement of tangible economic benefits of health services probably overwhelm any tendencies in the opposite direction, such as the failure to take account of all spillover effects of programs, the complementary relationship between health and education ([43], pp. 413–14; [63], p. 37), and the favorable shift in life attitudes that may result from the presence and awareness of health programs.

The views that follow on the probable users of services and on opportunities for realizing gains in mortality and morbidity are necessarily speculative at this time, and, therefore, sketchy.

To begin with, a trivial point must be made: housewives will not be excluded from new health services programs. Notwithstanding, their services are not reflected in the national income.

More important, society will not refuse additional services to the aged, the mentally retarded, and the permanently disabled. The effect of additional health services expenditures on economic growth will be smaller in amount than the improvement in health status.

Within the aged population, the more elderly (75 years and over) are increasing faster than the younger aged (65 to 74 years). The former use a good many more health services; one study puts their per capita use of institutional services at 2.75 times the amount used by the younger aged ([25], p. 29).

Some of the treatments yielded by medical research are quite costly. One example is hemodialysis for persons with chronic kidney failure who cannot live without this treatment or its alternative and supplement, kidney transplantation. The unique lifesaving feature of this treatment, combined with the public sponsorship and financing of its development, means that it cannot be denied to persons who cannot afford the $14,000 a year it costs at a hospital center or the $5,000 it costs at home ([5], p. 150).

Marked differences in health status exist between certain population groups and the rest of the population. The poor and the Negroes have not only the highest communicable disease rates but also the highest maternal and infant mortality rates and the highest rate of rejection under selective services ([11], pp. 173–74). Indeed, for the Negro the lag in health status may be worsening ([13], pp. 819–23). However, the relative efficacy of health services in reducing such differences is not known [54]. What is known, however, is that it requires repeated personal participation to use such services. To induce such use may take major efforts in attitudes beyond anything that has been attempted in the past ([7], p. 52; [23], p. 66).

In certain instances the opportunity for promoting economic growth through health are smaller today than formerly. The cost of syphilis, as well as of tuberculosis, has declined substantially and so have the potential benefits, compared to the costs, of a further reduction in incidence. In certain chronic disease, such as arthritis, alleviation of pain and greater comfort are the expected payoff, rather than cure and more productive employment.

Nor is it certain that even effective preventive health services will lead to a saving in the use of health resources. When preventive measures entail more frequent contact with a physician, the learning experience may lead to a heightened appreciation and cultivation of such services and also to the discovery of more illness ([25], p. 29–30).

Preventive services are not necessarily effective. We may not yet know what to do with respect to a given disease; the incidence of the disease may be so low as to induce a very high proportion of false positives in a screening examination, even when the test is quite reliable ([60] pp. 328–29); or the natural history of the disease may still be obscure, so that the effect of medical intervention is unknown ([30], p. 307).

Implications

Better health permits an increase in output for the economy; the converse holds true for poorer health. Being healthy is also a good thing in itself.

The potential increase in output can take place through improvements in the quality of labor and through an increase in the size of the labor force, about which more is known. Realizing the potential depends on the application of sound fiscal management policies.

It is widely accepted that a higher national income is conducive to better health. A higher income clearly permits the purchase of more goods and services, including health services. The extent to which expenditures for health services increase with income is a question of the elasticity of demand ([15], pp. 94–95). The extent to which an increase in such services enhances health status is a question of the particular diagnostic condition and circumstances.

Another way in which a higher income is believed to contribute to better health is through an improvement in the standard of living, including nutrition, housing, education, et cetera. This proposition is challenged by Fuchs, who has found a direct correlation between income and adult health status, expressed through the death rate. Although Fuch's findings cannot be said to be firmly established at this time, they may be more valid for age classes above 45 or 55 than for the younger ages, and point to intensified research efforts in this area.

To the extent that Fuchs' findings are validated, they signify that a wealthier economy, which can afford more health services than a poorer one, also requires more health services in order to achieve or maintain a given health status.

Whether more health services—and better health—mean greater output or merely more consumer satisfaction is again a question of concrete circumstances. My own view is that increasingly in the United States and for the foreseeable future increases in health services expenditures are much more likely to yield consumer satisfaction than economic growth. This does not deny that a large fraction of existing expenditures result in an enlarged labor force. Nor is it necessarily poor policy for an affluent society to choose to devote increasing amounts of resources to consumer well-being, rather than to further increases in the national output.

There is evidently much room for research on the consumption or growth effects of additional health expenditures, since the views expressed in this paper are largely speculative, without a firm foundation in fact or in systematic inquiry.

One is also led to conclude that a great deal remains to be learned about the relative importance of various factors that affect health status. Studies of individual diseases and diagnostic conditions are useful, as well as attempts at more general inquiries into the effects of health services on health status. Important questions are also outstanding concerning the appropriateness of various statistical techniques for studying these relationships.

REFERENCES

[1] American Medical Association Commission on The Cost of Medical Care, *Significant Medical Advances, Vol. 3 of Report*, Chicago, American Medical Association, 1964.

[2] Aaron Antonavsky, "Social Class, Life Expectancy and Overall Mortality," *The Milbank Memorial Fund Quarterly*, vol. 45,, No. 2 (April 1967, part 1), pp. 31–73).

[3] Richard Auster, Irving Leveson, and Deborah Sarachek, *The Production of Health, an Exploratory Study*, New York, National Bureau of Economic Research, 1966 (processed).

[4] Robin Barlow, "The Economic Effects of Malaria Eradication," *American Economic Review, Papers and Proceedings*, vol. 57, No. 2 (May 1967), pp. 130–148.

[5] Committee on Chronic Kidney Disease, *Report*, Washington, D.C., Bureau of the Budget, 1967.

[6] Ronald W. Conley, *The Economics of Vocational Rehabilitation*, Baltimore, the Johns Hopkins Press, 1965.

[7] Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us*, New York, Committee for Economic Development, 1962.

[8] Rene Dubos, *Mirage of Health*, Garden City, N.Y., Doubleday & Co., 1959.

[9] Otis D. Duncan, Ray P. Cuzzart, and Beverly Duncan, *Statistical Geography*, Glencoe, Ill., Free Press, 1961.

[10] Otto Eckstein, "A Survey of the Theory of Public Expenditure Criteria," in James M. Buchanan, ed., *Public Finances: Needs, Sources and Utilization*, Princeton, N.J., Princeton University Press, 1961, pp. 439–494.

[11] Robert L. Eichhorn and Edward G. Ludwig, "Poverty and Health," in Hanna H. Meissner, ed., *Poverty in the Affluent Society*, New York, Harper & Row, 1966, pp. 172–180.

[12] Rashi Fein, "Health Programs and Economic Development," in Selma J. Mushkin, ed., *The Economics of Health and Medical Care*, Ann Arbor, Mich., Bureau of Public Health Economies, the University of Michigan, 1964.

[13] Rashi Fein, "An Economic and Social Profile of the Negro American" *Daedalus*, vol. 94, No. 4 (Fall 1965), pp. 815–846.

[14] Martin S. Feldstein, *Measuring the Costs and Benefits of Health Services*, Copenhagen, World Health Organization, 1966.

[15] Paul J. Feldstein and W. John Carr, "The Effect of Income on Medical Care Spending," American Statistical Association, *1964 Social Statistics Proceedings*, Washington, D.C., The Association, 1965, pp. 93–105.

[16] Hugh Folk and Donald E. Yett, *Methods of Estimating Occupational Attrition*, St. Louis, Mo., Department of Economics, Washington University, 1965.

[17] Victor R. Fuchs, *Some Economic Aspects of Mortality in the United States*, New York, National Bureau of Economic Research, 1965 (processed).

[18] Victor R. Fuchs, "The Contribution of Health Services to the American Economy," *The Milbank Memorial Fund Quarterly*, vol. 44, No. 4 (October 1966, pt. 2), pp. 65–101.

[18a] "Discussion of Paper by Victor R. Fuchs," *The Milbank Memorial Fund Quarterly*, vol. 44, No. 4 (October 1966, pt. 2), pp. 101–103.

[19] Eli Ginzberg, "Health, Medicine, and Economic Welfare," *Journal of the Mount Sinai Hospital*, vol. 19, No. 6 (March–April 1953), pp. 734–743.

[20] Lillian Guralnick, *Socioeconomic Differences in Mortality by Cause of Death: United States, 1950, and England and Wales, 1949–53*, paper prepared for the International Union for the Scientific Study of Population, 1963, Washington, D.C., the U.S. Public Health Service (processed).

[21] Ruth S. Hanft, "National Health Expenditures, 1960–65," *Social Security Bulletin*, vol. 30, No. 2 (February 1967), pp. 3–13.

[22] Jack Hirshleifer, James C. De Haven, and Jerome W. Milliman, *Water Supply: Economics, Technology, and Policy*, Chicago, University of Chicago Press, 1960.

[23] George James, "Health Services and the Community's Needs," *Medical Care*, vol. 5, No. 2 (March–April 1967), pp. 62–69.

[24] John Jewkes and Sylvia Jewkes, *Value for Money in Medicine*, Oxford, Basil Blackwell, 1963.

[25] Herbert E. Klarman, *The Economics of Health*, New York, Columbia University Press, 1965.

[26] Herbert E. Klarman, "Socioeconomic Impact of Heart Disease," in Second National Conference on Cardiovascular Disease, *The Heart and Circulation*, vol. 2, Washington, D.C., Federation of American Societies for Experimental Biology, 1965, pp. 693–707.

[27] Herbert E. Klarman, "Syphilis Control Programs," in Robert Dorfman, ed., *Measuring Benefits of Government Investments*, Washington, D.C., The Brookings Institution, 1965, pp. 367–410.

[28] Herbert E. Klarman, "Discussion of Papers on Economics of Health," *American Economic Review, Papers and Proceedings*, vol. 57, No. 2, (May 1967), pp. 151–55.

[29] Herbert E. Klarman, "Present Status of Cost-Benefit Analysis in the Health Field." *American Journal of Public Health*, vol. 57, No. 11 (November 1967), pp. 1948–1953.

[30] E. G. Knox, "Cervical Cytology: A Scrutiny of the Evidence," in Gordon McLachlan, ed., *Problems and Progress in Medical Care*, London, Oxford University Press, 1966, pp. 279–309.

[31] Leonard A. Lecht, *The Dollar Cost of Our National Goals*, Washington, D.C., National Planning Association, 1965.

[32] Monroe Lerner and Odin W. Anderson, *Health Progress in the United States, 1900–1960;* Chicago, University of Chicago Press, 1963.

[33] Forrest E. Linder, "The Health of the American People," *Scientific American*, vol. 214, No. 6 (June 1966), pp. 21–29.

[34] Arthur D. Little, *A Report on the Social and Economic Benefits of the Pharmaceutical Industry in the United States*, Washington, D.C., Pharmaceutical Manufacturers Association, 1961.

[35] Ronald McKean, *Efficiency in Government Through Systems Analysis*, New York, John Wiley, 1958.

[36] Thomas McKeown, "Medicine and World Population," in Mindel C. Sheps and Jeanne Clare Ridley, eds., *Public Health and Population Change*, Pittsburgh, Pa., University of Pittsburgh Press, 1966, pp. 25–40.

[37] Thomas McKeown and R. G. Brown, "Medical Evidence Related to English Population Changes in the Eighteenth Century," *Population Studies*, vol. 9, No. 2 (November 1955), pp. 119–41.

[38] Thomas McKeown and R. G. Record, "Reasons for the Decline of Mortality in England and Wales during the Nineteenth Century," *Population Studies*, vol. 16, No. 1 (July 1962), pp. 94–122.

[39] Benton F. Massell and Richard R. Nelson, "Research and Economic Growth: The Role of Public Policy," in Edmund S. Phelps, ed., *The Goal of Economic Growth*, New York, W.W. Norton & Co.., 1962, pp 129–40.

[40] Ida C. Merriam, "Social Welfare Expenditures, 1965–66," reprinted from *Social Security Bulletin*, vol. 29, No. 12, (December 1966), pp. 3–15.

[41] Iwao M. Moriyama and Lillian Guralnick, "Occupational and Social Class Differences in Mortality," in *Trends and Differentials in Mortality*, New York, Milbank Memorial Fund, 1956, pp. 61–73.

[42] Selma J. Mushkin, "Health As An Investment," *Journal of Political Economy*, vol. 70, No. 5 (October 1962, pt. 2). pp. 129–57.

[43] Selma J. Mushkin, "Comments on Herbert E. Klarman's Syphilis Control Programs," in Robert Dorfman, ed., *Measuring Benefits of Government Investments*, Washington, D.C., The Brookings Institution, 1965, pp. 410–14.

[44] National Commission on Community Health Services, *Financing Community Health Services and Facilities*, Washington, D.C., Public Affairs Press, 1967.

[45] Mark Perlman, "On Health and Economic Development: Some Problems, Methods, and Conclusions Reviewed in a Perusal of the Literature," *Comparative Studies in Society and History*, vol. 8, No. 4 (July 1966), 433–48.

[46] A. R. Prest and R. Turvey, "Cost-Benefit Analysis: A Survey," *The Economic Journal*, vol. 75, No. 300 (December 1965), pp. 683–735.

[47] Dorothy P. Rice, *Economic Costs of Cardiovascular Diseases and Cancer, 1962*, Washington, D.C., Government Printing Office, 1965.

[48] Dorothy P. Rice, *Estimating the Cost of Illness*, Washington, D.C., Government Printing Office, 1966.

[49] Dorothy P. Rice, "The Economic Value of Human Life," *American Journal of Public Health*, vol. 57, No. 11 (November 1967), pp. 1954–66.

[50] Ffrangcon Roberts, *The Cost of Health*, London, Turnstile Press, 1952.

[51] Gerald D. Rosenthal, *The Demand for General Hospital Facilities*, Chicago, American Hospital Association, 1964.

[52] Gerald D. Rosenthal, "The Demand for General Hospital Facilities: A Principal Components Analysis," American Statistical Association, *1964 Social Statistics Proceedings*, Washington, D.C., 1965, pp. 84–92.

[53] Thomas C. Schelling, *The Life You Save May Be Your Own*, Washington, D.C., The Brookings Institution, 1966 (processed.)

[54] Samual Schwartz and John H. Vinyard, "Prenatal Care and Prematurity," *Public Health Reports*, vol. 80, No. 3 (March 1965), pp. 237–48.

[55] Anne R. Somers, "Some Basic Determinants of Medical Care and Health Policy: An Overview of Trends and Issues," *Health Services Research*, vol. 1, No. 2 (fall 1966), pp. 193–209.

[56] Mortimer Spiegelman, *Significant Mortality and Morbidity Trends in the United States Since 1900*. Bryn Mawr, Pa., The American College of Life Underwriters, 1966.

[57] Edward G. Stockwell, *Discussion of Papers on Social and Economic Differentials in Mortality*, Annual Meeting of the Population Association of America, 1966.

[58] John Storck, "Assessing the Country's Health in a Time of Change," *Public Health Reports*, vol. 81, No. 9 (September 1966), pp. 821–828.

[59] Jacques Thedie and Claude Abraham, "Economic Aspect of Road Accidents," *Traffic Engineering and Control*, vol. 2, No. 10, (February 1961), pp. 589–595.

[60] W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach*, Glencoe, Ill, The Free Press, 1956.

[61] Burton A. Weisbrod, *Economics of Public Health*, Philadelphia, University of Pennsylvania Press, 1961.

[62] Burton A. Weisbrod, "Investing in Human Capital," *The Journal of Human Resources*, vol. 1, No. 1 (summer 1966), pp. 5–21.

[63] Bernard M. Wolf, *The Economics of Medical Research and Medical Care: From the Point of View of Economic Growth*, Bethesda, Md., Resources Analysis Branch, National Institutes of Health, 1964 (processed).

[64] Barnes Woodhall and Seymour Jablon, "Prospects for Further Increase in Average Longevity," *Geriatrics*, vol. 12, No. 10 (October 1957), pp. 586–591.

THE DIRECT AND INDIRECT COST OF ILLNESS

BY DOROTHY P. RICE*

Public and private expenditures for health and medical care reached $47.3 billion in fiscal year 1967, representing more than a 12-fold increase since 1929.[1] During this same period, the Nation's gross national product (GNP) increased 7-fold. The ratio of health and medical care expenditures to the GNP increased from 3.6 in 1929 to 6.2 percent in 1967 (table 1, p. 484).

These expenditures for health and medical care are designated as "direct" expenditures and do not measure the full economic costs imposed upon the Nation by illness, disability, and premature death since they do not include the loss of output to the economy. These are the indirect costs, which are more difficult to measure but are important in assessing the burden of poor health on society.

This paper summarizes the direct and indirect costs of illness, disability and death by major diagnostic groups for a base year, 1963, including a brief discussion of some of the economic concepts employed.[2] Also included is a brief description of the application of these data to assess the benefits of a federally supported screening program to demonstrate control of cancer of the uterus.[3]

ANNUAL DIRECT COSTS OF ILLNESSES

Direct costs of illnesses comprise the expenditures for prevention, detection, treatment, rehabilitation, research, training, and capital investment in medical facilities. In terms of services or type of medical expenditure, direct costs include amounts spent for hospital and nursing home care, physicians' and other medical professional services, drugs, medical supplies, research, training, and other nonpersonal services.

For a decade or so the Social Security Administration, U.S. Department of Health, Education, and Welfare, has prepared estimates of public and private expenditures for health and medical care. Data are available on expenditures for health services under specified public programs and by object of expenditures or type of health service,

---

*Chief, Health Insurance Research Branch, Division of Health Insurance Studies, Office of Research and Statistics, Social Security Administration.

The author gratefully acknowledges the assistance of Barbara S. Cooper in the development of this report.

[1] U.S. Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, "Public and Private Expenditures for Health and Medical Care, Fiscal Years 1929–67," *Research and Statistics Note,* No. 21, Nov. 20, 1967.

[2] For a full description of the distribution by diagnosis of the direct and indirect cost of illness, disability and death in 1963 and the methodology used, see: Rice, Dorothy P. "Estimating the Cost of Illness." *Health Economics Series,* No. 6. U.S. Department of Health, Education, and Welfare. Public Health Publication No. 947–6. May 1966.

[3] For a full description of the benefits and cost of programs for control of cancers of five specified sites through early detection, see: U.S. Department of Health, Education, and Welfare. Office of the Assistant Secretary for Program Coordination. "Disease Control Programs—Cancer, October 1966." *Program Analysis 1966–3.*

that is, hospital care, physicians' services, drugs, medical research, construction, and so on.[4]

These annual expenditures for the various health services may also be distributed by type of illness or diagnosis. The estimated distribution of 1963 health expenditures for specified health services is presented here for eight diagnostic groups, categorized according to the International Classification of Diseases, Adapted (ICDA). The remaining 10 major diagnostic groups identified in the ICDA listing are included in "all other." The allocation of funds by disease is limited in this paper to four categories of expenditures: hospital care, nursing home care, services of physicians, and services of other health professionals. Expenditures for these services amounted to $22.5 billion, approximately two-thirds of the total outlay ($34.3 billion) for health and medical care in 1963.[5] The remaining third includes a variety of personal and nonpersonal expenditures which could not be allocated by type of illness or diagnosis.

Of the total $22.5 billion, expenditures for diseases of the digestive system (including dental conditions) ranked highest, totaling $4.2 billion, or 18.5 percent of the total (table 2). Expenditures of $2.4 billion for mental disorders ranked second, with 10.7 percent of the total. Hospital and nursing home care and professional medical services for patients with diseases of the circulatory system amounted to $2.3 billion, the third highest expenditure category. These three diagnostic groups—diseases of the digestive system, mental disorders, and diseases of the circulatory system—comprised almost two-fifths of the total expenditures.

Annual Indirect Costs of Illnesses

Direct expenditures, as shown above, do not measure the full economic costs imposed upon the Nation by illness, disability, and premature death since they do not include the loss of output to the economy. The estimated distribution of 1963 losses in output resulting from mortality and morbidity are presented for eight major diagnostic groups. The basic method for calculating the economic value of lost output was to apply prevailing average earnings to the productive time lost by sex and age groups for each major cause of death and major type of illness. Distribution by disease was based on the variety of sources including Vital Statistics, National Health Survey, Social Security Disability Allowances, and special studies.

The calculation of the annual loss of output resulting from mortality and morbidity involved the application of several principles of economics on which there is some consensus and which are summarized below.[6]

Economic Principles Employed

1. EARNINGS

The appropriate measure of output loss for individuals is the year-round, full-time earnings, which include wages and salaries before

[4] Reed, Louis S. and Rice, Dorothy P. "National Health Expenditures: Object of Expenditures and Source of Funds, 1962." *Social Security Bulletin*, August 1966, pp. 11–21.

[5] U.S. Department of Health, Education, and Welfare, Social Security Administration, Division of Research and Statistics. "National Expenditures for Health Care Purposes by Object of Expenditure and Source of Funds, 1960–63." *Research and Statistics Note*, No. 10, May 1, 1967, p. 1.

[6] Klarman, Herbert E. *The Economics of Health*. New York: Columbia University Press, 1965. pp. 162–173.

deductions. The proper measure of expected earnings is the arithmetic average or mean (not the median, which is frequently employed because it is available in published form). Adjustments were made to take cognizance of wage supplements—employer contributions for social insurance, private pensions, and welfare funds.

2. EMPLOYMENT

Not everybody would have worked or been productive if death or illness had not interfered. Some are too old, some too young, some unwilling to work, and some unable to find a job. The estimate of losses in output assumes that if it were not for these illnesses or causes of death, persons stricken would have had the same employment experience as persons in the same age and sex groups. Labor force participation rates (proportion of all civilians who are employed or looking for a job) were applied and further adjustments were made for the number who would have been employed under conditions of relatively high employment, defined as 4-percent unemployment. Without the assumption of high employment, losses due to mortality and disability cannot be isolated from losses due to unemployment.

3. HOUSEWIVES' SERVICES

Housewives' services are estimated at the average earnings of a domestic worker, which has been subject to criticism in some quarters.[7] Although the economic contributions of housewives are not included in the national income accounts, omitting the value of their services in calculations of indirect costs distorts comparisons of costs of illnesses striking primarily one sex. In 1963, earnings of a domestic worker amounted to $2,670.[8]

4. TRANSFER PAYMENTS AND TAXES

Transfer payments, such as pension and relief payments, are not included here among the costs of disease. Double or triple counting would be involved if a relief payment were added to the estimated loss of earnings by an individual. Further, the relief payments may also be used for medical care treatment, already accounted for in the direct costs. Similarly for taxpayments, it would be double counting to add income tax losses to loss of earnings, and triple counting if the tax receipts were used for public payments for medical care.

5. MEASUREMENT OF INTANGIBLES

Intangible or psychic costs of disease, such as pain and grief, are omitted. These costs do not directly involve a loss of output and are not readily measurable. Several economists feel that ignoring the intangibles may distort the overall economic and social costs because the

[7] Kissick, William L., M.D. "Planning, Programing, and Budgeting in Health," *Medical Care*, July–August 1967, vol. V, No. 4, p. 213.
[8] Rice, D. "Estimating the Cost of Illness." *op. cit.* p. 14.

implicit assumption is that the economic value of intangible losses is zero.[9]

Findings—Indirect Costs

1. MORTALITY LOSSES

Approximately 1.8 million persons died in 1963 from all causes. The losses in output resulting from these deaths are equivalent to more than 600,000 man-years, or a total value of $2.7 billion (table 3). As expected, losses for those who died from circulatory diseases rank highest; losses in output for those who died from all forms of neoplasms rank second; and diseases of the nervous system and sense organs, including stroke, are the third highest causes of deaths and resulting losses in output. Together, these three major diagnostic groups account for 71 percent of the deaths and a slightly higher proportion (74 percent) of the total losses in output.

2. MORBIDITY LOSSES—INSTITUTIONALIZED POPULATION

Approximately 1.5 million persons were in institutions in 1963. Their losses are estimated at approximately $5.1 billion, reflecting the value of output of 1.1 million man-years. Of the $5.1 billion institutional losses, $3.6 billion, or 71 percent, are attributed to mental disorders, reflecting the outpt losses for the 720,000 persons in mental hospitals who are classified as suffering from mental, psychoneurotic, and personality disorders.

3. MORBIDITY LOSSES—NONINSTITUTIONAL POPULATION

Morbidity losses for the noninstitutional population were estimated for the following population groups: (1) the currently employed, (2) those unable to work because of chronic illness or disability, and (3) women unable to keep house because of illness or disability. (Days lost from school are omitted as they are not considered losses to productivity.) For all the noninstitutional population, total morbidity losses amount to 3.8 million man-years, of which 2.9 million were productive years lost, representing a total of $15.9 billion lost to the economy in 1963.

The productivity losses for the currently employed group, as measured by work-loss days, accounted for approximately three-fifths of the total; in terms of the total years lost, the currently employed group comprises 44 percent of the total. Those unable to work represent almost half (47 percent) of the total years lost for the noninstitutional population, but many in this group are too old to work and would not be productive if they had been well (table 4).

---

[9] Klarman, Herbert E. "Syphilis Control Programs." In: *Measuring Benefits of Government Investments*, Ed., Robert Dorfman. The Brookings Institution, Washington, D.C., 1965. pp. 367–410.

U.S. Department of Health, Education, and Welfare, Public Health Service. "Economics of Health." *Proceedings of Conference on the Role of Economics and the Economists in Solving the Problems in Health Services.* Dr. Richard Musgrave, chairman, Jan. 29, 1965. pp. 42–57.

Kissick, "Planning, Programing, and Budgeting in Health." *op. cit.* p. 213.

Gorham, William. "Allocating Federal Resources Among Competing Social Needs," *Indicators.* Department of Health, Education, and Welfare. August 1966, pp. 9–10.

4. MORTALITY AND MORBIDITY LOSSES

Losses in output and indirect costs of mortality and morbidity in 1963 are summarized according to age and diagnosis in table 5. A total of 6.2 million man-years were lost in 1963 due to death or illness. Of this total, three-fourths (4.6 million man-years) were productive years lost, valued at $23.8 billion. Combining mortality and morbidity losses shows that mental disorders account for approximately one-fifth of the total costs, followed by circulatory diseases, representing 17 percent of the total. Respiratory illnesses ranked third. These three major diagnostic groups accounted for approximately half of the value of losses in output for those who died or were ill and disabled in 1963.

The distribution by age of mortality and morbidity losses shows some interesting relationships: of the total 6.2 million man-years lost in 1963 due to death, illness, and disability, those age 45 to 64 accounted for the highest proportion—36 percent of the total. The man-years lost for those age 65 and over ranked second, with 31 percent of the total. The distribution by age of productivity losses shows a different pattern: losses were also highest for those age 45 to 64, accounting for a larger share of the productivity losses—39 percent of the productive man-years lost and 45 percent of the indirect costs. As expected, productivity losses for the age group 65 and over represent a considerably smaller proportion of the total, reflecting the relatively lower productivity of this age group. These data are summarized in table 6.

5. LOSSES TO GNP

To measure the losses to GNP due to death, illness, and disability in 1 year, the value of output imputed to housewives must be excluded as the economic contributions of housewives are not included in the national income accounts. In 1963, a total of 1.1 million man-years lost are attributed to housewives who died or were too ill to keep house. The imputed value of their losses in output is estimated at $2.9 billion. Almost one-fourth of the productive man-years lost and one-eighth of the indirect costs of mortality and morbidity in 1963 are attributed to women prevented from keeping house as shown in table 7.

The value of the losses in output amounted to $20.9 billion in 1963 for the potentially productive members of the labor force (excluding housewives). On this basis, the GNP in 1963 would have been increased approximately 3.6 percent, if illness, disability, and death from all diseases had not interfered.

6. ANNUAL ECONOMIC COSTS

The annual economic costs—the sum of the indirect expenditures for medical care and indirect costs of mortality and morbidity—are summarized for each of the eight diagnostic categories in table 8. The total economic costs of illness, disability, and death are estimated at $58 billion in 1963, comprised of the following:

(*a*) $34.4 billion spent for medical care, services, and supplies. Of this total, $22.5 billion or two-thirds, were distributed among the major diagnostic groups. The remaining $11.7 billion includes expenditures for a variety of personal and nonpersonal services, which could not be allocated to specific diseases.

(*b*) $23.8 billion lost to the economy in 1963 due to premature death, illness, and disability from all diseases. Mortality losses account for 11 percent of the total, and the remaining 89 percent are morbidity losses.

Of the $46.3 billion total economic costs distributed among the major diagnostic groups, $7 billion, or 15 percent, represent the costs of mental, psychoneurotic, and personality disorders reflecting the high direct and indirect institutional costs for the mentally ill. Cost of diseases of the circulatory system rank second, totaling $6.4 billion. Costs of diseases of the digestive and respiratory systems rank third and fourth, respectively. Together, the above four major diagnostic groups accounted for $23.8 billion, representing more than half the total economic costs of illness, disability, and death from all causes.

Table 9 summarizes the economic costs in 1963 of illness, disability, and death, with and without the imputed value of output for housewives.

Economic Cost of Mortality

The direct and indirect costs of morbidity and mortality for the major diagnostic groups have been presented for a single year, 1963. From the economist's point of view, however, single-year cost estimates represent only a portion of the estimated losses in output resulting from illness, disability, and death, thereby seriously underestimating the economic costs to society. If an individual had not died in this year, he would have continued to be productive for a number of years. If he is ill and disabled this year and his disability continued into future years, his future productivity will be affected. It is the present value of these future losses that constitutes the appropriate measure of the cost of a disease.

For mortality, the estimated cost or value to society of all deaths is the product of the number of deaths and the expected value of an individual's future earnings with sex and age taken into account. This method of derivation must consider life expectancy for different age and sex groups, changing pattern of earnings at successive ages, varying labor force participation rates, imputed value for housewives' services, and the appropriate discount rate to convert a stream of costs or benefits into its present worth.

This section outlines the underlying concepts involved in the calculation of lifetime earnings, and presents, according to the major causes of death, the estimates of present value of productivity losses discounted at 4 percent for all those who died in 1963.[10] No attempt was made to estimate losses associated with future disability.

The estimating procedures for the development of the lifetime earnings as presented here are described in detail in the study by the author, "Estimating the Cost of Illness." The following is a brief review of the basic assumptions and economic concepts employed:

1. LIFE EXPECTANCY

The lifetime earnings data were developed on the assumption that each cohort will follow his or her pattern of life expectancy as re-

[10] For lifetime earnings discounted at a 4-percent rate and presented by age, sex, color, and education, see Rice, Dorothy P. and Cooper, Barbara S. "The Economic Value of Human Life," *American Journal of Public Health and the Nation's Health,* 57:1954-1966, 11 (November) 1967.

ported for 1963 at successive ages by the National Center for Health Statistics.

2. LABOR FORCE PARTICIPATION

The estimate of lifetime earnings takes into account varying labor force participation rates at different ages. The assumption is that an individual will be in the labor force and productive during his expected lifetime in accordance with the current pattern of labor force participation for his sex.

3. CROSS-SECTIONAL EARNINGS

Following the procedure used in estimating single-year output losses, the same adjusted estimated mean annual earnings in 1963 of male and female full-time, year-round paid civilian workers are also used here. In the application of these cross-section survey data to the estimates of lifetime earnings, the assumption is that the future pattern of earnings for an average individual will remain the same as that reported for the base year, 1963. This model recognizes that the average individual may expect his own earnings to rise as he ages and gains experience in accordance with the cross-section survey data for 1963.

4. HOUSEWIVES' SERVICES

Very few researchers have developed lifetime earnings data for women owing to the difficulties associated with measuring their earnings when they move in and out of the labor market.[11] To omit the value of services of housewives, however, would seriously underestimate the value of the lifetime productive contribution of females.

As for the single year estimates of losses in output, housewives' services are estimated in this report at the average earnings rate of a domestic worker—$2,670 in 1963. This imputed value is clearly on the low side for it makes no allowance for the housewife's longer workweek or the size of the household cared for.

5. DISCOUNTING

The value of money decreases with time. A dollar in earnings in the future is worth less than a dollar at present because of the productivity of capital and cost of postponing consumption. In order to examine the present monetary value of man in today's terms, his future expected earnings must be converted to their worth today. Banks and Kotz state that—

> A given sum is normally worth more today than an equal sum at some future date, because the money (or resources) can be profitably invested (or consumed) in the interval between today and the future. Interest is the premium paid to reflect the fact that any given sum or resources could be put to profitable uses over a period of time. * * * It follows

[11] See, for example:
Bridgman, D. S. "Problems in Estimating the Monetary Value of College Education." *Review of Economics and Statistics.* XVII:180–184, 3, Supp. (August) 1960. pt. 2.
Fein, Rashi. *Economics of Mental Illness.* New York: Basic Books, 1958.
Hansen, W. Lee. "Total and Private Rates of Return to Investment in Schooling." *Journal of Political Economy.* LXXI:128–140, 2, (April) 1963.

> that the value of money which is not currently available, but which will become available (or spent) some years hence must be discounted for the interest which could be earned in the interim, which is why the present value of a dollar to be received in the future is always less than 100 cents.[12]

For the valuation of public health or recreational programs, where the budgets are measured in terms of dollar values of expected lifetime earnings, the aggregate earnings must be discounted at an appropriate rate of interest to determine their value today.

Although there is general agreement among economists that discounting should be employed, there is no agreement on which discount rate, that is, rate of interest, to use. Yet the selection of the discount rate is most important since its effect is considerable. The higher the discount rate, the lower the present value of future earnings. With a high rate of discount, earnings far in the future yield only a small present value. Conversely, lowering the discount rate increases the present value of these earnings far in the future.

The selection of an appropriate discount rate in the valuation of public programs is especially important as the use of too low a rate for discounting future benefits may lead to uneconomic adoption of projects. Banks and Kotz deplore the lack of a central governmental authority to prescribe the discount rate to be used despite its importance in arriving at public investment decisions.

The author has prepared data on the present value of lifetime earnings employing four discount rates: 2, 4, 6, and 8 percent. Only the data based on the 4-percent discount rate are presented here; the data for the other three may be obtained from the author.

6. PRODUCTIVITY INCREASES

While future earnings must be discounted to reflect lost interest, average annual earnings must be increased to reflect gains in productivity. It is an understatement of lifetime earnings to assume that a person 10 years from now will earn the same amount as a person of the same age and sex, earns today. In order to adjust for the gain in productivity, an average annual gain can be projected and applied to the annual earnings. This rate of increase may be incorporated into the discounting calculations to obtain a net effective discount rate. For example, assuming a rise in productivity of 3 percent a year,[13] a discount rate of approximately 7 percent will be reduced to a rate of approximately 4 percent (1.07/1.03=1.039), the rate used in this report.

7. ALLOWANCE FOR CONSUMPTION

There is a diversity of opinion regarding the treatment of consumption. Insurance companies treat consumption as a deduction from a person's contribution to output. Dublin and Lotka and Weisbrod deduct consumption from total output in their calculations of the earning losses.[14] Miller's recent report presents lifetime earnings for men

---

[12] Banks, Robert L., and Kotz, Arnold. "The Program Budget and Interest Rate for Public Investment." *Public Administration Review.* XXXVI:283–292, 4, (December) 1966.

[13] Klarman, *The Economics of Health. op. cit.*, p. 168.

[14] Dublin, L. I., and Lotka, A. J. *The Money Value of Man.* The Ronald Press Co., New York, 1946, p. 70.
Weisbrod, Burton A. *Economics of Public Health.* University of Pennsylvania Press, Philadelphia, 1960, p. 49.

aged 18–64 by educational level and occupation, with deductions for personal maintenance shown separately.[15]

Fein and others on the other hand, make no such adjustment. Fein summarizes his views as follows:

> Certainly the net figure (gross value less consumption) derived by Dublin and Lotka to indicate the money value of a man to his family is correct for their purposes. It is not at all apparent, however, that the net concept is the correct one when we deal with the economic value of a man to society. It is true that man consumes partly in order to maintain himself, and in this sense some of his consumption may be considered as a gross investment to take care of depreciation; it is also true, however, that consumption is an end in itself and can be viewed as a final, rather than an intermediate, step in the creation of other products. The question involved concerns the purposes for which an economy exists.

On an individual's income,

> The individual enjoys life, and it is for this purpose that the social economy exists.[16]

In accordance with the above viewpoint and because we are measuring the economic value of man to society and not to his family, no allowance for consumption is made.

Findings

1. Discounted lifetime earnings

Table 10 shows the lifetime earnings for males and females in each age group, discounted at 4 percent. For a male infant under age 1, the expected lifetime earnings discounted at 4 percent are $59,063.

The present value of male lifetime earnings reaches a peak of $128,698 in the 25–29 age group and decreases steadily to $210 for those aged 85 and over.

The level of expected lifetime earnings for females is somewhat lower than that for males up to the age group 60–64. An infant female can be expected to earn $34,622 in her lifetime, or almost three-fifths the amount for males in the same age group. The highest expected female lifetime earnings, $67,960, are for the 20–24 age group. Peak female earnings are only about half as large as those for the male.

Beginning with the age group 60–64, female earnings are higher than male earnings. The higher expected earnings for females in the older age groups are due to relatively small number of males in the labor force and the larger number of females keeping house.

2. Application of earnings to 1963 deaths by diagnosis

Applying these expected lifetime earnings by age and sex to the 1.8 million deaths in 1963 results in a loss of nearly $50 billion to the economy at the 4-percent discount rate. These deaths represent a loss of 32.5 million years. Total years lost are estimated by multiplying

---

[15] Miller, Herman P. and Hornseth, Richard A. *Present Value of Estimated Lifetime Earnings.* U.S. Department of Commerce, Bureau of the Census, Technical Paper 16, 1967.

[16] Fein. *Economics of Mental Health. op. cit.,* pp. 18–19.

the expected number of years remaining (according to the 1963 Life Tables) to persons at the midyear of each age group by the number of deaths in each age and sex group.

Table 11 distributes the losses according to cause of death. Circulatory diseases account for the largest number of deaths in 1963—782,000 or 43 percent of the total. Neoplasms rank second, accounting for 16 percent. Diseases of the nervous system and injuries rank third and fourth, respectively. Together, these four major diagnostic groups account for almost four-fifths of total deaths (78 percent).

Total years lost show somewhat different relationships. Although circulatory diseases and neoplasms still rank first and second, their proportions are smaller, 28 and 15 percent, respectively. Diseases of early infancy (not shown separately) move to third place (14 percent) followed closely by injuries (13 percent). Diseases of the nervous system drops to fifth place (8 percent) in terms of the total years lost. Their diminished proportion of total years lost compared with deaths is due to the fact that about 78 percent of the deaths from this disease are among the aged who have a relatively short life expectancy.

When discounted lifetime earnings are applied, the picture changes again. Circulatory diseases and neoplasms again rank first and second, respectively. Injuries follow closely behind neoplasms and diseases of the nervous system rank fourth.

Table 12 presents the distribution by broad age groups of lifetime earnings for the major diagnostic categories, showing the effect of age on this measure of loss. Considerable variation is found among the diagnoses. Injuries account for the largest share of the lost earnings for the two younger age groups—under age 25 and ages 25–44. For the two older age groups diseases of the circulatory system account for the largest share.

3. TOTAL ECONOMIC COST OF MORBIDITY AND MORTALITY

When the annual direct expenditures for illnesses are combined with the annual morbidity and lifetime mortality losses, the total economic cost in 1963 amounts to more than $105 billion, distributed as follows:

(*a*) $34.3 billion spent for medical care, services, and supplies, of which $22.5 billion were distributed by diagnosis;

(*b*) $21.0 billion lost to the economy due to morbidity in 1963;

(*c*) $49.9 billion is the present value of lost output for those who died in 1963.

The distribution by diagnosis of the total economic cost of $93.5 billion shows the greatest losses are for those with diseases of the circulatory system, which account for more than one-fifth of the total. Injuries are second in rank, accounting for approximately one-eighth of the total costs. Neoplasms and diseases of the digestive system rank third and fourth respectively. Together, these four major diagnostic groups account for more than half of the total cost (table 13).

Application of Data to Cost-Benefit Analyses

A conceptual framework and systematic methodology for measuring the economic costs of all causes of deaths and illness, as presented here, is of considerable interest to public health planners, providing a quantitative basis for evaluating programs to prevent or eradicate

disease. Much interest has been generated in cost-benefit analysis—the measurement of the economic benefits derived from a program against the costs of engaging or investing in that program.[17] In the cost-benefit analysis applied to the health field, the total cost of the disease serves as the measure of benefits derived from preventing or eradicating the disease. However, a realistic assessment of the status of programs of disease control and prevention points to the fact that, except for a few infectious diseases, such as smallpox and diphtheria, control programs are likely to affect only a partial reduction in mortality or morbidity. Therefore, the appropriate measurement for allocating scarce resources among disease categories is the calculation of the value of the expected benefit of the partial reduction in mortality or morbidity and comparison with the estimated cost of accomplishing that reduction.[18]

To illustrate the application of cost-benefit analysis to a specific disease control program, this paper summarizes the results of a study made by the Cancer Control Branch of the Public Health Service (in which the author participated) to assess the potential direct benefits of a screening program to demonstrate control of cancer of the uterus.[19] The study summarized here was part of a recent effort by the Department of Health, Education, and Welfare to conduct analytical studies on selected segments of the Department's programs. Included were cost-effectiveness studies of motor vehicle injury programs, arthritis control programs, and maternal and child health programs.[20]

Economic Benefits of Uterine Cancer Control

The program that was analyzed consisted of projects to be conducted by hospitals, clinics, health agencies, and professional societies, with PHS grants support. The local objectives of these projects include the following purposes:

(*a*) Finding and curing early stage cases of cancer of the uterus (principally in situ cancer of the uterine cervix);

(*b*) Installing the cystological examination for this purpose as a routine examination in hospitals and outpatient clinics;

(*c*) Demonstrating methods of educating women who seldom visit physicians or hospitals to obtain periodic examinations;

(*d*) Studying and demonstrating improved examination methods, and the epidemiology of the disease; and

(*e*) Training cytotechnologists so that the work force can grow with the expanding use of cytological examination.

---

[17] Dorfman, Robert (Ed.). *Measuring Benefits of Government Investments.* Brookings Institution, Washington, D.C. 1965.
Prest, A. P., and Turvey, R. "Cost-Benefit Analysis: A Survey." *Economic Journal.* 75:683–735 (December) 1965.
Crystal, Royal A., and Brewster, Agnes W. "Cost-Benefit and Cost-Effectiveness Analyses in the Health Field: An Introduction." *Inquiry.* III:3–13, 4 (December) 1966.
Klarman, Herbert E. "Present Status of Cost-Benefit Analysis in the Health Field." *American Journal of Public Health and the Nation's Health.* 57:1948–1953, 11, (November) 1967.

[18] Klarman, Herbert E. "Letter to the Editor." *American Journal of Public Health and the Nation's Health.* 56:566, 4, (April) 1966.

[19] "Disease Control Programs-Cancer, October 1966" *op. cit.*

[20] See:
U.S. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Program Coordination. Disease Control Programs:
"Motor Vehicle Injury Prevention Program, August 1966." *Program Analysis, 1966–1.*
"Arthritis, September 1966." *Program Analysis, 1966–4.*
"Maternal and Child Health Care Programs, October 1966." *Program Analysis, 1966–6.*

Considering as basic costs in the appraisal of this program were the total anticipated project grant expenditures, including those for training and those applied to improvement of uterine cancer control procedures. Program costs over a 5-year period amounted to $73.8 million, or $68.1 million when discounted to present values. Added to these were the costs of hospital and medical care for the women expected to be found with cancer in early stages. It was estimated that of the 6.7 million examinations performed under this program, approximately 83,000 cases would be found and treated at the in situ stage. Early treatment costs amounted to $50.7 million discounted to present values. On this basis, 5-year program costs amounted to $118.7 million (discounted).

Considered as benefits were the savings in estimated earnings of these same women, including an imputed value for services of a housewife, and the savings in the medical costs that would have accrued had their cancers been allowed to progress to later stages. The assumption is that of the 83,000 women found to have cancer of the cervix, approximately 34,000 would be saved from death as a direct result of the program.

Although such screening projects customarily find numbers of cases of cancer that already have reached the later invasive stages, these were eliminated from the computation. Because of the nature of cancer of the uterus, most of these cases would have been discovered soon without the program. Therefore, the program could not be credited with the lives that are saved in this group.

Mortality savings are the present value of lifetime earnings for women who would be saved from death due to cancer of the cervix as a direct result of specific early cervical cancer detection programs. The general procedure followed for estimating mortality savings was to obtain the total value of lifetime earnings lost to cancer fatalities by applying estimated lifetime earnings described above by detailed age groups to the annual projected number of deaths of 1985 from cervical cancer. The projected mortality reductions resulting from this specific 5-year program was applied to the lifetime earnings losses to obtain the dollar savings from the program. The earnings were discounted to their value in 1968, the base year.

Savings of treatment costs were also estimated. The assumption is that if cancer of the cervix is found and treated at the in situ stage, savings would result from the fact that the higher treatment costs at the invasive stage would be averted. Included are expected reductions in the amounts that would be spent for hospital care, physicians' services, nursing home care, other professional services, drugs, and medical supplies, if the program would not be undertaken.

The final step in the analysis was a comparison of the present value of program costs in relation to the present value of estimated benefits from 1968 (the base year) to 1985. The ratio of discounted benefits to discounted program costs, is the benefit-cost ratio. It was found that the lifesaving effects of the case-finding activities of the program alone, could return after 5 years of operation, $9 for every dollar invested in the total program. The following summarizes the data developed in this study:

Undiscounted grant costs related to examinations, case findings, and deaths averted	
5-year grant program costs	$73,750,000
Examinations:	
Number performed	6,712,000
Number per case found	80.7
Grant cost, per examination	$10.99
Cancer cases found:	
Number	83,182
Grant cost, per case found	$887
Cancer deaths averted:	
Number	34,206
Grant cost, per death averted	$2,156
Estimated 1968–72 program costs and benefits, discounted to present values	
Program costs:	
Grant awards	$68,086,000
Treatment of cases found at early stages	$50,652,000
Total	$118,738,000
Program benefits:	
Earnings saved	$998,319,000
Treatment costs averted by program	$73,045,000
Total	$1,071,364,000
Benefit-cost ratio	9.0

Conclusions

The economic toll of illness, disability, and death in 1963 amounted to $105 billion when losses in output for those who died that year were taken into account. The figure would, of course, be higher for the current year. Direct expenditures for health and medical care amounted to $47.3 billion in fiscal year 1967.[21] Assuming approximately the same number of deaths and disability days as in 1963 and an annual 3-percent increase in productivity, the fiscal 1967 economic toll would amount to at least $125 billion. Dr. Philip R. Lee, Assistant Secretary for Health and Scientific Affairs in the Department of Health, Education, and Welfare states:

> It is clearly recognized * * * that the economic toll of illness is an intolerable burden on the American society, and that this society can no longer afford to neglect the disease and disability which afflict so many of its numbers.[22]

With limited resources, however, sound decisions must be made as to investments in specific health and medical care programs by the Government and priorities assigned to maximize the benefits from every tax dollar spent for these purposes. The recent emphasis on planning, programing, and budgeting (PPB) in the health field has emphasized the need for improved and systematic measurement tools of the effectiveness of health programs. The data presented in

[21] Merriam, Ida C. "Social Welfare Expenditures, 1929–67." *Social Security Bulletin.* December 1967. Table 5.
[22] Lee, Philip R., M.D. "Health and Well-Being." *The Annals of the American Academy.* Vol. II, September 1967, p. 203.

this paper provide a framework for measuring the monetary benefits associated with health programs designed to reduce death and disability.

When the expected savings in hospital days, physician visits, disability days, and deaths are translated into monetary terms, the program planner and administrator has one more tool to guide him in the decisionmaking process. Where alternative programs, or alternative levels of investment, are at issue, the relative dollar costs and benefits may be compared. All other things being equal, a program with a benefit-cost ratio of 9 would have a higher priority and would be preferable to one with a ratio of 3.

There is some difference of opinion among health planners and economists with regard to the translation of benefits into dollar terms and the calculation of the benefit-cost ratio to measure the effectiveness of alternative programs. Early in 1966, the Secretary of the Department of Health, Education, and Welfare directed that a series of program analysis groups be established to thoroughly analyze the objectives, benefits, and costs of existing and alternative programs. The varying approaches to this directive illustrated the divergence of opinion on the use of dollar values in the cost-benefit analysis.

The program analysis studies of selected disease control programs (motor vehicle injuries, cancer of specific sites, arthritis, syphilis, and tuberculosis) rigorously developed benefit-cost ratios based on identical lifetime earnings using identical discount rates.[23] Benefit-cost ratios and use of lifetime earnings were also used in the program analysis studies of human investment programs (adult basic education and work experience and training) and in a study of the benefits associated with vocational rehabilitation services.[24]

On the other hand, several studies measured the costs of alternative programs and enumerated the benefits in terms of various mortality and morbidity measures without translation of the benefits into dollar terms. For example, the Department of Health, Education, and Welfare study of maternal and child health care programs enumerated the savings in terms of preventing or correcting chronic handicapping conditions, reducing infant mortality, and reducing unmet dental needs.[25] The kidney disease program analysis study stated the benefits as reductions in mortality, prevalence, and morbidity days.[26] In the draft of the latest Department of Health, Education, and Welfare program analysis, "Delivery of Health Services for the Poor," benefits are expressed in terms of deaths averted and various indices of reduced morbidity such as hospital days saved, disability days averted, et cetera.[27] The recent paper "Cost-Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease," by Klarman, Francis, and

---

[23] U.S. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Program Coordination. "Disease Control Programs—Selected Disease Control Programs, September 1966." *Program Analysis, 1966–5.*

[24] U.S. Department of Health, Education, and Welfare, Vocational Rehabilitation Administration. *An Exploratory Cost-Benefits Analysis of Vocational Rehabilitation.* August 11, 1967.

[25] "Disease Control Programs—Maternal and Child Health Care Programs, October 1966." *op. cit.*

[26] U.S. Department of Health, Education, and Welfare, Public Health Service. *Kidney Disease Program Analysis—A Report to the Surgeon General.* July 1967.

[27] U.S. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Planning and Evaluation. *Program Analysis—Delivery of Health Services for the Poor.* October 1967.

Rosenthal also avoided translation of benefits into dollar terms and measured the benefits only in terms of life years gained.[28]

What are the issues and problems involved that result in these differences in analytical procedures?

The major issue is that of the intangible benefits which "are difficult, if not impossible to measure." [29] Several economists feel that ignoring the intangibles may distort the overall economic and social costs because the implicit assumption is that the economic value of intangible losses is zero.[30] Others are concerned that unless the ratio of measured benefits to nonmeasured benefits is the same in all programs—which it isn't—those programs with a higher proportion of nonmeasurables will have relatively lower benefit-cost ratios and will be unfairly compared.[31]

Another issue is that translation of benefits into dollars puts a higher value on one age group than another. The lifetime earnings of an aged person can scarcely be favorably compared with the lifetime earnings of a person in his twenties. Similarly, the lifetime earnings of an infant are less than those of the person in his twenties. Thus, if one program aims at reducing mortality or morbidity of one age group, and another program, of equal costs, reduces mortality or morbidity of another age group by an equal amount, the benefit-cost ratios will not be the same. The concern is whether one age group should be "saved" before another, solely because of its greater productivity. The choice between such programs seems cold and calculating if based on an economic value placed on human life especially when faced with the choice, for example, between programs to save children from future illness or programs to assist older people crippled by chronic diseases. "We are not dealing with pieces of hardware but with human beings whose needs are obvious." [32]

"It must be obvious that economic criteria are not the only criteria which should be applied to the allocation of resources and the distribution of program outputs," states Mr. William Gorham, Assistant Secretary of HEW for Program Coordination. "Nor do the analyses made as part of the PPB process constitute a prepackaged instant decisionmaker intended to replace judgment, commonsense, and compassion or turn resource allocation decisions over to computers." [33] Nevertheless, Gorham feels that the cost-benefit and cost-effectiveness analysis is an important tool for the decisionmaker as program objectives are clearly identified, the alternative methods of reaching these goals are defined, the costs over a period of time are measured carefully along with the benefits to be derived from the alternative programs.

In health programs, the savings or benefits may be in terms of such indexes as lives saved, disability days prevented, savings in hospital days, and physician visits. The translation of these tangible savings

---

[28] Klarman, Herbert E., Francis, John O'S., Rosenthal, Gerald D. *Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease.* Paper presented at the annual meeting of the American Public Health Association, Miami Beach, Fla. October 24, 1967.

[29] Gorham. "Allocating Federal Resources Among Competing Social Needs." *op. cit.*

[30] Klarman. "Syphilis Control Programs." *op. cit.*
Proceedings of Conference on the Role of Economics and the Economists in Solving the Problems in Health Services. op. cit.

[31] Fein, Rashi. *Problems of Assessing the Effectiveness of Child Health Services: Economic Aspects.* U.S. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Program Coordination. Occasional Papers—No. 1, May 5, 1967.

[32] "The ABC's of PPBS." *The Secretary's Letter.* vol. 1, No. 3, July 1967, p. 3.

[33] Gorham. "Allocating Federal Resources Among Competing Social Neds." ***op. cit.***

into one common denominator—the dollar benefits—and the calculation of the numerical ratio of such benefits-to-dollar costs are merely the final stages in the analysis. Where there are "unmeasureable" social benefits, these should be explicitly stated. But the existence of such "unmeasureable" benefits should not preclude the rigorous analysis and measurement of the expected tangible results of various alternatives. To this analyst, the final stage of placing dollar values on benefits is merely an extension of the various medical and vital statistical indexes into economic terms.

As succinctly concluded by Fein in a recent paper on the economic aspects of assessing program effectiveness:

> I do not say "don't measure effectiveness." Nor do I say "rely totally on the benefit-cost ratio." I suggest that the ratio tells us much. It is a powerful and useful tool. It should be sought—for even in the seeking we learn a lot, we think the problem through, we specify things more carefully. It is foolish to reject the attempts to measure and rely on vagueness.[34]

TABLE 1.—*Health and medical care expenditures in relation to GNP, United States, selected fiscal years 1929–67*

[Dollar amounts in billions]

Fiscal year	Gross national product	Total health and medical care expenditures	
		Amount	Percent of GNP [1]
1928-29	$101.0	$3.6	3.6
1934-35	68.7	3.1	4.5
1939-40	95.1	3.8	4.0
1944-45	211.1	7.9	3.7
1949-50	263.4	12.1	4.6
1954-55	379.7	17.9	4.7
1959-60	495.6	26.4	5.3
1964-65	655.5	39.0	5.9
1965-66	715.2	42.8	6.0
1966-67 [2]	763.1	47.3	6.2

[1] Calculated on basis of millions of dollars for health and medical care expenditures.
[2] Preliminary estimates.

Source: Ida C. Merriam, "Social Welfare Expenditures, 1929-67," *Social Security Bulletin*, December 1967, tables 2 and 5.

[34] Fein. *Problems of Assessing the Effectiveness of Child Health Services: Economic Aspects. op. cit.* p. 14.

TABLE 2.—*National health expenditures, selected categories: Estimated amount and distribution, by diagnosis and object of expenditures, 1963*

Diagnosis	Amount (millions)					Percent distribution				
				Professional services					Professional services	
	Total	Hospital care	Nursing home care	Physicians	Others [1]	Total	Hospital care	Nursing home care	Physicians	Others [1]
Total	$22,530.0	$11,579.0	$825.0	$6,867.0	$3,259.0	100.0	100.0	100.0	100.0	100.0
Neoplasms	1,279.0	1,006.1	27.2	206.0	39.7	5.7	8.7	3.3	3.0	1.2
Mental, psychoneurotic and personality disorders	2,401.7	2,059.7	29.7	281.5	30.8	10.7	17.8	3.6	4.1	.9
Diseases of nervous system and sense organs	1,416.4	684.0	178.2	508.2	46.0	6.3	5.9	21.6	7.4	1.4
Diseases of circulatory system	2,267.3	1,272.7	207.1	714.2	73.3	10.1	11.0	25.1	10.4	2.2
Diseases of respiratory system	1,581.1	751.0	----------	803.4	26.7	7.0	6.5	----------	11.7	.8
Diseases of digestive system	4,158.7	1,335.5	9.1	398.3	[2] 2,415.8	18.5	11.5	1.1	5.8	74.1
Diseases of bones and organs of movements	1,430.0	501.7	52.0	553.2	423.1	6.3	4.3	6.3	6.6	13.0
Injuries	1,702.8	995.4	72.6	604.3	30.5	7.6	8.6	8.8	8.8	.9
All other	6,293.0	2,972.9	249.1	2,897.9	173.1	27.9	25.7	30.2	42.2	5.3

[1] Includes nursing care and services of dentists, podiatrists, physical therapists, clinical psychologists, chiropractors, naturopaths, and Christian Science practitioners.
[2] Includes dental care.

TABLE 3.—*Annual mortality and morbidity losses—Indirect costs, by population group and diagnosis, 1963*

Diagnosis	Amount (millions)				Percent distribution			
			Morbidity				Morbidity	
	Total	Mortality	Institutional	Noninstitutional	Total	Mortality	Institutional	Noninstitutional
Total	$23,773.1	$2,731.0	$5,104.3	$15,937.9	100.0	100.0	100.0	100.0
Neoplasms	1,334.5	483.8	58.4	792.3	5.6	17.7	1.1	5.0
Mental, psychoneurotic and personality disorders	4,634.0	10.0	3,640.7	983.2	19.5	.4	71.3	6.2
Diseases of nervous system and sense organs	1,825.3	299.8	307.0	1,218.5	7.7	11.0	6.0	7.6
Diseases of circulatory system	4,145.6	1,225.9	328.9	2,590.7	17.4	44.9	6.4	16.3
Diseases of respiratory system	3,305.7	139.4	8.4	3,157.8	13.9	5.1	.2	19.8
Diseases of digestive system	1,343.6	123.4	20.5	1,199.6	5.7	4.5	.4	7.5
Diseases of bones and organs of movements	1,230.6	5.6	81.3	1,143.7	5.2	.2	1.6	7.2
Injuries	2,052.4	241.7	105.3	1,705.4	8.6	8.9	2.1	10.7
All other	3,901.4	201.4	553.8	3,146.7	16.4	7.4	10.8	19.7

TABLE 4.—*Morbidity losses, noninstitutional population: Number and distribution of total years lost and productivity losses, by population group, 1963*

Population group	Total years lost		Productivity losses			
			Man-years lost		Indirect cost	
	Number (thousands)	Percent	Number (thousands)	Percent	Amount (millions)	Percent
Total	3,838.6	100.0	2,883.6	100.0	$15,937.9	100.0
Currently employed [1]	1,695.6	44.2	1,695.6	58.8	9,764.1	61.3
Unable to work	1,820.0	47.4	865.0	30.0	5,311.4	33.3
Keeping house [1]	323.0	8.4	323.0	11.2	862.4	5.4

[1] All persons in this group are assumed to be productive.

TABLE 5.—*Annual mortality and morbidity losses: Estimated indirect costs, by diagnosis and age, 1963*

Diagnosis	Amount (millions)					Percent distribution				
	All ages	Under 25	25 to 44	45 to 64	65 and over	All ages	Under 25	25 to 44	45 to 64	65 and over
Total	$23,773.1	$1,113.4	$7,143.1	$10,733.5	$4,783.2	100.0	100.0	100.0	100.0	100.0
Neoplasms	1,334.5	40.5	265.9	678.2	349.9	5.6	3.6	3.7	6.3	7.3
Mental, psychoneurotic and personality disorders	4,634.0	244.3	1,821.0	1,898.7	670.0	19.5	21.9	25.5	17.7	14.0
Diseases of nervous system and sense organs	1,825.3	69.5	342.0	733.6	680.1	7.7	6.2	4.8	6.8	14.2
Diseases of circulatory system	4,145.6	41.3	540.7	2,065.5	1,498.1	17.4	3.7	7.6	19.2	31.3
Diseases of respiratory system	3,305.7	214.8	1,245.4	1,535.4	310.2	13.9	19.3	17.4	14.3	6.5
Diseases of digestive system	1,343.0	68.9	450.0	669.1	155.5	5.7	6.2	6.3	6.2	3.3
Diseases of bones and organs of movements	1,230.6	31.9	319.4	640.6	238.8	5.2	2.9	4.5	6.0	5.0
Injuries	2,052.4	166.7	897.7	772.1	216.0	8.6	15.0	12.6	7.2	4.5
All other	3,910.4	235.5	1,261.0	1,740.3	664.6	16.4	21.2	17.7	16.2	13.9

TABLE 6.—*Morbidity and mortality losses: Number and distribution of total years lost and productivity losses, by sex and age, 1963*

Sex and age	Total years lost		Productivity losses			
			Man-years lost		Indirect costs	
	Number (thousands)	Percent	Number (thousands)	Percent	Amount (millions)	Percent
Total	[1] 6,207.3	100.0	4,626.0	100.0	$23,773.1	100.0
Sex:						
Males	3,428.2	55.2	2,552.2	55.2	16,922.9	71.2
Females	2,779.1	44.8	2,073.8	44.8	6,850.3	28.8
Age:						
Under 25	571.0	9.2	319.7	6.9	1,113.4	4.7
25 to 44	1,461.2	23.5	1,258.6	27.2	7,143.1	30.0
45 to 64	2,237.4	36.0	1,817.8	39.3	10,733.5	45.1
65 and over	1,937.4	31.2	1,229.8	26.6	4,783.2	20.1

[1] Total includes 628 deaths, or 314 man-years lost during the year, for which the age at death was unknown.

TABLE 7.—*Morbidity and mortality losses: Number and distribution of productivity losses, by labor force status, 1963*

Labor force status	Productivity losses			
	Man-years lost		Indirect costs	
	Number (thousands)	Percent	Amount (millions)	Percent
Total	4,626.0	100.0	$23,773.1	100.0
Labor force	3,547.1	76.7	20,892.6	87.9
Keeping house	1,078.8	23.3	2,880.6	12.1

TABLE 8.—*Annual economic cost: Estimated direct expenditures and indirect costs of morbidity and mortality by diagnosis, 1963*

Diagnosis	Amount (millions)			Percent distribution		
	Total	Direct expenditures	Indirect costs	Total	Direct expenditures	Indirect costs
Total	$46,303.1	$22,530.0	$23,773.1	100.0	100.0	100.0
Neoplasms	2,613.5	1,279.0	1,334.5	5.6	5.7	5.6
Mental, psychoneurotic and personality disorders	7,035.7	2,401.7	4,634.0	15.2	10.7	19.5
Diseases of nervous system and sense organs	3,241.7	1,416.4	1,825.3	7.0	6.3	7.7
Diseases of circulatory system	6,412.9	2,267.3	4,145.6	13.8	10.1	17.4
Diseases of respiratory system	4,886.8	1,581.1	3,305.7	10.6	7.0	13.9
Diseases of digestive system	5,502.3	4,158.7	1,343.6	11.9	18.5	5.7
Diseases of bones and organs of movements	2,660.6	1,430.0	1,230.6	5.7	6.3	5.2
Injuries	3,755.2	1,702.8	2,052.4	8.1	7.6	8.6
All other	10,194.4	6,283.0	3,901.4	22.0	27.9	16.4

TABLE 9.—*Annual economic costs: Total amount with and without housewives' output, by type of cost, 1963*

Type of cost	Amount (millions)		Percent	
	Including housewives' output	Excluding housewives' output	Including housewives' output	Excluding housewives' output
Total	$58,036	$55,156	100.0	100.00
Direct expenditures	34,263	34,263	59.0	61.1
Indirect costs	23,773	20,893	41.0	37.9
Mortality	2,731	2,012	4.7	3.6
Morbidity	21,042	18,881	36.3	34.2

TABLE 10.—*Present value of lifetime earnings: Amount discounted at 4 percent, by age and sex*

Age	Males	Females
Under 1	$59,063	$34,622
1 to 4	64,989	37,938
5 to 9	79,333	46,289
10 to 14	96,736	56,422
15 to 19	114,613	64,936
20 to 24	126,688	67,960
25 to 29	128,698	66,826
30 to 34	122,904	64,389
35 to 39	111,956	60,998
40 to 44	97,301	56,603
45 to 49	80,325	50,896
50 to 54	63,027	44,371
55 to 59	45,948	37,467
60 to 64	28,387	30,164
65 to 69	15,043	23,579
70 to 74	9,264	18,118
75 to 79	5,344	12,888
80 to 84	2,935	6,916
85 and over	210	1,123

TABLE 11.—*Total mortality losses—Number of deaths, estimated total years lost, and lifetime earnings discounted at 4 percent, by diagnosis, 1963*

Diagnosis	Number of deaths	Total years lost (thousands)	Lifetime earnings (millions)	Percent distribution		
				Number of deaths	Total years lost	Lifetime earnings
Total	1,812,921	32,533.0	$49,928.1	100.0	100.0	100.0
Neoplasms	290,208	4,760.1	8,460.2	16.0	14.6	16.9
Mental, psychoneurotic, and personality disorders	4,651	121.5	250.9	.3	.4	.5
Diseases of nervous system and sense organs	215,648	2,570.9	3,853.5	11.9	7.9	7.7
Diseases of circulatory system	782,098	9,089.2	15,761.4	43.1	27.9	31.6
Diseases of respiratory system	105,235	2,174.0	2,665.4	5.8	6.7	5.3
Diseases of digestive system	71,700	1,445.3	2,458.5	4.0	4.4	4.9
Diseases of bones and organs of movements	3,772	76.6	127.7	.2	.2	.3
Injuries	130,665	4,155.9	8,297.1	7.2	12.8	16.6
All other	208,944	8,139.5	8,053.4	11.5	25.0	16.1

TABLE 12.—*Total mortality losses: Estimated present value of lifetime earnings discounted at 4 percent, by diagnosis and age, 1963*

Diagnosis	Amount (millions)					Percent distribution				
	All ages	Under 25	25 to 44	45 to 64	65 and over	All ages	Under 25	25 to 44	45 to 64	65 and over
Total	$49,928.1	$10,195.6	$9,753.8	$19,870.3	$10,108.4	100.0	100.0	100.0	100.0	100.0
Neoplasms	8,460.3	561.3	1,563.6	4,573.2	1,762.1	16.9	5.5	16.0	23.0	17.4
Mental, psychoneurotic and personality disorders	251.0	37.2	103.7	98.3	11.8	.5	.4	1.1	.5	.1
Diseases of nervous system and sense organs	3,853.5	318.9	567.8	1,535.8	1,431.0	7.7	3.1	5.8	7.7	14.2
Diseases of circulatory system	15,761.3	180.5	2,209.0	8,334.1	5,037.7	31.6	1.8	22.6	41.9	49.8
Diseases of respiratory system	2,665.4	885.0	370.8	894.9	514.7	5.3	8.7	3.8	4.5	5.1
Diseases of digestive system	2,458.5	325.7	611.6	1,174.1	347.1	4.9	3.2	6.3	5.9	3.4
Diseases of bones and organs of movements	127.7	39.0	22.3	44.4	22.0	.3	.4	.2	.2	.2
Injuries	8,297.1	3,132.5	3,228.1	1,650.1	286.4	16.6	30.7	33.1	8.3	2.8
All other	8,053.3	4,715.5	1,076.9	1,565.4	695.6	16.1	46.2	11.0	7.9	6.9

TABLE 13.—*Total economic cost: Estimated direct expenditures, indirect cost of morbidity, and present value of lifetime earnings discounted at 4 percent, by diagnosis, 1963*

Diagnosis	Amount (millions)				Percent distribution			
	Total	Direct expenditures	Morbidity	Total mortality	Total	Direct expenditures	Morbidity	Total mortality
Total	$93,500.3	$22,530.0	$21,042.2	$49,928.6	100.0	100.0	100.0	100.0
Neoplasms	10,589.9	1,279.0	850.7	8,460.2	11.3	5.7	4.0	16.9
Mental, psychoneurotic and personality disorders	7,276.6	2,401.7	4,624.0	250.9	7.8	10.7	22.0	.5
Diseases of nervous system and sense organs	6,795.4	1,416.4	1,525.5	3,853.5	7.3	6.3	7.2	7.7
Diseases of circulatory system	20,948.4	2,267.3	2,919.7	15,671.4	22.4	10.1	13.9	31.6
Diseases of respiratory system	7,412.8	1,581.1	3,166.3	2,665.4	7.9	7.0	15.0	5.3
Diseases of digestive system	7,837.3	4,158.7	1,220.1	2,458.5	8.4	18.5	5.8	4.9
Diseases of bones and organs of movements	2,782.7	1,430.0	1,225.0	127.7	3.0	6.3	5.8	.3
Injuries	11,810.6	1,702.8	1,810.7	8,297.1	12.6	7.6	8.6	16.6
All other	18,046.6	6,295.0	3,700.2	8,053.9	19.3	27.9	17.6	16.1

ON THE SOCIAL UTILITY OF ACCIDENTAL DAMAGE TO HUMAN RESOURCES

BY SIMON ROTTENBERG*

About 100,000 accidental deaths occur in the United States each year. There were 55 accidental deaths per 100,000 of population in 1965. Accidental injuries involving at least 1 day of restricted activity or requiring medical attention were suffered in 1964 by 53 million persons.

The most important causes of accidental deaths were motor vehicle accidents and falls. These two, together, produced almost two-thirds of all accidental deaths. The most important cause of accidental injuries was falls; other numerically important causes were being struck by moving objects, bumped into by objects or persons, lacerated by cutting of piercing instruments and handling or stepping on rough objects. Somewhat less than half of all injuries occur at home. The injury rate in a recent year was 286 per 1,000 of population. It is reported that an accidental death occurs in the United States every 5 minutes and an accidental injury every 3 seconds.

The cost of death is the sum of the values of the deceased person's expected contribution to the output of the economy, if he had not lost his life, perhaps net of his expected consumption; the resources employed in medical therapy, prior to his death; the administrative cost of relevant insurance schemes; the difference between the cost of early and later burial (earlier burial cost can be expected to be higher because later costs are discounted in calculating their present values); and the grief of surviving kinfolk.

The cost of injuries is the sum of the values of the lost output of the injured person while he is recovering from his injury or the lost output of permanently disabled or partially disabled persons for the remainder of their lives; the resources employed in therapy and prosthetic restoration; insurance administrative costs; and the pain suffered by the injured.

The prevention of accidents is also costly. Accidents can be prevented if more caution is exercised, other things being equal, or if hazardous activities are avoided, or if the environment is altered, so that given activities are caused to be less hazardous.

A simple illustration of the three classes of preventive strategies can be seen by reference to the use of a ladder. Some increment of care can be exercised in climbing it (as by mounting it at a correct angle or by climbing slowly), or the object sought to be fulfilled in the climbing of the ladder can be foregone, or the ladder can be fitted with rubber

*Duke University.

NOTE. I have been much helped in the preparation of this paper by discussions with my colleague, John S. McGee.

"feet" to diminish the likelihood that it will slip. None of these strategies come at zero cost.

Every activity has associated with it unwanted side effects to which a probability number can be assigned in principle. While we conventionally speak of these side effects as "accidents," they can really be foreseen in a probability sense. It is known, experientially, that some activities are more hazardous than others and that given activities are more hazardous in some sets of conditions than in other such sets.

It is known, for example, that the probability of contracting diseases of the skin is higher in activities that require the use of chemical agents and cutting oils and greases than in some other activities.

It is also known that, given the occurrence of the side effect, there is a probability distribution of injuries of different severities or magnitudes. Falling from a ladder will produce, say, skinned knees or broken wrists or broken backs or death; each outcome has some frequency associated with it and, by converting severity of outcome to some common measure (for example, number of days lost from work) expected values for the unwanted side effects of different activities can be constructed.

We shall assume, as a first approximation, that these magnitudes are known. Then every rationally calculating individual who decides that he will engage in an activity does so because the gain to him from doing so exceeds the cost to him of the activity even though the expected cost of unwanted side effects has been netted out of the gain. A rule which constrains his behavior—by requiring him to exercise more care than he would have done in the absence of the rule, by compelling him to forego an activity in which he would otherwise have engaged, or by requiring him to bear the cost of altering the environment to diminish the probability of side effects or to diminish the severity of outcome, should the side effects occur—will cause him to be less well off in some welfare sense. Safety, that is to say, is achieved at the cost of other things; the consumption of more safety is at the cost of the consumption of less of other things. The quantity of safety that rational individuals will consume will be a partial inverse function of its "relative price," or cost. The larger the gain foregone by avoiding hazardous gainful activity, for example, the more reluctant will one be to avoid such activity and the smaller will be the quantum of safety consumed. Rational individuals allocate their incomes between safety and other things so that their utility or welfare is maximized. A rule compelling the consumption of a different basket of commodities leaves them worse off.

We can now relax the assumption that the probability of the occurrence of an accident as a side effect of some activity and the distribution of severities of injurious outcomes are fully known to all individuals. Suppose them to be not known to some or all. For these individuals there is uncertainty and every such individual constructs an estimate of the relevant probabilities and distributions. These estimates can be expected more closely to approximate the true numbers, the greater is the quantity of information sought out and assimilated. But the acquisition and commanding of information is also not costless and rational individuals will usually not erase ignorance completely but will carry the information-generating process only to that scale where the cost of the last increment of information is equal to

the incremental gain produced by assimilating that quantitatively final unit of information. It does not pay to be better informed than this. Usually some quantity of ignorance and uncertainty is consistent with the optimal state.

If some individuals possess less than an optimum quantity of information about the degree of hazard associated with an activity in which they engage and if they mistakenly choose not to invest in the addition of an increment to their current stock of information, it does not follow that they should be further informed at the public expense. The loss they suffer will fall upon themselves and the gain to be derived from better information will be reaped by them.

If some accidents have third-party effects, however, the appropriate social policy is, for those accidents, different. If A undertakes an activity which can have unwanted side effects that injure B and if B is not identifiable *ex ante*, then a rule constraining A's behavior is indicated. Suppose that automobiles, if driven at 100 miles per hour in specified conditions of driver skill, road dimensions, road surface and traffic density will suffer a collision with other automobiles x times per 100,000 miles traveled and suppose the identities of the "other automobiles" are not known until *after* the collision takes place. If x is considered to be "large enough," it might pay to enforce a rule constraining speed of travel to some smaller number than 100 miles per hour.

What, however, if at 50 miles per hour the number of collisions per 100,000 miles is $x/3$? Shall speed be held by rule below that number?

Clearly the solution to this problem of the appropriate speed limit is not found in the rule of minimization of accidental injury and death, for if that rule were followed, the speed limit would be set at zero. The solution must be discovered in a calculus of cost and gain in which time and convenience is traded for physical assets, and human lives and limbs. The values of the variables are likely to be such that the rational society will accept some number of automotive accidental deaths and injuries. The achievement of a lesser number than this will come at too high a cost.

Constraint put upon the speed of automobile travel is not the only strategy available for forestalling a suboptimal quantity of damage done to some by accidents accompanying action by others; the allocation of the costs of damage can, in some circumstances, affect behavior and cause an appropriate quantity of care to be taken.

The automotive case is representative of a class of cases in which unwanted side effects are jointly produced and it is difficult to say how costs should be appropriately allocated to induce that change in behavior that will diminish the incidence of those side effects. An automobile being driven at 100 miles per hour will collide with another x times per 100,000 miles *but only if another automobile is in its path.* The side effects would have been avoided if *either* the first were driven more slowly *or* the second were not there. Given that the second *is* there, the severity of the consequences will be less if, say, passengers in the second automobile were protected by seatbelts. If the cost of the damage were assessed upon the speeding driver (and he knew this in advance), he would have an incentive to drive more slowly; if it is assessed upon the driver and passengers of the struck automobile (and they knew this in advance), they would have an incentive to

keep out of harm's way—perhaps by avoiding high-speed or densely traveled roads and they would have an incentive to use seatbelts. The principle on which the exercise of caution should be allocated among joint actors and, therefore, on which the incidence of the cost of accidental damage should be allocated is that of cost-minimization for achievement of the objective sought. If, on other criteria, it is known that the incidence of accidents, adjusted for severity, should be reduced by some fraction, then the cost of damage should be allocated among the joint actors so that the total and summed cost of the incremental caution of all joint actors will be less than that cost would be, if any other set of additional cautionary behavior of the actors were combined. The case of joint production of accidents is made difficult to solve because only some, but not all, of the costs of caution can be priced in markets. Seatbelts can be priced, but proxies must be found to value time lost by avoiding densely traveled roads, and it is more difficult still to value alertness and close attention.

In some cases of joint production of damage, the explicit allocation of damage cost may be irrelevant. Consider the case of occupational hazards to health. Damage done by accidents at work are the joint product of employers and workers. Employers offer employment which requires that workers perform certain operations under certain conditions with respect to materials, tools, machines, space, ventilation, etc., that are arranged by the employer. Employers act in two senses: they offer to employ and they arrange the conditions of work. Workers act also in two senses: they accept offers of employment and they perform the operations that define the relevant employment in some particular way (say with more or less caution). The occurrence and frequency of accidents and the severity of the injuries they cause requires that both act. It is in this sense that it is said that work-injuries are jointly produced.

Some occupations are more dangerous than others. The frequency of disabling injuries, adjusted for the number of employee-hours worked, was, in a recent year, about seven times as high in anthracite mining as in laundries and drycleaning plants. But hourly earnings in mining were more than twice earnings in laundries and part of the difference in earnings reflects the difference between the industries in the risks of accidental injury. In this way, the market sees to it that risks run are paid for. Occupations in which the probability of accidental damage to workers is relatively high pay more in wages to attract workers to them. Workers are paid wages part of which compensate them for the pure labor services they render and part of which pay for the risk of injury.

Just so, any given occupation may be made by the employer more or less safe as when he does or does not invest in warning lights and signals, safety guards on moving machine parts, ventilating systems, catch platforms, goggles, masks, hard hats, steel-tipped shoes, safety education campaigns among his employees, etc. As in the case of comparative hazard in different occupations, the larger the cost explicitly incurred by the employer in rendering his employment safe, the safer it will be; the safer is the employment, the lower will be the wage paid in it.

The number of workers employed in any occupation is partly determined by the cost of labor. In reckoning cost, the employer is indifferent to its source; it does not matter to him whether a 5-cents-per-hour

component of cost is derived from the installation of safety equipment or from cash wage payments made to workers. If there are no legal rules enforcing safety standards to some level, it can be expected that there will be diversity among employers with respect to the quantities of safety that they offer in some given occupational employment. Some mines will be more hazardous than others and there will be differences between these classes in the wages paid for work in them. Workers will then distribute themselves between the two classes of mines. Those who choose to accept employment in the more dangerous class may purchase their own safety gear with the higher wages paid them or they may choose the option of expending their whole wage for other objects (including the object of self-insurance for the cost of medical care and lost wages, if they suffer accidental injury) and run the larger risk of accidental damage done them.

Individuals are, of course, risk averse to different degrees of intensity. Some will, for gains of given magnitudes, not climb ladders at all; some will do so only to first-story heights; and some will climb higher still. Thus, if occupations are rank ordered by the the probabilities of accidental injury of given severity in them, workers will distribute themselves among occupations inversely to their distaste for risk. Those for whom the "cost" of a given mathematically expected injury is relatively low will more likely accept relatively dangerous employment than will other workers.

If a common rule is enforced by law that specifies the minimum quantity of safety that employers may offer, some wage workers will be required to consume more safety and less of other things than they want. Legally enforced safety standards can be expected to produce an "accident-prevention industry" that is too large. Accident-preventing activities should be carried only so far that the gain from those activities is equal to their cost at the margin. More than this produces a smaller number of accidents and a smaller quantity of accidental damage, but this is a suboptimal outcome.

The enforcement of legal safety standards could foreclose some occupations completely and systems for the assessment of the costs of repairing damage could be coupled with repair costs of sufficiently high magnitude to have the same effect. In this case, such standards or cost-assessment systems would affect the distribution of the working population among occupations such that society would consume a dispreferred basket of commodities. That is to say, the composition of output would be "wrong." It is only necessary for this outcome to occur that the price exacted by the State for permission to violate the legal standards (i.e., the value of the punishment visited upon offenders) be sufficiently high, or that the set of worker risk preferences be different from the set of employer risk preferences.

It is conventionally thought that safety, because its consequence is less damage, is a good thing and that hazard, because its consequence is more damage, is a bad thing and from this the conclusion is often wrought that "the more safety—and the less hazard—the better." While this is correct *ceteris paribus*, it is not correct once it is seen that safety is costly and that hazard is a substitute for other resources in the production of desirable objects. Then it becomes clear that the accidental damage and destruction of some quantity of human resources serves the social purpose and, therefore, that it is possible for safety legislation to diminish social welfare.

THE OPERATING STRUCTURE OF THE MEDICAL CARE SYSTEM—AN OVERVIEW

A Reconnaissance of the Economics of the Medical Care Industry

by Gerald Rosenthal*

CONTENTS

Introduction

1. This paper will attempt to present the broad overview of the operating structure of the medical care system. The operating structure

*Department of Economics, Brandeis University.

Note: An earlier version of this paper was prepared for a Task Force on Health, Education, and Welfare, September 1965. For a thorough and extensive bibliography of many of the topics discussed in this paper see: Klarman, Herbert E. *The Economics of Health.* New York: Columbia University Press, 1965.

is that dimension of the system within which the producers of medical care services and the consumer of services interact. The actual amount of medical care which will be produced in the operating sector is a result of the interaction between the demands of consumers and the supply which the producers are willing to provide. The demand for medical care reflects the interaction between the incidence of morbidity and the population and the social and economic structure of that population. The morbidity factors reflect somehow the *need* for medical care while the social and economic factors determine how these needs are translated into a *demand* for care. The supply of medical care is determined by the inputs of capital and labor into the production of these services and the technology which limits the way in which these inputs can be combined. What is of primary significance is the fact that all the elements in this process interact in a dynamic way in that dimension of the medical care system which we call the operating structure.

2. The other dimension of the medical care system is that which is encompassed in what might be called planning activities. Planning for this purpose is a process of intervention in the operating structure of the medical care system, the object of which is to affect the output of system operations. Much of the planning process consists of attempting to select policies which will yield desired changes in the output of the health care system by: (*a*) affecting the environment within which the system operates; (*b*) affecting the range of choices available to operators; and (*c*) affecting the mix of rewards for production and consumption of health services. The degree to which any policy will serve the planner's objectives will depend on the nature of the operating system which it is intended to affect.

3. This paper is a discussion of the basic operating dimensions of the medical care system. Its object is to make clear the interrelatedness of the operators in the system and to examine some of the things that are known about the way these operating dimensions respond to the environment within which they operate, the range of choices available within the operating structure in the system and the degree to which the actual organization of medical care system is responsive to the economic stimuli operating in the market.

4. If the operating system were structured in such a way so as to yield an "optimal" output of medical care services, there would be no need for planning activities as they have been defined here. Even the most cursory examination of the overall medical care system in the United States makes evident the degree of discrepancy between the operating results of the system as viewed by operators and the objectives for the medical care system as held by the many groups and individuals evolved in intervening with the system from outside the operations structure; i.e., engaged in the process of planning. Indeed, the interaction between the operating structure of the medical care system and the planning process itself is of such long standing that it would be difficult, indeed, to ascertain what the operating structure of this activity would look like in the absence of outside intervention. System objectives designed to increase the overall economic efficiency of the system as a whole, as well as those designed to maintain a qualitative level of care different from that which would be forthcoming in the absence of intervention represent but two of the many

discrepancies between the operator's view of the effectiveness of the system and the planner's view. While little will be said about the planning activities specifically in the course of this discussion, an effort will be made to point out those areas of intervention which reflect the impact of planning policies generated from outside the operating structure on the operating structure itself.

The Medical Care Operating System Defined

5. One of the major difficulties with the literature of the medical care field is the tendency to deal with physicians' services, medical care facilities, the drug industry, etc., as separate entities. From an economic viewpoint, it is far more realistic to consider these as subsectors of a medical care operating system. The basis for this argument lies in the fact *that, from the consumer's point of view, medical care is the product which is desired.* The demand for services of the physicians' subsector, the medical care facilities subsector, etc. are derived demands since they are here defined as inputs into the production of medical care and not as forms of final consumption. The relation between the output of medical care and these subsectors can be described by means of a production function in which the output of the medical care sector is expressed as a function of a combination of physicians' services, medical care facilities, other medical manpower and other inputs not elsewhere classified. The actual combination of inputs will depend on a number of things, including:

(*a*) the particular mix of morbidity in the population consuming the medical care,
(*b*) the tastes and preferences of the consumers of medical care,
(*c*) the available technology, and
(*d*) the relative costs of the various inputs.

6. The use of this broader concept of the medical care operating system makes explicit the existence of choices such as that between increasing the number of medical schools in order to provide more physicians and increasing the quantity of medical facilities to make those physicians already available more productive. Since what is desired is medical care, and not physicians or medical care facilities per se, such choices are basic to any subsequent decisions affecting the specific subsystems.

The Demand for Medical Care

7. Empirical studies of the demand for medical care have generally used one of two approaches.[1] Some of these studies have attempted to

[1] Barry, Charles. *Factors Affecting the Utilization of Physicians' Services.* The Brookings Institution (processed).
Feldstein, Paul. "The Demand for Medical Care," *Report of the Commission on Cost and Medical Care,* vol. 1. Chicago: American Medical Association, June 1964.
——— "Research on the Demand for Health Services," *Milbank Fund Quarterly,* pt. II.
Rosenthal, Gerald. *The Demand for General Hospital Facilities,* Monograph No. 14, Chicago: American Hospital Association, 1964.
——— "The Demand for General Hospital Facilities: A Principle Components Analysis," *Proceedings of the Social Statistics Section,* 1964, Washington: American Statistical Association, 1964.
Wirick, Grover, and Robin Barlow. "Social and Economic Determinants of the Demand for Health Service," in *The Economics of Health and Medical Care,* Ann Arbor: The University of Michigan, 1964.

relate various characteristics of the population being studied to aggregate expenditures on medical care, usually with some correction for price changes. Other studies have attempted to relate similar characteristics of the population to specific items of medical care consumption such as physicians' visits, patient days, general hospital facilities, et cetera, or on expenditures for these items. While the particular characteristics studied vary, they are of four general types:

(*a*) Characteristics reflecting morbidity in the population, a variable difficult to measure directly;

(*b*) characteristics reflecting the awareness of the desirability of consuming medical care;

(*c*) characteristics reflecting the existence of substitutes for medical care; and

(*d*) characteristics reflecting the ability to command resources in the market.

8. Characteristics such as the age and sex distributions of the population have frequently been used to indicate morbidity circumstances. In general, populations with higher proportions of older persons use more of all types of health services. While the age 65 has been the most used cutoff point between the aged and the rest of the population, there is evidence to suggest that the increase in the use of medical care starts around age 55 and increases sharply at age 75. Sex differences reflect both the use of medical care for childbearing and differences in the incidence of disease between men and women. In some studies, racial differences have also been associated with differences in morbidity.

9. The existence of illness does not necessarily show up in the form of demand for medical care. Other characteristics of the population, primarily social and economic, reflect elements that intervene in the process whereby medical "needs" are translated into a demand for services. Much of the literature in the medical care field concerns itself with the problem of meeting "needs" for medical care without noting the difference between need and demand. This often leads to a situation where the solution to filling unmet needs is simply held to be an increase in the supply of the needed service. The economic realities are such, however, that without also directly altering effective demand, the additional supply will not be utilized to fill these "needs." The low occupancy rates in many hospitals located in areas of high medical need are a good example of the failure of this simple kind of solution.

10. It is essential that the distinction between the medical notion of need and the economic notion of demand be understood. Not only is medical care which is "needed" not demanded, but often, medical care which is not needed, by some medical technical standards, *is* demanded. There is no value judgment implied in this distinction. Medical care in this respect differs little from most other types of consumption. The evident difference between the need for transportation and the demand for automobiles serves as another example of this dichotomy. If policies are to be directed at minimizing the difference between the need for medical care as determined by medical opinion, in the light of current technology, and the care actually available and utilized, explicit examination of the level and structure of demand for medical care and the factors which determine them is necessary.

11. Regardless of the morbidity pattern of a given population, a

wide range of responses in terms of demand for medical care may be found. The characteristic most often associated with the awareness of the desirability of medical care (perception of need) is the educational level of the population. In general a more highly educated group tends to take symptoms of illness more seriously and to seek medical care more readily. (There is some difficulty in demonstrating this since higher educational levels are associated with higher economic levels which also tend to encourage consumption of medical care.) Disagreement exists regarding the implications of this greater readiness to seek care for the overall quantity of medical care consumed. One argument maintains that because care is sought earlier, and since disease responds to treatment more readily in its earliest stages, less total medical care will be consumed. The opposite argument is that the increase in physicians' services sought more than offsets whatever reduction of medical care usage occurs as the result of early detection.

12. Another set of characteristics which has been related to the demand for medical care is that associated with family status and housing conditions. For the most part, these characteristics have been thought to reflect the ability of a household to serve as a locus for medical care activity. These aspects of the population show the greatest association with the use of hospital facilities (the overall use of facilities is greater for single, widowed, and divorced than for married persons) and less association with other elements of medical care. The impact on hospital use shows up as a shorter length of stay rather than in lower admissions rates. This indicates that the household may serve as substitute for hospitals during that period of an illness episode which does not require complex medical services. It is possible that shorter length of stay reflects greater economic pressures to return to work generated by family responsibilities but the same relation between marital status and hospital use holds in Great Britain where job security and disability pay are such as to minimize the economic arguments. It is likely that a genuine substitution relation exists between convalescent days spent in the hospital and in the patient's own household if there is another family member to provide basic care. While few experiments have been carried out to ascertain the degree to which an intensive homemaker program could be usel to provide a substitute for a household member, this would be a fruitful area for study.

13. The ability to command medical care services in the market has been measured primarily by three variables: price level, income level, and insurance coverage. Attempts to measure the price elasticity of demand for medical care services encounter two major difficulties. First, there is rarely a single price for each service or a single uniform package of services. Second, there is great heterogeneity with respect to the quality of a given unit of service. This makes it difficult both to determine how the price is to be measured and to establish an empirical relationship between price and demand.

14. Traditionally, the pricing mechanism for medical care services is held to conform to the economists' model of the discriminating monopolist in which each consumer is charged according to his ability to pay.[2] A number of factors have reduced the use of this basis for

[2] Kessel, Reuben. "Price Discrimination in Medicine," *Journal of Law and Economics.* Chicago: University of Chicago Law School, October 1958.

pricing. The rise of prepayment, coupled with the associated rise in the development of fee schedules, is one major factor. (This trend toward fixed fee schedules is by no means limited to insured services since in many cities, dentists post a schedule of charges for services in their offices. Even in these communities, there is little or no insurance for dental services.) A pricing mechanism truly based on the individual's ability to pay would eliminate the utility of possessing insurance. Further, since much of hospital insurance represents a form of bulk purchase from the institutions by a third party on behalf of large groups of users, determination of the individual consumer's ability to pay is impossible. In addition to the increase in insurance coverage, more and more of the free care traditionally provided by medical care subsectors is provided through welfare institutions outside of the medical care industry. These agencies represent bulk purchasers (as do private insurers) and require the establishment of reimbursement rate schedules prior to the act of consumption. The establishment of payment of "customary charges" for participating physicians under part B of medicare has also served to compress further the range of prices charged by an individual physician. The net effect has been to raise the lower end of the physician's price schedule and, thereby, raise his average price. In brief, the growth of insurance has reduced the possibility of charging higher prices of the wealthier patients, (although income limits for fee schedules permit some variable pricing.) The institutionalization of charity outside of the medical care industry has also diminished the amount of free care to the poor (studies indicate this is less true for hospital services). These influences narrow considerably the range of prices for a given service at any moment.

15. While the growth and development of insurance has diminished the influence of ability-to-pay pricing, it has made the determination of price elasticity of demand more difficult. The existence of insurance obscures, to a great degree, the price for medical care as perceived by the consumer of these services. (Research is currently underway pertaining to the relation between out-of-pocket expenses and the length of hospital stay, medical considerations being held equal.)[3] In spite of the difficulties, attempts to measure the price elasticity of demand independent of differences in income, insurance, and so forth, have found a negative price elasticity. Other things being equal, more medical care services are consumed at lower prices than at higher prices for the same services. In addition, there is reason to believe that demand is becoming *more* price elastic over time. Such a response follows from the changing nature of expenditures on medical care. At least some of the increased expenditure represents less urgent consumption over which the consumer exercises a greater degree of choice and is, therefore, more likely to respond to price changes. (Consumption of contact lenses and elective surgical procedures are cases in point.)

16. The problem of adjusting for quality changes is quite difficult to resolve. There is little doubt that the price indexes in use today overstate the amount of the price rise over time. The use of the daily charge as a reflection of the price of hospital care ignores the declining length of stay for many illnesses. The price of an illness should reflect both

[3] Burgess, Alex M., Carr, W. John, Osler Peterson, and Gerald Rosenthal. *The Economics of Hospitalization* (in progress).

the cost per unit of input and the amount of inputs. The development of a medical care price index based on cost of illness has been underway for some time; its use is limited at this time.[4] On a comparative basis, areas with higher price levels *seem* to be those with more well-developed medical services and, therefore, a higher quality of care. While this requires a more objective comparative study of quality, at least some of the differences in price observed at a given moment in time also reflect differences in quality and are, therefore, overstated.

17. The relation between the demand for medical care and income levels has been subjected to much analysis.[5] Both budget studies and cross-sectional analyses among geographic areas show a positive association between these two variables. While efforts to estimate income elasticities have yielded varying results, indications are that the income elasticity is between 0.3 and unity. *This means that a 10-percent rise in income implies a 3- to 10-percent increase in expenditures on medical care.* At least some of this increase in expenditure will represent an increase in quality and nonmedical consumption such as the use of private rooms and television sets. Some will probably reflect a shift to greater use of the physician in lieu of self-diagnosis and treatment. This positive income elasticity of demand in a period of rising real incomes should lead to a continuation of the increase in relative share of medical care expenditures. It is not, however, expected that the distribution of these expenditures among the subsectors of the medical care industry will remain constant. Policy decisions based on the current mix are apt to generate misallocation of resources.

18. The association between insurance coverage and the demand for medical care has been well substantiated although disagreement exists as to the reasons for this association. At least some of the disagreement stems from a disinclination to acknowledge the degree of price elasticity of demand on ethical grounds. To consume more medical care at a lower price than at a higher is considered somehow immoral. It is argued that the positive association between insurance coverage and demand stems from the association of insurance coverage with other characteristics such as educational level, the educational impact of the insurance itself, and the self-selection that biases insured groups toward higher users—not from the insurance itself. There is no economic basis for doubting the real impact of medical insurance upon the price to the consumer. At least some of the positive insurance elasticity represents a special case of the negative price elasticity mentioned earlier (par. 15). Other aspects of insurance coverage make measurement of this insurance elasticity difficult. The actual benefit structure in terms of services and amounts of coverage differs from policy to policy. The percentage of the population with insurance coverage tells little about the actual "amount" (potential benefits) of insurance. Since most empirical studies have had to use aggregate measures, few estimates of the "real" impact of insurance have been developed. Such detailed information is necessary to determine the impact of Government programs on demand. Evaluation of the impact of the program of medical insurance for the aged, for example, would have benefited from such

[4] Scitovsky, Ann A. "An Index on the Cost of Medical Care—A Proposed New Approach," in *Economics of Health and Medical Care.* Ann Arbor: School of Public Health, University of Michigan, 1964.

[5] Feldstein, Paul J., and W. John Carr. "The Effect of Income on Medical Care Spending." *Proceedings of the Social Statistics Section,* 1964. Washington: American Statistical Association, 1964.

knowledge. For some of the aged, this kind of program provided the only insurance coverage, for others, change from the previous level of benefits, and for others, only a change in the insuring mechanism. The lack of solid estimates of the potential increase in demand expected to occur as a result of this legislation reflected the absence of "hard" analysis.

19. The benefit structure varies not only among individuals, but among the different types of medical services. Medical insurance coverage has been most extensive for hospital services rather than, for example, outpatient facilities. Such uneven coverage distorts the relative prices of the components of care and leads to proportionately higher use of hospital services which use more resources. Some early studies found that when all services are insured, a shift occurs from hospital to outpatient care. (This will not necessarily result in the use of fewer resources since such an insurance program is likely to result in higher overall consumption.) *The uneven nature of the structure of insurance benefits in the United States may create a situation where sensible economic behavior on the part of the consumer results in an inefficient and excessive use of resources for the industry as a whole.* This will also significantly constrain the degree to which a more efficient use of inputs into the production of medical care services can be encouraged from outside the operating system (that is, by the planners).

The Supply of Medical Care Services

20. The supply of medical care available, or forthcoming, depends on two things: the amount of resources devoted to its production and the technology which determines the ways in which these resources can be combined. The technology of medical care, and of most production processes, is such that identical outputs can be produced by a number of alternative combinations of inputs. Every change in technology widens the range of choice. Ideally, the particular combination of inputs used to produce a given output should be the "least cost" combination. To the extent that less costly inputs can be substituted for more costly ones, the process of producing medical care can be made more efficient. Since each geographical area of production does not necessarily face the same costs, there may be, at the same time, many different ways of producing medical care without any inefficiency. It must be emphasized that there may be considerable difference between "least cost" to an individual operator and "least cost" over the system as a whole. This difference provides much of the "raison d'etre" for planning activities in the United States.[6] This section of this paper (paragraphs 20 through 52) will examine the ways in which a number of the most basic inputs in the production of medical care are themselves produced and used, the organization of the subsectors of the medical care industry, the ways in which changes in technology have affected their role in the production of medical care, and the impact of planning policies on the efficiency of the system.

MEDICAL CARE FACILITIES

21. Perhaps the most significant aspect of the development of modern medicine has been the increasing importance of the hospital in

[6] Rosenthal, Gerald. *Planning, Policy Tools, and the Implementation of Health Care System Objectives.* Paper presented to annual meeting of the American Public Health Association, Octotber 1967. (Originally titled *"Health Planning and Financing of Medical Care."*)

the treatment of disease. Much of this importance results from the nature of technical change in medicine. At one time, all of the technology of medical care was encompassed by the physicians' bag and the role of the hospital was primarily custodial. Today, the practice of medicine requires the use of large quantities of capital equipment for diagnosis, treatment, and even prevention of disease. More and more, this capital has been centered in the hospital, as has more of the manpower needed to produce medical care.

22. This increasing capital usage in the production of medical care has not, however, been accompanied by a decrease in the use of manpower. Actually, the reverse has been the case. In the past decade, a significant proportion of all of the total increase in employment in the United States occurred within the hospital sector. During the 1960's, employment in hospitals has been greater than total employment in the basic steel and automobile industries combined. On a per-unit-of-output basis, it takes more manpower and more capital to produce 1 patient-day of care each year than it did the previous year. The simplest reason for this is that most of the technical change in this industry has been directed toward increasing quality of the output rather than its quantity. This increasing resource intensity of the hospital has led to more rapidly rising costs for hospital services than for any other element in medical care, and has generated great interest in the efficiency of producing hospital services.

23. One aspect of efficiency relates to the desirability of having each institution self-contained. When a number of institutions in the same area are accumulating equipment to provide certain complex and expensive services, there is a likelihood of unnecessary duplication of services. In economic terms, duplication is unnecessary when the same total services could have been produced for the same area with fewer units of capital equipment. All duplication is not *per se* unnecessary, and the fact of duplication is not *prima facie* evidence of inefficiency. It remains to be shown whether the same services could be produced with fewer units of equipment.

24. There are two sets of costs which must be considered in determining the necessity for medical capital equipment. First, there are those costs associated with having more than is needed. Most medical equipment is expensive to purchase and maintain. Underutilized equipment may encourage less discriminating use. This holds true for hospital beds as well as for more complex elements of hospital care such as laboratory facilities. On the other hand, there are real costs associated with not having the facilities of equipment available when they are needed. Overuse of capital can result in a decrease in the quality of care and excessive depreciation of available equipment.

25. If the demand for medical facilities were constant, or changing in a consistent direction, it would be possible to plan facilities for use at optimal intensity. However, the demand for all medical services, including those of facilities tends to fluctuate from day to day. Most facilities are built with a capacity to provide service which is geared to the peak demand on the ground that the costs of not having needed facilities in times of high demand are greater than the costs of having underused facilities the rest of the time. *If this is so, than any change which reduces the range of fluctuation of demand will make the operation of medical facilities more efficient.*

26. A considerable amount of research has been directed at this aspect of producing hospital services suggesting a number of points of significance.[7] First, it was found that facilities with a greater output have relatively less fluctuation in demand. For the hospital, this observation is reflected in the fact that *larger facilities (serving under markets) are able to maintain higher rates of occupancy (intensity of use)*. This relationship is not supurious but, rather, reflects the fact that larger facilities serve broader markets in which the range of fluctuation in demand is *proportionately* smaller than that of smaller markets. Thus, increasing the market served by individual facilities will enable these facilities to be used more intensively, and thereby increase their efficiency.

27. Increasing the market size will allow an increase in the size of institutions rather than construction of a greater number of smaller institutions, each serving a smaller market. Increasing the size of the institution reduces the relative fluctuation in demand, which leads to more intensive use of facilities. In addition, there is a growing body of evidence demonstrating the existence of true economies of scale in the production of hospital services.[8] Some studies suggest that, *for a given mix of output*, a larger scale of operation will enable the production of output at a lower cost per unit. This relation is not immediately evident since larger facilities tend to have higher average costs. However, recent research has related these higher costs to the fact that larger facilities generally produce a more complex and diverse set of services, possibly at a higher level of quality. The existence of true economies of scale in hospital operation provides an additional reason for expecting expansion of markets served by facilities (increasing the scale of operation) to yield increases in the economic efficiency of production of hospital services.

28. Economies of scale are particularly important to general hospital services. These services are the most costly and tend to be housed in considerably smaller facilities than mental, tuberculosis, and chronic disease services. At least part of the reason for the size differential stems from the fact that use of the above specialized services does not generally contain the same element of immediacy from the viewpoint of the consumer. In addition, mental and tuberculosis care have been largely provided by the State to serve geographic areas considerably larger than single communities. (This same observation holds true for Veterans' Administration facilities.) The specialized facilities do not experience a wide fluctuation in use due to their much longer average length of stay and low patient turnover.

29. While the above discussion has highlighted the costs of having small facilities, there are also costs associated with expansion of the market served by a given facility. The consumer of medical care places a fairly high value on its accessibility. In part, this value is due to

---

[7] Berry, Ralph. *Competition and Efficiency in the Market for Hospital Service.* Unpublished doctor of philosophy dissertation, Harvard University, September 1965.
Blumberg, Mark. "DPF Concept Helps Predict Bed Needs," *Modern Hospital,* December 1961.
Long, Millard. "Efficient Use of Hospitals," in *Economics of Health and Medical Care.* Ann Arbor: University of Michigan, 1964.
Rosenthal, Gerald. *Demand for General Hospital Facilities, op. cit.* Chapter 6.

[8] Berry, Ralph. *Competition and Efficiency, op. cit.*
Feldstein, Paul. *An Empirical Investigation of the Marginal Cost of Hospital Services,* Graduate program in hospital administration, University of Chicago, 1961.
Fitzpatrick, Thomas, *et al. The Nature of Hospital Costs.* Ann Arbor: Bureau of Hospital Administration, Graduate School of Business Administration, University of Michigan, 1964.
Ingbar, Mary Lee and Lester D. Taylor. *Hospital Costs in Massachusetts: An Econometric Study,* Harvard Press, 1968.

medical necessity and, in part, to the convenience of either the consumer or the physician. For emergency and urgent admissions, the cost associated with limiting the immediate accessibility of facilities will have a medical basis. The use of the phrase "convenience," should not imply that these costs are not important. From the patient's point of view, less convenient medical service tends to discourage use which might yield medical benefit. The convenience of the physician may be even more significant. If physicians' services are the scarcest input in the production of medical care, increasing convenience may increase productivity since more patients can be seen in the setting of the hospital than outside. This is particularly relevant when considering the higher admissions rate and greater length of stay where the population is thinly spread over broad geographic areas. The hospital often serves as a locus in which many patients living far apart can be served by a single (or few) physician(s). The standards of adequacy in the Hospital Survey and Construction Act (Hill-Burton) incorporated the assumption that areas of less dense population would require more facilities (of smaller size) to provide the same level of care as could be provided by fewer facilities elsewhere. This reflected both the less intensive use of smaller facilities and the "convenience" aspects mentioned above.

30. The costs associated with expanding the scope of a hospital's market are not identical for all elements of the output of medical facilities. For many forms of medical care output, the market is perhaps already too broad. For example, cost of not having immediate access to a blood bank is probably much greater in a community than is the cost of not having access to facilities for performing kidney transplants. For this last mentioned service, existing facilities now serve patients from a large geographic area. There is a lack of competent physicians to staff great numbers of such facilities. Moreover, from an economic point of view the desired more intensive use of resources requires a higher level of demand than most small markets are likely to generate.

31. Implicit acknowledgment of scale economies, intensity of use, and the costs of inaccessibility are embodied in the notion of regionalization and areawide planning. The regionalization concept organizes care around a major teaching hospital which provides a full range of highly complex and costly services. The central hospital provides these services for a broad geographical area which is also served by other institutions producing less complex care. These less complex institutions serve smaller geographic areas and fulfill the accessibility requirement while attempting to maximize the degree to which more complex and expensive services can be utilized. Patients enter the hospital system at the level of the local hospital. Those requiring more complex services are referred to intermediate or primary facilities. In principle, there is much evidence to suggest that such an arrangement would make possible a more efficient use of medical facilities although the usual reason for regionalization is given as improvement in the quality of care.

32. There are a number of constraints which make such a reorganization difficult. Experiments in planning, such as those in Michigan, suggest two major problems.[9] One factor involves the control of most

[9] McNerney, Walter J. and Donald C. Reidel, with Darwin O. Finkbeiner and Edward M. Dolinsky, *Regionalization and Rural Health Care: An Experiment in Three Communities*, University of Michigan, Ann Arbor, 1962.

institutions. The community in which a hospital is located may desire and be willing to pay for, a broader range of services than is consistent with the planning goals. Individual facilities with a community may not wish to play a secondary role in the provision of medical care. Each hospital in a community may decide to provide a range of services which could be provided in sufficient amounts by a single hospital. Perhaps the most single factor which renders this part of the operating system for medical care services indifferent to the costs involved in such decisions, has to do with the reimbursement structure under Blue Cross and title XVIII of the Social Security Act. These two sources of payment for hospitalization represent the large and growing share of all of the income to the general hospital. Both of these reimbursement structures have incorporated within them the commitment to pay the actual costs of providing care "however widely they may differ among institutions." This cost reimbursement structure renders the producer indifferent to the cost of production both in determining the kinds of goods and services which it will produce in a manner in which the production will take place. Considerable effort has been spent in recent years attempting to develop reimbursement systems which do not incorporate within their structure this automatic "cost passthrough." Experience in the areas of military procurement and in more recent years specifically for hospital services has led to a general recognition of the economic inefficiency which is stimulated in policies of this kind. To the extent that such cost reimbursement structures are not adapted to the economic realities of the marketplace, these incentives will continue to provide stimulus for ignoring the objectives of economic efficiency in the operating structure for medical care services.

33. Another deterrent to a more efficient organization of facilities is the institution of staff privileges for the physician. Under current practice, few hospitals allow physicians other than those on their staff to admit or treat patients. Under such a system, in order for a physician to transfer a patient to a more complex facility where he has no privileges, he must give up his patient to another physician. The physicians in the top level institutions argue against opening up staff privileges to physicians in smaller communities on the grounds that they are not as technically competent or that the staff of their institution is large enough. The physicians in the smaller communities press their local institutions to provide facilities in which they can produce more complex care.

34. The economic implications of such behavior are evident. For the individual institutions, duplication of equipment may result in underuse of each piece of capital and thereby raise the average cost in each hospital. Since these costs have to be borne, to a large extent, by *all* of the users of the facility and by the *community as a whole* (through community rated insurance, welfare payments, etc.) the *overall* cost of care to the community is increased. In addition, the increased demand for such facilities will divert facilities from other areas and divert resources away from other uses and into the production of these *unnecessary* services. If the production of these services requires the use of other personnel (i.e., lab technicians, radiological assistants) which are in scarce supply, the result might be a diminution of quality. The reallocation of resources stemming from one com-

munity's increase in demand will serve to make services more costly to *other* communities.

35. Those costs which fall on communities and groups in the communities outside the one which is making the decision to acquire resources can be called external costs. While the individual institutions and communities involved in the decision to provide for themselves services which might have been provided elsewhere consider the costs *to them* of this decision relative to the benefits they derive, the relevant costs include the external costs. In this case, as in the case in paragraph 19, efficient behavior for the individual institution may be inefficient for the community as a whole. The failure to consider external cost when making decisions about allocation is typical of hospital planning. To the extent that planning can be made to take place on a broader basis (community, areawide, etc.) more of the external costs may be considered and it is more likely that economically efficient decisions will be made. As part of the reimbursement structure under title XVIII, individual institutions are entitled to fund depreciation of capital equipment. This would create within each institution a source of capital funding which is indifferent to the system's objectives in the same manner that the cost reimbursement structure pays for whatever costs the institution incurs. However, the suggestion has been made that such funded depreciation be pooled over a broader section of the medical care system and, to the extent that this suggestion becomes incorporated into policy, should provide a stimulus, at least with regard to capital expenditures, for evaluating the needs of the system as a whole before it is time to vote additional resources to increase capitalization.

36. One form of organization which has received attention is *progressive patient care* in which the hospital itself is divided into intensive care, intermediate care, and ambulatory care units. Under such an organization, most of the resources are organized around the intensive care units and the fewest resources and services are provided to the ambulatory care units. Since this arrangement only applies within a single institution, little potential exists for expanding the scale of operation and thereby achieving some economies of scale.

37. Another, potentially more significant approach, deals with the relation between the hospital facility and other medical care institutions which provide a *less complex* form of care such as convalescent care facilities, nursing homes, and facilities for chronic diseases. Such relations have been stimulated as a matter of policy as part of the benefit structure of the title XVIII hospital insurance. For the period of illness during which the patient does not need the extensive facilities of the general hospital, it should be possible for these other medical care facilities to provide less costly and equally satisfactory care. This reasoning has more relevance in the light of increasing incidence of chronic diseases with long periods of low level care interspersed with brief periods of intensive care. It must be emphasized that the distinguishing feature of the various medical care institutions should *not* be the length of stay. What is relevant is the *intensity of care.* There is little basis on which to differentiate between 4 months of nursing home care and 5 years.

38. The question basic to the discussion here is whether or not the creation of facilities to provide care of low-level complexity is likely to result in a more efficient use of resources. The answer depends on

a number of things. If there is an increasing demand for more facilities which can be met either by building more complex facilities to be used in the customary way or by building less complex facilities at lower cost and thereby freeing the existing facilities to meet the increased demand, the total cost of medical care to the community will be less under the latter arrangement. On the other hand, if building convalescent facilities results in a less intensive (and thereby more costly) use of the existing complex facilities together with an increased substitution of convalescent facilities for the home (rather than for the hospital) the net result may be to increase greatly the cost to the community. The fact that convalescent care has a lower average cost than general hospital care conceals the fact that, if the general hospital facilities are not already fully used, the cost of consuming the extra days of convalescence in the hospital may be much lower than the cost of building a new (albeit less expensive) convalescent home. Certain hospitals are in the process of experimenting with establishing relations with nursing homes whereby instant reentry to the hospital is guaranteed by the institution. In this manner, the individual hospital hopes to utilize its equipment over a greater number of patients at a higher level of intensity.

39. Ambulatory patient facilities within the hospital have increased. Such facilities encourage the more intensive use of capital equipment for diagnosis and treatment by enabling the hospital to spread their use over more patients than can be contained within the institution. In many institutions, the emergency room has evolved into a substitute for the physician's office and the hospital has become the family doctor. Many of these changes are related to the far-reaching alterations that have taken place in the organization and distribution of physicians' services and it is this aspect of the supply of medical care which must now be discussed.

PHYSICIANS' SERVICES

40. The central figure in the organization and distribution of medical care has always been the physician. Because of this central role, it is easy to fail to appreciate the radical and far-reaching changes that have taken place, both in the nature of what constitutes physicians' services and in the way these services are used to produce medical care.[10] An increasing trend toward specialization may be seen as a narrowing of each physician's area of competence with respect to the individual patient as a whole in order to maintain a high level of competence for that part of the patient's medical requirements which fall within the specialist's jurisdiction.

41. The fact of nonsubstitutability among physicians implies that, for the purposes of supplying medical care, the number of available physicians tells little. Much of the discussion of the past few years calling attention to shortages of physicians has totally ignored the possibility that even very great increases in the number of doctors might not yield an increase in desired physicians' services. (See paragraphs

[10] There are a wide range of issues of medical manpower which deal with the elasticity of supply which are not pursued in this paper. They are, however, of importance for policy. For a discussion of career choice, job attachment, and supply elasticity in medical manpower markets and their implications for policy see: Rosenthal, Gerald. *Economic Dimensions of Health Manpower Planning.* Paper presented to annual meeting of the American Public Health Association, October 1967.

5 and 6). This emphasis on training specialists has led to repeated calls for medical schools to produce more general physicians and one medical educator has called for separating out the potential generalists at an earlier stage in medical training.

42. The above mentioned trend toward specialization of physicians' services is closely related to the increasing amount of capital equipment centered within the hospital. (See paragraphs 21 and 22). The degree to which specialities are dependent on capital equipment varies from almost totally, for specialities such as neurosurgery and radiology, to very little, for specialities such as phychiatry and dermatology. The varying level of capital-intensity of the production of physicians' services has a number of implications for the allocation of resources in medical care and for the productivity and pricing of physicians' services.[11]

43. While not conclusive, there is evidence to suggest that, in terms of patients seen per unit of time (as a rough productivity measure), specialities with a higher degree of capital intensity (as measured by relative dependence on the hospital) have had greater increases in productivity than have nonhospital related specialities. Since the price per unit of output has not declined proportionally, the incomes of capital-intensive specialities have tended to be higher than those of less capital-intensive specialities and the absolute differential has increased over the past few decades. At the same time, as a result of the increasing differences in incomes among the specialities, the prices of non-capital-intensive physicians' services have tended to rise more than other physicians' services. While these observations must be corroborated by more intensive research, early evidence suggests that these differences are not accounted for by differences in the length of training (i.e., psychiatry has a long training period, yet falls on the low end of the income scale among specialities), hours worked, or any of the usual explanations. The new physician, by choosing a more capital-intensive specialty, can expect a higher income without additional investment, *since the community, through the hospital, provides the equipment at little or no cost to him.*

44. In a perfect economic market, the specialities in shortest supply relative to demand should show the largest return. This would encourage a shift in the proportions of physicians' services from the least scarce to the most scarce (i.e., from least lucrative to most lucrative). However, the differential impact of productivity described above (par. 43) has created a set of returns to the physician that make hospital-based specialities far more economically rewarding. The result has been to distort the supply of physicians' services. While more documentation is needed, the evidence suggests that surgeons are in relative abundance while general practitioners and psychiatrists are relatively scarce. In some areas, surgeons have adjusted to this by becoming more flexible and providing some general medical services, although another likely effect would be to increase the use of expensive hospital facilities in order to use acquired skills.

45. The preceding discussion dealt with the relation between increased specialization and capital intensity. The productivity of physicians was seen to be centered around the hospital as the locus of

[11] Penchansky, Roy and Gerald Rosenthal. "Productivity, Price, and Income Behavior in the Physicians' Services Market—A Tentative Hypothesis." *Medical Care.* October–December 1965.

capital. (Pars. 40–44). The hospital also has an impact on the productivity of physicians through the provision of a single locus of service. As mentioned earlier (par. 29) in areas where the physician is a scarce resource, centering his practice around the hospital may enable him to provide considerably more service by minimizing nonmedical activities such as travel time. In this case, the patient substitutes his travel time for the physician's. Even in urban areas, the physician may find considerable saving of time by using the hospital and, thereby, increase his productivity. This can take two forms. He might tend to keep patients in the hospital longer or admit them more often in order to make followup care easier by concentrating his responsibilities. In addition, he may substitute the hospital's emergency service for his night calls and other medical activities outside of more convenient working hours. This tendency is accentuated by the structure of prepayment and health insurance (par. 18).

46. While there is little question that the use of a hospital setting makes the physician more productive, considerable attention has been directed at other forms of organization of physicians services which can have the same impact. Of primary importance is the effect of group practice. In organizational terms, the group represents two changes in medical practice. It is, first, a formalized set of referral patterns which is held to maximize the communications among the various physicians needed to provide for the total illness experience. To the extent that such an organizational form reduces the duplication of information gathering and retesting often associated with the handing on of patients from specialist to specialist, a real increase in efficiency is gained. It should be noted that such an increase in efficiency does not require the formation of a group. The rising trend toward physicians' offices within the hospital, and even the rise of the medical arts building, providing a central locus for physicians, has contributed to increased communication and referrals among different specialists. This tendency is particularly noticeable among dentists. It is not rare to find a general dentist, a dental surgeon, and other dental specialists sharing the same offices and auxiliary personnel.

47. The group often provides a setting for the location of capital outside of the hospital (and not owned by the community). Much of this is associated with the increased intensity of use discussed earlier. (Pars. 21 and 22). Particularly with respect to diagnostic equipment, the group practice involves a number of physicians who, in the aggregate, may provide sufficient use of equipment to merit its purchase. In addition, the group structure enables the pooling of funds with which to acquire such capital. This may encourage a more efficient practice of medicine by placing the capital equipment, the various specialities, and a single patient in a consistent relationship.

48. The degree to which this capital accumulation outside of the community facility, and for the use of only a small part of the physicians, contributes to the overall efficiency of medical care depends on a number of factors. If the intensity of use made of capital equipment by the group is less than that which would be made of such equipment if available to the whole community, it may be a less efficient use of capital. More important, however, is the impact on the intensity of use of the community facilities which occurs as a result of the group's decision to acquire their own equipment. This same argument has relevance for the trend, in some areas, of physicians to invest in their own

hospital facilities. The result can be to increase the efficiency of some medical care while decreasing the efficiency of other medical care. The net impact of such a tendency is not automatically one way or the other.

OTHER MEDICAL PERSONNEL

49. One of the most significant changes in the productivity of physicians over the past decades stems from the radical changes that have taken place in the use of other medical personnel. The increase in both number and diversity of other medical personnel has made the medical care industry one of the largest sources of employment in the U.S. economy. There is great variation in the levels of skills involved in these jobs. It is estimated that approximately 50 percent of the personnel in a nonteaching general hospital can be trained within 6 weeks, while many jobs, such as those of registered nurses, require years of training.

50. At least some considerable segment of this employment represents a transfer of functions from higher to lower skill levels directed at making the higher skills more productive. Many practicing physicians can remember the time when it required a doctor to take a blood pressure, a job which, today, is routinely performed by nurses, technicians, and other less skilled (and less costly) personnel.[12] This trend is most observable in dental practice where in many sections of the country, all prophylactic dental care is provided by the hygienist. It is interesting to note, however, that in many areas of the country such a practice is accepted neither by the patient nor the dentist. This downward transfer of functions is observable at all levels. Bedside care, which was once the sole province of the nurse, is now often provided by practical nurses, nurses aides, and even nursing students. Even among physicians, such a downward transfer of functions occurs. For many intricate surgical procedures, the final, less difficult, stages of the procedure are done by a general surgeon or a resident while the more difficult part of the procedure is done by a more highly skilled surgeon specialist.

51. From an economic point of view, any substitution of less costly for more costly inputs that does not result in an offsetting reduction in either the quality or quantity of output is an increase in efficiency. Merely looking at the production of units of hospital service, one finds a considerable increase in inputs, both in terms of capital per patient day and in personnel per patient day. However, in the absence of the considerable lessening of the skill embodied in many of these personnel, the increase in costs would have been very much greater, if, indeed, the resources necessary to produce medical care according to the old formula would have been forthcoming at all.

52. Much of the foregoing discussion with respect to physicians' services has relevance for other medical manpower. Not only have physicians and dentists been made more efficient by the use of paramedical personnel, but at each level of skill, less scarce resources have tended to be substituted for more scarce resources. This type of change yields a more efficient use of resources and increases the product which can be obtained from a given quantity of resources. The observed economic

[12] Weiss, Jeffrey. *The Job Structure of Health Manpower.* Unpublished Ph. D. dissertation, Department of Economics, Harvard University, 1965.

behavior of the submarkets in which the supply of medical care is produced demonstrates a dynamic set of responses to changes in demand and changes in technology. It is essential that the dynamic nature of the medical care industry be understood and acknowledged when evaluating the performance potential of the industry.

Public Policy and Medical Care Systems Operations

53. The preceding discussion of medical care systems operations has made clear the degree to which the operators in that system respond to a wide range of influences generated both within and outside of their sphere of control. The process of planning was described earlier (pars. 2–4) as being deliberate attempts to affect the behavior in the system by generating such influences from outside the operating sector. These influences are called planning policies and may take many forms including legislation, regulation, or encouragement. In general, planning policies may be described along a continuum ranging from prohibitive and restrictive policies on the one hand to stimulative and coercive policies on the other.

54. The extent of planning activities, as described above, which currently operate in the medical care system, would be difficult to overestimate. At every point in the system a wide range of deliberate influences exist to affect the operators' behavior. Often, these influences are not consistent in their direction and may represent attempts to serve conflicting objectives. The general discrepancy between system's objectives and operators' objectives is replicated throughout the system at every level of decisionmaking. The use of policy tools to implement the Nation's objectives with regard to the medical care system requires extensive understanding of the directions and strengths of the existing influences. The growing public involvement in medical care at all levels coupled with shifting expectations makes increasing interest in and reliance on formal planning structures inevitable. Such a shift places a much higher premium on explicit setting of objectives and effective planning, as measured by the degree to which the objectives are secured. The selection of policy tools to serve these ends will rely heavily on an understanding of the current operating structure of the medical care system. The purpose of this discussion has been to contribute to this understanding.

88-744—68—vol. 2——13

THE BASIC FORCES INFLUENCING COSTS OF MEDICAL CARE

BY VICTOR R. FUCHS*

INTRODUCTION

My aim is to indicate how an economist goes about analyzing the basic forces influencing costs of medical care. This is not just another way of saying that I will give you my opinions—I will do that also—but it is an attempt to take explicit note that economics is, above all else, a way of looking at questions. In Lord Keynes' words, "The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions."

To be sure, even among economists there is not always just one way of looking at things. Winston Churchill used to complain that whenever he asked Britain's three leading economists a question, he received four different answers—two from John Maynard Keynes. Nevertheless, there is a common fund of concepts, a common core of analysis, that nearly all economists use. When, in the course of applying these concepts, my own value judgments or empirical estimates appear, they will be appropriately labeled.

The basic analytical approach is a consideration of those factors affecting the demand for medical care, and those affecting the supply. Demand and supply, the two magic words. Some of us, when visiting hospitals, have discovered that by putting on a white coat and talking rudely to nurses, it is easy to pass as a physician. To be mistaken for an economist is often even simpler. All one need do is nod gravely and say "demand and supply."

DEFINITIONS OF TERMS

Demand for and supply of what? I shall assume that medical care refers to the services rendered by physicians, dentists, and other health professionals, plus all the goods and services consumed in connection with their work, or upon their direction. Thus, the costs of medical care include the costs of hospitals, drugs, and the like. This lumping of diverse health services is a concession to convention and to the limitations of time. Ideally, one should apply the demand-supply analysis separately to hospitals, dentists, drugs, and so on because the forces that influence the cost of one type of health service are often different from those that influence another.

---

*Associate Director of Research, National Bureau of Economic Research, New York. NOTE: This address was given at the National Conference on Medical Care Costs, Washington, D.C., June 27, 1967.

What is meant by costs? At least three possible meanings can be distinguished. It could mean price, or cost of production, or expenditures. When people speak of the rising costs of medical care, they frequently are referring to rising expenditures, and this is the way I shall use the term this morning.

EXPENDITURE TRENDS

We all know that these expenditures have been growing rapidly. In round numbers, expenditures for medical care have risen from under $4 billion in 1929 to over $40 billion in 1965 and probably close to $50 billion in 1967. Even as recently as 20 years ago, expenditures were only $10 billion. Of course, expenditures for most other goods and services have also risen; it is therefore more meaningful for some purposes to look at the share of total spending allocated to medical care. This too has risen, from under 4 percent in 1929 to about 6 percent in recent years. Nearly all of this relative increase has occurred since 1947.

Before examining the factors responsible for this trend, it is worth noting that there is nothing wrong *a priori* with changes in industry and sector shares of gross national product. Indeed, such changes seem to be a natural concomitant of economic growth. For instance, the relative importance of agriculture has declined precipitously in most Western countries. During the last half of the 19th, and the first half of the 20th century, there was a significant rise in the relative importance of manufacturing. Now we are witnessing in this country the growth of what I have described elsewhere as the "first service economy." [1] If agriculture's share of GNP falls from over 9 percent to under 4 percent, as it did in the United States between 1947 and 1965, some other industries must show increases. There is no magic in the 4 percent figure for medical care; it is now 6 and it could be 8 or 10.

REASONS FOR CONCERN ABOUT COSTS

Why then should there be a national conference on the costs of medical care? Let me suggest three reasons for concern.

First, questions arise in my mind concerning the contribution that these increased expenditures make to health. Although we spend much more per person for medical care than any other country, the blunt truth is that we do not enjoy the highest health levels. On the contrary, many European countries have age-specific death rates considerably below our own. The relatively high infant mortality rate in this country is disturbing, and difficult to explain. The disparity in death rates for middle-aged males is even more shocking, and has more serious economic implications. In the United States of every 100 males who reach the age of 45, only 90 will reach 55. In Sweden the comparable figure is 95. During this critical decade when most men are at the peak of their earning power, the U.S. death rate is double the Swedish rate, and higher than that of almost every Western nation. It certainly seems legitimate to ask why. This is not necessarily with a view to spending less for medical care—I doubt if anyone can foresee a decline—but with a view to developing more effective use of the resources that we are now devoting to health.

---

[1] Victor R. Fuchs, *The Growing Importance of the Service Industries*, Occasional Paper 96, National Bureau of Economic Research, Columbia University Press, New York, 1965.

A second reason why we should be concerned about medical care costs is the peculiar structure of the medical care industry. Most industries in the United States consist of profit-seeking firms actively engaged in competition with one another. The fundamental rationale of the American economic system is that the hope of profit (and the fear of loss) under conditions of open competition are the best guarantees of efficiency, and appropriate price and rate of output, and a fair return to the various factors of production.

The medical care industry is organized along radically different lines. Nonprofit operations are the rule in the hospital field; there are severe restrictions on entry and competition in medical practice, and advertising and patent control dominate the market for drugs. Thus, there is no *a priori* basis for believing that the prices and quantities of medical care approach those that would be socially optimal.

A third reason, it seems to me, is that a large and increasing portion of the cost of medical care is paid by third parties. In particular, the taxpayer is being called upon to pick up a substantial share of the bill. Because payment for medical care is increasingly regarded as a collective responsibility, it is natural and appropriate that there should be collective expressions of concern, such as this conference reflects, about the quantity and quality of medical care, and about its price.

These quantities and prices are determined by demand and supply. Let us consider each side of the equation in turn.

Demand for Medical Care

Economists say that the demand for any good or service depends upon relative prices, income, and tastes.

PRICE

How does price affect expenditures? Perhaps the most firmly established proposition about the demand for medical care is that it is relatively inelastic with respect to price. If the price rises relative to other prices, the decline in the quantity demanded will be proportionately less than the increase in price. The result is an increase in medical care expenditures. If, other things remaining unchanged, price rises by 10 percent and quantity demanded falls by only 5 percent, expenditures will rise by approximately 5 percent. Some studies suggest that the price elasticity of demand for medical care may be as low as 0.2, i.e., quantity demanded declines by only 2 percent when price rises by 10 percent. But present knowledge does not permit fixing a specific value other than to say that the elasticity is surely below unity.

An aspect of the price of medical care that is not widely recognized, is that it really has two components. One is the nominal price charged by the physician or hospital; the other is the value of the patient's time.[2] For instance, the nominal price of a visit to a physician might be $10, but the trip to and from his office, the wait, and the actual examination will probably take an hour or more. This time might be worth more or less than $10 depending upon the alternatives available to the patient.

---

[2] Gary S. Becker, "A Theory of the Allocation of Time," *Economic Journal*, 75, No. 299 (September 1965), pp. 493–517.

Once it is understood that the price of medical care includes both components, a number of interesting implications become apparent. Even when a sliding fee scale is not used, the total price of medical care tends to vary with earning power. The price is lower for retired people and the unemployed than for those with jobs, is generally lower for women than for men, and so on. Also, even when the nominal price is reduced to zero, as under prepayment plans or socialized medicine, the true price is not zero.

INCOME

One of the factors to be considered in any demand study is real per capita income. During the past 20 years this has risen by over 50 percent, and there is no doubt that the demand for medical care increases with income. What is less clear is whether the demand for medical care is elastic or inelastic with respect to income, i.e., does a given percentage increase in income lead to more than, or less than, the same percentage increase in medical care expenditures, other things remaining the same. This question is only gradually yielding to attack as more and better data become available and analytical techniques are sharpened. Some recent studies suggest that the elasticity may be significantly below unity, and few investigators believe that it is greater than unity. At most, the demand for medical care seems to increase approximately in proportion to income. If this is true, we cannot attribute any of the increase in the share of total expenditures accounted for by medical care to rising income.

INSURANCE

A special factor that complicates the analysis of the demand for medical care is the growth of insurance and prepayment plans. Once a person is covered by such a plan, the effective price to him of additional units of medical care depends only upon the value of his time. It seems to me that this may explain a large part of the increase in the quantity of medical care demanded, and may also help explain the apparent insensitivity of insured consumers to increases in the nominal price of medical care. It is worth noting that hospital care has shown the most rapid rate of increase in expenditures, and it is hospital care that has been most thoroughly covered by insurance and prepayment.

The curious behavior of dental expenditures also offers support for this hypothesis. All the available evidence suggests that at any point in time the demand for dental care is more elastic with respect to income than is the demand for physicians' services. Nevertheless, during these recent decades of sharply rising real income, expenditures for dental care have increased less than have expenditures for physicians' services. One possible explanation is the very small role played by insurance and prepayment plans in the dental field. Expenditures for eyeglasses and appliances, and for drugs, two other components of medical care that are typcially paid for directly by the consumer, have also risen much less rapidly than have expenditures for hospitals or physicians.

This should not come as a surprise. The advocates of insurance and prepayment had something like this in mind. They wanted to remove any financial barriers to obtaining medical care. But it is a basic law of economics that if you lower the price, the quantity demanded will

increase. A critic of the British National Health Service put the matter cogently, albeit a bit strongly, in a recent issue of *The Lancet*. He wrote, "If taxi fares and meters were abolished, and a free national taxi service were financed by taxation, who would go by car, or bus, or walk * * * the shortage of taxis would be endemic, rationing by rushing would go to the physically strong, and be more arbitrary than price, and 'the taxi crisis' a subject of perodic public agitation and political debate."[3]

I am not suggesting that insurance and prepayment should be abandoned. But we do need to discover techniques, possibly such as coinsurance, deductibles, or experience rating, to check prices and expenditures without interfering with essential health services.

TASTES

All factors other than income or price that affect demand are put by economists in a catchall category called taste. In the case of medical care, these would be the factors that affect the health levels of the population, and those that affect attitudes toward seeking medical care at any given level of health. Taste for medical care, therefore, would be related to: (1) Demographic variables, (2) education, (3) environment, (4) ways of living, and (5) the genetic stock of the population.

Research on these matters is only in its infancy, and there are few reliable findings to report. We know that an increase in the proportion of elderly people in the population tends to increase the demand for medical care, other things remaining the same. The effect of increased education is unclear. It probably leads to improved health levels, and thus less need for medical care, but may also lead to greater demand for medical care at any given level of health.

Most observers believe that recent environmental changes, particularly the increase in real income per capita, have contributed to better health status. I think that this inference is incorrect. Some tentative findings from my research suggest that the environmental and lifestyle changes of the past two decades have had either a neutral or negative impact on health for most of the population. One piece of evidence in support of this hypothesis is the stability of age-adjusted death rates in the United States in the face of large increases in medical care and improvements in medical science.

All these questions, however, are in need of more study. The National Center for Health Statistics is now developing vast new bodies of relevant data. I believe that a combined assault on these data by health experts and social scientists will yield information comparable in importance to that emerging from the laboratory in our continuing efforts to understand and improve the Nation's health.

ACCOUNTING ILLUSION

In concluding this discussion of demand, it should be noted that part of the observed increase in medical care costs is an accounting illusion. It does not involve any increase in real costs—only money costs. It is the result of an increase in the proportion of medical care

[3] Arthur Seldon, "National or Personal Health Service," *The Lancet*, March 25, 1967. No. 7491, vol. 1, p. 675.

produced and sold in the market, and a decline in the proportion provided outside the market by family, friends, and neighbors. Only the former is included in the GNP. A generation ago, a considerable amount of bed care, and associated services for the sick, were provided for at home. Surely there is relatively less of this today.

Some of the reasons for this shift other than increases in income and insurance coverage are: (1) Urbanization, (2) the fragmentation of the family, and (3) the increased labor force participation of women. We do not know how much of the increase in observed medical care costs can be attributed to this shift; I believe that the amount involved is substantial. One corollary is that "home care" programs and other current plans to transfer costs back out of the hospital will reduce the money costs of medical care by more than they will reduce real costs.

Supply of Medical Care

I turn now to the supply of medical care. In studying the supply side of an industry there are three main elements to be looked at. The first is the supply of the factors of production—labor and capital—flowing into the industry. The second is changes in productivity, and the third is the degree of monopoly control, or other market imperfections that may influence the supply actually available to consumers.

SUPPLY OF PRODUCTIVE FACTORS

With respect to the supply of labor *to* the health industry, the crucial question is whether the industry has to pay inordinately high wages in order to attract an increasing fraction of the total labor force. In my judgment, the answer to this is "No." In technical terms, the supply of labor to the medical care industry is very elastic. This is true, incidentally, of most other industries as well. Except in the extreme short run, the U.S. labor force is highly mobile and adaptable; studies of interindustry differences in earnings consistently refute the hypothesis that expanding industries must pay unusually high wages to bid away labor from other industries.

Between 1950 and 1960, medical care employment rose by 54 percent—compared with only 14 percent for total employment. Throughout the postwar period the annual rate of increase has been about 5 percent for medical care employment compared with a little over 1 percent for the economy as a whole. Despite this rapid expansion, wages for medical care personnel seem to have been rising at about the same rate as in many other industries. This last point has not been thoroughly documented, but is the most reasonable inference from the data available.

An analysis of the supply of capital to the medical care industry is much more difficult to undertake because most capital is used in hospitals, and most hospitals are nonprofit. Thus, the flow of capital is not determined by the rate of profit (as it is in most industries), but by Government decisions and fortuitous philanthropy. It is possible, however, to devise methods of financing and reimbursing hospitals that would make the flow of new investment more responsive to market-type mechanisms. The Soviet Union and other Socialist nations have been attempting to do precisely this with substantial portions of their "nonprofit" economies.

PRODUCTIVITY

Changes in the supply of any goods or service, in the sense of changes in the price-quantity relationships, depend primarily on changes in productivity. It is commonplace to argue that productivity in medical care has advanced less rapidly than in the economy as a whole; but in the absence of reliable measures of the output of medical care this must remain a matter of speculation.

The development of such measures is an extremely difficult task because of our ignorance concerning the precise contribution of medical care to health. In addition, output is not limited to improvements in health but takes other forms including validation services and the hotel aspects of hospital care.[4]

There is some reason to believe that the available measures understate the true output of the medical care industry. A visit to a physician today is surely more productive than one 20 years ago, and this is even more true of a patient-day in a hospital. On the other hand, it is possible that many of the expensive procedures that are now part of "best practice" techniques are really not worth the money in the sense that their marginal contribution is small and the same amount of resources used in other ways would yield more utility to the consumer.

The common practice of reimbursing hospitals on the basis of their costs, as under medicare and many other public and private programs, appears to be an open invitation to inefficiency. At best, the ability of hospital management to improve productivity is imperfect because of the independence of the attending staff. Under present arrangements, almost no one has any incentive to be concerned with the efficiency of the hospital as a whole.

Another weakness in the hospital supply picture is that, with few exceptions, each hospital is independently "owned" and managed. Unlike other industries where an exceptionally able manager gradually comes to exercise supervision over an increasingly large pool of resources through the growth of his firm, through mergers, and through establishment of branch plants, this pattern is absent in the hospital field. Also, it is much easier for inefficient management to remain in charge for long periods of time.

PHYSICIANS

The physician plays a key role in the supply of all medical care; his decisions and behavior affect almost everything else. Physician supply is now more specialized than formerly. This growth of specialization is often attributed to exogenously determined advances in medical science, but such an explanation ignores the role played by changes in demand. Two hundred years ago, Adam Smith observed that the division of labor is limited by the extent of the market. The relevant market for any one physician's services has grown tremendously because of the growth of income and population, the increased concentration in urban centers, and improvements in transportation. All these trends would lend to increased specialization, even if medical technology remained static. Moreover, given an increase in real income, people want to buy more medical service for any given health

[4] Victor R. Fuchs, 'The Contribution of Health Services to the American Economy," Milbank Memorial Fund Quarterly, vol. 44, No. 4, pt. 2, October 1966, pp. 65–102.

condition. One way of buying more service would be to visit several different general practitioners, or to visit the same one several times. Alternatively, one can buy more medical service in each visit through the use of specialists. The specialist in medicine usually has more, not merely different, training than a general practitioner. The more valuable the patient's time, the greater will be the demand for "high powered" doctors. This demand-induced growth of specialization is thus a cause as well as a result of advances in medical science. Without a specialized practice, without the demand for specialized equipment and procedures, these advances would probably come more slowly.

Physicians have frequently been critized because of their high earnings and their alleged desire to restrict their numbers. Such criticism, it seems to me, does not go to the heart of the matter. Most of the difference between the earnings of physicians and those of other occupations should not be attributed to their control over entry and competition, but to the long hours that they work, the lengthy period of education required, and the absence of pensions, paid vacations, and other fringe benefits. Moreover, physicians' earnings account for less than 20 percent of total health expenditures, and to the extent that they enjoy some monopoly return it could only be a small part of this fraction.

A more valid criticism, it seems to me, can be directed against physicians for their opposition to changes in the methods of producing and financing medical care. The medical profession, or at least a significant and articulate portion of it, seems to believe that there can be rapid and far-reaching technological change without disturbing the traditional organization of medical practice. This belief is irrational. One clear lesson from economic history is that technological innovation means organizational change.

Possibly the most harmful aspect of physicians' market control is the extremely narrow range of options available for someone seeking personal medical care. One bit of evidence is the size distribution of earnings in the entire medical care industry which can only be described as unnatural. Nearly all American industries have a distribution which reflects a fairly smooth vertical hierarchy of personnel. There are usually large numbers performing routine functions; and relatively fewer persons at each successive stage of increased power and responsibility. For instance, the *1960 Census of Population* shows that in nearly every industry the number of persons with earnings from $7,000 to $10,000 in 1959 far exceeded the number with earnings above $10,000. Only in the medical care industry do we find almost a void in the $7,000 to $10,000 category; those above $10,000 are three times as numerous. Today the void is in the $10,000 to $15,000 range.

Whether consumers would use less expensive medical care personnel, if they were available, would depend upon a number of factors—the institutional setting and supervision, whether there is a financial incentive to do so, and so on. That it is technically possible for professionals with fewer than 10 to 12 years of training beyond high school to render useful medical care has been repeatedly demonstrated in a variety of settings.

As some of my earlier remarks suggested, patients with high incomes, and patients with acute conditions would undoubtedly continue to seek the highest possible level of training and experience. But the demand for something less might be large in cases of chronic illness,

or in isolated communities, or among those with low incomes. To say that everyone should get "highest quality care" is a counsel of perfection that presently deprives many people of the opportunity of getting even moderately good care. The natural conservatism of doctors, allied with the strong egalitarian drives of some social reformers, has served to limit the supply of medical care below that which would be available in a freer market setting.

NEW MEDICAL TECHNIQUES

One special feature of the supply of medical care is the appearance of radically new medical techniques and procedures. Normally, when economists speak of the supply of a commodity they assume that the quality of the commodity remains unchanged. This is almost never strictly true, even for such staples as coal or wheat, but frequently the change in quality comes gradually, can be objectively measured, and an increase in quality can be thought of as a decrease in price.

In the case of medical care, some of the new procedures such as renal dialysis and open heart surgery are so radically different from anything previously available that they cannot conveniently be analyzed in this manner. Part of the increased expenditure for medical care is undoubtedly attributable to the appearance of these new techniques for treating conditions that simply could not be treated before.

CONCLUSION

SUMMARY OF THE DEMAND-SUPPLY ANALYSIS

What conclusions emerge from this analysis of demand and supply? By now it should be clear that cost is the result of many forces, that rising costs are not necessarily bad (or necessarily good), and that economists have some interesting questions to ask, but are far from being able to supply all the answers. Many of the estimates have a large range of uncertainty, but sustained scientific investigation can reduce that range and increase understanding.

If we take as our analytical task the explanation of why medical care now accounts for 6 percent of gross product instead of 4 percent, as it formerly did, the following developments all seem to have played a role:

1. An increase in medical care prices relative to other prices facing a relatively inelastic demand. These price increases are probably related to the institutional rigidities that surround the organization and production of medical care.
2. The growth of insurance, prepayment, and other forms of third party payment.
3. An increase in the proportion of elderly people in the population.
4. A shift from nonmarket to market production. If we measured all costs, the increase for medical care would not be as great as the GNP accounts indicate.
5. The introduction of radically new medical techniques and procedures to treat conditions that formerly could not be treated at all.
6. More tentatively, I have suggested that there may be greater need for medical care now to offset changes in the environment and in ways of living that are detrimental to health.

Two possible explanations that I believe the available evidence rejects are the rise in real per capita income, and the alleged rapid rise in wages for medical care personnel.

SUGGESTIONS FOR DISCUSSION

Before concluding, I would like to offer a few suggestions—suggestions that perhaps merit further discussions today and tomorrow. They deal with matters that affect the costs of medical care, but involve judgments about objectives and methods in fields where I can only claim to be an interested bystander. Perhaps they can be justified on the grounds that health, like war, is too important to be left to the "generals."

1. I begin with medical education. This priority is deliberate. Given the key role that physicians play in personal health care, in determining expenditures for hospitals and drugs, and in directing the work of other medical care personnel, it seems unlikely that significant changes can occur in medical care without changes in medical education. At the risk of some exaggeration, I would argue that the biggest shortcoming of medical schools is indicated by their *name*. If they were to transform themselves into *schools of health*—with all that such a transformation implies for attitudes, objectives, personnel, and curriculum, many of the other goals that will be discussed at this conference would be much closer to attainment.

A "school of health" would have the twin objectives of training personnel and advancing knowledge to meet the health needs of the community. It would define these needs broadly, would be concerned with the future as well as the present, and would want to meet health needs at various levels. There would be several different "educational tracks" with some possibility of moving from one to another, and there would be a strong interest in continuing education. The aim would be to provide a continuum of trained personnel to deal with a continuum of health problems. Some students would be ready for professional careers with less preparation than is currently required. Others might receive even more training than physicians now do. Members of the last group would truly be "captains" of health teams—not only in name, but in spirit and in practice.

A physician emerging from such a school would take a broad view of his duties and responsibilities. It is not likely that a man so trained and so motivated would *want* to be a solo practitioner. Similarly it is unlikely that he would *want* to be paid on a "piecework" basis. He would expect, and would deserve, a good salary, but he would also want time to read and study, time to think and plan, and time to maintain his own physical and mental health. Far from regarding auxiliary medical personnel and new technologies as threats to his status or financial position, he would welcome developments that would permit the delivery of more medical care at lower cost per unit. Finally, his yardstick of success would not be the number of cases in which he personally was able to alter the course of events, but improvement in the health levels of the population that he and his colleagues serve.

2. My second suggestion concerns hospitals and has two parts. First, I would like to see widespread adoption of reimbursement plans that provide incentives for efficient operation. Such plans could establish target rates for each hospital, or establish fixed rates for groups of hos-

pitals providing comparable service. These reimbursement rates would probably be related in some way to average costs. Inefficient hospitals, therefore, would be under strong pressure to bring their costs down, while efficient hospitals would find themselves with extra funds which they could spend for improving the range and quality of service offered. Such a system might well enlist the support of attending physicians. If the medical staff realized that by holding down costs the hospital would be able to buy new equipment, or make other improvements, the hospital administrator would be in a much better position to obtain their cooperation. Reimbursement along these lines would facilitate another useful change—the development of hospital systems that include many separate and diverse types of establishments under common management. This would permit more able managers to exercise control over a larger range of resources and would also permit more efficient utilization of these resources within the system.

3. My third suggestion is to arrange for a minimum guarantee of medical care to every citizen through some sort of insurance or prepayment plan or plans. Most people can afford such plans; the minority that cannot should be subsidized by the Government. Presumably this minimum guarantee exists now, because we are often told no person need go without necessary medical care. This may be true, but there is little merit in having this guarantee rest on the judgment and benevolence of physicians or hospital administrators. It should be a common charge against the total society. We have been moving in this direction, but on a hit-or-miss basis. We are now trying to care for the aged, for some of the poor, and compulsory insurance has been proposed for all wage and salary workers. This piecemeal approach is likely to be highly inefficient and inequitable. It is time this Nation faced, in an adult way, its responsibility to assure some minimum level of medical care to all persons as a matter of right.

4. Having said this, I would also urge that we declare a moratorium on misleading talk about complete equality of medical care. This is technically not a realistic possibility, and in my view it is not a desirable objective as long as there is substantial inequality in the distribution of other goods and services. It would be most extraordinary if poor people, given an income subsidy, were to choose to bring themselves up to some common high level of medical care rather than to increase their consumption of a variety of goods and services. To arbitrarily impose this equality at the taxpayer's expense is to redistribute income in an extremely inefficient way. Moreover, insistence on equality may very well impede the development of the quantity and quality of medical care.

We can take a lesson from the field of higher education. In recent years there has been a tremendous increase in the demand for higher education, far greater than the increase in the demand for medical care. What has been the response of the higher education industry? On the whole, it has shown more adaptability and flexibility than has the medical care industry. There has been some crowding of existing institutions. There has been better use of facilities and personnel through trimester and quarter plans, and more intensive use of summer sessions. Some schools have expanded capacity. Perhaps most important of all, new institutions of considerable variety and scope have been created. In particular, one should note the growth of junior colleges.

One reason why higher education was able to respond so quickly is that the industry is not fooling itself with slogans about "equality" and "high quality for all." No one in education pretends that Small-town Junior College offers as good an education as does Harvard. There is an attempt to assure that everyone who wants higher education can get it is some form, and to some degree, and this guarantee is probably more reliable than the one currently offered for medical care.

5. My final suggestion—almost plea—is for us to remember that what we are really concerned with is health—not costs as such, and not medical care as such. My reading of the health literature leaves me with the impression that the greatest potential for improving the health of the American people is not to be found in increasing the number of physicians, or in forcing them into groups, or even in increasing hospital productivity, but is to be found in what people do and don't do, to and for themselves. With so much attention given to medical care, and so little to health education and individual responsibility for personal health, we run the danger of pandering to the understandable urge to buy a quick solution to a difficult problem. "Eat, drink, and be merry" runs the refrain. "Smoke two packs a day." "Engage in every physical and emotional excess known to man, for tomorrow you can come to Dr. Squash and have it all taken care of in two easy visits." (Some would add "at government expense.") Do we really believe that if only there are enough Dr. Squashes, or if only they practice in groups, everything will turn out all right? Let me express some doubts. I am impressed with Douglas Colman's recent observation, "Positive health is not something that one human can hand to or require of another. Positive health can be achieved only through intelligent effort on the part of each individual. Absent that effort, health professionals can only insulate the individual from the more catastrophic results of his ignorance, self-indulgence, or lack of motivation." [5]

By all means let us discuss medical care. By all means let us discuss the cost of medical care. But above all, let us discuss them in relation to the more fundamental objective—better health.

---

[5] J. Douglas Colman, "National Health Goals and Objectives," speech presented at the National Health Forum, Chicago, Ill., Mar. 20, 1967.

THE ECONOMIC STRUCTURE OF AMERICAN HOSPITALS

BY RALPH E. BERRY, JR.*

INTRODUCTION

Hospitals in the United States have traditionally been classified by ownership or control. They are also classified by the type of service rendered and the average length of stay. The emphasis on ownership is useful for certain purposes but for much economic analysis it would be advantageous to systematically classify hospitals by industry. In fact, analysis of hospitals within an industry framework can throw light on various aspects of the production and distribution of hospital services which would otherwise remain somewhat obscure.

The purpose of this paper is to summarize such a classification and subsequent analysis.

THE FOUR HOSPITAL INDUSTRIES

Within the hospital sector there appear to be four distinct subsectors or industries. It should be noted that exact boundaries are somewhat arbitrary, as they are in any classification of an economic sector by industries. The concepts of cross-elasticity of demand and cross-elasticity of supply are often useful and are applied in this attempt to classify the hospital sector. The variables generally considered crucial in classification of hospitals are the type of service rendered to the typical patient and the length of time the average patient remains hospitalized. Thus, hospitals are usually classified as psychiatric, tuberculosis, or general and other special, and as short term or long term according to the average length of stay—less than 30 days or 30 days or more, respectively. Although some general hospitals may treat psychiatric patients and some short-term facilities may have long-term patients, and vice versa, each hospital is classified by available data sources under one of the broad headings given above.

Given available data, it is convenient to consider the hospitals of the United States as each being members of one of four industries: (1) All psychiatric hospitals; (2) all tuberculosis hospitals; (3) all long-term general and other special hospitals; and (4) all short-term general and other special hospitals. These groupings seem appropriate in terms of plausible cross-elasticities of demand. In the eyes of the consumer, the hospitals produce basically different and, more significantly, nonsubstitutable products. Most consumers expect to consume psychiatric treatment in psychiatric hospitals and treatment for tuberculosis in tuberculosis hospitals. The chronically ill seek out long-term general or other special hospitals; and persons with acute conditions usually seek treatment in short-term general or other special hospitals.

*Harvard University.

On the supply side the boundaries are less sharply defined. Although facilities are specialized, many of the resources utilized in the production of hospital services of one kind can be adapted to the production of hospital services of another kind.[1] Given complementary resources, hospital beds produced to provide a specific service can be used to provide almost any kind of service. General nursing service is usually just that—general. Paramedical personnel are normally qualified to perform services required in the treatment of a variety of illnesses. Finally, the nonmedical resources, such as administration, maintenance, and various custodial and janitorial resources, can be quite easily adapted to changes of product.

The cross-elasticity of supply is far from infinite, however. The state of medical technology requires a certain fairly rigid specialization of resources. The cobalt bomb, representative of a class of fixed capital with a low or zero elasticity of substitution and reasonably unadaptable, and the radiologist, representative of highly specialized human capital, are examples. On balance, however, the cross-elasticities of supply are probably significantly higher than the cross-elasticities of damand. This conclusion is reinforced by consideration of those hospitals which would be most properly classified as multiproduct firms according to the four-industry format outlined above. For example, rapidly changing medical technology has made feasible the treatment of psychiatric disorders not only with short-term therapy, but, in some cases, also on an outpatient basis. Some of the larger, urban, short-term general hospitals are therefore active in more than one industry. On the demand side, the cross-elasticity is zero—an appendectomy is no substitute for psychiatric therapy. On the supply side, however, the cross-elasticity may well be significantly greater than zero—certain resources may be assigned to either the psychiatric ward or the surgical ward.

Thus far the focus has been on the product boundaries of the industries within the hospital sector of the economy. Although not without problems, this somewhat simplified approach has certain merits as well as valid credentials. In the literature of industrial organization, an industry is usually defined as "the sellers of a particular product, one side of the market in which buyers and sellers arrange their transactions." [5]

Four industries within the hospital sector have been defined. Although data constraints require that most empirical analysis be confined to this four-way classification, it is essential to consider a number of points which should not be neglected.

First, a single firm may be a member of more than one industry. Thus, the short-term general hospital which maintains a psychiatric ward is a member of two industries. In this sense, the hospital is analogous to General Motors, which functions as a member of the automobile industry and of the refrigerator industry (and of many others) simultaneously.

Second, the breakdown is not only not mutually exclusive among the four basic products included but it also neglects other, different products entirely. Although treatments for psychiatric disorders and tuberculosis are fairly well-defined products, "general and other special hospital services" is a heterogenous classification at best. A

[1] Of course, adaptability has a time dimension; and the examples given are ones on which the time constraint is least binding.

more precise delineation would identify many of the "other special" services as separate products, such as maternity orthopedic, or pediatric hospital services. Since most general hospitals have separate services for these "other special" products, the second problem is closely related to the first; to define more products and, hence, more industries would increase the number of hospitals which were members of more than one industry.[2]

Third, there are many nonhospital medical institutions, such as nursing homes, which perform functions that are substitutes for some of those performed by many hospitals. The use of cross-elasticities of supply and demand implies a certain approach to industry definition. These cross-elasticities are measures of the degree of responsiveness on the part of certain economic agents to changes in the terms of exchange in other product markets. The closer the two are to being substitutes for one another in consumption, the higher will be their cross-elasticity of demand. Theoretically, perfect substitutes will have an infinite cross-elasticity: i.e., they are the same good and must have the same price. The slightest deviation between the prices of perfect substitutes will evoke a mass movement to the lower priced good. The degree of responsiveness to price differences and price changes diminishes as the product substitution becomes less perfect. In this context, nursing homes, convalescent homes, rest homes for the elderly and the like, although treated here as beyond the boundary, would be properly classified at the periphery of the industries in the hospital sector.

Finally, the role of spatial markets in industry definition should be noted. Certain products are sold in national markets or even international markets; other products are sold only in regional or local markets. When regional markets are the rule, there is always the question of whether there is actually a single (national) industry or a group of closely related regional industries. Obviously, geographical boundaries, like product boundaries, are somewhat vague. While a few services of a few hospitals are actually products within a national or even international market, such as the Mayo Clinic or the kidney transplant facility of the Peter Bent Brigham Hospital, most hospitals, like gas stations, are only in the relevant market if they are within reasonable driving distance.

The Special Nature of Enterprise Ownership in the Hospital Sector

Before the salient characteristics of the four hospital industries are outlined, some of the unique characteristics of ownership and control in the hospital sector, their probable causes, and their implications should be noted. Only a relatively small number of hospitals are owned and operated by individual entrepreneurs or managed under the corporate structure familiar in other sectors of the economy. Many hospitals are publicly owned and operated by government agencies at the Federal, State, and local levels. The remaining hospitals are nonprofit enterprises.

[2] An alternative classification would include the "other special hospitals" and the relevant service of all general hospitals as an industry. Thus, for example, all maternity hospitals and the maternity wards of general hospitals would make up a "maternity industry." Although this would solve some problems (not many), data limitations presently preclude its use as a feasible alternative.

Profit motivated hospitals account for less than 1 percent of the average daily census in both the psychiatric hospital and tuberculosis hospital industries. In the long-term general hospital industry approximately 2 percent of the inpatient population is in proprietary institutions. In the short-term general hospital industry the share of the market serviced by profit motivated firms is approximately 6 percent.

Public institutions dominate the psychiatric, tuberculosis, and long-term general hospital industries. Public hospitals service more than 95 percent of the psychiatric inpatients, more than 90 percent of the tuberculosis inpatients, and approximately 70 percent of the long-term general and other special inpatients.

Nonprofit enterprises dominate the short-term general and other special hospital industry. More than 70 percent of all patients suffering from an acute illness are treated in voluntary nonprofit hospitals. The relative roles of public and private hospitals in the several industries may be associated with the public good/private good characteristics of certain treatments.

The treatment of disabling psychiatric disorders, especially those requiring exceedingly long-term or permanent institutional care, is basically a public good. Although the good need not be consumed collectively and the exclusion principle may be enforced (witness the existence of private producers), mental illness is treated in publicly owned facilities. An explicit choice has been made to change the pattern of consumption. The treatment of mental illness, like public education, is a merit want.

As with any public good, various alternatives are available to increase the quantity of resources allocated to the production of psychiatric services. The Government could purchase the services from private producers and distribute them, subsidize private production (or consumption), or merely commandeer the necessary goods and/or resources. In the treatment of mental illness, the choice has usually been public production. Government agencies hire the services of resources in the factor markets, produce psychiatric hospital services, and adopt a pricing policy that essentially removes price and income constraints from consumption.

Tuberculosis is a communicable disease. In the care of transmissible diseases such as tuberculosis, there are very definite external effects of treatment and care. The treatment of a communicable disease benefits not only the particular consumer in the form of cure, but also benefits other members of society in the form of preventive protection. Tuberculosis treatment is, thus, a classic example of a public good.

The dominance of public hospitals in the tuberculosis industry implies that, as with psychiatric care, the Government agencies responsible for tuberculosis treatment have chosen to produce the quantity deemed optimal by some social welfare value judgment rather than to use the alternative methods of increasing the level of effective demand. By financing public production out of general revenue, society has succeeded in eliminating the normal constraints of price and income from individual demand considerations.

If the overwhelming dominance of public facilities in the psychiatric and tuberculosis hospital industries can be attributed to the public good character of treatment for psychiatric disorders and tuberculosis, the different role played by public institutions in the long-term general

hospital industry and the larger share of the market serviced by private institutions can be attributed in part to the relative importance of the public good and private good characteristics of the services rendered by facilities in the industry.

The treatment dispensed by certain hospitals in the long-term general and other special hospital industry is characterized by significant external effects. Hospitals which specialize in the treatment of alcoholic and/or addictive diseases provide social benefits which exceed private benefits, for example.

In addition to providing those goods which have significant external effects, society has decided that certain services provided by long-term general and other special hospitals are merit wants. Most patients with chronic diseases, and all patients in the terminal stages of disease, are members of that part of the population which is permanently unproductive. The treatment of these patients, and of aged patients who constitute a disproportionate share of the inpatient population of long-term hospitals, is a merit want. As in the case of psychiatric treatment, society has chosen to alter the patterns of consumption that would prevail in a "free market."

All treatment provided by long-term general and other special hospitals is not characterized by external effects or merit want considerations, however. A significant part of the total output of the long-term general and other special hospital group falls into the category of private goods. Many orthopedic patients, for example, are treated, cured, and returned to the productive labor force by long-term hospitals. The private benefits from the treatment of long-term illnesses are sufficient in many cases to stimulate a sufficient level of output.

Unlike the three long-term industries, the short-term general, and other special hospital industry is dominated by private facilities. Something in excess of 75 percent of the total demand for short-term hospital services is satisfied by private institutions. The large part played by private hospitals and the pricing policy of public hospitals in this industry suggest that private benefits have usually been considered sufficient to generate a reasonably satisfactory allocation of resources in terms of the implicit social welfare function. This is not to say that public good characteristics are nonexistent, but only that they do not dominate the product mix of the short-term general and other special hospital industry as they do those of the other industries in the hospital sector.

The function of certain public hospitals and the various Government programs designed specifically to alter consumption patterns of many consumers, especially of the medically indigent, are indicative of the existence of a well-defined system of merit wants. In addition there are unambiguous external effects of both consumption and production of many of the services provided by the hospitals in this industry. On balance, the significantly different roles of private and public hospitals in the short-term general hospital industry can be traced in part to the basically more private good character of much of the output of this industry and in part to a choice by society to employ means other than public production when it desires to alter consumption patterns. Further, these two factors are probably not unrelated.

The special nature of enterprise ownership in the hospital sector must be kept in mind when analysis of economic structure, conduct, or performance is undertaken. There are many economic concepts

which can be readily applied with much advantage in an analysis of hospitals, but there are certain concepts, especially those that rely on an implicit assumption of profit maximization on the part of firms, which should be viewed as suspect when applied to the hospital sector. This problem will be dealt with more directly in the last section of the paper.

It was said that analysis of hospitals within a framework of four industries could throw light on various aspects of the production and distribution of hospital services that otherwise might remain somewhat obscure. In the following pages the findings and conclusions of such an analysis will be summarized.[2]

Secular Trends in Demand and Technology

Secular trends in demand have not been similar in the four hospital industries. The most marked increases in demand have occurred in the markets for services provided by short-term general and other special hospitals. Whether utilization is measured by total admissions or by the average daily census, the short-term general hospital industry has experienced the most significant pressure from increased demand. From 1956 to 1966 total admissions to short-term general and other special hospitals increased by 33 percent and the average daily census increased by 38.3 percent. The supply of short-term general hospital services has increased, but probably not as much as has demand for them. The number of short-term general and other special hospitals increased by 10 percent from 1956 to 1966. More significantly, the total number of beds in short-term general and other special facilities increased by 31 percent. An incease in the overall occupancy rate, in spite of a slight increase in the average length of stay, has partially compensated for the difference between the rates of increase of the average daily census and the number of available beds. The overall occupancy rate of the short-term general hospital industry was 72.4 in 1956 and 76.5 in 1966. The average length of stay increased slightly from 7.7 days in 1956 to 7.9 days in 1966. On balance, excess demand has undoubtedly contributed in part to the inflationary trend in prices for short-term general hospital services.

In contrast to the marked upward shift in utilization of short-term general hospitals, the utilization of long-term general and other special hospitals has declined significantly in the past decade. Admissions to long-term general and other special hospitals declined by 17.7 percent from 1956 to 1966, and the average daily census of the industry was 9.3 percent lower in 1966 than it was in 1956. Apparently in response to the decline in demand, the number of hospitals in the long-term general hospital industry declined by 26.3 percent and the number of beds declined by 11.7 percent from 1956 to 1966. The difference between the rates of decline in the number of hospitals and the number of beds implies that a significant number of the smaller facilities have been leaving the industry. The decline in demand for the output of the long-term general hospital industry can undoubtedly be traced in part to a high cross-elasticity of demand between certain custodial care services produced in long-term hospitals and similar services available in facilities, such as nursing homes, external to the industry as presently defined. Although there is no way to measure the total demand for long-term care with available data, it is reasonable

to conclude that aggregate demand for long-term care did not decline as markedly as the demand for care produced by the long-term general hospital industry, and perhaps did not decline at all.

The total demand for tuberculosis hospital services has declined significantly in the past decade. Admissions to tuberculosis hospitals declined by 40.8 percent, and the average daily census declined by 60.2 percent from 1956 to 1966. In response to the decline in demand for tuberculosis treatment over the period, the number of tuberculosis hospitals declined by 50.3 percent and the total number of beds in tuberculosis hospitals declined by 53 percent. In addition, the overall occupancy rate in the industry fell from 80.3 to 68.2 percent over the same period.

Reduced demand for tuberculosis treatment is consistent with improved environmental conditions. Furthermore, given the communicable nature of tuberculosis, widespread testing techniques which result in early discovery of the disease have undoubtedly had a cumulative effect on the total demand for tuberculosis treatment. Early discovery and improved drug technology account for the significantly shorter average length of stay implied by the difference in rates of decline of admissions to and average daily census of tuberculosis hospitals.

The total demand for psychiatric treatment as measured by the number of patients admitted to psychiatric hospitals, increased almost as much as the demand for short-term general hospital services. Admissions to psychiatric hospitals increased by 31.5 percent from 1956 to 1966. If the number of psychiatric patients treated in nonpsychiatric hospitals were included, the rate of increase would be greater. Unlike the marked upward shift in demand for short-term hospital services, however, the increased demand for psychiatric treatment did not result in increased pressure on facilities. In fact, consonant with an 11.5-percent decline in the average daily census of the psychiatric hospital industry, there were 9.1-percent fewer hospitals and 8.1-percent fewer beds in 1966 than there were in 1956. The introduction of new technology, specifically drug and shock therapy, made it possible for psychiatric hospitals to treat an increased number of patients with a significant reduction in the quantity of facilities by drastically reducing the average length of stay.

The secular trends in demand and technology have not been similar for the four industries in the hospital sector. The present structural similarities and dissimilarities can be summarized briefly in terms of product differentiation, size distribution, economies of scale, and barriers to entry.

Product Differentiation

In addition to the differences among hospital services which were used to establish product boundaries for the four hospital industries, product differentiation of varying degrees exists within each of the four industries. In the psychiatric hospital industry, for example, there are two distinct product groups. Certain psychiatric hospitals specialize in the production of services intended to provide curative treatment. Other psychiatric hospitals are engaged in the production of essentially custodial care services. In general, shorter term psychiatric care is more likely to be oriented toward curative treatment, longer term care is more likely to be oriented toward custodial care. There is

certainly a distinct difference between those patients that are treated and released, whether the cure is complete or partial, and those that are committed on essentially a permanent basis. New technology, cited above, has led to an increase in the proportion of psychiatric hospitals that specialize in curative treatment. A similar dichotomy of custodial care and curative treatment product groups exists in the long-term general and other special hospital industry.

Product quality differences certainly exist among hospitals producing the same service in all four industries. The appendectomy performed in a large urban teaching hospital and that performed in a small rural hospital are differentiated products, for example. In each of the four industries, the spectrum of facilities and services available in different hospitals can be cited as evidence of product differentiation.[3] A hospital with more facilities and services available is more likely to produce a given product of higher quality than a hospital with fewer facilities and services available.

The range, or scope, of services produced by a hospital is also related to the availability of facilities and services. Since the unit of output produced by hospitals, a patient day (admission, or discharge), is an implicitly weighted average, the spectrum of facilities and services available is an indication of product differentiation in the form of varying complexities of services produced. Pure quality differentiation exists in all four industries. The complexity of the scope of services is especially differentiated in the two general and other special hospital industries.

Size Distribution

Certain similarities of size distribution exist among the four hospital industries. The size distribution of hospitals in each of the four industries is highly skewed. There are also similarities among the industries in the size-ownership relationships which are probably a function of capital requirements and other potential barriers to entry.[4] In all four industries proprietary hospitals tend to be the smallest type of hospital. In three of the four industries public institutions are larger, on the average, than private institutions. In the fourth industry, the short-term general and other special hospital industry, voluntary hospitals tend to be slightly larger than public hospitals.

The average size of hospitals is very dissimilar in the four industries, however. Psychiatric hospitals tend to be very large. The average (mean) size of psychiatric hospitals in 1963, for example, was 1,440 beds. The short-term general and other special hospitals tend to be the smallest type of hospital; their average (mean) size in 1963 was 122.8 beds. Although much smaller than psychiatric hospitals, tuberculosis hospitals and long-term general and other special hospitals are almost twice as large as short-term general and other special hospitals. In 1963 their average (mean) sizes were 213.9 beds and 227.6 beds, respectively. Since the distributions are highly skewed, the median sizes of hospitals in each industry are more appropriate measures of average size. The median sizes in 1963 were: tuberculosis hospitals, 143.4 beds;

[3] The American Hospital Association [1] lists a number of facilities and services actually available within, and reported by the institution each year. Not all hospitals have all nor even most of the facilities and services. In 1966, for example, 83.4 percent of the reporting hospitals had an emergency department, but only 57.6 percent reported having a pathology laboratory and 9 percent had the capacity to provide cobalt radiation therapy.

[4] See below.

long-term general and other special hospitals, 129 beds; and short-term general and other special hospitals, 74.6 beds. Unfortunately, the median size of psychiatric hospitals could not be determined since it was in an open-end class in the primary source of data. [1]

Economies of Scale

Analysis suggests that economies of scale exist in the production of hospital services in each of the four hospital industries. The average size of the institutions in each of the four industries has shown a tendency to increase over time. Although a variety of factors might have caused this growth, it is certainly consistent with the existence of economies of scale. Analysis of the relationships between the level of output and the average cost of producing that output suggests that economies of scale exist in each of the four industries. Input/output analysis also suggests increasing returns to scale in the production of all four types of hospital service.

The factor which tends to obscure the relationship between the level of output and average cost, especially in the two general and other special hospital industries, is the interaction of hospital size, hospital cost, and product differentiation. In a number of recent studies [3, 4, 6, 7] attempts have been made to allow for this interaction and to resolve the question of whether or not hospital services are produced subject to economies of scale. Albeit each study may have certain shortcomings, the weight of evidence supports a strong conclusion that hospital services are produced subject to economies of scale.

Barriers to Entry

Many and varied barriers to entry exist in each of the four industries. Economies of scale, capital requirements, product differentiation or the interaction of all three can be cited as potential barriers to entry. The pricing policies of public institutions inhibit entry especially in the three long-term industries where the services produced are perhaps classified most properly as public goods or merit wants.

Proprietary hospitals in each of the four industries face an additional barrier to entry. Sources of capital are available to public institutions and to voluntary nonprofit hospitals which are not available to proprietary hospitals.

In the markets for complex general hospital services, the lack of qualified personnel may also inhibit entry.

Barriers to exit are more relevant than barriers to entry in those industries which have been faced with a declining demand. High fixed costs in the form of fixed capital may inhibit exit, especially in the short run. In addition, political barriers may inhibit exit of public institutions for longer periods.

Summary and Conclusions

Hospital services are not homogeneous and economic analysis of the hospital sector can benefit from a multiple industry classification. The identification of hospital industries by means of plausible cross elasticities of demand and supply can be used to some advantage in the establishment of such a framework for analysis.

When separate industries are distinguished a different picture of the relative pressure on hospital facilities comes into focus. There have been marked differences in the secular trends in demand for the several types of hospital services. The dramatic increase in pressure is unique to the facilities providing essentially short-term services.

Product differentiation exists within each of the four hospital industries, but it is most pronounced and presents the greatest problems in the case of the short-term general and other special hospital industry. Since medical care is becoming increasingly hospital oriented, the whole issue of the quality of medical care can be viewed within this context.

There is a wide range of quality of care available in the United States. Quality of care is associated with size, complexity, proximity of educational and research programs, and a host of other factors. While there is widespread sympathy for "Cadillac only" medical care, there is no doubt that if it exists it is practiced in a limited number of hospitals. Since resources are scarce, there are many policy questions concerning the appropriate quality of care and the promotion of quality improvements which remain to be answered. The role of third parties in medical care markets is complicated in large part both by their concern for and their effect upon the quality of care.

While the quality of a given service may vary with hospital complexity, complexity itself is another dimension of the quality problem. Hospital size, complexity of services available, and the cost of care are highly interdependent. Most evidence suggests that hospital services are produced subject to economies of scale. The higher per diem cost observed in larger hospitals is a phenomenon of the relationship among hospital cost, hospital size, and the complexity of the scope of services produced. A given hospital may expand its output to take advantage of economies of scale, to increase its scope of services, or both. Questions about the optimal size of hospitals are premature until answers are available concerning the optimal mix of scopes of services.

Questions concerning hospital cost and the quality of care, especially those related to public policies designed to promote efficiency within the hospital sector, are the more complex because of the unique role of nonprofit enterprises. In the short-term general hospital industry external intervention has traditionally been nongovernmental with the consequence that the majority of hospital services in the United States are provided by nonprofit enterprises. Little is known about the exact form of the objective function of nonprofit enterprises. Do the managers of nonprofit enterprises act differently than their profit motivated counterparts when faced by certain market data or is the form of the intervention such as to change the market data they face? If incentive reimbursement schemes or other market oriented stimuli are to be successful policy tools for promoting efficiency in the hospital sector, answers must be sought to these and similar questions.

REFERENCES

[1] American Hospital Association. *Hospitals*, Journal of the American Hospital Association, Guide Issues, Vols. 30–41, 1956–1967.

[2] Berry, R. E. *Competition and Efficiency in the Market for Hospital Services: The Structure of the American Hospital Industry*, Cambridge: Interfaculty Program on Health and Medical Care, Harvard University, 1965.

[3] ———— "Returns to Scale in the Production of Hospital Services", *Health Services Research*, Summer 1967, pp. 123–139.

[4] Carr, W. J. and Feldstein, P. J. "The Relationship of Cost to Hospital Size", *Inquiry*, June 1967, pp. 45–65.

[5] Caves, R. *American Industry: Structure, Conduct, Performance*, Englewood Cliffs: Prentice Hall, 1964, p. 6.

[6] Feldstein, P. J. *An Empirical Investigation of the Marginal Cost of Hospital Services*, Chicago: Graduate Program in Hospital Administration, University of Chicago, 1961.

[7] Inghar, M. L., and Taylor, L. D. *Hospital Costs: A Case Study of Massachusetts*, Harvard University Press, 1968.

THE RATIONING ABILITY OF PRICE IN THE MARKET FOR HOSPITAL SERVICES

BY HAROLD A. COHEN*

The structure of the hospital industry is vastly different from the model of pure competition.[1] The major divergencies are the absence of the profit motive, the absence of competition, and presence of externalities. In such an industry, the invisible hand cannot be relied upon to bring forth an efficient use of resources. If such efficiency is to to attained without a wholesale reconstitution of motivating forces and organization within the entire medical industry, then a conscious effort (or visible hand) is required to steer the industry, i.e., to channel private aims for the public good.

This conscious effort to achieve efficiency has always been needed in the hospital industry. In recent years, however, the human life span, the ability to treat sickness, disposable income, and, thus, the demand for hospital services have all greatly increased. We have witnessed a tremendous upsurge in the relative and absolute amount of resources being devoted to health and hospital care. In turn, the importance of attaining an efficient use of hospital resources has increased.

This paper will examine those aspects of the hospital industry which are most likely to contribute to *inefficient* resource use. This examination will serve the dual purpose of demonstrating the need for a special conscious effort and identifying those areas which are most likely to hinder the visible hand in its attempt to equate social benefits and social costs at the margin.

Efficiency in the use of resources, in this study, is synonomous with the achievement of a Pareto Optimum.[2] It identifies the goal as producing the items which members of a community want most while using the least amounts of scarce resources. There are three elements or steps to achieving this goal which may be described as follows:

1. Technical efficiency is defined here as producing whatever amount the firm is producing at the lowest possible average cost for that level of output. The attempt to achieve technical efficiency follows logically from the profit motive. No entrepreneur will deliberately waste resources (operate at a higher cost than necessary for a given output) for that would lessen profits. While all entrepreneurs try to attain technical efficiency, skill is required to be technically efficient. The force of competition is expected to eliminate those who are incapable of being reasonably efficient and, in the long run, the firms that remain are those best able to efficiently produce the level of output desired.

---

* University of Georgia. The author is indebted to Alfred E. Kahn, Richard A. Stanford, and especially L. Aubrey Drewry for their helpful comments. This work was partially financed by the National Institutes of Health.

[1] The hospitals we are interested in are more precisely referred to as "non-Federal Government and voluntary short-term general hospitals."

[2] See John Krutilla and Otto Ekstein, *Multiple Purpose River Development*, pp. 15–17.

2. Industrial efficiency is the condition of the firm's producing at that level of output which will enable it to achieve its lowest average cost for the product. Entrepreneurs do not attempt to achieve this goal but, given the attainment of technical efficiency and sufficient competition, the invisible hand will lead them to it in the long run.

3. Allocative efficiency is the condition of having that number of industrially efficient firms which will produce the amount of the item the community wants at the product's lowest average cost. Again, entrepreneurs do not attempt to achieve this goal but, given the profit motive, pure competition, and a correctly working market or price system, the invisible hand leads to this end.

If, then, the profit motive, competition, and the market are all working correctly, efficiency in the use of resources will be attained. If any of these is not functioning in the prescribed manner then the use of the market as an allocative mechanism must be questioned and the relative costs and benefits of alternative allocative arrangements must be examined.

We shall now turn our attention to the hospital. We will be particularly interested in the lack of a profit motive, the general lack of competition, and the existence of externalities.[3]

1. Technical Efficiency and the Hospital

There are two questions to which we must turn our attention: (1) Does anything take the place of profit maximizing and make hospital authorities attempt to achieve technical efficiency?[4] and (2) does anything take the place of competition and eliminate (in the long run) hospital authorities incapable of attaining a reasonable degree of technical efficiency?

A. The Incentive to Technical Efficiency

In answer to the first question it is suggested that service or output maximizing replaces profit maximizing as the logical motivation for eliminating deliberate waste and encouraging the hospital director to use as few resources as possible in the production of any one service. By being technically efficient in the production of each individual service, the hospital has more funds available to provide additional output (a new service or better units of the old).

This conclusion is reached by Sidebotham and Page who suggest that:

> In a variety of services * * * including * * * the erection of hospitals, the financial problem is one of * * * providing the most comprehensive service within the limits of the resources available.[5]

---

[3] Externalities refer to the failure of marginal cost to include all social costs and of price to reflect all social benefits.

[4] Insuring an attempt to attain technical efficiency is a major problem faced by all societies in which all or many goods are produced by any institution which does not have profit maximization as its main goal. All such economies take strong measures to assure such an attempt.

For an examination of Soviet controls see Barry M. Richman, *Soviet Management*, especially chapter II. Also see Charles Montecute, *Soviet and Efficiency in Hospitals*, with Montecute's conclusion, on p. 259, being: "It was a natural consequence of the inception of the National Health Service * * * that attention should be directed to costing as an aid to efficient hospital management. * * *"

[5] Sidebotham and Page, *Accounting for Local and Public Authorities*, pp. 12, 13 quoted in Montecute, *Op. cit.*, p. 31.

Montecute suggests that the use of costing in hospitals is so that the profit motive can be replaced by what might be called the optimal use of resources motive.[6] This means, of course, that the hospital will strive to attain technical efficiency as is shown in figure 1.

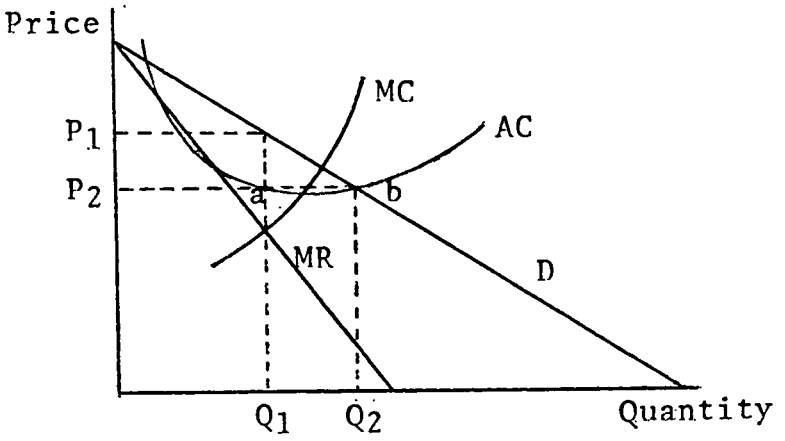


FIGURE 1

A profit maximizing firm would clearly operate at P_1 and Q_1 (MC =MR) but note that Q_2 is the greatest output that can be attained subject to the revenue constraint represented by the demand curve and so our hypothetical firm would operate at P_2 and Q_2. In neither case does P=MC; the profit maximizing firm above underproduces (P >MC) while the output maximizing firm overproduces (P<MC). Since "b" (like "a") is on the firm's average cost curve, the firm is being technically efficient.[7] Since neither is the low point of the average cost curve, we see that either competition or deliberate planning (or an accident is required to attain industrial efficiency.

While this shows that output maximization subject to a revenue constraint allows the firm simultaneously to determine price and quantity and attain an equilibrium position, it is a very simplified view of a nonprofit oriented hospital's decisionmaking process. The basic problems are that it overly simplifies the nature of hospital revenues and also suggests that quantity is measured in the same way by profit and nonprofit organizations. These two problems are discussed below in the section on industrial efficiency.

As suggested above output maximization includes consideration of both quality and quantity. These have to be weighted somehow by each hospital. If quality was to be held constant, the hospital would determine the average cost curves for various hospital services and offer those services in which a hospital of its size could expect to be

[6] Montecute, *op. cit.*, p. 31.

[7] Technical efficiency as defined above (p. 2) refers to operating on the LRAC curve but figure 1 is valid for both the short and long run. Also, for those public services in which the revenue constraint is a matter of Government appropriation it makes most sense to speak of a fixed total revenue constraint. In the case of hospital services, where revenue is partly determined by a market mechanism, it is more reasonable to speak of a total revenue constraint which varies with price or average revenue. The same results would be shown with AR a rectangular hyperbola as would be shown with a fixed total revenue curve.

most efficient.[8] Quality is not held constant and, indeed, is sometimes measured by the number of different services available. This might lead hospitals to add services in which they cannot attain average costs near the industry average. This quality consideration may also be part of the reason why non-profit-oriented hospitals offer a different output mix than proprietary, or profit-oriented, hospitals. Nevertheless, whatever output mix the hospital does offer, it will attempt to produce that output at the lowest possible total cost.

A further reason for attempting to be technically efficient is that, *ceteris paribus*, it reduces the deficit or creates a surplus in Government hospitals; this outcome has a definite utility value for the Government control group. Technical efficiency also reduces, *ceteris paribus*, the amount of philanthropy required by the voluntary hospital; the collection of philanthropy frequently has disutility.[9]

The attempt to achieve technical efficiency is further illustrated by the institutional framework within which certain types of hospitals prepare their budgets. This includes the Mother House method by which Catholic hospitals are prodded to keep their costs down and the auditing requirements of many State enabling acts.[10] However this desire to achieve technical efficiency does not guarantee its achievement. We must briefly examine the prospects of a hospital's attaining this degree of efficiency.

B. THE ACHIEVEMENT OF TECHNICAL EFFICIENCY

Certain hospitals are put under more pressure to achieve technical efficiency than others. In the Mother House procedure and in the United Hospital Fund procedure several hospitals using similar accounting techniques submit budgets or costs to a central agency. This agency should be able to spot significant waste.

Most hospitals undergo detailed auditing by Blue Cross, Government welfare agencies, and private accountants. At least one of these examinations should uncover any significant waste. While these audits are designed to verify cost figures used in determining the level of Blue Cross or welfare payments, the use of at-cost reimbursement formulas by these agencies may lead to inefficiency in the hospital.[11] The revenue constraint would give no incentive to efficiency just as the profit motive may give no incentive to efficiency under a similar type of pricing formula for utilities.[12]

---

[8] This statement presupposes a measure of output where the units of different hospital services are summed after being weighted by an estimate of their industry-wide average cost.

[9] Certain UHF hospitals report fund raising costs of over $50,000.

[10] For example Georgia's Hospital Authorities Act requires that hospitals, established under its codes, be audited at least once a year; this is in addition to the semiannual audit of State accounts by Georgia, the impending audit of medicare accounts (which will be done by Blue Cross as a fiscal intermediary for the Federal Government), and the cost analysis done by Blue Cross on behalf of its patients. Articles in the July 1967 *Social Security Bulletin* discuss Medicare's requirements relative to hospital efficiency.

[11] The importance of these sources and the use of cost-related reimbursement formulas give the hospital an incentive to allocate as much of their cost as possible to services covered by such formulas. Audits by these agencies must determine if allocations are made on an economic basis.

[12] Wilcox, in arguing that utility earnings should be related to efficiency, suggests raising the percentage of return as rates are cut and lowering the return when rates are raised which, he says, would require the commissions "to develop standards of efficiency in physical terms, together with methods of measurement by which performance can be appraised." (Clair Wilcox, *Public Policy Toward Business*, p. 298). A similar criterion of reimbursement may be desirable for Blue Cross even though the nonprofit hospital would receive more than cost for Blue Cross patients. It may be hard to convince a Blue Cross subscriber that this is equitable, but it would be to his advantage if this type of reimbursement caused hospitals to lower their costs. Surely the subscriber would rather pay 105 percent of $50 than 100 percent of $55. Another obstacle to instituting such a reimbursement plan is the difficulty in measuring the quality of medical care. An increase use and refinement in making medical audits can largely overcome the present inability to make useful evaluations of relative quality.

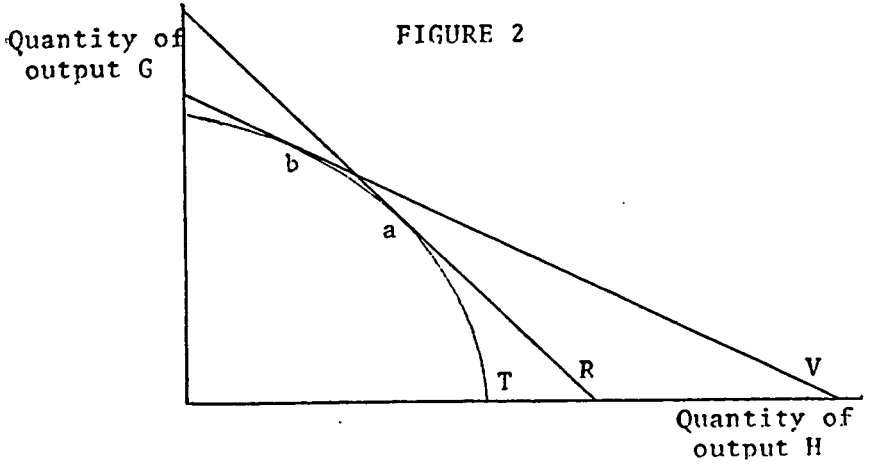
In addition to the numerous audits, the American Hospital Association publishes statistics showing costs and services offered for virtually all hospitals in the United States. The hospital trade journals also publish much information concerning both suggestions for lowering costs and the records of achieved cost savings.

With all this information available, it is likely that hospital authorities, in their attempt to achieve technical efficiency, can come very close to doing so. We shall, then, proceed on the assumption that, given the nature of their physical plant, hospitals produce their output at the lowest possible average cost. We must now ask whether the services or output which the hospital chooses to perform are those it can perform efficiently and see to what extent, if any, there may be a systematic influence which would cause the physical plant to be inefficient.

2. Industrial Efficiency and the Hospital

Hospital size may be measured in various ways just as is the case with manufacturing enterprises. In this paper hospital size is conceived as measured in units of services performed (weighted by average costs). Therefore, discussing the nature or composition of the output mix is as intimately related to a discussion of hospital size as a discussion of the amounts of each service. Clearly any unit of service added represents an increase in size unless, of course, an equal amount of service is subtracted. However the determination of whether or not services are of equal value differs between firms which maximize profits and those which do not. Profit maximizing firms assign relative values to their outputs according to the ratio of their prices. This is not necessarily true for a not-for-profit firm.

FIGURE 2



Let us assume the two outputs (G and H), the related transformation curve (T), and the isorevenue line (R) as shown in figure 2. A profit maximizing firm would produce at point "a" but a nonprofit, service maximizing firm would produce at "a" only if it assigned weights to the services, i.e., assigned values to G and H, that were equiproportional to their market prices. The firm would operate at point "b" if

for some reason it decided that the relative values are reflected by the isovalue line (V).

There are several reasons why the hospital might not evaluate a service at its market price. The following is a partial list of these reasons.

1. The hospital authorities may be aware of external economies or diseconomies—it may give away inoculations against smallpox and have a high valuation of the service at the margin.

2. Hospital authorities may believe that certain services bring forward more philanthropic contributions or the possibility of Government grants—this may be true of occupational therapy or cobalt units.

3. The hospital may be required by law to provide certain services, such as ambulance service and outpatient department, and with a certain high level of activity—in this instance the hospital is similar to a common carrier—making it difficult to operate at point "a" even if so desired.

4. The hospital authority (including the medical staff) may simply believe it can better judge the needs of the patient and cannot use the price mechanism to induce a correct decision by the patient: For example, many hospitals require chest X-rays upon admission. Another example may be the expensive chapels in many hospitals run by religious authorities.

5. Prices, or revenues, may be collected in such a way that the composition of auxiliary services performed may not affect revenues: For example, a welfare contract based upon expected patient loads.

6. Hospital authorities may believe that a wider range of services may attract physicians into the region or may lead physicians to affiliate themselves with the particular hospital when there are alternative hospitals in the region.

7. A wider range of hospital services may also bring additional prestige to the hospital authority. Reder points out that due to prestige motives, attraction to top administrators, and attraction to prestige medical staffs "there is likely to be a strong tendency toward duplication of facilities and generation of excess capacity. This tendency has been recognized and has prompted suggestions for overall regional planning of hospitals." [13]

8. Hospital authorities may believe the hospital is the best place to carry on certain services and do not trust the market to perform such service well; this may be the case for medical research and hospital blood banks.[14]

9. The hospital may tend to offer more of services covered by Blue Cross or Medicare are due to the lower percentage of bad debts.

10. The hospital may carry on services for which Blue Cross and welfare agencies are willing to pay a disproportionate amount of cost; these services include medical research and education.[15]

Of course, the hospital, in its desire to maximize service, will not overlook market price completely. Increased revenues or increased efficiency are required to increase service; and patient revenues are an

[13] Melvin Reder, "Some Problems in the Economics of Hospitals," *The American Economic Review*, vol. 55, May 1965, p. 480.

[14] See, for example, the Federal Trade Commission's Final Order (8519) *In the Matter of the Community Blood Bank of the Kansas City Area, Inc.* in FTC, *News Summary*, No. 38, Oct. 27, 1966.

[15] See Joint Economic Committee, *Federal Programs for the Development of Human Resources*, pp. 616–655, 658–679, 731, 769.

important source of income. The greater importance of insurance sources including medicare for the aged, however, may have a significant effect on the nature of services offered by hospitals. There are two reasons for this: First, the hospital will tend to perform services most liberally covered by hospitalization (points 9 and 10 above), and second, patients covered by hospitalization insurance will tend to request a larger percent of their medical treatment in the hospital. This latter point will be discussed more fully later in this paper when we examine the price mechanism.

Thus, even should the hospital produce its desired output at the lowest possible cost (i.e., be technically efficient), the composition of that output is not likely to be the same as that which a profit oriented firm would choose.[16] The reasons listed above for a difference in valuation of output between profit oriented firms and nonprofit hospital authorities are also reasons for a divergence in the composition of output as was shown in figure 2. Since some of these reasons do reflect the presence of significant social externalities, their being considered by a hospital administrator may well lead to a better use of hospital resources than a reliance upon competitive enterprises—although tax and subsidy policy might be used to accomplish some of the same effects. We shall return to the matter of pricing after more explicitly considering the question of hospital size.

Some of the factors in the list above (such as items 6 and 7) suggest that subtractions of services will not offset additions. Additional reasons why the hospital may be larger than the market would induce (or the plant too costly) are: First, the practice whereby organized hospital auxiliaries donate specialized equipment to the hospital; second, the availability of Hill-Burton and Appalachian funds, due to their income effects; and third, rules regarding minimum size and service availability established by accrediting agencies or third party payers.[17]

Many of the factors discussed above imply the inadequacy of the pricing mechanism. We shall now consider the market for hospital services with a view toward seeing if marginal cost pricing is likely to bring about that amount and type of services such that marginal social benefits are equal to marginal social costs.

MARKET OR ALLOCATIVE EFFICIENCY

Market efficiency is achieved when marginal social benefit equals marginal social cost. The market satisfies this equation when it also satisfies three other equations. This system of equations may be summarized as requiring marginal cost pricing and an absence of externalities and may be listed as follows:

1. Marginal social benefit=Price.
2. Price=Longrun marginal cost.
3. Longrun marginal cost=Marginal social cost.

$\therefore$ Marginal social benefit=Marginal social cost.

Given a competitive market, the first equation is fulfilled if there are no neighborhood effects. In the long run pure competition is a suffi-

[16] For example, proprietary hospitals almost never offer ambulance service.

[17] The latter include rules established by the AMA for accrediting internship programs, the Joint Commission for Accreditation, and Blue Cross and medicare for participation. See JEC, *op. cit.*, pp. 683–688, 698–701, and William H. Stewart, "The Positive Impact of Medicare on the Nation's Health Care Systems," *Social Security Bulletin*, July 1967, pp. 9–12.

cient, but not necessary, condition for the satisfaction of the second equation. (Perfect competition is sufficient in the short run.) It is possible for prices in a regulated industry to be set equal to LRMC. For the third to be satisfied requires the absence of external economies.[18]

The hospital industry is not purely competitive; moreover it exhibits externalities on both the supply and the demand side. The hospital industry is not unique in any of these characteristics. All of these characteristics have been examined in detail in the abstract by economists, and also examined in connection with various other industries.

The majority of externalities associated with hospital services appears to be beneficial. Neighborhood benefits are sizable in treatments for communicable disease, medical education, and medical research; that is, members of the community other than present or future patients are benefited by these services. Perhaps in our society an individual with communicable disease can be isolated from potential victims and the disease brought under control without the person being confined to a hospital; this possibility is not, however, highly likely. Therefore, people other than the sufferer should pay something toward the cost of confinement rather than have him "air his germs in public." The hospital-based education of much medical personnel benefits the patient who uses their services but is never hospitalized. Such a patient should "pay" the hospital for its teaching function.[19] Medical research also benefits many people who are never admitted as patients: especially, for example, vaccine-oriented research.[20] The hospital may also offer microeconomic benefits to the community: it is a force for attracting industry and it gives "insurance" value to citizens who never need the hospital but must pay to have it there in case they do need it.[21] Neighborhood costs occur when the occupancy rate is such that an additional patient leads to a lessening of the quality of care given to other patients on the floor due to their having to share fixed capacities.

The external economies generated by hospitals, or by improved medical treatment, are related to shortening of periods of incapacity and, conversely, elongation of working life. Whether such economies exist depends largely on the relation between an individual's wages and his productivity. If an employee is being paid his marginal revenue product, that is his gross value to the firm as measured by the additional revenue the firm realizes from his productive activity then his incapacity does not affect his employer beyond the effect on the worker. If, however, the employee's wage is less than his MRP, or if he collects some form of "sick pay," then his employer loses more in revenue than in costs and should be willing to pay the hospital to hasten the return of the employee.

If the MRP of workers in training is less than their training pay; and also if elongation of the working lifespan permits less training for replacement, then a longer working life has a value to the firm

[18] In this study, the existence of neighborhood effects refers to cases where a purchase involves utility (+ or −) to people other than the demander (positive utilities are referred to as neighborhood benefits and negative utilities as neighborhood costs). External economies, here, refer to cases where a sale involves costs (+ or −) to persons other than the supplier (as usual, increases in cost are referred to as external diseconomies).

[19] Perhaps it might lead to better allocation if all medical students were charged for their education and then passed these costs on to all who benefit from their education. But as long as so much other education is being subsidized, that of teachers for example, this might lead to an underenrollment in medical education.

[20] See JEC, *op. cit.*, pp. 601–605, 826–829.

[21] Burton A. Weisbrod, "Collective-Consumption Services of Individual-Consumption Goods," *Quarterly Journal of Economics*, vol. 78, No. 3, p. 274, August 1964.

above its value to the employee.[22] But an elongation of lifespan beyond retirement increases pension payments and thus raises costs for the firm more than it raises revenues. The net effect on longrun average costs of elongated lifespans is, thus, indeterminable on an *a priori* basis.

The most important externalities associated with hospital treatment appear to be positive and on the demand side, i.e., neighborhood benefits. The policy suggestion which flows from this is not original or solely applicable to the hospital industry. People in the hospital's community should be recognized as "users" of the hospital and be incorporated into the revenue structure through the tax-subsidy mechanism. This has the effect of making these externalities internal and represents a direct attempt to satisfy the first equation.

The second equation signifies that all those who are not willing to pay the long run marginal cost associated with a product will not use the product. That is, price is supposed to play the rationing role in an exchange economy. But, in the case of hospital services, society, on equity grounds, is unwilling to let this situation prevail except in extreme examples such as the kidney machine—and, of course, there is controversy there.[23] Virtually all Government hospitals and many voluntary hospitals are unable (or, in practice, choose not) to refuse service on the basis of inability to pay.[24]

Alchian and Allen have examined the rationing function and the setting of prices for nonprofit institutions. They state that "an administrator of goods has less incentive to use the marketplace and to get the best possible price if he is not responsible to someone who owns the goods as private property" and they expect that prices in nonprofit institutions will be below the "market clearing" price for two basic reasons. First, the lower price will be available to the administrator in his consumption of the good; and second, the administrator will be able to ration the product among the mass of demanders and gain good favor with those he chooses.[25] Both these reasons are effectively irrelevant in most hospitals. Unlike Rose Bowl tickets and concert tickets (the examples used by Alchian and Allen) the members of the hospital board and the administrator do not usually expect to avail themselves of the hospital's services; and if they were to avail themselves of those services, most hospitals either provide free hospitalization or provide care at reduced rates to the administrators. Most board members already have insurance through their regular occupations.

Second, if shortages develop the rationing of beds is primarily per-

[22] Presumably employers hire workers up to the point where the discounted cost to the firm over the worker's years of service equals the discounted revenue he will bring to the firm. If discounted cost exceeds discounted revenue during the training period then the reverse must be true later. This assumes either some immobility or some training that is fully applicable only to the training firm otherwise the trained individual could command his MRP by moving.

[23] The kidney machine enables people whose kidneys do not function to remain alive. Since it costs thousands of dollars for each patient-year of treatment (the treatment continuing until death) and there are tens of thousands who would be able to benefit by the machine, a complete public program would be both highly redistributional and very expensive. The Public Health Service has some machines now (with short-term appropriations) but not nearly enough machines for all who could use them. The few other machines in existence are allocated on the basis of ability to pay.

[24] The treatment accorded to paying patients may not be the same as that accorded the subsidized patient; but some treatment is accorded to those who cannot pay the going price (or the marginal cost) of their treatment. In this way a hospital differs from a common carrier; a common carrier may refuse service only on the basis of the customer's inability to pay the going rate for his class.

[25] Armen A. Alchian and William R. Allen, *University Economics*, pp. 158–161.

formed by the medical staff and not the board or administrator.[26] Furthermore, the question of whether "queues," "shortages," and "low prices" do occur is empirical in nature and we see that several hospitals attain percentage occupancy rates in the nineties (showing that such rates are technically feasible) and yet during the period 1946–64 the occupancy rate for non-Federal Government hospitals ranged from 63.2–79.5 and that of voluntary hospitals ranged from 72.2–79.5. Proprietary hospitals had a range of 59.5–68.3 which may show that nonprofit institutions have higher rates of utilization than profit-oriented ones, but the figures do not show that rates cause large queues.[27]

Alchian and Allen do not address themselves to the interesting question of whether nonprofit institutions price above, below, or at marginal cost. The examples they choose are of monopolists who would, in a profit maximizing situation, price above marginal cost. To say that price is below this monopoly price does not relate it to MC. Hospital authorities have recognized that prices should be "reasonably related" to average cost. Several State and local hospital associations have ratified the statement on pricing policy of the American Hospital Association which says in part: "Established rates for individual patient services and accommodations should be based upon and be reasonably related to the full cost which is incurred in providing these services and accommodations as computed in accordance with recognized cost-finding procedures." [28]

This policy does not set any definite relation between price and marginal cost unless one assumes a precise relationship between marginal cost and "full cost" as well as a precise meaning to "reasonably related." Thus no *a priori* relationship between price and marginal cost can be determined. This is the same conclusion we got from assuming output maximizing subject to a budget constraint as was shown in figure 1.

To this point we have been discussing the market for hospital services as if it were fairly similar to the typical commodity or service market where exchange involves a confrontation of a potential seller and a potential buyer. The market for hospital services is complicated by the presence of other parties to the transaction. These other parties are insurance companies (with Blue Cross most important), government (with the Federal Government likely to become most important), and the medical professions (with the AMA and the individual hospital's staff most important). There are two basic types of medical insurance: (1) where a stipulated amount of cash for a given treatment is paid to the insured who then, presumably, pays the hospital, and (2) where the insurance company pays a stipulated amount of cash for a given treatment to the hospital which accepts this, usually, as payment in full for the insured.

Insurance of the first type affects the height of the demand curve, but, except for the availability of funds, would not alter the basic

[26] For an analysis of physician rationing see Martin S. Feldstein, "Hospital Bed Scarcity: An Analysis of the Effects of Inter-Regional Difference," *Economica*, vol. XXXII, No. 128, November 1965, pp. 393–409.

[27] American Hospital Association (AHA), *Hospitals, Journal of the AHA*. See also Herman M. Somers and Anne R. Somers, *Medicare and the Hospitals*, p. 58.

[28] Board of Trustees of the AHA, *"Guiding Policy for Hospitals in Pricing Their Services to the Public, Point III,"* quoted in Lem E. Hay, *Budgeting and Cost Analysis for Hospital Management*, p. 6.

workings of the market.[29] This insurance affects allocation by creating a divergence between price and marginal expenditure. Klarman writes:

> (I)mportant in the demand for hospital care is one of the effects of health insurance; namely, the creation of a differential between the price of a good and the out-of-pocket expenditures by the consumer. It is possible for the prices of medical care to rise relative to the prices of other goods while the ratio of the out-of-pocket charges to other prices is declining. If so, the tendency would be to substitute medical care for other goods.[30]

The trouble with this kind of insurance from the hospital's standpoint is that the money goes to the insured party, who may or may not use the money to pay his bills. This explains the intricate enforcement policies that many hospitals have adopted in attempting to assure receipts of insurance funds—and why the hospitals established Blue Cross as a third-party payer.

If people consume hospital services rationally (equate utilities and disutilities on the margin) then the demand schedule for hospital services of those with type 1 insurance will rise by approximately the amount of the insurance payment. There are two reasons why the curve may not rise by the full amount: (1) the demander may recognize that any increase in hospital utilization will raise his insurance costs; and (2) the demander will have less funds at his disposal because he has bought insurance. The first consideration will be of little significance for group insurance holders because they recognize that any increase in cost due to their stay will be spread among many insurance purchasers. If the insurance policy is drawn up on an individual basis, the insured will place more weight on this first factor. The second factor is related to whether there is a change in the potential consumer's marginal utility of hospital service relative to money (and the other goods money could buy) caused by the income effect. The extent of this effect differs between individuals, but we can predict that because of it the rise in demand will be less than the insurance payment. The net effects of type 1 insurance would appear to be, *ceteris paribus*, an increase in the price and quantity of hospital services, an increased demand for complementary services such as those of physicians, an increase in producer's surplus, and an increase in the consumer's surplus realized by those insured who utilize the hospital.[31]

The second type of insurance company, referred to in hospital literature as third-party payers, greatly alters the nature of the market transaction whenever their payments to the hospital are accepted as payment in full for the patient.[32] The prices or rates which are determined by bargaining between the hospital and the insurance company

[29] For an interesting account of the effect of insurance funds and other special factors on price see Jessica Mitford, *The American Way of Death*, especially ch. 2.

[30] Herbert E. Klarman, *The Economics of Health*, p. 28.

[31] Physicians have been "charged" with raising their prices for those with Blue Shield protection. The likelihood of such an action is probably the main reason for the "payment-in-full" contracts which Blue Shield has developed for people below certain income levels. Producer's surplus is the summation of the differences between the lowest price at which individual suppliers would sell the product and the actual sale price. Consumer's surplus is the summation of the differences between the highest prices at which individual demanders would buy the product and the actual purchase price.

[32] Misallocation also stems from insurance which pays in full for one service (semiprivate rooms) but in part for a substitute service (private rooms). This problem is dealt with in Burton A. Weisbrod, "Some Problems of Pricing and Resource Allocation in a Nonprofit Industry—The Hospital" *Journal of Business*, 1965, pp. 18–28.

are only of indirect importance to the consumer; i.e., through the effect of hospital prices on insurance rates. The mechanics of the market transaction when direct payments are not accepted as payment in full are somewhat altered from the type 1 situation, although the main results of the market transaction are quite similar (except for an expectation of fewer bad debts).[33]

When direct payments from insurance companies to hospitals are accepted as payment in full for services rendered, the rationing or allocating effect of the price of the good is virtually eliminated. Demand would appear to be limited by three basic considerations: (1) the direct disutility of hospital treatment, such as pain and confinement; (2) the indirect disutility of hospital treatment, such as loss of income; and (3) the necessity of finding a doctor to certify that hospital treatment is necessary. In these cases where human psychology eliminates the first obstacle and insurance, or lack of income, etc., eliminates the second, the rationing function of the medical profession becomes vital. This is due, in part, to inability of the typical hospital service consumer to discern the probabilities associated with the various outcomes of a given hospital service and the utilities or disutilities associated with these outcomes. For those with both hospitalization insurance and loss-of-income insurance (whether through private insurance directly or as an employee benefit) the direct and indirect marginal expenditure for hospitalization may be zero, or negative due to income tax policy. A negative marginal expenditure may also result from multiple insurance coverage.

The net effect of type two insurance with a payment-in-full feature is an even greater increase in the demand for hospital services and their complements and a greater substitution effect causing people to want more insured treatments done inside the hospital rather than, say, in their doctor's office.[34] This increase in demand, however, is not allowed to cause a price rise in the short run unless there is an increase in average costs and then price rises only to the extent of the increase in cost.

This underutilization of the price mechanism is then spread to other types of insurance because for several reasons hospitals, unlike physicians, have been reluctant to charge different prices for the same service. The price mechanism can only work in a roundabout manner—through increases in the price of complements such as physician's services and uninsured hospital services. This is not the most efficient way for the price system to work.

This near abandonment of price as the rationing mechanism has led to great dissatisfaction among physicians in every instance that such abandonment has taken place.[35] The main argument physicians have

[33] The supply schedule shifts to the right by the fixed amount (less costs of negotiating with the insurance companies including extra bookkeeping expenses); i.e. the payment acts like a reverse excise tax. The complicating factor is that due to the income effect mentioned above, the demand for hospital services will shift downward. But, it is most unlikely that all this reduction of income will be taken from hospital expenditures. Rather some will take the form of reduction in the consumption of other goods and services. Therefore the movement of supply is likely to be more than the movement of demand. This results in a lowering of price which is less than the insurance payment and probable rise in consumer surplus. Any increase in consumer surplus, though, is less than the insurance payments. This, of course, does not mean that the purchasing of this type of insurance is irrational. The insurance serves its main function of lowering the cost of sickness to those who get sick. In dollar terms this lowering is less than the cost of insurance to those who do not get sick but the reverse may well be true in terms of utility. As in the case of type one insurance, the more elastic the supply schedule, the more the insurance will benefit consumers.

[34] The fulfillment of these expectations is documented in Seymour E. Harris, *The Economics of American Medicine*. See the section on hospitals.

[35] "Europe's Advice to United States on Medical Care," *U.S. News & World Report*, Jan. 23, 1967, pp. 64–66.

with Medicare is that they must certify that a patient requires hospitalization in order for the patient to be eligible for benefits. This writer expects that relying upon the physician to refuse certification, thus running the risk of not being paid or losing the sufferer and his family as patients, is not the best way to ration scarce hospital resources; the complete abandonment of the price system should be avoided if at all possible.[36] A partial reliance upon the price system can be accomplished by requiring nonindigent medicare patients to pay at least part of their costs (say 30 percent) and perhaps by wider use of deductibles and co-insurance.

Summary, Conclusion, and Guides for Policy and Research

The main theoretical argument of this paper has been that, due to the absence of the profit motive and competition, the invisible hand cannot be expected to lead the hospital industry to an equilibrium position characterized by a socially efficient use of hospital resources. Indeed, we have seen that if the motivating force of profit maximization is replaced by output maximization subject to a budget constraint, an equilibrium position characterized by either overproduction or underproduction is possible. We conclude that a conscious effort, or "visible hand," is required to achieve industrial and allocative efficiency and that the importance of this conscious effort has grown and will continue to grow along with the increasing percentage of our national resources which is devoted to medical care.

The absence of competition as a regulator and the not-for-profit nature of activities in many parts of the medical system lessen the importance to management itself of adequately considering alternative costs; i.e. the benefits forgone by not having the resources available for a different use. This leads to our most important conclusion: Hospital authorities must be somehow persuaded to see the social importance of alternative costs and taught to use this tool of analysis in their pricing and other decisions.

Two further problems in the way of achieving efficiency are the limited extent to which hospital users are capable of making rational *medical* decisions and difficulties in relaying allocated costs to decision-makers. Patients frequently leave decisions completely in their physician's hands. It is often not a matter of accepting or rejecting the doctor's advice but of having previously decided to "do whatever the doctor says." The price mechanism does not perform its rationing function effectively in this circumstance, and the cost must be relayed to the physician who, in turn must advise or decide for the patient on the basis of his much better *medical* estimate of probable benefits and costs. Since the doctor's income rises more when he suggests use, a completely impartial decision cannot be expected; perhaps salaried

[36] Two studies of the National Health Service show physicians are not performing the rationing function adequately. Numerous unnecessary days of care were described in Gordon Forsyth and Robert F. L. Logan, *The Demand for Medical Care*, pp. 78–104. The form of rationing is criticized by Feldstein, who writes "First, there is the hospital physician's feeling of greater responsibility for the patient in his own personal care than for people who are waiting for admission. Medical training, medical ethics, and the law of negligence all emphasize this asymetry of of responsibility. 'The best possible care for my patient' forms the basis of the clinician's decisionmaking. This is, of course, an attitude to be admired in the doctor and one that, as potiential patients, we all hope exists. But it does offer an explanation of why doctors allow the number of patients treated to vary more in response to differences in bed scarcity than they allow the care given to the patients already admitted to hospital." Feldstein, *op. cit.*, p. 404.

physicians or diagnostic specialists would better perform this function.

One decision left almost entirely to the physician, is whether a given patient should be discharged. (Medicare *makes* discharging the doctor's responsibility.) This decision, should be based upon the marginal social cost involved in retaining the patient; these costs vary with different degrees of hospital utilization, type of malady, and patient, and employee scheduling practices. The relevant marginal cost needs to be relayed to the physician before he can base his decision on a comparison of probable cost and benefits.[37]

The problems associated with the price mechanism are further compounded by the existence of nonpatient sources of revenue (philanthropy, welfare) and by insurance companies (including medicare). Their effects should be structured, where possible, to foster the efficient use of hospital resources. For example, Medicare provides "a lifetime limit of 190-days of covered services in psychiatric hospitals. Psychiatric care in general hospitals, however, does not count against the 190-days lifetime limit".[38] (Psychiatric care in a general hospital is treated like any other illness.) The obvious effect of this provision is to induce people to seek psychiatric care in general hospitals rather than in psychiatric hospitals—at least after a while. This effect would make social sense only if there were diseconomies of scale in psychiatric treatment. If, however, large psychiatric hospitals increase the productivity of the extremely scarce U.S. supply of psychiatrists, then this Medicare provision is detrimental to the development of an efficient hospital system. If the rationing effect of price is to be altered in governmental programs, it should be altered to do a better, and not a worse, job.

From the point of view of the Nation as a whole, the matter of lavishing funds on large hospitals as a form of intercommunity competition for industrial plants and for physicians may be wasteful, unless it is argued that the main effect of a large marginal growth in hospital capacity is on the total number of physicians and industrial establishments rather than upon their location.[39] Such a competition becomes ludicrous in U.S. conditions since it is financed largely from central government taxes (Hill-Burton). This argument suggests that Hill-Burton administrators should consider only such factors as needs for patient care and option demand in determining the size hospital which a community needs (giving due consideration to area growth) and contribute funds only toward the construction of that size hospital (subject to the cost considerations related to optimal hospital size). While the local community might wish to build a large hospital for its qualities as a lure, none of the additional construction costs would be paid from the taxes of those citizens who stand to lose in the competition.

---

[37] It is the hospital's job to ascertain the probable level of utilization and, using this knowledge, to inform the physician of the relevant level of cost to be considered in a discharge decision. The hospital can require physicians to give estimates of how many of their patients they expect to be dischargeable in the near future. These estimates, perhaps taken from the patients' daily charts, can be compared with scheduled admissions. Statistical theory can then be used to estimate the level of excess capacity that will exist for the next few days. This estimate, along with the relevant cost findings for the hospital, should be relayed to physicians or made available to them before a final discharge decision must be made. (It is assumed the hospital would have previously explained the desirability of considering these cost findings to its house and attending staff.) Government grants might be used both to find the information and later, to put it into use.

[38] Howard West, "Health Insurance for the Aged: The Statistical Program," *Social Security Bulletin,* vol. 30, No. 1, January 1967, p. 16.

[39] This situation is similar to that surrounding the discussions concerning the social benefits derived from resources devoted to firm advertising.

Another source of difficulty in determining alternative costs, and thus establishing prices, is the widespread use of cash accounting in hospitals despite the AHA's endorsement of accrual accounting.[40] Perhaps accrual accounting could be required as a condition for a government grant or for medicare accounts.

Another important area in which alternative cost information is almost totally lacking is the quality-quantity dichotomy: In many decisions involving medical care, quality and quantity are substitutes for each other. Data relating to quality differences among hospitals are also grossly lacking. But these quality differences must be taken into account when investigating the average cost curves.[41] Several hospitals have attempted to determine their quality through a medical audit, i.e., an investigation of their medical records by impartial medical experts or, less desirably, by their house staff. These audits are not sufficiently homogeneous to be used in most cross section analysis; in addition, these audits have not been made in a wide enough range of hospitals. Perhaps a Government research grant could be given to establish a sound medical auditing procedure which would then be required of hospitals participating under certain programs.

Another grant might examine methods by which the price system could be utilized to induce a more efficient usage of hospital facilities. This research might examine the long-run marginal costs involved in maintaining 7-day operation of all parts of the hospital. Perhaps test hospitals could be established where doctors would be charged for various services after the fashion of peak and offpeak pricing of public utilities.

These suggested studies are not merely exhaustive. They show that much additional cost information is required before the visible hand of hospital managers, trustees, doctors, city councilmen, county commissioners, Federal and State officials, insurance and drug executives, and so forth, can achieve an efficient use of hospital resources. It is hoped this paper and the symposium will help lead to both a search for that information and an increased recognition of its importance.

---

[40] American Hospital Association, *Uniform Chart of Accounts and Definitions for Hospitals*, p. X.

[41] Caleb Smith, "Survey of the Empirical Evidence on Economies of Scale," in National Bureau of Economic Research, *Business Concentration and Price Policy*, pp. 220–223.

BIBLIOGRAPHY

Alchian, Armen A. and William R. Allen. *University Economics.* Belmont, Calif.: Wadsworth, 1964.

American Hospital Association, *Hospitals, Annual guide issue:* vol. 39, No. 15, pt. 2. (Aug. 1, 1965).

——— *Uniform Chart of Accounts and Definitions for Hospitals.* Chicago; American Hospital Association, 1959.

Board of Trustees of the American Hospital Association. "Guiding Policy for Hospitals in Pricing Their Services to the Public, Point III." in Lem E. Hay. *Budgeting and Cost Analysis for Hospital Management.* Bloomington, Ind.: Pressler Publications, 1963.

"Europe's Advice to U.S. on Medical Care." U.S. News & World Report: vol. 52, No. 4 (Jan. 23, 1967) pp. 64–66.

Federal Trade Commission. "Final Order (8519)" *In the Matter of the Community Blood Bank of the Kansas City Area, Inc.* in FTC, *News Summary:* No. 38. (Oct. 27, 1966). p. 1.

Feldstein, Martin S. "Hospital Bed Scarcity: An Analysis of the Effects of Inter-Regional Difference." *Economica.* vol. XXXII No. 128 (November 1965), pp. 393–409.

Forsyth, Gordon and Robert F. L. Logan. *The Demand for Medical Care.* London: Oxford University, 1960. 154 pp.

Harris, Seymour E. *The Economics of American Medicine.* New York: Macmillan, 1964. 526 pp.

Joint Economic Committee, Congress of the United States. *Federal Programs for the Development of Human Resources,* vol. 2, pt. III (partial). Washington: U. S. Government Printing Office, 1966.

Klarman, Herbert E. *The Economics of Health.* New York: Columbia University, 1965.

Krutilla, John V. and Otto Ekstein. *Multiple Purpose River Development.* Baltimore: Resources for the Future, John Hopkins Press, 1958.

Mitford, Jessica. *The American Way of Death.* New York: Simon & Schuster, 1963.

Montecute, Charles. *Costing and Efficiency in Hospitals.* London: Oxford University, 1962.

Reder, Melvin. "Some Problems in the Economics of Hospitals." *The American Economic Review:* vol. LV, No. 2 (May 1965), pp. 472–480.

Richman, Barry M. *Soviet Management.* Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Sidebotham, R. and C. S. Page. "Accounting for Local and Public Authorities" in Charles Montecute, *Costing and Efficiency in Hospitals.* London: Oxford University, 1962.

Smith, Caleb, "Survey of the Empirical Evidence on Economies of Scale" in National Bureau of Economic Research, *Business Concentration and Price Policy.* Princeton, N.J.: Princeton University, 1955.

Somers, Herman Miles and Anne Ramsay Somers. *Medicare and the Hospitals.* Washington: The Brookings Institution, 1967.

Stewart, William H. "The Positive Impact of Medicare on the Nation's Health Care Systems." *Social Security Bulletin,* vol. 30, No. 7, (July 1967), pp. 9–12.

Weisbrod, Burton A. "Collective-Consumption Services of Individual-Consumption Goods." *Quarterly Journal of Economics,* vol. 78, No. 3 (August 1964), pp. 471–477.

——— "Some Problems of Pricing and Resource Allocation in a Non-Profit Industry—The Hospitals." *Journal of Business,* 1965, pp. 18–28.

West, Howard. "Health Insurance for the Aged: the Statistical Program." *Social Security Bulletin,* vol. 30, No. 1 (January 1967), pp. 3–16.

Wilcox, Clair, *Public Policy Toward Business* (3d ed.). Homewood, Ill. Richard D. Irwin, 1966.

AN ANALYSIS OF ALTERNATIVE INCENTIVE REIMBURSEMENT PLANS

BY PAUL J. FELDSTEIN*

CONTENTS

I. INTRODUCTION

The purpose of this paper is to present a framework for analyzing the effects of several suggested methods for hospital reimbursement under Medicare. Although no attempt is made to discuss all possible reimbursement mechanisms, it is hoped that the framework and criteria offered are sufficiently general so that additional reimbursement methods may be considered as modifications in design and effect of the models discussed. In this way, it is hoped that the paper can serve as a background document for future evaluation of reimbursement systems.

A common problem with all reimbursement plans is the measurement and effect on quality of care. Since this problem is not unique to any one plan, a discussion of quality, its measurements, and incentives for it has been postponed to the latter part of the paper.

Comments are requested regarding the general models presented and their effects, as well as regarding modifications of these models. Suggestions are also requested on empirical indices for measuring the effects of alternative systems.

II. OBJECTIVES OF HOSPITAL REIMBURSEMENT SCHEMES

It is possible to choose from among two objectives in selecting a hospital incentive reimbursement system. The first objective would be to minimize the cost of *hospital* care for given levels of care, while the second would be to minimize the cost of an *illness* espisode. Under various reimbursement schemes, either one, both, or neither of these objectives may be achieved. Ideally, it would be desirable for a reimbursement method to achieve both goals; however, since this may not occur, there should be some discussion regarding the objective of a hospital reimbursement scheme.

*Bureau of the Budget and University of Michigan.

For purposes of definition, a payment scheme which seeks to reduce institutional costs by means of cost saving techniques and equipment, is defined as having the minimization of hospital costs as its goal. A payment scheme that results in the transfer of patients to less costly institutional settings as is medically possible, (e.g., from the hospital to the nursing home to home care) has the minimization of the total cost of an illness as its goal. In the first case, the emphasis is on *efficiency*, whereas in the second it is on *effective* use of facilities.[1]

While it would be desirable and sometimes possible to achieve both of the above goals, certain reimbursement systems may, in achieving one of the goals, have no effect or a negative effect on the achievement of the other goal. For example, a payment system that seeks to transfer patients to less costly institutions may have no effect on increasing efficiency in any of the institutional settings. Similarly, a payment system that results in a tendency toward hospital efficiency and also toward early discharge may increase the total costs of illness and shift them from the hospital to the patient. This shift may be either strictly financial or it may be social, i.e., increasing the patient hardships if there is no one available to care for him at home. Such a shift can take place if alternative institutions are not available or if care in such alternative institutions is not covered by insurance.

Therefore, in order for hospital costs to be minimized, there must be a tendency toward *efficiency*—to produce the same product, with the same level of quality, at a lower cost. In order for total illness costs to be minimized, there must be a tendency toward *effective* use of all the different institutional settings, i.e., patients are placed in or transferred to that setting which can provide the necessary care at the lowest cost. However, in examining this latter alternative, care must be taken to ascertain whether (a) alternative facilities are available, (b) the payment mechanism covers use in alternative facilities (the home as another institutional setting should not be disregarded), and (c) that the administrative, social, and medical costs of transfer are less than the gain from transferring.

Since these two objectives of efficiency of hospital care and effective use of institutions are not necessarily related, the incentives for achieving them also may not be the same. Therefore, the discussion of these alternative incentive systems—increasing efficiency within a particular institutional setting and increasing effective use of alternative institutions—will be handled separately in part IV of this paper.

III. Process Versus Output Incentives

Generally, two major types of incentive systems have been suggested to attain either or both of the above objectives. One is to promote incentives on what may be called the "process" for providing care, while the other is to develop incentives toward achieving the specified output directly. Examples of process incentives include rewarding hospitals

[1] Both of the above goals are problems in achieving greater efficiency. In the first case, the hospital is the firm combining a series of inputs to minimize the cost of a given output. In the second, some other unit may be the firm (physician, patient, or even hospital) combining the inputs, which in this case are institutions, in such a manner so as to minimize the cost of an illness to the patient. Hospital costs in the first case become an input into the decisionmaking process in the second. The payment systems, incentives operating, and assumptions about the purchaser in these two cases may not be the same and are, therefore, treated separately. The word "effectiveness" was used since it is believed to connote these differences and is familiar to persons in the health field.

if they subscribe to data collecting agencies such as HAS, PAS, and further to credit them if they develop certain institutional structures within the hospital such as utilization review commitees. Such process incentives have as their expectation that if these tasks are undertaken, then reductions in cost for specified output will result. However, rewards for process incentives may lead to the production of a great many of the processes without any necessary change in the cost of care.

Direct output incentives, on the other hand, base their rewards and penalties on the final output itself, i.e., on a given quality of care at minimum cost. These are similar to the incentives under which the private sector of the economy operates, where rewards are related to prices and to the costs of production of the output. Such incentives not only encourage desirable processes to be undertaken, but also encourage a greater search for applicable information and technology from other fields.

Therefore, the assumption made in this paper is that the incentives are directed toward the final product rather than toward the process. If, on the other hand, process incentives were desired, then a list of them would have to be developed, supported by empirical evidence demonstrating that process incentives do in fact work.

IV. Alternative Reimbursement Systems

Alternative reimbursement systems involving various incentives directed toward the final product, are considered in the following sections. Section A deals with alternative hospital reimbursement systems, which are aimed at efficiency of operation. Detailed attention is given to reimbursement systems that measure the performance of hospitals in relation to each other. Then payment systems based on the hospital's individual performance are dealt with. Section B sets forth alternative payment systems designed to minimize the cost of an entire illness episode and to increase the effective utilization of different kinds of institutions.

A. Alternative Hospital Reimbursement Systems

In discussing alternative hospital payment systems, reference is made either to a hospital's relative cost curve (compared to other hospitals) or to its own average and marginal cost curves. As shown in figures 1, 2, and 3, the hospital's average cost curve is represented as being U-shaped, as is suggested by economic theory. The average cost curves shown represent the average cost per unit, e.g., a patient day, at different levels of output, e.g., number of patient days. In different cases, this average cost curve may represent the average cost of caring for all patients in the hospital, the average cost of caring for a particular type of patient, or the average cost of providing a given service to patients. With regard to the different payment systems to be discussed, what the average cost curve represents is specified in each of the payment systems considered.

Fig. 1: A Long Run Average Cost Curve

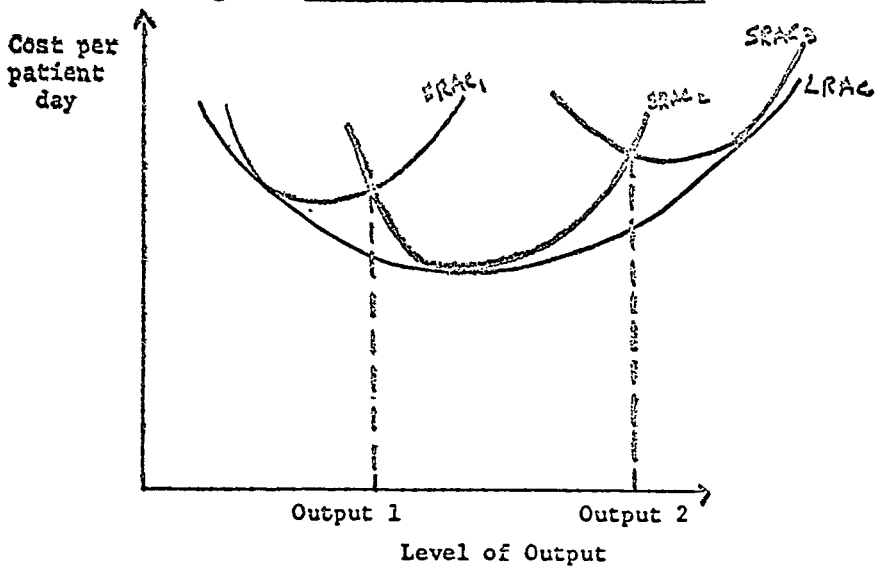


Fig. 2: Two Short Run Average Cost Curves

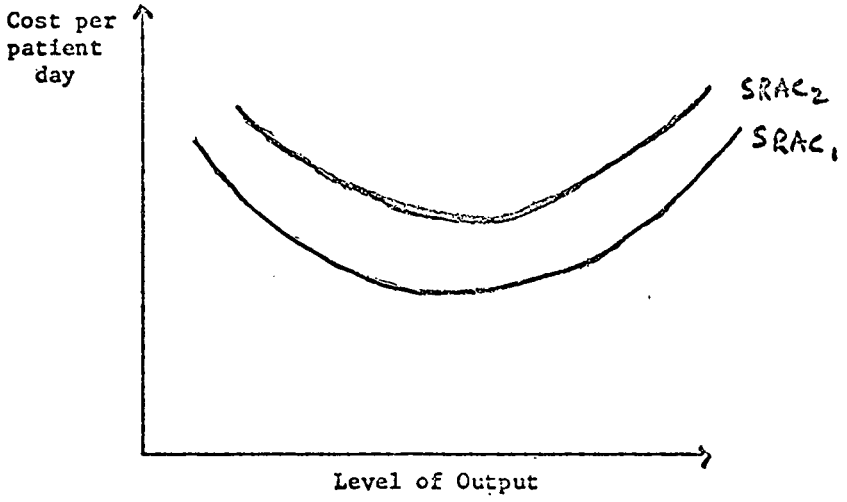
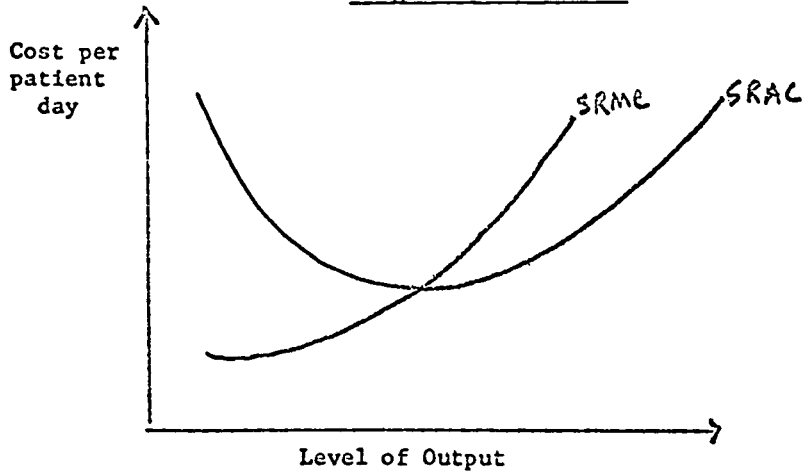


Fig. 3: A Hospital's Short Run Average and Marginal Cost Curves



Reference is also made in the following sections to both long-run (LRAC) and short-run (SRAC) average cost curves. The difference between these curves is that, in the short run, not all the components of hospital costs can be changed, e.g., scale of plant, while in the long run the time period is long enough for all components, including size of hospital, to be varied. There are also long-run (LRMC) and short-run (SRMC) marginal cost curves, which are similar to the LRAC and SRAC.

These represent the additional costs (increments to total costs) incurred by increasing output, e.g., patient days, by an additional unit. LRMC include all the additional costs for handling an additional patient such as an increase in bed capacity and staff, etc., while SRMC represent those additional costs that will be incurred—given the existing hospitals' facilities and staff—for caring for an additional patient day. This analysis can be made for a particular type of patient, e.g., aged, or for producing an additional unit of service in the laundry or in any other department.[2]

1. *Payment Based on Relative Performance*

Given this background on cost curves, the observed differences in hospital costs may be said to result from one or more of the following:

(a) The hospitals may be operating on different portions of their longrun average cost curves (i.e., they may have different ranges of output). This is shown in figure 1 where $SRAC_1$, $SRAC_2$, and $SRAC_3$ represent hospitals of different sizes. The hospitals differ only in the scale of their operations. (For outputs up to output 1, hospital 1 has lowest average costs, and above output 2, hospital 3 has lowest average costs.)

(b) The hospitals may be operating on the same point on their LRAC, i.e., at the same level of output, however, their LRAC are

[2] For a more complete discussion of economies of scale in hospitals and their measurement, see W. John Carr and Paul J. Feldstein, "The Relationship of Cost to Hospital Size" *Inquiry*, July 1967 and also Ralph Berry, "Returns to Scale in the Production of Hospital Services", *Health Services Research*, Summer 1967.

either higher or lower than the other hospitals because they are producing different products, e.g., differences in quality, caring for a different mix of patients, etc. If it were possible to adjust for these differences in product then their cost curves would be similar.

(c) The hospitals may be operating at the same level of output but their LRAC are either higher or lower than the other hospitals because of differences in efficiency. Figure 2 shows that hospital 1 has lower average costs than hospital 2 for all ranges of output.

(d) The hospitals may be operating at the same level of output and have similar SRAC but due to their accounting systems, their SRAC for a given type of patient or service may appear different. (Under these circumstances, a reimbursement system based on the average of the costs of all hospitals provides an incentive for hospitals to find out what their real costs are.)

With these reasons for the differences between hospital costs in mind, different reimbursement systems may be analyzed according to the relationship that the reimbursement price has to the hospital's LRAC or SRAC relative to the costs of other hospitals in the community.[3]

(a) *Reimbursement according to mean average costs of all hospitals in the community.*—One basic method of relative reimbursement is to set the reimbursement price according to the average of the average costs of all the hospitals in the community, thus rewarding hospitals operating below the mean and penalizing those operating above it.

If hospitals are operating on different portions of their LRAC under a mean average cost reimbursement scheme (see fig. 1), then the lower cost hospitals will presumably increase the provision of care, until their costs eventually begin to rise. On the other hand, if hospitals are high cost because they are operating on the increasing portion of their LRAC, then a lesser payment for their services will cause them to contract their services—in which case their cost per unit may fall. If the hospitals are high cost because they are operating on the declining portion of their LRAC, then they will either have to cease production of their service, change their services, or stop offering it at the reimbursement price paid. In the case of hospitals whose costs are high because they are operating on the declining portion of their LRAC, eventual increases in demand for such services may enable them to expand their output sufficiently so that their cost per unit falls.

If differences between a hospital's costs are not a result of differences in their position on the same LRAC curve (i.e., their level of output), then such cost differences may be a result of relative efficiency in producing the same product or similar degrees of efficiency in producing different products. (With reference to fig. 2, the different average cost

[3] One of the controversies under the present reimbursement system and certainly to be one in any proposed system is the extent to which third-party payers and private-pay patients should pay part of the costs of activities other than patient care, e.g., education and research. The argument being that the benefits from these activities accrue to everyone and therefore they should be subsidized. The opponents of this position would suggest that the most equitable form of subsidization of these activities should not be borne by the present sick; that if these activities should in fact be subsidized then the funds for these activities should come from separate sources (preferably general tax funds) and that there be some determination made as to which institutions could most efficiently undertake these activities. In any case, the reimbursement systems discussed assume that all differences in hospital costs are a result of differences in costs of patient care. Hospital cost systems should be able to separate these other costs from patient-care costs and that payment for these other activities can be handled/discussed separately.

curves could represent two different products or differences in efficiency for the same product.) In such cases, rewarding the low-cost hospital and penalizing the high-cost one should cause the low-cost hospital to expand its services in the long run. The high-cost hospital should either reduce the number of aged patients it serves, reduce the number of its services (quality), subsidize these patients from other types of patients, or cease offering service to Medicare patients at the going price.

In summary, rewarding hospitals whose operating costs are below the mean, and penalizing those above it would result in the total amount expended on hospital reimbursement being less than if total costs were reimbursed. It is also hoped that it would result in the expansion of the more "efficient" hospitals, while the less "efficient" ones would be forced to become more efficient or to contract their services.[4]

There could also be modifications to the payment systems for those hospitals above and below the mean. For example, hospitals whose costs are above the mean could be penalized by either receiving less funds, some percentage of the difference in their costs and the mean cost, or to just subject them to more intensive auditing procedures. For hospitals operating below the mean, the rewards may be personal (e.g., bonus for administrative and/or rest of staff) or institutional (e.g., paying them some percentage of the difference in their costs and the mean costs). The effect of any of these modifications for rewards and penalties (if which there are many more than mentioned here) will still be similar in effect to the simple model discussed; that is, reward hospitals below the mean while penalizing those above it.[5]

(b) *Modifications of the system of payment by mean average cost.*—There are many modifications of the system of reimbursing hospitals on the basis of mean average costs. In this section, several of these modifications are considered.

(i) Hospitals may be classified by size, service, control, etc., before determining the mean average cost.

(ii) Payment may be based on the mean average cost of *intermediate products* that are considered to be subject to economies of scale (e.g., X-ray, laundry, etc.). The effects of this policy would be similar to reimbursement based on the mean average cost of the final product, patient care. However, instead of being forced to cease production of patient care or participation in Medicare, a higher cost hospital would have to seek to purchase the intermediate output from another hospital or institution in the community.

---

[4] There has been some discussion as to whether reward payments should be restricted in their use (e.g., to construction or modernization funds) or whether they should be lump sums for use as the hospital sees fit. In my opinion, unrestricted rather than restricted funds are preferable. A hospital may have a greater need to introduce cost-saving equipment, to hire better trained staff, to provide more services, or to educate its administrative staff in the latest management techniques than it does to provide additional facilities. Thus, the restriction of reward payments to limited purposes is less efficient than allowing the hospital to spend such payments on the things it needs most.

[5] There has been some concern that decreasing payments to a high cost hospital would result in hardships to the patients still using that hospital. The suggestion has therefore been made that as long as the hospital's intentions are good, its payments should not be decreased. However, it would appear that under the present "cost-plus" system of reimbursement, hospitals that are high cost and with good intentions already have had the time to change—if they are going to.

Since it is not clear that patients in these hospitals are better off before the change, the changes desired should be made as rapidly as possible rather than as slowly as possible. One method of achieving this would be to pay high cost hospitals less and inform their boards of trustees of this, in order to encourage a change in the management of the institutions.

(iii) Hospitals are reimbursed according to their individual costs, but they only receive the mean increase in costs of all the hospitals over the year or over some other time period. (A further modification of this is to first classify the hospitals and then to calculate the mean increase in costs for each group.) The individual SRAC curves (whether it is based on all patients or particular categories) are rising at different rates and reimbursement would be according to the average increase.[6]

The incentive in this case is on inhibiting rising hospital costs. Hopefully, hospitals will continually substitute the inputs they use so that those inputs whose costs are increasing less rapidly will be substituted for inputs whose costs are rising faster. For example, there might be more capital investment to offset rising labor cost and there could be a greater use of lesser skilled persons in place of highly skilled persons. If the calculations of the mean rise in hospital costs were based on all patients and services, then there might be a tendency to change the mix of kinds of patients and kinds of services. This would result in an emphasis upon those patients and services whose input costs were rising at a slower rate. However, this payment system could also provide an incentive toward collusive arrangements between hospitals with regard to cost increases and prices paid for inputs (e.g., collusive arrangements with regard to nurses salaries, as described by D. Yett).[7]

The effects of such a payment system would be similar to those described for the mean average cost reimbursement system. The lower cost hospitals would be rewarded, and would be able to expand, while the higher cost hospitals would in some manner be penalized. If their costs were higher because of product differences (e.g., because of higher quality) then these quality differences could be cut back. Another problem with this method of reimbursement is that the higher cost hospitals will start from a more advantageous position than the lower cost, more efficient hospitals. This is because those hospitals with higher costs will presumably have more room to reduce costs.

In another related reimbursement proposal, hospitals would be paid on a delay basis. This would mean reimbursing hospitals for their individual average costs, but not reimbursing them for their rising costs for 6 months or a year. The effects of this proposal would be equivalent to considering all hospitals as being the high-cost hospitals when the payment price is based on mean average costs or as being those hospitals whose rise in costs is greater than that for the average and payment is based on the average increase in costs. (In this case, price is always below a constantly rising SRAC curve.)

(iv) The proposal of paying all hospitals so that the reimbursement price is always lower than their average costs is similar to placing them in the position of the hospitals in the earlier case whose average costs were above the mean and payment was based on the mean average costs of the hospitals in the community. Unfortunately, under

---

[6] It is assumed that the average *dollar* increase is used rather than the average *percent* increase. The average percent increase would penalize the lower cost institutions while providing the higher cost ones with an amount in excess of their cost increase.

[7] Donald Yett, "The Supply of Nurses: An Economist's View," *Hospital Progress*, February 1965.

this proposal the contraction of service and the reduction in quality would not be met by an increase elsewhere.[8]

(v) Another modification of the general approach of setting price in relation to mean average costs is to set price according to the average of all the hospital's short run fixed costs ("readiness to serve" costs) and pay actual additional costs (SRMC). Here again the effects would be the same as relating price to mean average costs; however, an additional effect would be to decrease short run fixed cost by having higher variable costs, e.g., bonuses, etc., in order to achieve greater short-run flexibility. Since these variable costs would be fully reimburseable, this program could also result in overall higher program cost.[9]

(vi) A final example of a payment system based on relative performance is to reward hospitals in a reverse way. Pay the higher cost hospitals more than the low cost ones so that they can presumably become more efficient, e.g., with more funds for modernization of plant and equipment or with higher salaries for better trained staff. The effect of such a proposal is the reverse of the earlier case. The higher cost hospitals would expand in those areas where there are more funds—they would not necessarily become more efficient. However, an incentive may have been created for the low cost hospital to become high cost!

2. Payment Based on Individual Performance

The analysis of reimbursement schemes based on an individual hospital's performance is also done with reference to the hospital's average and marginal costs, shown in figure 3. As in the previous diagrams, the theoretical SRAC curve decreases as more units of service are provided within a given size of hospital. Then, after some point, it begins to rise. Empirical evidence on the shape and magnitude of the SRAC and SRMC curves, however, found SRMC to be small, approximately 25 percent of SRAC.[10]

Over the range observed, on the other hand—which included quite high occupancy rates—SRMC did not increase but remained constant.[11] These findings, which are consistent with prevailing opinion, have important implications for the payment schemes discussed below.

(a) *Payment of full costs.*—The first of the payment schemes based on individual performance is one involving payment of full costs. Under this payment system, the hospital would be reimbursed according to its actual average cost per unit, regardless of whether the hospital were operating on the low or high point of its average cost curve. Thus, there would be no incentive to reduce costs, either by changing its rate of output (i.e., moving to the low point on its cost

[8] The scheme of setting the reimbursement price lower than average costs for all hospitals seems analogous to the case of the farmer that decided to reduce his expenses by feeding his horse less and less each day. However, just when he had the horse trained to get along without food, it died.

[9] A variation of this approach might be to pay actual short run fixed costs ("readiness to serve" costs) and the average of all of the hospital's additional costs per patient day (SRMC). The incentive would then be to substitute SR fixed costs for SR marginal costs, e.g., hire more staff on a full-time basis, etc., so that SRMC would decrease still further as a percent of total costs. Such a policy would also probably increase rather than decrease total program costs. It would also be quite difficult to measure readiness to serve and short-run marginal costs.

[10] Paul J. Feldstein, *An Empirical Investigation of the Marginal Cost of Hospital Services*, graduate program in hospital administration, University of Chicago, Chicago, 1961.

[11] It is believed that rising SRMC were not observed because if hospitals were operated at occupancy levels greater than 100 percent for any period of time, there would be additional social costs to the patient which the hospital would not want to incur.

curve) or by becoming more efficient at all levels of output (i.e., shifting the entire cost curve down). The only incentive under this payment system would be that of enabling the hospital to achieve whatever goals it might desire, e.g., to add additional services and activities, become a larger hospital, etc. As long as the reimbursement price were based on average cost per patient-day and the marginal cost of an aged patient were less than average cost, the hospital would receive more than the cost of providing that care. Since the proportion of patient days used by aged patients (presently 25 percent) is expected to increase, this contribution to the hospital's other activities would not be insignificant.[12]

There are various modifications to this payment scheme. For example, the reimbursement rate may be set either slightly above or slightly below average cost. This is comparable to existing and proposed modifications to the present reimbursement system. In either case, there would be little incentive toward efficiency in such a payment system.[13]

Another modification is to contract with a hospital to provide so many patient-days, e.g., 10,000 at so many dollars per patient-day. The intent of such a negotiated fee is to provide the hospital with an incentive to produce care for less than the negotiated price. The problems with this proposal are similar to those with proposals based on a relative cost (relative to other hospitals) payment system. Unless quality can be measured, there might be a tendency to decrease quality. Further, there might be a tendency toward longer lengths of stay, since presumably the marginal cost of an additional day would be lower than the marginal cost of a new patient, while the price paid to the hospital per day would be the same. Also, to the extent possible, a hospital would choose patients with less costly diagnoses in order to reduce its cost.[14]

(b) *Payment Based on Fixed ("readiness to serve") and Incremental Costs.*—Under this proposal, payments to hospitals are based

---

[12] The study by Edward Kaitz, "Pricing Policy and Cost Behavior in the Hospital Industry," unpublished doctoral thesis, Graduate School of Business Administration, Harvard University, February 1967, attempted to analyze the effect on hospital costs and expansion of services of a payment system based upon "reasonable costs." One of the conclusions of the study (based upon personal interviews with hospital administrators in Massachusetts) was that "* * * cost based reimbursement systems provide what may be regarded as the most crucial incentive for the development of ancillary services" (p. 5.14).

[13] The idea of a negotiated rate is really a modification of payment according to individual hospital costs. Depending upon the purchasing power of the buyer, the negotiated rate could be anywhere from an amount less than average costs (but not less than variable costs) to an amount that may be much greater than full costs. The question of possible quality reduction if the negotiated rate is less than what the hospital requests is always present. Further, small purchasers may be subsidizing the larger purchasers.

In negotiating a rate the incentive is for the hospital to negotiate harder; the negotiator for the purchaser does not have as much at stake unless there are personal incentives to him. Situations such as this generally provide an incentive toward bribery for those negotiators that have no personal stake involved.

Unless payment is made according to some formula basis, e.g., relative reimbursement, etc., then negotiated rates may not be much different than the present system of payment of costs. Further, some inequities will be involved, e.g., the rewards going to those that can bargain better. A negotiated rate provides some incentive to reduce costs (assuming quality is not reduced instead) because the hospital can retain some percent of the difference in its costs and the negotiated rate. However, this may not be a very strong incentive since "* * * the hospital may find it difficult to obtain a subsequent increase in charges until its projected costs provide a justification" (Irwin Wolkstein. "The Legislative History of Medicare Hospital Cost Reimbursement." a paper prepared for the Secretary's Advisory Committee on Hospital Effectiveness. Nov. 1, 1967).

[14] If the contract were for cases rather than for patient days, then the incentive for the hospital would be to have shorter lengths of stay. Unless there are alternative facilities available and funds to pay for them, early discharge might result in certain hardships to the patient. Also, if readmissions were to result, then there would be an incentive for the hospital to define them as *new* cases.

upon two sets of cost data: the first set are those costs which do not vary in the short run, e.g., the usual fixed costs, plus costs such as salaries, which the hospital cannot vary during a given time period (e.g., a month); the second set are based upon incremental costs, i.e., SRMC per patient day.

There are several variations to this payment scheme. One is to pay actual costs presented by the hospital in the above manner. In this case, the intended effect of such a payment approach would be to have the hospitals develop improved methods for budgeting and prediction of costs, which might then lead to greater emphasis on cost control techniques. However, other than this hoped for effect, a payment system based on actual costs does not provide any further incentive to reduce costs.

A modification of the payment scheme based on fixed and incremental costs is to pay actual short run fixed expenses ("readiness to serve" expenses) plus a specified incremental cost per patient day (SRMC). The intent here is to place some control on additional patients admitted.[15] For instance, if the hospital were operating near capacity under this system, then it would be faced with the prospect of accepting a small payment (SRMC) for additional patient-days compared to their *opportunity cost* of discharging the patient earlier and admitting a different patient. In the latter case, it would receive an amount in revenue much greater than its SRMC and perhaps equivalent to its SRAC per patient-day. Therefore, a hospital operating at or near capacity would have an incentive under this system for early discharge of aged patients and the admission of new patients in order to maximize the net revenue it receives.[16]

For a hospital operating at a low occupancy rate, then its opportunity cost of discharging a patient whose continued stay results in a payment approximately equal to its SRMC is its SRMC. In this case, as long as the payment was in excess of its SRMC the hospital would have a tendency to keep the patient on—hence increasing its occupancy rate. There would seem to be little incentive toward increased efficiency under this type of payment system. Instead, it would appear to influence whether or not facilities are being *effectively* used.

In the long run, that is, that period when the hospital can vary all its resources, including its size of plant or number of beds, a payment system based upon some specified level of LRMC and payment of actual LRAC (if it can be determined), would have the following effect: since payment is based on actual LRAC, we would expect to observe greater economies of scale; hospitals would have a tendency to add additional services, build in greater flexibility, etc., thus increasing their average costs but decreasing their LRMC. This would lead to fewer hospitals serving an area and as a result to an increased

---

[15] An example of such a plan is the one used in Saskatchewan, see Malcolm Taylor, *The Administration of Health Insurance in Canada*, Oxford University Press, Toronto, 1956, p. 163.

[16] If there are conditions attached to receiving a large sum in order to cover readiness to serve costs, such as the handling of so many aged-patient-days, then the tendency would be to have longer lengths of stay. This is because the marginal costs are lower for longer term patients than for new stays. Once the requirement of number of patient-days has been met, the tendency would be to admit nonaged patients, since the incremental profit would be greater. (The differences in new revenue are the SRMC of an aged patient and the price per aged-patient-day—which would approximate SRMC compared to the SRMC of a new patient and the price charged—which would be in the neighborhood of SRAC.)

problem of longer travel times for patients, their visitors, and physicians.

However, if the number of hospitals did not decrease to take advantage of larger economies of scale, and if demand did not increase sufficiently to enable the expected larger hospitals to operate at that level of output where their costs are at a minimum, then it would be possible that the overall costs of the program would increase. (In this case, LRMC would be less than LRAC at the level of output where the hospital is operating, whereas other payment systems might reimburse the hospital at the point where LRAC is at a minimum.)

If the reimbursement price were set in relation to what the optimum scale of plant would be for a given expected rate of output and if payment were made for actual LRMC, then the effect would be similar to the examples given earlier of setting price according to a hospital's cost relative to the cost of other hospitals in the community. For example, if the price for the individual hospital were set either higher or lower than its LRAC, then it would be equivalent to being either a low- or high-cost hospital in the earlier relative reimbursement examples.

With all of these reimbursement approaches, however, estimation of long- and short-run "readiness to serve" and incremental costs would present large problems both conceptual and empirical for use in pricing.

(c) *Declining reimbursement price with increased length of stay.*— Under this proposal, an initial price per patient-day is set which is equal to or greater than SRAC per patient-day, and then as length of stay increases, the price paid per patient-day decreases. The intent here is to provide an incentive toward early discharge.

This payment scheme is similar to the earlier one where payment is based on a short-run fixed and incremental cost. In both cases, the decision to discharge will be based on the opportunity cost of a longer length of stay. For example, as length of stay increases and the price paid falls, the opportunity cost to the hospital of a longer length of stay is whether another patient can be admitted (even another aged patient) or whether there is additional capacity, e.g., low occupancy. In a situation of low occupancy, there is an incentive for the hospital to keep the patient as long as the price per day exceeds the hospital's SRMC of caring for the patient.[17]

An example of such a scheme where the relevant costs weren't given proper consideration was the payment system once used by a certain Blue Cross plan. Under this plan, price per day decreased with increased length of stay. However, the lowest price paid for additional days of stay was greatly in excess of the hospital's SRMC. Unfortunately, there was excess capacity in the area, so instead of reducing length of stay, the scheme merely provided an incentive for increasing it.

---

[17] Although the costs per day of a shorter stay are greater than the costs per day of a longer stay, the important factor is the "incremental net revenue" per patient-day. A hospital operating near capacity, therefore, under a decreasing payment system would probably find it could increase its net revenue by having shorter average lengths of stay. In this way, the hospital would be able to receive a greater incremental net revenue, since the price per day of new patients would probably be proportionately greater than the difference in costs for short versus long stays.

B. ALTERNATIVE ILLNESS (OR CAPITATION) PAYMENT SCHEMES

In addition to the reimbursement systems described above, there are other reimbursement schemes available that are quite different in both their intentions and their effects. These schemes involve lump-sum payments to either a purchaser or provider of care. They are similar to plans under which a patient receives indemnity insurance when he becomes ill or a physician receives funds to provide care for a patient's specific illness (or all his medical care needs) or an institution receives funds to provide care for a prospective number of patients. This reimbursement scheme is essentially a form of insurance coverage. Payment is based upon some average expected expense, and it is reimbursable either to the patient, to the physician, or to a hospital (if either the physician or hospital choose to undertake this role).

Theoretically, if the purchaser of care were knowledgeable regarding alternative forms of treatment for particular illnesses, and if he were aware of the alternative prices and benefits of these alternative forms of care, then the patient (or his purchaser) would allocate the available funds in such a way as to provide a maximum amount of care for a given budget. If, for example, a physician undertook this role for a patient (the physician either insuring the patient or receiving the payment for care for a specific illness), then the physican would provide his services and contract out for other services, e.g., hospital care, when he thought it was necessary. We would expect the same to occur if another agency played this role, e.g., the hospital.[18]

This approach would promote *effective* use of alternative forms of institutional care, since the purchaser would switch to less expensive forms of care as it becomes possible. Also, this form of care could develop incentives toward *efficiency* within individual institutional settings, since in selecting, for example, among alternative hospitals, the purchaser (e.g., the physician) would be selecting according to both the price and quality of care of that institution. (In this case, this payment scheme becomes similar in its effects to the case of setting price equal to the mean average cost of the hospitals in the community.)

Incentive systems could also be built into these alternative illness or capitation payment schemes. For example, if comprehensive prepaid group practice results in lower use of the hospital and thereby a lower total cost of patient care compared to a payment system which results in greater use of the hospital and hence higher patient illness costs, then a comprehensive prepaid group plan could receive some of the difference in their costs and the costs of the higher priced plans. This incentive should encourage the expansion of such plans.

Payment to a purchaser, either physician or hospital, for providing care for a given illness episode has the following problems which must be worked out:

1. What is the incentive system to the purchaser, e.g., a fixed fee per illness episode? The problems with this are the same as when relative reimbursement rates are set, e.g., what happens to high cost providers, how to insure that quality is not reduced, and so forth.

[18] An example of a plan where the hospital would provide this role on a capitation rather than illness basis is the plan proposed by Thomas Tierney and Robert Sigmond, "Could Capitation Ease Blue Cross Ills?" *Modern Hospital,* August 1965.

2. What is the definition of an illness episode? There might by a tendency for the provider to define an illness, when possible, as one which carries with it the highest fixed fee possible. Further, there may be a tendency to define a long illness episode, with multiple hospital admissions, as separate illness episodes, if payment is based on a fixed price per illness episode.

3. If the fixed fee is on a capitation basis rather than for an illness episode, then there is still the same type of problem as with relative reimbursement rates, i.e., what if the cost of providing care exceeds the capitation fee paid per person? Does the provider suffer the loss, do other patients subsidize it, is quality reduced?

V. Criteria for Evaluating Alternative Reimbursement Systems

A number of possible consequences will occur as a result of implementing any of the alternative payment systems discussed in this paper. Of these effects, some decisions must be made with regard to which ones are to be included in the evaluation of alternative reimbursement systems and the relative importance that should be given to each. The following are some criteria the author believes should be considered in evaluating reimbursement plans:

A. HOSPITAL EFFICIENCY

Comparisons of alternative incentive systems would have to consider whether hospital efficiency has increased.

Increase in efficiency may be empirically defined as the change in average cost per patient day (or stay), either for all patients or for that class for which the hospital is being reimbursed. In the dynamic hospital situation, increased efficiency may not result in a decrease in costs but merely in a slower rise in costs.

When the objective is to decrease total costs per illness episode, then data on total costs would have to be collected. The interpretation of such data would be somewhat more difficult, since data on transfers as well as length of stay and costs in each institutional setting, including the home, would have to be collected. (The problem of ascertaining quality of care for this objective and for the objective of hospital efficiency is discussed below.)

Another way of observing whether an increase in efficiency has occurred is to see whether there is an increase in the number of institutions that now share facilities and equipment. Data measuring the rate at which this is occurring could be easily collected and the problem of quality would not be as great as it is in measuring hospital or illness costs.

A problem which may occur when payment is based on relative performance is collusion among hospitals. It will certainly be in the interest of the higher cost hospitals, or hospitals with larger than average increases in cost, to promote such behavior. However, it should be relatively easy to detect and deal with such behavior. (If necessary, personal financial incentives could also be instituted.)

If the more efficient institutions are rewarded, we should be able to observe an expansion of their services and either a contraction of the services of higher cost hospitals or an attempt to lower their

cost per patient day. If such a tendency could be observed, it would improve our knowledge about the effects of reimbursement systems. On the other hand, if change in unit cost and patient loads did not occur as expected, then hospitals may be operating with other goals in mind that would have to be understood if the efficiency of the system is to be affected.

B. QUALITY OF CARE

In evaluating alternative reimbursement systems, their effect on quality of care cannot be disregarded. Are the more "efficient" really efficient or really lower in quality? Do relative reimbursement methods result in administrators taking the easy way out—reducing quality rather than seeking cost-saving methods? When considering incentives toward early discharge, the "quality of care" in that institution will be different than an institution that provides a longer length of stay, especially if inputs are used as the measure of quality. However, in order to test quality in this case it is necessary to measure the care received in all the institutional settings providing care for a particular illness episode—not only the hospital that provides early discharge. It is only in this way that differences in quality of care under different systems can be effectively compared.

Any experimentation on incentive reimbursement rates should also provide for experimentation on incentives for different levels of quality. Quality incentive systems should be such that the payment for quality should be greater than the cost of achieving or maintaining that level of quality. An important question is how much quality is to be paid for? It is possible to increase quality (if it can be defined) by paying more than the cost of providing it. However, at some point the marginal return to increased quality becomes too expensive (in terms of its value and the price necessary to achieve it). Further, if different levels of quality can be distinguished, should an incentive be provided for every hospital to have the highest level of quality—should every hospital be a Massachusetts General? If the payment system is on the basis of relative costs with an additional amount for quality, then what is the incentive to the patient to go to a lower quality institution (even though it may have the proper level of quality for the type of care he needs)? One possible solution to this problem is to set quality levels by diagnosis rather than by hospital.

Examples of methods for determining different levels of quality to be used as the basis for quality reimbursement would be:

(a) peer group establishment of criteria for input services into different patient diagnoses.

(b) peer group review of a random sample of all patient records in the community (not just the low-cost or high-cost hospitals). This group would then rate the hospitals according to level of quality provided.

(c) shift the quality criteria problem to the industry itself; i.e., the establishment of a quality payment schedule if high-cost hospitals can develop quality measures that show they are appreciably different from low-cost hospitals—and if these measures are agreed to by all the hospitals.

In any case, the introduction of any relative reimbursement system for producing increased efficiency should carry with it a second payment schedule for reimbursement for different levels of quality.

C. HOSPITALS DROPPING OUT OF THE SYSTEM

Another question to be considered in evaluating reimbursement systems is: How important is it if some payment system results in hospitals choosing not to participate? If a hospital dropped out, then there would be the problem of increased travel for both the patient, his visitors, and his physician to other institutions. Further, if a hospital dropped out of the payment system voluntarily or involuntarily (by going bankrupt), there would be the problem of staff appointments for the patient's physician. Would a patient have to seek a physician who had an appointment or was willing to be on the staff of an "approved" hospital?

D. SHIFTING OF COSTS

A good incentive system should not enable a hospital to shift (either medically or financially) part of the cost of hospital care to the patient or to other patients in the hospital. Such shifting of costs could occur, for example, if the hospital found that it had to increase its charges to other patients or for other services in order to continue providing care to aged patients.

Also, the promotion of early discharge should not impose costs on the patient that are not covered by insurance for care he needs after he leaves the hospital.

E. IMPROVED COST ACCOUNTING SYSTEMS

A desired byproduct of an incentive system should be an improved accounting system. The hospital should develop a cost system according to its product lines, e.g., by type of patient, by service, etc. In this way, the hospital would be better able to predict its costs, as well as more likely to be able to determine the advantages, if any, of shared facilities, contracted services and benefits of expansion.

F. ADMINISTRATIVE COSTS OF ALTERNATIVE SYSTEMS

Lastly, the administrative costs (including increased auditing, etc.) of proposed payment systems, should be considered so that alternative reimbursement systems can be evaluated on the basis of all their relative costs and benefits.

VI. Preferences

Since the purpose of this paper is to serve as a background document for analyzing alternative reimbursement methods, its ending is deliberately inconclusive. However, after thinking of the problem, one naturally begins to develop certain preferences. Although still tentative, the following are my preferences:

1. If substantial reductions in medical care costs are desired, then reimbursement experiments should consider major innovative schemes for reimbursement, e.g., relative reimbursement, capitation methods, prepaid group practice, etc., rather than merely experimenting with minor modifications of the present reimbursement formula.

2. Incentives should be directed toward the final product rather than the process of achieving it, for the reasons stated earlier.

3. The reimbursement system should contain some element of relative performance. Further, I believe such performance should be judged on the basis of particular product lines, e.g., type of patient, type of service, rather than on the overall level of costs or on the yearly dollar increase for the whole hospital. If such data are not available, given present cost accounting systems in hospitals, then payment could be made on a more general basis. However, it would be in the interest of the hospitals concerned, especially the higher cost ones, to cost out their product lines. Then, payment could be based upon the relative costs or dollar increase of particular patients or services, at least for those hospitals that have broken down their accounts in this way.

4. Any experimentation on relative reimbursement rates should also provide for experimentation with reimbursement systems for different levels of quality. Such quality reimbursement incentives could be based upon levels of quality developed by peer group judgment, by peer group review of patient records, or by letting the higher cost hospitals develop such quality indices which are acceptable to other hospitals in the community. This would promote among the hospitals themselves the incentive to develop and justify quality measures to the other hospitals in the community and hence it would improve decisionmaking with regard to what is being purchased for a given price.

5. Lastly, it would appear that the largest reductions in cost may occur from a reimbursement system that seeks to achieve effective use of health facilities, not just hospitals.[19]

Experimentation with new payment systems therefore should also be concerned with the establishment of various types of comprehensive prepaid plans that include incentives to either the purchasers or providers to minimize the total cost of care for given levels of quality.

[19] For an excellent review of the literature in this area, see Herbert E. Klarman, "Effect of Prepaid Group Practice on Hospital Use," *Public Health Reports*, November 1963.

THE USE OF AN ECONOMETRIC MODEL FOR HEALTH SECTOR PLANNING

BY MARTIN S. FELDSTEIN *

The growing policy interest in the health care sector reflects both an inherent concern with the Nation's health and an awareness that the price mechanism is an inadequate regulator of resource allocation in this area. A prerequisite of appropriate Government policy is an understanding of the aggregate behavioral relations of producers and consumers whose decisions determine the effects of Government policy on the supply and use of health care services. The purpose of this paper is to show how this information could be estimated and analyzed in the framework of an econometric model of the health care sector.

A fully developed model would show how hospitals and other institutions, doctors and paramedical personnel, governments at all levels, insurance agencies, and patients interact to determine, the pattern of services that are provided; the resources used in their production; the patients who consume these services; the payments received by institutions and individuals; and the prices paid by consumers: It would moreover be dynamic, indicating the reaction lags and time paths along which variables respond to each other.

The estimation of a full-scale dynamic model of the health care sector presents a number of difficult economic and statistical questions. Moreover, the currently available data are not adequate to provide all the information necessary for policy purposes. These problems are not insuperable; substantial progress should be possible during the next few years. The current paper presents a preliminary prototype model. It is small and its dynamic properties are severely limited. But it does suggest a core for a larger model and demonstrates the properties and policy uses of an econometric model of the health care sector.

CONDITIONAL PREDICTION PLANNING

Federal Government policy in the health care field [1] currently concentrates on subsidizing the construction of hospital facilities and, through the Medicare and Medicaid programs, providing health in-

* Department of Economics, Harvard University. This paper draws heavily on my "An Aggregate Planning Model of the Health Care Sector," *Medical Care*, 1967. For an example of an econometric model of the British health care sector, see ch. 9 of M. S. Feldstein, *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service*, Amsterdam: North-Holland Publishing Co., 1967.

NOTE.—The presentation in this paper recognizes that the reader may not be familiar with the current use of econometric models in macroeconomic forecasting and policy. The concepts and methods are therefore explained as they are required: for further discussion, readers may consult the excellent expository paper by Suits (1962) as well as fuller treatments in Marschak (1953), Tinbergen (1952), Theil (1964), Goldberger (1959), and Goldberger (1964).

[1] Although the analysis in this paper is in terms of Federal health policy, the econometric model approach would be equally useful at the State and metropolitan level.

surance for the aged and those in low-income groups. Future developments may involve the Government more directly in the supply of nursing home facilities, the training of medical personnel, and the care of individual patients. Through *each* of these activites, the Government influences, both directly and indirectly, *all* aspects of the provision and use of health care services. For example, providing additional support for hospital building in one State would not only influence its pattern of hospital admissions but would also affect the building of private nursing homes, the attraction of doctors to the State, the wages of nurses and paramedical personnel, et cetera. All of these effects are relevant to determining the optimal level of such support for hospital construction.

More generally, associated with each possible Government health sector policy are: (1) a set of *available* facilities—both those directly influenced by the Government (for example, hospital beds) and those indirectly influenced (for example, private nursing homes); (2) the *costs* incurred by the Government and by others; (3) the pattern of *utilization* of facilities (for example, hospital admission and duration of stay by diagnosis); and (4) the ultimate effect of this care on the *health* of the Nation. In principle, the Government should select that policy which maximizes a "welfare function" (in which the variables are measures of the Nation's health, the costs incurred by the Government, and other costs) subject to the constraints imposed by the behavioral and technological relations between Government policy variables, total availabilities, costs, utilization, and health. In practice, this approach to health sector policymaking is far from attainable. The behavioral relations linking Government policy to the overall availability and use of services are almost completely unexplored. Technological relations between the use of health services and the resulting improvements in community health are known only for a quite limited range of activities. Because of the extreme difficulties in estimating these technological relations, a less demanding approach to health sector policy must be sought.

It is nevertheless important to preserve, as much as possible, the idea of choosing among policies in terms of their *effects*. If the ultimate impact which a Government action will have on the Nation's health cannot be assessed, it is at least possible to use estimated behavioral relations to predict its overall effects on the pattern of availability and use of health care services. The making of such conditional predictions would be the primary use of an econometric model of the health care sector. The importance of using a multiequation model is that this method can capture the complicated interactions and feedbacks of the health care system. In this way, the model permits calculating the indirect as well as the direct effects of Government policy. This will be elaborated below.

Monitoring and Explanatory Information

Although such conditional prediction planning requires using the entire model, each individual equation can by itself provide information which could aid policymakers. It is helpful to distinguish two types of information, which may be called monitoring information and explanatory information.[2]

[2] Both structural and reduced form equations are relevant here. The distinctions between these two forms of equations and between the types of information they contain will be made in sec. 2.

Monitoring information permits assessing individual aspects of the current operation of the health care system. More specifically, it answers the question: How do differences in variable x affect some other variable(s) in the health care system? An example will clarify this. It is known that areas differ in the number of hospital beds available per thousand population. Anyone responsible for health sector policy should know what effects this has on the types of cases treated, the mean duration of stay per case, et cetera. Because clinicians serve in a single area, they will not be aware of these differences between areas in the patterns of admission and treatment. Similarly, a crude statistical comparison of admission rates or mean durations of stay in different areas would not distinguish the effects of bed availability differences from other factors which vary among areas. In contrast, the equations of an econometric model can conveniently indicate how the health care system responds to differences in bed availability, demographic characteristics, income, et cetera. The current model indicates that interstate differences in the number of short term beds per capita have substantial effects on the number of cases admitted to hospital and the average duration of stay per case. A 10-percent difference in the number of beds available per capita causes a 6.4-percent difference in the admission rate and a 2.3-percent difference in mean stay.

Explanatory information relates to specific suspected problems and to the causal relations which must be understood as a prerequisite of a general appraisal of health sector operations. It answers questions of the form: What are the reasons for the differences between areas and through time in variable y? Do these imply any malfunctioning of the health care sector? For example, officials might start with the observation that hospital admission rates differed substantially between States and then ask whether this was due to differences in population age-sex structure and medical characteristics or whether it also reflected differences in income, insurance coverage, hospital availability, et cetera. The current model suggests that the primary factors influencing admission rates are the availability of hospital beds, the proportion of urban residents and married females in the population and, to a smaller extent, the proportion of low-income families in the population.

The Nature of Structural and Reduced Form Equations

Before looking at the specification and estimates of the current model, it will be useful to review the general concepts and terminology of complete-system econometric models.[3] An econometric model is set of interdependent statistically estimated equations. The model *as a whole* explains the values taken by one set of variables (the *endogenous* variables) in terms of the values taken by the remaining variables (the *predetermined* variables). The number of equations in the model is equal to the number of endogenous varaibles. Some equations may be merely definitional; that is, true as an identity, and therefore need not be statistically estimated. Each nondefinitional equation describes some behavioral or technological relationship. The dependent ("left-hand side") variable of such an equation is one of the endogenous var-

[3] See note at beginning of this study for references to further discussion.

iables; some of the explanatory ("right-hand side") variables may also be endogenous. The estimated coefficients of the explanatory variables in a particular equation indicate the *direct effects* of each such variable on the dependent variable of that equation. The set of behavioral, technological, and definitional equations together constitute the *structural form* of the model.

If the equations are all linear, it is easy to solve the set of equations for the endogenous variables; that is, to express each endogenous variable as a linear function of all of the predetermined variables. This new set of equations is known as the *reduced form* of the model. Each coefficient in a reduced form equation indicates the total effect that predetermined variable has on an endogenous variable. An example will clarify this very important distinction between the direct effect coefficients of the structural form equations and the total effect coefficients of the reduced form equation. Assume that the structural equation with the number of private nursing home admissions as dependent variable has among its explanatory variables the number of beds in Federal Government hospitals (a predetermined variable) and the number of beds in private hospitals (an endogenous variable). Both variables should have negative coefficients, indicating that an increase in the availability of either type of hospital bed would decrease the number of persons admitted to nursing homes. Now, consider the effects of an increase in the number of Federal hospital beds. The direct effect is to decrease the number of nursing home admissions by the amount indicated by the structural equation coefficient. But the increased number of Federal hospital beds would decrease the building of private hospitals; fewer private hospital beds would increase the demand for nursing home admissions. The *total* effect on nursing home admissions of the change in the number of Federal hospital beds, reflecting both its direct effect and its indirect effects such as that through the availability of private hospital beds, would be indicated by the coefficient of the Federal hospital bed variable in the reduced form equation for nursing home admissions.

Because policy depends on *total* effects and not just direct effects, the reduced form equations are of great importance. The virtue of a multi-equation model of the health care system is that it can capture the interdependencies and feedbacks in the system's behavior and transmit this information to the reduced form equations.

The Specification

The endogenous variables of the model are: number of persons with health insurance (INS);[4] number of general practitioners (GP); number of medical specialists (SPEC); number of available short-term general hospital beds (BA); number of admissions (ADM); and mean duration of stay per case (MS). All variables except the last are expressed as rates per thousand population. The predetermined variables are of two types. The first consists of past (1950) values of the endogenous variables.[5] The second group is the exogenous variables, that is, those which are "causally prior" to the variables de-

[4] The variables are represented by a symbol of one to four letters that can act as an obvious mnemonic device. To reduce the amount of notation, this symbol will signify both the name of the variable and the logarithm of its value.

[5] The lagged value of the proportional utilization of hospital capacity, $P = (ADM) \cdot (MS) / 365\ (BA)$, is included among the lagged endogenous variables.

termined within the health sector; population age structure as measured by the percentage of persons aged 65 and above (AGE); income distribution as measured by the percentage of families with incomes below $2,000 (INC); urbanization, the percentage of persons living in communities designated as urban by the census (URB); and the percentage of married females in the population (MAR).

The selection of variables and the specification of the individual equations is based on previous work by others,[6] on the author's study of the British health care sector [Feldstein (1967), chs. 7 and 9], and on experiments with the current data. But it would be wrong to claim that the specification finally selected is a "correct" picture of the health sector or even a reasonably accurate one. The model is too small; too many variables are omitted; the measurement of the exogenous variables is too crude. It is proffered only as a rough first approximation.

The variables included in each equation will now be discussed.[7] The current (1960) value of the proportion of persons with health insurance (INS) was posited to depend on the population's age and income distributions, urbanization, and the previous (1950) proportion of persons with health insurance (INS_{-10}). Including the variable INS_{-10} implies that changes in the exogenous variables do not have their full direct effect on INS immediately, but rather that the current value of INS reflects a gradual adjustment from previous values. This is a very plausible assumption for a variable which measures a type of behavior (being insured) which once acquired is very likely to be maintained.[8] A similar dynamic specification is also used for the numbers of general practitioners (GP) and specialists (SPEC) and for the availability of hospital beds (BA). The number of general practitioners in the State was related in the model to insurance, income, and the previous number of general practitioners. The number of specialists was assumed to depend directly on insurance, bed availability, age, income, and the previous number of specialists. The bed availability equation included insurance, the number of general practitioners and specialists, income, urbanization, the previous stock of beds, and the previous utilization of available beds (P_{-10}). The two hospitals use

---

[6] A summary of research published before 1965 is given by Klarman (1965). Other useful sources include: Anderson (1956, 1963), Auster (1966), Axelrod (1964), McNerney (1962), Roemer (1959), Rosenthal (1964, 1965), Somers (1961), Stageman (1962), Weisbrod (1961) and U.S. Department of Health, Education and Welfare (1962, 1964).

[7] All equations in the model were assumed to be linear in the logarithms of the variables. This has two advantages. First, the coefficients are estimates of constant elasticity responses. Second, certain identities linking the variables are linear in logarithms (for example, $P = MS - BA$ — log. 365).

[8] This lagged response process can be easily formalized. Using each variable symbol (for example, INS) to stand for the logarithm of the value of that variable, the structural insurance equation may be written:

(1) $$INS = b_0 + b_1 AGE + b_2 INC + b_3 URB + b_4 INS_{-10}$$

If there were no delay in the response of INS to changes in the three exogenous variables, the equation of the (fully adjusted) insurance variable (INS*) could be written:

(2) $$INS^* = \beta_0 + \beta_1 AGE + \beta_2 INC + \beta_3 URB$$

where β_1, β_2 and β_3 are the "long run" elasticities. Instead of assuming an immediate response, we posit

(3) $$INS - INS_{-10} = \lambda\ (INS^* - INS_{-10})\ :$$

that is, the decade change in (the logarithm of) the insurance variable is proportional to the difference between the new, fully adjusted variable (INS*) and the past actual value of (INS_{-10}). The value of the response elasticity ($0 < \lambda < 1$) measures the reaction speed. Substituting equation (2) into equation (3) yields:

(4) $$INS = \lambda\beta_0 + \lambda\beta_1 AGE + \lambda\beta_2 INC + \lambda\beta_3 URB + (1 - \lambda)\ INS_{-10}.$$

This provides an interpretation of the coefficiencies of equation (1). The value of b equals $1 - \lambda$ or, $\lambda = 1 b_4$. The higher the value of b_4, the slower the response of INS to changes in the other variables. Moreover, the values of b_1, b_2, and b_3 are the first-decade ("short-run" or "impact") elasticities; to obtain the corresponding long-run elasticities (β's). each b value must be divided by $1 - b_4$.

variables (ADM and MS) were not related to their own lagged values on the assumption that, since these did not represent stocks, they would not be influenced by their own distant past. Admissions were assumed to depend upon bed availability, age, urbanization and the number of married women. The mean duration of stay per case was assumed to be influenced by insurance and the number of general practitioners, as well as by urbanization and the number of married women.

TABLE 1.—*Health sector model: Estimated structural form parameters**

Dependent variable	INS	GP	SPEC	BA	ADM	MS	AGE	INC	URB	MAR	P_{-10}	INS_{-10}	GP_{-10}	$SPEC_{-10}$	BA_{-10}	Constant	R^{-2}
INS	−1						0.322	−0.209	−0.224			0.283				5.715	0.711
							(.090)	(.093)	(.090)			(.048)					
GP	.258	−1						.225					0.972			−1.882	.776
	(.128)							(.089)					(.102)				
SPEC	−.288		−1	−0.157			.159	−.327						0.712		2.934	.905
	(.159)			(.113)			(.118)	(.097)						(.052)			
BA	.249	.127	−.187	−1				.081	−.112		.159				0.616	−1.397	.742
	(.128)	(.115)	(.091)					(.102)	(.099)		(.210)				(.084)		
ADM				.641	−1		−.053		−.118	.561						2.249	.639
				(.085)			(.073)		(.040)	(.337)							
MS	.186	.026		.229		−1			.190	−1.559						5.955	.600
	(.109)	(.116)		(.132)					(.079)	(.440)							

*Standard errors shown in parentheses.

NOTE: −1 indicates the dependent variable of the equation.

KEY FOR TABLE 1

Each row of this table corresponds to 1 equation of the structural model. The −1 in the row indicates the dependent variable of the equation; this is also indicated by the symbol in the column headed "Dependent variable." The other numbers not in parentheses are the coefficients of the explanatory variables indicated at the tops of the corresponding columns. The numbers in parentheses are the estimated standard errors of the coefficients. The column headed R^{-2} gives the square of the adjusted multiple correlation coefficient.

The variables are defined as follows (see text for fuller definitions): INS=insured population; GP=general practitioners; SPEC=specialists; BA=hospital bed availability; ADM=admission rate; MS=mean stay; AGE=persons 65+; INC=families with income below $2,000; URB=persons in urban communities; MAR=married females; P=percentage utilization of available beds. Whenever appropriate, variables are expressed as rates per population. All unsubscripted variables refer to 1960. The subscript $_{-10}$ indicates that the variable relates to 1950.

The coefficients of the three exogenous variables in the INS equation indicate that the proportion of the population with health insurance increases with the proportion over age 65 and decreases with the proportion of low-income and urban families. The values of the coefficients are estimates of the impact elasticities, i.e., effects within 10 years of a change in the exogenous variable. The low value of the coefficient of INS_{-10} indicates a relatively high speed of response of INS to changes in the other explanatory variables; approximately 70 percent of the ultimate response of INS to a change in an explanatory variable occurs in the first decade. More specifically, the response speed may be measured by 1 minus the coefficient of the lagged dependent variable: $1-0.283=0.717$. This implies that the longrun elasticities are only $(1/0.717)=1.39$ times the size of the impact (first decade) elasticities.

It is interesting to compare the response speeds of general practitioners, specialists, and hospital beds. All three respond much more slowly than INS to changes in the determining factors. As might be expected, specialists respond more quickly than general practitioners and the availability of beds more quickly than either type of doctor. Also interesting is the relation between income and the availability of general practitioners and specialists. The signs of the INC coefficients in the GP and SPEC equation indicate that general practitioners become relatively more abundant than (are "substituted for") specialists in States in which low-income families are more common.[9]

The bed availability (BA) equation indicates that the current supply of beds in a State reflects differences in the supply of doctors in the area, the proportion of the population insured and urbanization as well as the inherited stock of beds. The effects of interstate differences in bed availability on bed use is the most important aspect of the ADM and MS equations. The "direct" elasticity of the admission rate with respect to bed availability (0.641) is more than twice the mean stay elasticity (0.229), indicating that in areas of relative bed scarcity the number of cases treated is reduced substantially more than the average stay per case.[10]

The Reduced Form Equations

Table 2 presents the coefficients of the reduced form equations which express each endogenous variable as a linear function of the predetermined variables. As explained above, these coefficients measure the total effects (both direct and indirect) of the predetermined variables and thus complement the monitoring and explanatory information of the structural equations. Although the reduced form coefficients differ from those of the structural equations (except the coefficients of the INS equation and of certain lagged dependent variables), the signs and magnitudes are approximately the same. Of course, some predetermined variables have indirect but not direct effects on given endogenous variables; in such cases, the structural coefficient is zero and the reduced form coefficient nonzero.

[9] This statement reflects not only the coefficient values but also the fact that there are on average nearly twice as many general practitioners as specialists.

[10] Note that the sum of these elasticities (0.870) is the elasticity of the number of beds used with respect to the number of beds available. Thus, percentage utilization of capacity falls, but only slightly, as availability increases. For a fuller discussion of the implications of the effort of bed availability on use, see Feldstein, 1964, 1965, 1967b.

TABLE 2.—*Health sector model: Reduced form parameters*

Dependent variable	Predetermined variable									
	AGE	INC	URB	MAR	P_{-10}	INS_{-10}	GP_{-10}	$SPEC_{-10}$	BA_{-10}	Constant
INS	0.322	−0.209	−0.224	.....	.....	0.283	.....	.....	.....	5.715
GP	.083	.171	−.058	.....	.....	.073	0.972	.....	.....	−1.001
SPEC	.054	−.283	.095	.....	−0.026	−.097	−.020	0.734	−0.100	1.792
BA	.081	.104	−.193	.....	.164	.098	.127	−.137	.635	.186
ADM	−.001	.066	−.242	0.561	.105	.063	.082	−.088	.407	2.427
MS	.081	−.107	.103	−1.559	.038	.077	.054	−.031	.145	7.253

KEY FOR TABLE 2

Each row of this table corresponds to 1 equation of the reduced form model. The symbol in the 1st column identifies the dependent variable of the equation. The numbers in each row are the coefficients of the predetermined variables indicated at the tops of the corresponding columns.

The variables are defined as follows (see text for fuller definitions): INS=insured population; GP=general practitioners; SPEC=specialists; BA=hospital bed availability; ADM=admission rate; MS=mean stay, AGE=persons 65+; INC=families with income below $2,000; URB=persons in urban communities; MAR=married females; P=percentage utilization of available beds. Whenever appropriate, variables are expressed as rates per population. All unsubscripted variables refer to 1960. The subscript $_{-10}$ indicates that the variable relates to 1950.

This estimated reduced form can be used to predict the values which the endogenous variables would take in each State in 1970 if the structural parameters of the health care sector remained unchanged. A prerequisite of such a prediction is a forecast of the 1970 values of the exogenous variables: AGE, INC, URB, and MAR. The lagged endogenous variables (e.g., INS_{-10}) are merely the 1960 values and are therefore already known. Substituting the exogenous and lagged endogenous variables into each equation yields estimates of (the logarithms of) the 1970 variables. The process can be repeated to predict the endogenous variables for 1980 by using the 1970 predictions as forecasts of the lagged endogenous variables.

The structural and reduced form equations together can also be used to predict the effects of a Federal Government policy which causes a once-for-all change in the value of one of the endogenous variables. To be more specific, assume that in 1960 the Government causes an increase in the proportion of the population with health insurance in one State. This also affects the values of the other endogenous variables in 1960. To calculate these effects, the Government policy can be represented as an increase in the constant term of the INS structural equation and the modified system solved for the reduced form equations. The new reduced form coefficients, which would be the same as those of table 2 except for the constant terms, permit calculating the revised values of all of the endogenous variables. Of course, a once-for-all Government policy in 1960 will continue to have effects in the future. If the revised 1960 endogenous values are used as the lagged variables for predicting the 1970 values, a revised set of 1970 predictions will be obtained. The Government can therefore associate with any single current action a time path of effects on all variables.

A different type of conditional prediction problem is raised by a Government policy that changes the structural relations of the health care sector. For example, if the supply of all hospital beds were made an instrument of Government policy, the BA structural equation would become irrelevant and incorrect. Nevertheless, the other structural equations (estimated with BA treated as endogenous, for the period before the policy change) could still provide the basis for con-

ditional predictions of the effects of different Government bed-supply policies. These five remaining structural equations could be solved for the corresponding endogenous variables as functions of the old predetermined variables and the now exogenous BA. The five reduced form equations could be used to make conditional predictions of the effects in 1960 and future years of different supply policies, i.e., different time paths of BA.

Table 3 presents the coefficients of such reduced form equations. Because the new bed-supply policy changes the structure of the health sector (and not merely the value of one variable), the reduced form coefficients of the predetermined variables are also changed. For example, in the original structure a higher value of URB caused both a higher MS and a lower BA. The lower BA in turn caused a lower MS. Thus, the total effect of URB on MS (0.10) was less than its direct effect (0.19). The substitution of a Government bed-supply policy stops the indirect effect of URB on MS through BA and thus causes the new total effect (0.15) to be higher than the old.

TABLE 3.—*Health sector model: Policy reduced form*

Dependent variable	Predetermined variable								
	BA	AGE	INC	URB	MAR	INS_{-10}	GP_{-10}	$SPEC_{-10}$	Constant
INS	.....	0.322	−0.209	−0.224	.....	0.283	.....	.....	5.715
GP	.....	.083	.171	−.058	.....	.073	0.972	.....	−1,001
SPEC	−0.157	.066	−.267	.065	.....	−.082	.....	0.071	1.591
ADM	.641	−0.53	.....	−.118	0.561	.....	.....	.....	2.306
MS	.229	.062	−.034	.147	−1.559	.055	.025	.....	7.585

KEY FOR TABLE 3

Each row of this table corresponds to 1 equation of the reduced form model. The symbol in the 1st column identifies the dependent variable of the equation. The numbers in each row are the coefficients of the predetermined variables indicated at the tops of the corresponding columns.

The variables are defined as follows (see text for fuller definitions): INS=insured population; GP=general practitioners; SPEC=specialists; BA=hospital bed availability; ADM=admission rate; MS=mean stay; AGE=persons 65+; INC=families with income below $2,000; URB=persons in urban communities; MAR=married females; P=percentage utilization of available beds. Whenever appropriate, variables are expressed as rates per population. All unsubscripted variables refer to 1960. The subscript $_{-10}$ indicates that the variable relates to 1950.

CONCLUDING REMARKS

During the past 20 years, economywide econometric models have been developing as important tools of research and policy. Complete models should be able to play a similar role in individual sectors of the economy where public policy influences, but does not fully control, the sector's behavior. The current paper has indicated the ways in which an econometric model of the health care sector could be used to make conditional predictions of the effects of alternative policies and to provide monitoring and explanatory information.

The prototype model presented here is obviously too crude to be of practical use. A more detailed specification must be developed on the basis of careful studies of individual behavioral relations. Additional variables should be included to extend the model to other institutions and personnel, to out of hospital care, and to costs and prices. Currently included variables such as hospital admission and duration of stay should be disaggregated. The measurement of variables such as income distribution or age composition should be improved. Time-series information for each variable and area should be used to estimate

the time patterns of lagged responses and the characteristics of individual areas.

Although the State and standard metropolitan area data on utilization, expenditures and prices is still fragmentary, the growing recognition of the urgent need for information to use in health sector planning should soon lead to a correction of this deficiency. When such data becomes available, an econometric health sector model could provide a useful framework for conditional prediction planning.

REFERENCES

Anderson, O.W. and J. J. Feldman (1956). *Family Medical Costs and Voluntary Health Insurance:* A Nationwide Survey. New York: McGraw-Hill Book Co.

Anderson, O. W., P. Colette and J. J. Feldman (1963). *Changes in Family Medical Care Expenditures and Voluntary Health Insurance: A 5-Year Resurvey.* Cambridge: Harvard University Press.

Auster, R., I. Leveson and D. Saracheck (1966). "The Production of Health. An Exploratory Study," National Bureau of Economic Research, mimeographed.

Axelrod, S. J., ed., (1964). *The Economics of Health and Medical Care.* Ann Arbor, Michigan: University of Michigan.

Feldstein, M. S., (1964). "Hospital Planning and the Demand for Care," *Bulletin of the Oxford University Institute of Economics and Statistics,* 26, pp.. 361–68.

Feldstein, M. S., (1965). "Hospital Bed Scarcity: An Analysis of the Effects of Inter-Regional Differences," *Economica,* pp. 393–409.

Feldstein, M. S., (1967a). *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service.* Amsterdam: North-Holland Publishing Co.

Feldstein, M. S., (1967b). "An Aggregate Planning Model of the Health Care Sector," *Medical Care.*

Goldberger, A. S., (1959). *Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model.* Amsterdam: North-Holland Publishing Co.

Goldberger, A. S., (1964). *Econometric Theory.* New York: John Wiley & Sons.

Klarman, H. E., (1965). *The Economics of Health.* New York: Columbia University Press.

Marschak, J., (1953). "Economic Measurement for Policy and Prediction," in Wm. C. Hood and T. C. Koopmans, eds., *Studies in Econometric Method.* New York: John Wiley & Sons.

McNerney, W. J. *et. al.,* (1962). *Hospital and Medical Economics,* (2 vols.). Chicago: Hospital Research and Education Trust.

Roemer, M. I. and M. Shain, (1959). *Hospital Utilization Under Insurance.* Chicago: American Hospital Association.

Rosenthal, G. D., (1964). *The Demand for General Hospital Facilities.* Chicago: American Hospital Association.

Somers, H. M. and A. R. Somers, (1961). *Doctors, Patients and Health Insurance.* Washington, D.C.: The Brookings Institution.

Stageman, A. and A. M. Barrey, (1962). *Hospital Utilization Studies: Selected References Annotated.* Washington, D.C.: U.S. Government Printing Office.

Suits, D. B. (1962). "Forecasting and Analysis with an Econometric Model," *American Economic Review,* 52, pp. 104–32.

Theil, H. (1964). *Optimal Decision Rules for Government and Industry.* Amsterdam: North-Holland Publishing Co.

Tinbergen, J. (1952). *On the Theory of Economic Policy.* Amsterdam: North-Holland Publishing Co.

U.S. Department of Health, Education, and Welfare, (1962). *Research in Hospital Use: Progress and Problems.* Washington: U.S. Government Printing Office.

U.S. Department of Health, Education and Welfare, (1964). *Conference on Research in Hospital Use.* Washington: U.S. Government Printing Office.

Part V

Housing and the Quality of Man's Environment

HOUSING AND NATIONAL URBAN GOALS: OLD POLICIES AND NEW REALITIES

BY BERNARD J. FRIEDEN*

INTRODUCTION

Federal housing programs in the last 30 years have helped millions of American families to improve their living conditions. National urban goals today, however, encompass many issues other than housing standards. Programs designed to achieve one set of goals may block progress toward other equally important goals. National housing policy needs to be evaluated in the broader context of national urban goals, for it is becoming clear that past solutions to housing problems are intensifying other urban problems. Housing goals remain important: several million families are still living in slum conditions that should no longer be tolerated in the United States. But new housing strategies are needed that will recognize more complex goals: freedom of residential choice, the elimination of racial disparities, and a chance for all families to have access to good schools, adequate public services, and job opportunities.

This paper reviews the changes that are taking place in American housing, the conflicts that are developing between housing strategies and other goals, and some possible directions for new housing policies to follow. It is an abridged version of an article that appears in *The Metropolitan Enigma*, edited by James Q. Wilson and published in 1967 by the Chamber of Commerce of the United States.

Although "the housing problem" has taken on a wide variety of meanings over time, it has persisted as a major policy issue in America's cities. In its most basic form, the problem is one of assuring an adequate minimum living environment for people too poor to pay for decent housing at market prices. Other groups than the poor have also advanced claims to a housing problem, however. Considerable Government assistance has in fact been extended to middle-income people whose problem is not that of reaching minimum living standards but rather of obtaining a wider range of choice in the housing market or meeting the terms of homeownership. Current controversies over housing policy center on still unmet needs of low-income groups and of Negroes and other minorities who are victims of discrimination in the housing market. The issues at stake have important implications for urban renewal and the future of central cities, metropolitan development, civil rights, and the war on poverty. Serious conflicts over neighborhood change, open occupancy and fair housing laws, and

*Department of City and Regional Planning, Massachusetts Institute of Technology.

This paper was presented to the Task Force On Economic Growth and Opportunity, United States Chamber of Commerce, by the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University.

Federal urban programs arise directly from the shortage of decent low-cost housing in metropolitan areas.

This paper will focus on housing problems confronting the poor and minority groups—the nature and dimensions of these problems, the progress that has been made in dealing with them, current housing mechanisms and their future prospects, and strategies that may bring faster results. Current concern over low-income housing may at first seem exaggerated, in view of the substantial improvement in American housing since World War II, but there are disturbing signs that the problem is becoming more difficult to cope with, that progress to date has exposed a hard core of remaining issues that call for new approaches. In addition, housing policies that have worked in the past are now coming into conflict with other national objectives. Thus a review of past experience and a redefinition of goals and strategies are in order.

Effects of Slum Housing

To earlier generations of urban reformers, the elimination of slum housing was an essential step to remedy a host of social evils. Slums were thought to cause family breakdown, personal disorganization, crime, disease, and poverty. Contemporary social scientists are skeptical of such claims, and current research findings do not bear out simple causal relationships between physical conditions and social behavior. Experience with public housing also provides grounds for skepticism; families who move from rundown tenements into sanitary apartment projects do not necessarily change their way of life. In rejecting earlier simplistic views, however, there is a danger of falling into the equally questionable conclusion that housing conditions have no impact on social life. It is clear that housing interacts with other elements of the social and physical environment in influencing patterns of human behavior.

Current research suggests that crowding, inadequate plumbing and ventilation, and insufficient wiring and lighting clearly contribute to the spread of respiratory and digestive diseases and to home accidents.[1] Morbidity and mortality rates are sensitive to housing conditions; studies indicate that improved housing reduces the incidence of illness and death. Evidence on the psychological effects of crowding and the extent of stress resulting from dilapidated and poorly maintained housing is less clear cut. Lack of space and of privacy appear to interfere with children's study and with parental control. There is evidence that housing is an important factor in self-perception. Houses that convey an obvious image of neglect and inferior status have been found to influence the self-evaluation and motivation of people who live there, leading to feelings of pessimism and passivity.

This quick sketch of the research evidence on effects of substandard housing is subject to two important qualifications. First, the amount of research focusing directly on housing conditions has been very limited. Much of what is known is deduced indirectly or as a byproduct of studies focused on different subjects. It is possible that stronger

[1] For a recent review of evidence concerning the effects of substandard housing, see Alvin L. Schorr, *Slums and Social Insecurity* (Washington: U.S. Department of Health, Education, and Welfare, 1963). The following discussion is based on Schorr's findings.

connections do exist between housing and behavior but that research to date has not been adequate to identify them. Second, such factors as morale, self-perception, and motivation appear to be related to a broader environment than the housing unit alone. Neighborhood surroundings, both the social and the physical environment, are extremely important—possibly more so than the living quarters. In recognition of the importance of the neighborhood, efforts to improve slum conditions have given increasing attention to providing better community facilities—schools, parks, clinics—and to improving public services. Although this paper is concerned primarily with housing, it should be understood that improvements in housing must be accompanied by related community development if they are to have significant impact on the lives of people who are now in inadequate housing.

One unmistakable aspect of the low-income housing problem is a simple matter of economics. Families living in poverty are often compelled to spend a substantial share of their income—one-third or more—for housing, thus cutting heavily into their budget for other necessities, such as food or health care. Other dissatisfactions resulting from inadequate heat and maintenance, rats and insects, and the many hazards and discomforts of slum housing have been compelling enough to lead to political action and direct pressure for improvement. Rent strikes, demonstrations, and protest marches focused specifically on housing conditions have attracted national attention. A recent poll of Harlem residents put the need for better housing near the top of the list of "worst problems" (second only to dope addiction).[2] The Office of Economic Opportunity has found poor housing so frequent a concern in poverty areas that it has modified its antipoverty programs to give more attention to housing. Thus the people who live in slums have made it clear that they are not there by choice and that they want help in improving their living conditions.

The shortage of sound housing affects the people who live in slums most directly, but its indirect effects have also reached the public at large. Governmental activities that demolish housing—urban renewal, highway construction, other public works—have met with increasing resistance because of the difficulty of relocating displaced families. Needed facilities have been blocked or delayed as a result of opposition from people who cannot find other decent housing. These protests provide still further evidence of dissatisfaction with limitations of the housing market, as well as an additional reason for public concern with housing.

Dimensions of the Problem

Defining and measuring housing quality is a problem in itself. Notions of what constitutes a minimum level of decent living conditions change over time and depend upon some degree of social consensus that varies from place to place. Studies of housing in American cities have made use of fairly consistent standards since World War II, although society at large has undoubtedly upgraded its concept of adequate housing since that time. The application of these standards thus tends to yield, by current standards, a rockbottom estimate of the extent of substandard housing, but it does make possible the measure-

[2] "Negroes in Poll Ask More Police," *New York Times*, Sept. 4, 1966, p. 1.

ment of change in the supply and distribution of housing from a single set of benchmarks.

Four aspects of housing enter into this set of standards: structural condition, presence or absence of plumbing facilities, crowding, and cost in relation to the income of the occupant. Surveys by the U.S. Census Bureau provide the basic data on all these subjects. National housing censuses were taken in 1950 and 1960 and national housing inventories, based on a more limited sample, were conducted in 1956 and 1959. Substandard structural condition is reported as "dilapidated" housing, which according to Census definition "does not provide safe and adequate shelter and in its present condition endangers the health, safety, or well-being of the occupants." To be classified as dilapidated, a house must have one or more "critical defects"—holes over large areas of the foundation, walls, roof, floors, or chimney; substantial sagging of floors, walls, or roof; external damage by storm, fire, or flood—or a combination of lesser defects sufficient to require considerable repair or rebuilding; or must be of inadequate original construction, such as shacks with makeshift walls or roofs or dirt floors. The absence of a private toilet, bath or shower, and hot running water is reported separately. In interpretations of the Census data housing is generally classified as substandard if it is dilapidated or lacks one or more of these plumbing facilities.[3]

The other measures concern crowding and cost. The usual measure of overcrowding is an occupancy ratio of more than one person per room. The point at which the cost of housing becomes an unacceptable burden depends upon a judgment of the proportion of a family's income that can be allocated to housing without creating other deprivation. Welfare administrators and others concerned with low-income family budgets make widespread use of a norm of 20 percent of income as the maximum that can be spent for housing without cutting into expenditures on other necessities.

Condition, crowding, and cost are all related as measures of housing welfare and guides to public policy. Discussions of housing are often distorted when one or more of these factors are neglected. Thus slum clearance may be advocated as a way of eliminating the problem of substandard housing; instead, it may replace the slum problem with an overcrowding problem. Or strict enforcement of local housing codes may reduce the number of substandard and crowded housing units, while raising the price of housing so that high costs create fresh problems.

Taken together, information on substandard conditions, crowding, and cost provides a picture of current housing deficienies in the United States. According to the 1960 Census of Housing, 8.5 million families (including one-person households) were living in substandard housing—in dilapidated conditions or lacking one or more plumbing facilities. This number constituted 16 percent of all families in the United States. At the same time, 6.1 million families—12 percent of the total—were overcrowded; we do not know how many were both overcrowded and in substandard housing.

---

[3] For a fuller discussion of Housing Census definitions and their interpretation, see Leonore R. Siegelman, "A Technical Note on Housing Census Comparability, 1950–60," *Journal of the American Institute of Planners*, XXIX (February 1963), 48–54.

Substandard housing is clearly a problem of low-income families. Households with annual (1959) incomes of $4,000 or less constituted 38 percent of all households in 1960, but this low-income group accounted for three-fourths of the total families in substandard units and for 42 percent of all families in overcrowded housing. Nor is this prevalence of inadequate housing among the poor a result of insufficient spending for housing. Among families who earned less than $4,000 and lived in rental dwellings, more than 80 percent were paying 20 percent of their income or more for rent. Nearly half (46 percent) paid 35 percent or more of their income for rent, clearly an excessive amount to spend. It is sometimes alleged that the majority of people who live in slums are there because they place a low value on housing and do not spend much for it, and that many slum dwellers can actually afford to live in better housing. The information at hand indicates, to the contrary, that low-income people strain their budgets to pay higher rents than they can afford, but that there is not enough decent cheap housing available to meet their needs.

Contrary to popular impression, there is more substandard housing in rural than urban areas. About 40 percent of the families in substandard housing in 1960 were living within metropolitan areas. Overcrowding is more an urban than a rural problem, however: almost 60 percent of the overcrowded families lived in metropolitan areas.

These figures add up to a very conservative estimate of inadequacies in American housing. They exclude a large amount of deficient housing reported in the census as "deteriorating"; that is, having complete plumbing facilities but needing more structural repairs than would be provided in the course of normal maintenance. In 1960, 4.1 million families were living in housing of this type, 2.5 million of them inside metropolitan areas. Many other families live in housing that lacks central heating; still others live in housing that is sound but located in substandard surroundings, near noxious industry or truck terminals for example. Several million families live in quarters that are not dilapidated by census standards but that are in violation of the higher standards set by local housing codes. In sections of Philadelphia, the number of units in violation of the building code has been found to exceed census counts of deficient housing by as much as 2 to 1.[4] Thus it has been estimated that at least one-fourth of the American people are living in inadequate housing or an inadequate environment.[5]

Recent Trends

Housing conditions in the United States improved substantially in the 1950's. The improvement was not only in percentage terms, but also involved a reduction in the absolute number of families living in substandard and overcrowded housing.

Since the extent of improvement in American housing is still debated and not yet widely acknowledged, it is useful to review the evidence gathered by the Census Bureau in 1950, 1959, and 1960. Because of certain changes in census definitions, technical difficulties in making census comparisons have opened the way for misinterpretations that

[4] William Grigsby, *Housing Markets and Public Policy* (Philadelphia: University of Pennsylvania Press, 1963), p. 253, footnote 2.
[5] *Ibid.*, p. 253.

tend to obscure the substantial changes that have occured. I shall cite data from the housing census of 1950 and from two separate surveys at the end of the 1950's: a 1959 census survey on a sampling basis (covering approximately 180,000 units) which used exactly the same definitions as the 1950 census; and the 1960 housing census, which counted for the first time a number of single-room housing units and which changed certain definitions of housing condition. Data for 1960 have been adjusted to approximate the same categories of housing condition used in 1950 and 1959. The three surveys show declines in both substandard and overcrowded (1.01 or more persons per room) housing: [6]

[In millions]

	1950	1959	1960
Families in substandard housing	14.8	9.1	8.4
Families in overcrowded units	6.6	6.0	6.1

(A further postenumeration survey differed only slightly from the 1960 census results, estimating 8.5 million occupied substandard units. A later restudy of the 1960 data undertaken by the Census Bureau found that the national trends in substandard housing from 1950 to 1960 were measured accurately, despite widespread errors in block statistics: measurement errors at the block level tended to cancel out at the tract and city levels, and errors in structural condition classification were corrected by plumbing facilities data.[7]) The various sources of data are not far out of line with one another. They all indicate that the number of families living in substandard housing was reduced by about 40 percent in the 1950's, and the number of overcrowded families was reduced by 8 or 9 percent.

Comparable national surveys of housing conditions have not been undertaken since 1960, but FHA housing market analyses of some 65 cities in 1965–66 indicated continuing reductions in substandard housing.[8] On the basis of preliminary data, the Census Bureau has estimated that the national total of families in substandard housing dropped from 8.5 million in 1960 to 5.7 million by 1966.[9]

The experience of the 1950's shows considerable progress in improving housing welfare in the United States, but the performance of the housing market was less than satisfactory in some important respects. In urban areas, the improvement that did take place was not sufficient to reduce the number of Negro families in substandard housing, and in the country at large an increasing number of families at all economic levels were forced to spend more than a reasonable share of their income for rent in order to better their housing conditions.

---

[6] For the method used in making census comparisons, see Leonore R. Siegelman, *op. cit.* The relevant sources of data for 1950, 1959, and 1960 are: *U.S. Census of Housing: 1950,* vol. I, *General Characteristics,* pt. 1, U.S. summary, tables, 7, 11; *U.S. Census of Housing: 1960,* vol. IV, *Components of Inventory Change,* final report HC(4), pt. 1A–1, 1950–59 components, United States and regions, table 1; *U.S. Census of Housing: 1960,* vol. I, *States and Small Areas,* U.S. summary, final report HC(1)–1, table 9.

[7] U.S. Bureau of the Census, *Measuring the Quality of Housing: An Appraisal of Census Statistics and Methods,* Working Paper No. 25 (Washington, 1967).

[8] U.S. Congress, Senate Subcommittee on Executive Reorganization of the Committee on Government Operations, *Hearings, Federal Role in Urban Affairs,* 89th Cong., 2d sess., 1966, pt. 1, pp. 148–149.

[9] U.S. Bureau of the Census, *Current Population Reports,* Series P–23, No. 24, "Social and Economic Conditions of Negroes in the United States" (October 1967), p. 55.

Problems With the Turnover Process

In devising strategies for meeting the national housing goal expressed in the Housing Act of 1949—a decent home and a suitable living environment for every American family—one approach would be to accelerate the forces that operated successfully in the 1950's. Measures can be taken to stimulate a higher volume of new construction, more rapid turnover of existing units, and greater investment in maintenance and renovation of low-cost housing. But exclusive reliance on this turnover process creates other problems and leads to some bewildering dilemmas for public policy.

The pattern of metropolitan growth in the United States has been generally outward from the core of the central city, with new housing added in a series of rings wrapping around the central city and encompassing more and more suburban territory. Each ring of development tends to contain housing built at about the same time. Thus the old housing that middle-income families will abandon when they move to new quarters is generally concentrated in the inner part of the metropolitan area. Some years ago, these areas of declining middle-class occupancy constituted only a small part of the central city just beyond the downtown business district. By the 1950's, with more housing considered obsolescent, and with a larger number of middle-income people able to move to newer housing, the area of declining middle-class occupancy blanketed very large parts of the central cities and even a number of the oldest suburbs just beyond the city boundary. With low-income people relying almost entirely on older housing abandoned by more affluent groups, it was inevitable that the economic and racial composition of the central cities and some adjoining suburbs would undergo drastic changes.

As a result, the same process that brought about a striking improvement in national housing conditions yielded a plentiful harvest of acute social and economic problems in the central cities. The combination of new suburban development and housing turnover in the older cities led to a massive dispersal of the population along racial and economic lines. By the end of the decade, the suburbs had succeeded in attracting a concentration of white, young, middle- and upper-income families with children; while the central cities were left with higher proportions of the elderly, broken families, Negroes, low-income workers, and the unemployed. These social and economic disparities between central cities and suburbs do not hold true throughout the country, but they do apply to virtually all the large metropolitan areas and to urban areas of all sizes in the Northeast.[10]

The consequences have been severe, both for the people involved and for city governments. One result has been a growing social and cultural isolation of the poor from the rest of society. Workers in the central cities are increasingly cut off from blue-collar jobs in expanding suburban industrial centers, while manufacturing and related lines of employment continue to decline in the central cities. Transportation is difficult to the new industrial parks, and central-city workers are not sufficiently in touch with suburban firms to learn about job opportunities.

---

[10] For documentation and analysis of central city-suburban differences, see U.S. Advisory Commission on Intergovernmental Relations, *Metropolitan Social and Economic Disparities: Implications for Intergovernmental Relations in Central Cities and Suburbs* (Washington, D.C.: ACIR, 1965).

For Negroes, these problems are intensified. Even many of those who can afford to move to new housing in suburbia are blocked by discrimination or hostility, and are forced to stay in central-city ghettoes. The combination of economic and direct racial segregation produces widespread de facto segregation in housing and in the public services that are provided on a neighborhood basis: schools, recreation facilities, clinics.

The link between residential patterns and public services creates one of the most fundamental problems posed by current urban development: the gap between local needs and local resources. Disadvantaged groups in the central cities are highly dependent upon public services. They need a variety of educational programs for both children and adults, they have special needs for health and welfare assistance, recreation facilities, and police and fire protection. Central cities face high service demands, but many of their prosperous taxpayers have left and their commerce and industry are also moving to the suburbs. The cities have been struggling to cope with this gap between service needs and local tax resources. One strategy has been to press for greater State and Federal assistance; the recent proliferation of Federal-aid programs for the cities is a direct consequence. Another typical city strategy has been to develop urban renewal programs intended to win back (or hold onto) middle-income families, retail stores, and industrial firms. Despite the increase in Federal aid and the more limited effects of renewal programs, public services in the cities have had to be held far below the levels that are needed to enable disadvantaged groups to compete on equal terms with citizens of more prosperous communities.

In this sense, it is fair to say that our solution to the national housing problem has been creating a national urban problem. Federal housing policies seem to be at odds with other Federal objectives. Some programs—chiefly FHA mortgage insurance, Federal aid for highways that facilitate suburban growth, and aid for suburban water and sewer systems—accelerate the pace of new housing construction in the suburbs and the turnover of central-city housing. Others—principally urban renewal and aid for mass transit—aim at reviving the central cities and stemming the flight to the suburbs. Urban renewal in turn depletes the supply of low-cost housing and thus slows the turnover effects stimulated by national housing policy. The programs that favor suburban growth promote population dispersal along racial and economic lines, setting the stage for race segregation between central cities and suburbs, which runs counter to national civil rights objectives. This same dispersal handicaps the central cities in their efforts to provide adequate services for low-income groups, countering other national goals in the war on poverty and in providing equality of educational opportunity.

Continued dependence upon the turnover process to accommodate low-income groups is also producing open social and racial conflict in many central-city neighborhoods. In the 1950's, the growing Negro population of the central cities was able to take over a great deal of housing left behind by mobile white families who chose to move to suburbs. Even in the 1950's, however, this turnover process failed to reduce the number of urban Negro families living in substandard and overcrowded housing. The continued growth of Negro population in the cities means that many Negro families have nowhere to go but into

whatever vacancies turn up in white areas. But as the more mobile white families have departed, a core of deeply rooted people have remained in old neighborhoods—people who cannot afford to move to the suburbs or who want to stay where they are. Many established ethnic neighborhoods remain in the central cities, where people are tied to friends, family, churches, clubs, and the other loyalties that develop in a close-knit community. Often they are fearful of invasions by Negroes or other newcomers. Some neighborhoods have tried to stabilize the situation by supporting urban renewal programs designed to remove pockets of low-cost housing and price out potential invaders.[11] Others express their fears in a white backlash reaction, with open hostility to Negro newcomers and the threat of violence just below the surface.

The time has come to diversify our mechanisms for meeting the housing needs of low-income groups. The turnover system is working, but it is producing too many objectionable byproducts. These byproducts in turn have generated new policies that will interfere with the future turnover of old housing. In particular, many central cities have found their changing population composition unacceptable and have developed urban renewal programs to stabilize existing neighborhoods or to attract back middle-income families— both of which will sacrifice low-income housing needs in order to diversify the city's population.[12]

Still another reason for supplementing the turnover process with other approaches lies in the nature of national housing goals. The goal of a decent home and a suitable environment for every family is too limited for the needs of urban life today. In our spreading metropolitan areas, mobility is a prerequisite for equal opportunity. Families need to be able to move to keep up with changing work locations, to have access to specialized services or institutions, and to have some measure of freedom in choosing a desired type of housing or type of community. Most middle-income families now enjoy this freedom of movement through the operation of the housing market, though many middle-income Negroes still do not. Most low-income families have very limited freedom of movement: the turnover process restricts their choice to areas where old housing is concentrated. A more adequate statement of national housing goals would go beyond decent shelter and surroundings, and would include diversity of choice in housing and freedom of movement throughout metropolitan areas.[13] New strategies are needed to achieve this more complex goal.

Diversifying Housing Strategies

Widening the choices available in the urban housing market does not necessarily mean rejecting the turnover process. The continued use of old housing that is in sound condition is surely desirable in itself; further, it does provide for one significant choice by enabling many people with low and moderate incomes to live in the central cities and older suburbs. Although some families now in the cities are living

[11] "Citizen Participation in Urban Renewal," *Columbia Law Review*, LXVI (March 1966), 594.
[12] See Bernard J. Frieden, "Toward Equality of Urban Opportunity," *Journal of the American Institute of Planners*, XXXI (November 1965), 320–330.
[13] See *ibid.*

there reluctantly, the majority are probably there by choice. Middle- and upper-income groups can generally afford to move elsewhere if they want to. (At all income levels, however, Negroes have little freedom of movement.) Poorer families are more constrained by the shortage of suburban housing at prices they can afford, but many would be reluctant to leave in any case. Recent research and relocation experience in low-income neighborhoods demonstrate clearly that a high proportion of the people living there want to stay nearby in order to keep up their ties with friends, relatives, storekeepers, churches, and with a style of life that can hardly be found in the newer suburbs.[14] The growth of urban renewal and highway construction is forcing many of them to leave their old neighborhoods and to look for other housing in the city. Others, however, are potentially mobile and want to leave the city. If they have to stay where they are, they may be unable to take advantage of suburban job possibilities or to achieve their aspirations for a more middle-class style of life for themselves and their children.

Widening the housing choices available, therefore, means maintaining the turnover process and the choices it allows, but developing additional strategies that will open other opportunities. Three general approaches are worth exploring: (1) reducing the cost of new construction through improvements in building technology, so that more people can enter the market for new housing; (2) extending the possibilities for housing turnover, to facilitate wider choices of location, cost, and ownership as well as rental; (3) direct subsidies for new housing, managed so as to counter imbalances resulting from the turnover system.

1. REDUCING THE COST OF NEW HOUSING

In the past, efforts to reduce costs and widen the market for new homes have focused on financial mechanisms to assure a flow of investment funds into home mortgages and to finance some housing at below-market interest rates. More attention should be given to fostering cost reductions in construction, through the introduction of new technology, simplification of building controls, and the removal of obstacles to large-scale production and land development. The cost of new housing today prices most American families out of the market. By the end of 1965, the median sales price of new single-family houses had risen to $20,000. Sixty-five percent of all new homes sold for $17,500 or more, and only 9 percent for under $12,500.[15] Ownership of new homes is effectively limited to the top quarter of American families, who earn $8,000 a year or more.

The homebuilding industry, despite its significant role in the national economy, has lagged far behind other industries in sponsoring research and development, in applying technology, and in exploiting the benefits of mass production and large-scale operations. Several factors contribute to this lag: the structure of the industry, with a predominance of small firms; consumer reluctance to accept innovation; resistance to change from manufacturers of building components who

[14] See Marc Fried, "Grieving for a Lost Home," in *The Urban Condition*, ed. Leonard J. Duhl (New York: Basic Books, 1963), pp. 151–171; Herbert J. Gans, *The Urban Villagers* (New York: Free Press of Glencoe, 1962); Chester Hartman, "The Housing of Relocated Families," *Journal of the American Institute of Planners*, XXX (November 1964), 266–286.

[15] U.S. Department of Housing and Urban Development, *Housing Statistics*, XIX (May 1966), table A-22.

fear new materials; and resistance from labor unions that fear a loss of jobs. Other obstacles are more likely to respond to public policy, and to pave the way for changes within the industry itself. Builders and parts manufacturers have little incentive to develop a technology suited to mass production as long as the housing market is fragmented by separate building codes and land development controls in every local community. Some 5,000 separate jurisdictions in the United States have building codes. Many of these localities have adopted one of several national model codes or a State or county code, but typically with enough local variation so that there is little overall uniformity. The manufacturer of a new building component or system will have to design his product to meet the most stringent local requirements in whatever market he intends to enter, or he will have to produce costly variations to meet each local regulation. Code variations thus add to manufacturing cost and inhibit mass production.

Various Federal measures that have been proposed range from modest research programs in support of private efforts to more active use of Federal resources to create new markets and stimulate changes in local building controls. Virtually all the proposals presuppose the establishment of a clear and consistent Federal policy to promote technological advances in the housing industry as a means of widening the market for new housing. As a first step, current Federal activities in housing and construction should be alined in support of such a policy. Federal agencies responsible for direct construction of housing, for example, should develop a uniform set of performance standards for their own projects.

Congress has begun to define Federal policy in this area by establishing a Federal study of housing and building codes, zoning, tax policies, and development standards. The Housing and Urban Development Act of 1965 directs the Secretary of Housing and Urban Development to study "State and local urban and suburban housing and building laws, standards, codes, and regulations and their impact on housing and building costs, how they can be simplified, improved, and enforced, at the local level, and what methods might be adopted to promote more uniform building codes and the acceptance of technical innovations including new building practices and materials."

To make a significant impact on the housing industry, the Federal Government will clearly have to go far beyond putting its own house in order. Studies, code development, and experimentation will have to be supplemented by the use of Federal incentives that can offset some of the deep-seated resistance to change within the housing industry and within local government. The creation of new markets with Federal assistance may be able to stimulate substantial private investments in new housing systems and components. Creating markets beyond Government-built housing, however, does mean achieving greater uniformity in building and land development regulations, and here the obstacles are considerable. Many local building codes contain provisions that are not designed to protect the public as a consumer of housing, but rather to protect manufacturers of particular components or craftsmen with particular skills.[16] Further, local building codes and land development controls often reflect deliberate policies to limit the

[16] U.S. Advisory Commission on Intergovernmental Relations, *Building Codes*, *op. cit.*, p. 7; Martin Meyerson, Barbara Terrett, and William L. C. Wheaton, *Housing, People, and Cities* (New York: McGraw-Hill, 1962), pp. 130–131.

volume of new housing by raising its cost. This situation is particularly true in the suburbs of metropolitan areas, and it seems to be more widespread in zoning and subdivision controls than in building codes. For the creation of a mass market, greater consistency is needed in all these development controls.

Local resistance to new development is a growing threat to national goals that stress support of urban development and a high volume of housing construction. Much of this resistance is a result of local tax considerations. New low- and moderate-cost housing brings with it demands for new local services—particularly schools—that typically cost more than the amount of local property tax produced by the new housing. For suburban governments coping with tax increases to pay for new schools and other service increases occasioned by population growth, preferred development strategies favor some mixture of luxury housing, "clean" industry, and shopping centers—all of which tend to yield more in local taxes than they cost the local community to service. "Fiscal zoning"—requiring large minimum lot sizes for new housing in order to discourage low-cost subdivisions—is one means to this end. Excessive subdivision requirements and building code provisions also add to the cost of whatever housing may be built, thus slowing the rate of new development and improving the prospects for high assessed valuations. In addition to the tax motives, local policies to discourage development often contain a strong measure of determination to exclude unwanted people—principally those of low status or low income.[17]

To overcome these complex sources of resistance, several measures are needed that have little to do with building code reform directly, but represent other facets of Federal urban policy. One important strategy consists of easing local reliance on the property tax to pay for service costs; the growth of Federal and State aid to local communities has worked in this direction and will undoubtedly continue. Another approach involves metropolitanwide consideration of development controls, such as review of local regulations by metropolitan planning councils. A combination of local interest and State and Federal policy has been instrumental in establishing metropolitan planning agencies in most metropolitan areas of the country. Federal policy in this area can be extended to strengthen the role of metropolitan planning agencies in working toward more uniform development controls and high levels of new construction.[18]

2. Improving the Turnover Process

The turnover process itself would produce fewer objectionable by-products if it enabled more people with limited incomes to move into a wider variety of existing housing, including older housing in the suburbs. With about a fifth of American families moving every year, a large supply of existing housing becomes available for occupancy. The problem is to secure a larger share of this housing for people whose choices are now very limited, principally Negroes and low-income families.

---

[17] See Bernard J. Frieden, "Toward Equality of Urban Opportunity," *op. cit.*, pp. 323–324.

[18] See Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, *The Effectiveness of Metropolitan Planning*, prepared in cooperation with the Subcommittee on Intergovernmental Relations, Committee on Government Operations, U.S. Senate (Washington, D.C.: Government Printing Office, 1964), especially pp. 99–123.

Eliminating racial discrimination in the sale and rental of housing would contribute immeasurably toward improving living conditions for Negroes as well as countering the patterns of segregation that now result from the turnover process. Because Negroes are not free to move throughout urban areas, they are unable to take full advantage of vacancies that would meet their pressing housing needs. Racial segregation between suburbs and central cities has attracted a great deal of attention, but Negroes are also segregated *within* central cities. the sound housing that becomes available in white city neighborhoods often makes no contribution to solving the housing problems of Negroes who live in the ghettoes. Further, even Negroes who can afford used suburban housing are kept out of the suburbs by discriminatory practices.[19]

Both governmental and private activities have developed to combat housing discrimination. Government approaches have involved Federal action to prohibit discrimination in the sale or rental of federally aided housing, and a number of State and local laws to prohibit discrimination in private housing. A dozen States now have laws covering some categories of private housing; several others have laws covering publicly assisted housing.[20] The coverage of these laws and Federal policies together is far from complete.

More can be done to secure better enforcement of existing laws, however, and to combine governmental and private efforts to combat discrimination. Private groups have played a major role in stimulating Government action, and they have also worked directly at the local level. Citizen groups and private developers have sponsored a number of housing developments planned for racially mixed occupancy from the outset. Neighborhood associations have worked to prevent panic selling and to stabilize areas of racial transition. More than a thousand local fair housing organizations have been established to find homes for Negro families in white neighborhoods on an individual basis. Though these groups have focused mainly on helping middle-income Negroes, there is growing evidence that a broadening of their strategy to include families with somewhat lower incomes might yield more impressive results.[21] In addition, governmental support for these private efforts—such as staff assistance—should be considered. Measures to stimulate greater cooperation between public and private agencies can be particularly important in making existing laws more effective. Private organizations have been able to test whether real estate brokers comply with State laws against discrimination and to furnish public agencies with the evidence they need to enforce the laws. Government measures can strengthen the work of private groups by providing a needed context of public policy and leadership, while private groups can often operate best in sensitive local situations.

---

[19] For an analysis of potential Negro housing demand at middle-income levels, see U.S. Housing and Home Finance Agency, *Potential Housing Demands of Non-White Population in Sclected Metropolitan Areas* (Washington, D.C.: HHFA Office of the Administrator, 1962).

[20] American Jewish Congress, "Summary of 1962 and 1963 State Anti-Discrimination Laws" (New York, 1964).

[21] See George B. Nesbitt and Elfriede Hoeber, "The Fair Housing Committee: Its Need for a New Perspective," *Land Economics*, XLI (May 1965), 97–110; and George and Eunice Grier, *Equality and Beyond: Housing Segregation and the Goals of the Great Society* (Chicago: Quadrangle Books, 1966), pp. 68–82.

3. Subsidizing New Housing

In addition to accelerating the turnover of existing housing, another strategy for widening low-income housing choices calls for subsidizing new construction. Federal aid for new low-income housing began in the 1930's, but 30 years of effort have produced a total of only 600,000 units. This output must be compared with the figure of 8.5 million families who remained in substandard housing in 1960. Current Federal programs produce only a trickle of new low-income housing. The major subsidy programs—low-rent public housing, moderate-income projects (under section 221(d)(3)), and housing for the elderly—together are providing about 50,000 new units a year. The rent supplement program will reduce rents for low-income families living in new moderate-income developments, most of which are counted in the above total. Through January 1967, contracts had been allocated for 18,000 families to receive rent supplements.

The small size of these subsidy programs means that they can make only a limited contribution toward providing the millions of units that are needed. Further, they operate primarily in the central cities and older suburbs. Thus they tend to reinforce the turnover process by providing improved housing for the poor, but mainly in communities where the poor are already concentrated. Strategies to promote the use of these programs in the newer suburbs would be similar to those already discussed in connection with subsidizing existing housing: offering additional incentives, establishing State or metropolitan housing agencies that can cross local boundaries, or conditioning other community development aids on local participation in low-income housing programs.

The strategies discussed here all involve changes in public policy and increased attention to the Nation's housing needs. Enlarging the scale of present operations, mounting new programs, and accelerating technological change in the housing industry will be impossible without increased political support for such measures. Though the politics of housing policy are beyond the scope of this paper, it is evident that the full implications of present policies are not widely appreciated by public officials or by the American people. Much work needs to be done to clarify the nature of present policies in order to lay the groundwork for change. The issue of housing subsidies provides a simple illustration. Many middle-income voters oppose low-income housing aid because they consider it a unique and unwarranted subidy for the poor. Few people take special note of the enormous Federal housing subsidies given to middle-income groups by means of income tax deductions on mortgage interest payments. On balance, this country supplies much greater housing subsidies for middle-income people than for the poor, but in ways that seldom strike the public consciousness.[22]

A major task for political leadership is to make known the connections between present housing policies and national social and economic goals. The country pays a high price for maintaining present housing mechanisms, in terms of social tensions resulting from segregation, denial of equal opportunities, perpetuation of poverty, and

[22] Alvin L. Schorr, "National Community and Housing Policy," *Social Service Review*, XXXIX (December 1965), 433-443.

indefinite delay in achieving the simple goals of the Housing Act of 1949. The fact that housing policy intersects with other areas of social and economic policy is a source of confusion and conflict, but it can also be a source of strength. It means that a series of national goals are functionally interrelated, even though we have usually considered them separately. Increasingly, policies that are now separate will have to be brought into line if progress is to continue in any of these areas. But as harmonious policies are developed, public action can yield multiple benefits in achieving a wide range of national purposes.

JOBS AND HOUSING: A PROGRAM

BY FREDERIC MEYERS*

I start this short paper with the premise that our urban society is indeed in crisis, and that since ours is essentially an urban society, our total society is in crisis. Surely we can tolerate few more summers like the last several—they are symptoms of advanced social disintegration. I take it further as a premise that the fundamental value of a society is self-preservation. Hence, we seek perhaps the least expensive way of achieving a fixed goal; but cost, in this as in other wars of self-preservation, is not a critical decision criterion.

The spreading disintegration of urban society has its source in the urban slum where abominably housed people are huddled together sharing experiences of unemployment rates that would be intolerable if national in scale, and poverty and alienation that half-way measures bolster by promises unfulfilled for all but a miniscule number.

Among the source evils are exclusion of slumdwellers from opportunity to participate in the same world of work and income which is the center of the lives of most of us, and their exclusion from conditions and locus of abode which might permit them to share the values that television so insistently tells them are primary.

If jobs and housing are source evils, then let us attack the problems of jobs and housing. They are interrelated, since restriction to limited geographic areas for housing has produced a significant barrier to access to jobs. The loci of employment are often extraordinarily inconvenient or inaccessible to the slums. The problems then resolve themselves into finding jobs and finding housing convenient to jobs of a kind appropriate and acceptable in the majority view to people who hold jobs and are integrated into rather than excluded from the society and its values.

We have approached the problem of jobs, apart from general full employment policy, essentially through training. We have held to the shibboleth that people should only get jobs after they can show qualification for them. But the uncertainties of subsequent employment have reduced the willingness of the disadvantaged to participate, probably lowered the real effectiveness of the training programs (and perhaps increased their cost).

The problem in part is, I think, the accumulating disbelief of ghetto young people that implicit promises will be fulfilled; that participation in one of the many training programs will really lead to a job at wages customary for the employment, and that it will open a way out of the life of discrimination, of scrambling and hustling and repeated disillusion and bitterness. And even if training should lead to job offers, housing segregation may reduce their value so sharply that rationality suggests they be declined.

*University of California, Los Angeles.

If these general statements are descriptive, the solutions lie in adequate and believable assurances of jobs, and of housing in quality and location commensurate with the job and the income it should produce.

I suggest that the only way to make this promise believable is to reverse the usual approach. Instead of promising jobs after training, that we rather find assured jobs for people, place them in them at the usual wages, and *then* train them. It might be noted that the only major difference between such a program and history is in paying the rate for the job during the training period. Except for the professions, most skills, in our economy, were acquired on the job.

But the wage problem is critical, to give the program credibility and to begin, as rapidly as possible, the transformation in attitudes and values which is essential. And security during the training period and until the worker acquires the confidence that he can stand on his own in the labor market is equally essential.

To achieve these goals will take more than persuasion—mere persuasion has helped but little. And equal opportunities statutes do not provide equal opportunities to people who have been excluded from the job holding, employing society, and hence are not qualified.

The device with which I propose we experiment at least initially is in the letting of Federal contracts and the purchase of goods or services on Government account. These (and other Federal programs such as guaranteed loans of various kinds) have long required equal opportunity clauses, which have been constructive but patently insufficient—they also exclude the unqualified.

I propose that initially all tenders of contract and purchases of goods or services above a minimum dollar amount contain a requirement that an appropriate ratio of new hires both of the prime contractor or manufacturer or provider of services, and all subcontractors be selected not by the contractor but by an agency of the Federal Government.[1]

The choice of these new hires will be made not on the basis of present qualification, except where discrimination on the grounds of ethnic origin can be shown, in which case qualified workers would be hired.

However, most new hires chosen by the Federal agency would be trainees for entry jobs. Selection would be made principally for the purpose of reducing above average rates of unemployment among demographic groups experiencing persistently high unemployment—young people and especially Negro young people.

The employees thus chosen would be paid the rate for the job for which they are training, from the time of hire. Complete control of tenure would remain for, say 6 months, in the hands of the Federal agency, after which it would revert to the contractor who could dismiss for any good cause except incompetency. Good cause would be tested by the standards and procedures common to collective bargaining agreements, specially devised if no such agreement existed, or no clause adjudged to be adequate in its protection.

The contractor would be required to undertake a training program for these employees, with standards and procedures required to be

[1] This might later be extended as a condition of certain kinds of loans, e.g. small business. However, if this were done the subsidy element in the loan, in the usual form of lower than market interest rates would have to be increased in some fashion.

approved by the Government, and policed by the Government. The employee would be introduced into productive labor as rapidly as good training practice permitted. For persons with insufficient formal education to meet the usual qualifications for the job for which they were training, it would be required as part of the training program that they attend part-time education, on shift time where possible, without loss of pay. If education during working hours were not available, there should be release, with pay, for a number of hours equivalent to his classroom and travel time.

So far as competency is to be a condition of continued tenure, all final judgments should remain in the hands of the Federal agency for an extended period of time, say 4 years. The private employer could make recommendations, but decision would remain in governmental hands, and the criterion should be whether or not the employee is making that progress that should be expected of him given all the circumstances.

The cost of the program would, of course, be an element in the costing of the contract, and of subcontracts. For bidding purposes, all contractors might be required to make a set-aside of equal numbers of jobs, with a provision that the contract price be reduced by a specified amount for each job unclaimed for training for appropriate reason. This device could be used to equalize bidding conditions for contractors with access to persons eligible for the program and those not. Or it might be used as an indirect means of directing contracts to those contractors whose actual or potential locations are such as to employ large numbers of eligible trainees.

Housing, as we have already noted, is a major barrier to job access for qualified as well as for unqualified slumdwellers. Conventional public housing has not proved an adequate solution, and I have the impression that the commitment to find adequate alternative housing for persons dispossessed in the process of slum clearance (urban redevelopment) has more often than not been a mirage.

I propose another approach to the housing problem. Instead of constructing new public housing projects, the Government should acquire by purchase (exercising the power of eminent domain where necessary) conventional existing housing units, single or multiple family dwelling units as these may be customary in the neighborhood, spotted and dispersed in the areas of access to employment opportunities.[2] Where possible, single family units should be acquired as they become vacant. Multiple units should be acquired where desired, but units in them should become available only as existing tenants leave voluntarily. Units at all levels of price in which persons disadvantaged in the housing market could be expected to live should be purchased.

Placement officers in employment services, and the agency administering the employment and training program described above would be empowered to assign these housing units to persons placed, either through their agencies or otherwise, who were also, by reason of ethnic origin, discriminated against in the housing and job markets. Rentals would be at or near the market, and might include an option to buy not to be exercised until after a reasonable long minimum period, so as to assure that the unit would remain devoted to the public purpose to which it was intended.

[2] This should not seem unprecedented. My university a State university, recently acquired an existing apartment block, by threat of condemnation, for use of married students. The present purpose is at least as public.

No single or small set of ideas will provide a complete solution to the problem of the cities and of discrimination and the ghetto. Nor will any complex solution resolve the pressure immediately. But if, as I believe, jobs and housing are central to the problem, then I believe jobs and housing of the kind required can be provided with the devices I here propose. While they may seem to some, Draconian, the problems are critically dangerous. And no long-term impairment of our system of values or institutions is implied. Indeed, there are long-established precedents for both of the two measures. The power of eminent domain has been exercised for all kinds of objects certainly less critical than these and there really is little essential difference between condemning a piece of land on which to build public housing, and condemning a piece of land on which there is a house vacant or for sale. And since the Wagner Act and the Equal Employment Opportunities Act agencies of Government have stood ready to require an employer to hire someone he did not wish. Combine this with training subsidies which are an established instrument and the suggested program results.

Done on a sufficiently large scale, these programs, added to what we are now doing, could, I think, make a major contribution to the solution of the problems of poverty, segregation, discrimination, and social disintegration in the city.

THE QUALITY OF THE HUMAN ENVIRONMENT

BY ROGER REVELLE*

In thinking about human resources, we need to remember that human beings do not exist in a vacuum. They live in and are an integral part of a social and natural environment. The quality of the environment affects many aspects of human life and, in turn, is largely determined by human actions. I shall discuss two questions of environmental quality: the control of water, air, and land pollution, and the provision of facilities for outdoor recreation.

THE GROWTH OF POLLUTION

Pollution is the harmful alteration of our natural environment by our own actions. Pollutants are either unwanted byproducts of our activities, or the obnoxious residues of things we have made, used, and thrown away. Man has always produced pollutants, but until recently natural processes were sufficient to change most of them into harmless or beneficial substances. In the modern world, however, pollutants are produced in such large quantities that the capacity of air, water, and land to absorb them is often exceeded, and pollution results. Man is increasing in numbers, productivity, and technology, but the rest of nature remains fixed. As men clump themselves together in cities, the part of the environment that receives most of their pollutants is actually diminishing.

In 1890, some 22 million Americans were city dwellers. Today there are about 130 million, six times as many, and by the year 2000 the number will probably double again. With the growth of cities, we are placing ever heavier pollution pressure on a small fraction of our environment—about 2 percent of the area of our country. In the megalopolises of the future—the continuous strip cities like that which will soon extend from Boston to Richmond—the problems of waste removal and avoidance of pollution will be on the same scale as the problems of water, air, and food supply, and of human transportation.

In many ways, the quality of our environment has deteriorated with each new advance of the gross national product. Increases in electric power production means the burning of more coal and fuel oil, and hence the discharge of more sulfur dioxide into the air. The growth of the paper industry has brought a vast increase in trash. The production of new automobiles and the discard of old ones has resulted in unsightly piles of hulks. The growth of urban automobile transporta-

*Center for Population Studies, Harvard University.

NOTE. This essay is largely based on two previously published articles: "Pollution and Cities," in the *Metropolitan Enigma*, U.S. Chamber of Commerce, 1967: and "Recreation in a Hyperproductive Society," in *Daedalus*, American Academy of Arts and Sciences, pp. 1172–1191, 1967. Ideas contributed by John Calhoun, Dan Ogden, Ann Satterthwaite, Andrew Scheffey, Athelstan Spilhaus, Robert Teeters, Harold Thomas, and Nathaniel Wolman are gratefully acknowledged.

tion is choking both the mobility of our cities, and the lungs of the city dwellers.

Not only are we making and consuming more things, and consequently producing a larger quantity of residues, but our advancing technology is producing new kinds of poisons. President Johnson has said, "The uncontrolled waste products of our technology are menacing the world we live in, our enjoyment, and our health" [1]. The marvelous advance of American agriculture has depended largely on chemical fertilizers and pesticides, substances that were never before present in nature and against which our fellow creatures, the birds and fishes, have no defense mechanisms. Lead and other metallic additives in gasoline accumulate along our roadsides. New products of chemical industry poison our streams. Technological changes have reduced the kinds and amounts of materials that can be economically reused. Aluminum cans and plastic containers are worthless as salvage but virtually indestructible as litter.

At the same time that we are producing more and worse pollutants, our aspirations for a better environment are rising. We now see pollution where before we were able to ignore it. As the national income increases, people are less willing to trade off environmental deterioration for lower costs of goods and services. The "third parties" who do not benefit from these tradeoffs are more aware of their position and more impatient to see it remedied by the abatement or elimination of pollution.

In its report, "Restoring the Quality of Our Environment," the President's Science Advisory Committee has urged that—

> The public should come to recognize individual rights to quality of living as expressed by the absence of pollution, just as it has come to recognize rights to education, [and] to economic advances. * * * The responsibility of each polluter for all forms of damage caused by his pollution should be effectively recognized and generally accepted. There should be no "right" to pollute [2].

In dealing with pollution, we need to consider natural resources in two senses, both as raw materials or sources of energy, and as those parts of the environment that can be depleted or worsened by misuse. In our lifetimes and those of our children, the availability of energy and raw materials will not put any serious limits on American society, but we will be limited in human fulfillment—the quality of life will be lessened and dulled—by destructive changes in our environment. And future generations will suffer because of our prodigal methods of waste disposal. The phosphates we dump into streams and lakes are permanently lost to our farmlands; the rusting auto hulks in the countryside are a drain on our metal resources.

With growing pollution, our concepts of natural resources must broaden. As more and more cities are assaulted by increasingly severe air pollution, for example, we can no longer think of urban air as a "free good." Instead, it has become a natural resource—that is, a part of the environment for which the demand is liable to outrun the supply, and to which a cost can be attached. It is not a resource in the older sense of a raw material for manufacturing, but rather one of the "new" renewable resources—a limited part of the environment whose quality has great economic and social importance. As a common

property resource which must be used by all citizens, it requires public regulation or management. Preloff [3] has pointed out that—

> Invasion of the [common] domain [of city air] by polluters then may be regarded as trespass, and the cost of control assigned to the polluter.* * *

Costs and Benefits of Pollution Control

Like education and economic opportunity, controlling pollution is expensive; we need to remember that the quality of life depends only in part on the quality of the natural environment [4]. Economic abundance and civilized diversity are also ingredients. We must use our environment in a variety of ways, and limited environmental resources must be allocated optimally among different uses. Hence, the basic issue in considering problems of pollution is to define our goals [4]. In order to do this we need to find ways of balancing esthetic, recreational, and related aspects of the quality of human life against other values. Stated in economic terms, how much will people pay, and how much should they pay, for these intangibles [5]? How clean should this stream be, at what cost, for what purpose, and at what future time? It is meaningless simply to say that the stream should be "clean."

To reduce the costs and maximize the benefits of pollution control, we need both to consider our choices for different courses of action within the environment as a whole, and to take full advantage of its natural processes. For example, garbage disposal by incineration, land fill, or household grinding adds pollutants to the air, the soil, or the waters. The choice of which methods to use should depend on the character of the local environment and the total burden of pollutants we are adding to different parts of it. In windy areas that are steadily ventilated, a higher proportion of waste products can be disposed of to the air than in regions of light winds and frequent inversions. Different streams may be devoted to different purposes. Some can be kept clean and sparkling, others used to oxidize realtively large quantities of organic matter.

At present, we are unable to strike a meaningful balance between the costs and benefits of preventing pollution because we cannot measure adequately the psychic and physiological effects of a dirty environment. Consequently, our decisions are generally based on estimates of immediate monetary expenditures rather than on evaluations of the total benefits and costs of different courses of action. Yet even if all measures were taken to meet existing as well as future needs for control of air, water, and land pollution, the total expenditures per year on the average during the next 15 years would represent only a small fraction of the national income—less than 3 percent (table 1). As incomes continue to rise, it should be possible to increase the absolute amounts spent on waste disposal, even if the fraction of income for these purposes is not increased. With growing affluence, moreover, our people may be willing to allocate a larger fraction of their resources to maintain the quality of their environment, just as the fractions for education, health, and recreation are increasing.

TABLE 1.—*Estimated annual costs of pollution control in U.S. cities during the next 15 years*

[Annual costs in billions of dollars]

	Capital	Operating	Total
Air pollution:			
Automobile afterburners	$1.5	$1.0	$2.5
Sulphur dioxide removal from stack gases [1]	.3	1.0	1.3
Industrial control equipment [2]	.3	([3])	.3
Total	2.1	2.0	4.1
Water pollution:			
Reservoirs for seasonal equalization of river flows for waste oxidation	.4	----------	.4
Municipal sewage collection and treatment	.9	.6	1.5
Industrial effluent treatment	1.1	.5	1.6
Separation of combined sewers and storage of storm waters [4]	.5	----------	.5
Electric utility cooling towers [5]	.1	.6	.7
Total	3.0	1.7	4.7
Solid waste disposal:			
Collection of municipal wastes [6]	----------	2.6	2.6
Incineration of municipal wastes [6]	----------	.7	.7
Land fill of municipal wastes [6]	----------	.3	.3
Junk auto disposal [6]	----------	.2	.2
Demolition waste disposal [7]	----------	.9	.9
Total	----------	4.7	4.7
Total	5.1	8.4	13.5

[1] Assuming that all present and future fossil fuel generating plants will install equipment for removal of sulphur dioxide.
[2] Los Angeles County expenditures from text reference 6, projected to 1980 and multiplied by 20.
[3] Not available.
[4] Based on an estimated population of 40 million people served by combined sewers, and an assumed cost of $200 per person for sewer separation and/or storage of storm waters. See text reference 7.
[5] Assuming that all present and future electric utilities will construct cooling towers.
[6] Assuming that the amount of municipal wastes and junk autos will increase by 80 percent during the next 15 years.
[7] Assuming that amount of demolition wastes and unit costs of disposal are each half those of municipal wastes

NOTE.—Operating costs are averages over the next 15 years. Capital costs for water pollution control are annual investment costs for construction and equipment, assuming a discount rate of 4 percent and a 25-year useful life for the structures.

GOVERNMENTAL PROBLEMS IN ABATING POLLUTION

Pollution raises several problems of government in the United States.

Water, air, and land in part are a common resource, and in part are subject to individual possession or ownership. Traditionally in this country, individual rights to water and air have been tied to the ownership and use of land. For pollution, as for other aspects of land-use control, planning, zoning, and standards have become the most commonly employed tools to attain a reasonable compromise between private property rights and the public interest [8].

Under our system, the authority to implement these land-use controls is firmly lodged in local governments—counties and municipalities. This fragmentation of land-use controls is a hard fact of our political life, and it means that the Federal Government, the States, and regional authorities concerned with pollution must develop cooperative devices to deal with local governmental jurisdictions.

A second problem arises from attempts to assign the costs of control of pollution to the industries that produce it. A charge on polluters for use of the public air supply as an atmospheric sewer has been widely advocated, for example, but this must be national or at least

regional in scope, otherwise polluting industries may simply move from one State or municipality to another one which uses its carelessness about pollution as a recruiting inducement for industry.

Air and water pollutants do not recognize conventional political demarcations. They move across city and county lines and ignore State boundaries. In the eastern megalopolis, the prevailing winds blow contaminants from New Jersey's oil refineries, smoke-belching factories, and smoldering trash dumps right over New York City. To limit or prevent pollution, therefore, control measures are needed over regions determined by meteorological and hydrologic realities, rather than by legal artifacts. We need to treat the entire urban-suburban-rural complex that constitutes a modern metropolis as an entity. Existing levels of government need to cooperate in new ways, and it may be necessary to invent and test new types of governmental organizations.

Some precedents already exist for regional governmental organizations. Regional water management has a traditional standing in the United States—from the early, single-purpose sewage, irrigation, or drainage districts to such modern multipurpose organizations as the Ohio River Sanitary Commission (Orsanco) and the Delaware River Basin Commission (DRBC) [8]. In Germany, the Ruhr Valley Administration has shown that regional pollution control based on systems analysis and economic optimization can be carried very far in practice, through cooperative organizations, wise engineering, and ingenious use of workable but theoretically sound rules of thumb. Orsanco and DRBC have shown that impressive progress toward basinwide planning and management for abatement of pollution can be attained in the United States.

A regional agency needs a competent technical, planning, administrative, and legal staff. In addition, it must have adequate authority and resources to deal effectively with its responsibilities [8]. These include:

(1) Investigations of actual or potential pollution problems and development of effective plans. To do this, the agency needs to obtain and interpret physical, biological, and economic data to forecast the occurrence and extent of pollution, and to analyze the economic and other consequences of alternative courses of action.

(2) Control and regulation of pollutant emitters—for example, indsutrial discharges into a river system or combustion exhausts into the atmosphere.

(3) Taxation and assessment, both for control purposes, through charges on polluters, and for raising funds to carry out the purposes of the agency. The effluent charges levied in the Ruhr Valley are an example. Besides such revenues, the agency may also need to raise money by issuing bonds or by obtaining grants from other branches of government.

(4) Establishment and operation of treatment or disposal facilities.

(5) Cooperation with and advice to other governmental agencies, especially municipalities and counties, which have authority over land use.

A pollution control unit, to be successful, needs the support of a constituency that can help it establish and enforce sanctions against pollution, particularly voluntary sanctions based on publicity and public opinion. An effective constituency must be "grown" through a continuing program of education. The public needs to be supplied, in readily understandable form, accurate and objective information on problems and opportunities. A nongovernmental organization can be especially effective within "problem sheds" that include a number of political units. An example is the Water Resources Association of the Delaware River Basin, which works with industry and community advisory groups to support the work of the Delaware River Basin Commission [8].

Human Resources for Pollution Control

The problem of pollution has many aspects—public health, economic, sociological, political, and institutional. But in all aspects it is related to human beings. It would not exist except for man's activities. All living creatures are affected by pollution; only man has the ability to control or eliminate it. The sciences that deal with man, with his behavior, and his institutions are therefore central [2].

Many kinds of social scientists and social engineers, as well as chemists, biologists, statisticians, and sanitary engineers, must be involved in handling the pollution problems of the future. Economists are needed to determine the apparent and real costs of pollution and its abatement, and the economically "efficient" allocation of these costs among different elements of society. Urban and regional planners should keep the goal of minimal pollution constantly in mind. Specialists in public communication are needed to arouse public concern and impel reasoned action, to inform, yet not needlessly alarm, their fellow human beings. Political scientists, public administrators, and lawyers will have the task of devising new forms of governmental organization, intragovernmental cooperation, and public-private interaction. They must find both incentives for right action and legal sanctions against wrong actions—carrots and sticks—and ways to apply them.

The Role of the Federal Government in Pollution Control

Because the problems of air, water, and land pollution have little regard for political boundaries, are a national concern, require large expenditures for their solution, and affect the quantity and quality of our natural resources, the Federal Government must take a leading role in pollution abatement. Among the things it can and should do are:

(1) Avoid being a polluter in all operations it conducts, supports, or controls [2].

(2) Provide tax incentives for pollution abatement by industry—for example, allowances for rapid amortization of waste treatment equipment, and credits for sums spent on research or development [4].

(3) Establish environmental quality criteria and standards for maximum allowable amounts of pollutant discharges. Among other benefits, such standards may effectively create markets for advanced abatement equipment and processes [4].

(4) Carry out and support those kinds of pollution research that are not likely to be undertaken by private industry because they will not provide a competitive advantage to individual firms, including determining the effects of pollutants on human beings, animals, and plants, studying dispersion processes in the environment, and developing techniques for measurement and analysis [4].

(5) Underwrite industrial research and development on abatement devices that may later be widely sold. By establishing control regulations, the Government creates a market for the successful developer, and it should be able to recoup its share of the costs [4].

(6) Test and demonstrate advanced control methods in its own operations. An Army base could be used to test tertiary sewage treatment and complete recycling of water; the Tennessee Valley Authority might install a promising new device for sulfur dioxide removal in its coal-burning electric plants [2].

(7) Underwrite large-scale demonstrations of new control systems. Many developments in pollution control cannot be adequately tested or demonstrated on a small scale. It is not sufficient for a method to serve satisfactorily a single dwelling or neighborhood. New systems must be demonstrated in a complete community which has a complex mix of waste problems. The average city administrator feels that he cannot use the local taxpayers' money to take a chance, and consequently he is likely to recommend only economically proven systems. Hence demonstrations and full-scale trials of new systems need to be insured or partly paid for with Federal funds [2].

(8) Support the advanced training of technicians, engineers, economists, and scientists, for many kinds of tasks in pollution control. Federal fellowships, training grants, and institutional support are needed to insure that persons of ability and imagination are drawn into this broad field and trained in its intricacies [2].

(9) Provide meeting grounds for discussion and agreement on pollution problems and abatement measures among State, county, and city officials, industrial managers, and public representatives [8].

(10) Stimulate formation of special governmental or quasi-governmental agencies designed to handle metropolitan or regional waste disposal and pollution control on an integrated basis [2].

(11) Provide information and educational materials that can be used by all governmental and private agencies to create better public understanding of pollution problems [8].

In recent Federal legislation, the Congress has expresed its conviction that programs for pollution control should be progressive and continuously developing, and that they should be collaborative, engaging all levels of government and both public and private institutions. There is a strong reliance on both carrots and sticks: authorizations to establish Federal standards and appropriations for programs of enforcement; subsidies to State and local governments for construction of abatement facilities and provisions for the conduct or support for research, technical assistance, and training. Formal lines of re-

sponsibility from the Federal Government down through State government and local bodies of general jurisdiction are emphasized [8].

At the moment, Federal legislation may have outrun the development of administrative machinery for action and the knowledge on which action must be based. This legislation is directed almost entirely at the abatement of pollution after it has occurred. There has been little attention given to the prevention of pollution before it happens.

Some of the different possible modes of Federal action for pollution abatement are illustrated by the problems of junk automobiles. About 5 million motor vehicles, each containing on the average a ton of salvageable scrap steel, are discarded each year. Although nominally a source of valuable metal, junk cars are now surplus commodity. The cost of collecting, processing, and transporting auto hulks is more than the price of the No. 2 steel scrap into which they can be converted. Not only is this kind of scrap much less desirable than other grades, but changes in the steelmaking process have resulted in a smaller demand for scrap of any kind [2]. At the same time, the rapid increase in motor car production has been followed by a steady rise in the number of automobiles junked each year. Great piles of auto hulks now litter the country side and constitute a national eyesore.

To reduce the number of stored or abandoned automobile hulks, the Federal Government might support a price differential that would assure the recycling of hulks to the steel furnaces. In small or remote communities, this could include Government financing of a "cleanup train" [2], which would collect and compress junked automobiles and other outsized steel objects.

A tax could be devised to provide an incentive for quicker salvage of spare parts and disposal of the hulks. An annual Federal or State license might be imposed or a personal property tax placed on junk cars. Possibly a purchase excise tax could be devised which would provide a sinking fund for ultimate disposal of each automobile at the end of its useful life. An excise tax equal to 1 percent of the purchase price of a new car would amount to about $20 per ton of metal, and this would be ample to insure the reuse of automobiles as raw materials by the steel companies.

Federally conducted or sponsored research is needed on means for producing more uniform scrap from auto hulks, on methods of storage of hulks in excess of current market demands, and on uses other than for scrap steel. For example, it has been suggested that auto hulks might be combined with taconite ore, in order to produce a magnetic material of improved characteristics.

The Need for Outdoor Recreation

Do human beings need outdoor recreation? What values are there for modern man in the world of nature outside the cities? These questions cannot be answered scientifically, because of the irrational nature of play as it comes to us from our remote ancestors, and because man is the most adaptable of animals. If the rat and the sparrow can learn to live for endless generations in the cities, why not man? We must look not to the scientist for answers but to the prophets and the poets. Moses talked to God in the solitude of the mountains. Jesus found His mission in the wilderness and taught His disciples to look at the lilies of

the field. Guatama sought peace, and became the Buddha, under a spreading Peepul tree. St. Francis preached to the little birds, and "in spiritual ecstasy they flapped their wings and chirped rejoicing" [9]. Walt Whitman said:

> The passionate tenacity of hunters, woodsmen, early risers, cultivators of gardens and fields—all is a residence of the poetic in an outdoor people * * * Now I see the secret of the making of the best persons. It is to grow in the open air and sleep with the earth. I think all heroic deeds were conceived in the open air [10].

Until recently in the United States, recreation has generally been thought of as an adjunct to work, necessary only for physical and mental health, economic stimulation, or improved social behavior, a kind of patent medicine to cure juvenile delinquency and other mysterious social diseases.

These attitudes were understandable, and possibly desirable, when there was a continent to be won and a gross national product to be multiplied. But they are less justifiable, and maybe even nonsensical, in the new hyperproductive society. In a time of service industries, when most people make a living by selling things to one another or have jobs in which they are faceless components of superhuman organizations, recreation may represent the only chance many have to find themselves as unique individuals.

Under these circumstances, recreation becomes a human need and should be recognized as a human right, in the same sense that we have suggested the right to be free from pollution. To help the individual find uniqueness, his recreation may have to include challenges, demand skills, and provide a sense of risk and excitement, a feeling of adventure, a chance to gain individual excellence. These qualities are most easily attained when human beings are in contact with the natural world. But to insure a sufficient range of outdoor recreational opportunities, we need to consider the special requirements of different groups of human beings—children, teenagers, old people, students, young unmarried persons, married couples with and without children, various income groups, people at various levels of education, persons who work at night or in the daytime, manual workers, sedentary workers, and so forth.

In outdoor recreation, as elsewhere, there should be a happy combination of activities for both mind and body. One of the best features of 19th-century America was the tented chatauquas that sprang up across the country in the summertime. Centers like Aspen are the modern version. A hundred Aspens scattered throughout the country would provide places where people could talk to one another about exciting things in the morning and climb the local mountains in the afternoon.

National Parks and Other Federal Recreational Areas

In city parks and and in the suburbs, we ask for variety, novelty, and change. Just the opposite is true for the national parks. Although these treasures of the continent are in part sites for active recreation, skiing, mountain climbing, fishing, and camping, they are in essence great natural wonders, things of joy and beauty, places for an individual to

lose himself in contemplation. While their values cannot be enhanced by human action, they can easily be destroyed by it. They are the heritage of a people, for which each living generation is the trustee. Our aim must be to keep Yosemite close to what it was when John Muir first visited it, and Yellowstone as near as possible to the "national park and pleasuring ground" that Congress established in 1872.

National park attendance has exploded from the 3 million visits a year of 1930 to the 137 million of 1966. More people came to Yellowstone from June to September 1966, than visited the nearly 200 national parks, monuments, and historic areas 35 years ago. The heights of experience in the national parks are incompatible with crowds, and many of their wonders are too fragile to withstand much trampling. But within the next few decades, the number of Americans will increase by 30 to 40 percent, and average personal incomes will double, while the average working week will probably decline. Some people will work less hours per day, others fewer days per week, and almost all wage earners will have a paid vacation. From an average of 1 week per year, vacations may well lengthen to 4 weeks or more. The average person will travel about twice as many miles each year; perhaps half this travel, about 5,000 miles, will be in search of recreation.

These changes will combine to produce an enormous increase in the demand for outdoor recreation facilities at a distance from people's homes. The effects will probably be almost multiplicative: more people, times twice as much income per person, times twice as much travel, times fourfold lengthening of vacations. If suitable accommodations were available in national parks, forests, and wildlife refuges, the annual number of visitor-days could grow to more than half a billion by the end of the century. Excluding Alaska, the present area of Federal recreational preserves is somewhat over 40 million acres. If these preserves are not extended, there will be more than 12 visitor-day per acre in the year 2000.

At first thought, this would seem to be a low number. But the natural wonders of the parks occupy only a small fraction of the total park area. In Yosemite, the Mariposa Grove and the valley can be measured in tens of square miles, while the entire park covers 1,200 square miles. Yellowstone is more than 90 percent wilderness. Its sights can all be seen on a 150-mile drive (along which nearly a million automobiles now travel during the summer season).

Should we then seek merely to multiply facilities to accommodate the growing crowds? More parking lots, more campsites, more picnic tables, and wider roads would mean more crowding and the progressive destruction of the parks' most cherished values. New patterns of use must be developed that will allow the number of visitors to be increased without serious damage. A desperate but perhaps necessary step would be to ration visits among applicants and to stagger visiting times, but it would be better to diversify the environments and the uses. Some areas can be kept inviolate, except to the lonely few who are willing to work hard for solitude. These are the wilderness or roadless regions where John Muir found a good practical sort of immortality[11]. In other areas, secluded spots can be preserved where visitors will find those frogs in springly places, elysian springs and aromatic groves that once delighted the naturalist, William Bartram [12]. Other areas can be toughened with wear-resistant paths and various guidance devices, so that they can accommodate fair-sized crowds.

The areas of lesser scenic value in or near a national park could be developed for camping and overnight lodging, auto parks, picnic sites, and other services. The campsites could be well spaced to afford privacy and cleanliness. Souvenir, restaurant, and other commercial facilities, as well as ample parking space, could be installed. A variable fee system might encourage visitors to use such areas.

To reduce the use of private automobiles in the parks, a system of buses might be introduced. If necessary, this service could be offered free of charge, but it might be possible to recover capital and operating costs through fees. These buses could have glass or plastic tops to permit wide-ranging vision and a sense of being out-of-doors. Multiple doors would enable people to get in and out easily. Tours could be scheduled at frequent intervals, and stops would be planned so that visitors could see the principal scenic and historic attractions. Drivers could point out unexpected sights such as flowering plants or a bear and her cubs. Tape-recorded commentaries on the park's geology, animal and plant life, history, and other interesting aspects could be an important part of the bus ride, and the drivers could provide additional explanations for unusual attractions. Monorails or narrow-gauged railways should also be considered as a means of transportation in scenic areas. These would make it possible to keep automobiles out altogether.

To induce a more even flow of visitors to the parks a system of variable entrance and user fees might be devised that would tend to price people out at peak holiday periods and encourage visits on days or hours when the load is light.

Broadening the Base

It is evident that the areal extent of national recreation lands will need to be increased in the future, but it will be equally desirable to seek greater diversity, for example, through development of national seashores, riverways, wildlife refuges, wild rivers, undersea parks, and other special environments.

During this century, the United States has been transformed from a predominantly rural society into a predominantly urban one. As a result, we are becoming, even more than in the past, a land-surplus nation. More than half the counties in the United States lost population during the last 20 years. In the future, much of our land area will find its greatest use as recreational space. Land taken out of agriculture can be employed in this way; other, more obvious candidates for such use are the marshy islands of the Mississippi Delta and the rocky barrens of the Nevada desert.

The Shoreline: A One-Dimensional Resource

One resource must be given special consideration: the shoreline. This essentially one-dimensional boundary zone between the land and the sea, is one of the most limited and yet heavily used parts of our environment. Nearly half of all Americans live within a hundred miles of the ocean or the Great Lakes; this proportion will probably increase in future decades. By the end of the century, some 150 to 200 million people may be struggling for places on the beaches and in the narrow coastal waters. Even if the country's entire coastal strip were converted

into public beaches, this would mean about two people per foot. Actually only a fraction of the shoreline is available for public recreation. Long stretches are owned by industrial, shipping and military interests, or private individuals, and much of the remainder is too polluted to be safe for swimming or water sports. Bays and estuaries which are at the lower ends of river basins and receive their waste products are especially threatened by pollution. For existing shorelines, three sorts of action are urgent: The areas available for recreation should be retained, improved, and extended; the routes of public access to these areas should be increased; and measures to curb pollution should be taken wherever possible. In addition, the shoreline can be stretched by building spits and peninsulas, offshore bars and islands, and by dredging or improving estuaries. Chicago has already led the way by constructing a series of curving peninsulas along its lake front. Mission Bay, a great aquatic park near the heart of San Diego, was created by radical modification of a previously marshy estuary.

Many coastal areas are too cold for comfortable swimming. We should explore the possibility of using the waste heat from thermal powerplants or other similar sources to warm these waters. Offshore barriers to retard the mixing of cold and warm water might be required to achieve a satisfactory increase in temperature. The powerplants would not need to be built on the coastline, where they might spoil the view, but inland, and the necessary pipelines constructed.

Exciting possibilities exist for underwater recreation in the coastal zone. Recreational submarines of all sizes with viewing ports and automatic manipulating arms can be envisioned. Large underwater chambers, 100 to 200 feet deep, connected to the shore by metal tubes off steeply sloping coastlines, could be used as a new kind of tourist resort.

Manmade Rivers

Manmade lakes find some of their best uses in human recreation. New "rivers" might also be created for recreational and scenic purposes. Irrigation canals, which are, in effect manmade rivers, are commonplace. What modifications would have to be introduced if waterways were to be built not just for irrigation, but for their recreation and scenic qualities? The new rivers could be designed to intersect natural watercourses that now carry too little water—arroyos or streams in suburban areas where flow has been reduced by changes in land use. The scale of the projects could extend over a wide range, and they could be designed for multiple use. For example, if the water were carried through several channels instead of one, the new rivers would serve as flood-control devices. The range of possibilities would become vastly greater if water supplies in Canada could be tapped. New rivers would be an asset to the eastern as well as the western parts of the country, although they are likely to be most feasible among the eastern slopes of the Rockies and in the Great Plains.

Multiple Use of Federal and State Lands

In recent years, multiple-use planning and management of government land areas and development projects have been increasingly employed as a means of stretching resources to meet growing recreation demands. Ideally, a variety of activities can be made compatible

with the primary purpose of a development, and relatively small changes can produce large recreation opportunities. Multiple-use planning and management have required substantial alterations in the traditional outlook of engineers accustomed to single-purpose developments.

Continued progress along these lines is essential, and today it is almost taken for granted. But, like other valuable concepts, multiple use has limitations and pitfalls. Some uses are not possible in a multiple-use area. Allowing such use everywhere is ultimately equivalent to denying beautiful, lonely places to lovers of undisturbed nature. The danger is most serious, of course, where the supply of resources for this "fragile" use is limited, and where multiple use will destroy the resources irreversibly. Such preemptive uses as hunting and water skiing prevent other possible uses, but are not, in themselves, necessarily undesirable. If, however, they are uncontrolled in a laissez-faire multiple-use plan, they become single-use purposes by default.

Careful consideration, but not necessarily undue favoritism, must be given uses that are denied by any other use, and reasonable restraints should be placed on preemptive use. Those areas that will not bear the gamut of use must be determined early in the planning process. If the supply of any class of distinctive areas is clearly limited, these should be retained in single-use status. If, on the other hand, some classes are in abundant supply, a portion can be returned to the multiple-use pot.

Problems of Land Acquisition

Meeting national needs for recreation and preservation of scenic beauty will require public acquisition or other public control of much acreage now in private hands. There will be problems of equity for three parties—the property owners who must lose their holdings for the greater public interest, the local authorities that depend on property-tax revenues, and the Federal and State taxpayers who must foot the acquisition bill.

Property acquisition for Federal use often leads to difficulties for local taxing authorities. These have been largely ignored in the past. Recently, however, Congress authorized payments, based on a small percentage of land value, to local authorities for Federal wildlife refuges. The administration's proposal for a redwoods park includes economic adjustment payments over a limited period of time to local communities to counteract the initial loss of tax base and economic activity.

Just as Federal acquisition of property should involve neither niggardliness nor indifference to the problems of individual landowners and local taxing authorities, it should not be a source of reward for speculative dealing in land. But the present land-acquisition process unfortunately provides ample opportunities for inordinate price rises. These result from several factors. Among them are the longtime lapse between the area study and congressional authorization for purchase; piecemeal and strung out acquisition due to budgetary restraints; the availability, long in advance, of detailed information on specific acquisition proposals through the justification of budgetary requests; the unavoidable inclusion in appraisals of land in the vicinity of a proposed park of value elements created by the expectation of the

park itself; and the difficulty of getting detached judgments of values by local juries when condemnation cases go to courts. But one must ask whether the property owner who forfeits his land to condemnation should not share in, equally with neighboring owners whose land is not taken, the enhancement of values created by a public development.

Condemnation cases involving large estimated values—say $50,000 or more—should be adjudged by panels of disinterested experts under an administrative proceeding, with right of appeal to courts. Juries are often not sufficiently equipped to provide fair judgments in such cases.

In any case, the problem of escalating prices must be ameliorated if an adequate public acquisition program is to be carried through. Otherwise, land costs will rise beyond the capacity of politically feasible commitments of public funds, even though the "real" cost, in terms of resource use, of acquisition of previously little used lands, is often less than the money cost, because property owners, by and large, will save and reinvest the greater portion of payment received.

New approaches, some of them radical, are called for. As a general principle, acquiring authorities should be empowered to use with great flexibility the widest variety of tools at hand. This means, *inter alia*, minimizing full acquisition, for example, by such devices as purchase of easements, and authority for option and lease-back arrangements.

A Federal land agency should be established with authority to acquire options or to purchase before specific park authorization. Financed with an initial capitalization to be used as a revolving fund and with authorization to issue bonds, this agency would act as soon as plans indicated the desirability of beginning acquisition. It would pay local taxes on all its holdings. When a land-using agency obtained the necessary appropriations, it would reimburse the Federal land agency for its outlay, including accumulated interest and taxes.

Values and Costs

As indicated above, the creation of resources for public outdoor recreation sometimes requires only a kind of transfer payment. For example, establishment of a national seashore may chiefly involve acquisition of private property already being used for recreation. But in other cases, we must be concerned with the allocation of scarce resources among alternative uses: A flourishing sport fishery may call for severe restraints on commercial fishing; a national park may mean the cessation of lumbering or mining; land for a city recreational area could be used instead for offices or factories. Creation of the Indiana Dunes National Lakeshore was vigorously and successfully opposed for decades by local interests who believed that the area should be given over to industry. Recreation advocates must be able to convince a majority of their fellow citizens that the value of a recreational activity will be greater than the values it would displace, or else make such a nuisance of themselves that the line of least resistance is to placate them. In essence, the problem is to compare tangible market values with intangible "human" ones. This is a problem that can only be solved politically. Here as in many other human affairs, passions spin the plot.

The irrational nature of recreation and the impact of natural beauty, and the uniqueness of many public sites compound the difficulties. No

private entrepreneur can offer a Mount Rainier, a Grand Canyon, or a Crater Lake. Nor can the technique of cost effectiveness be used to balance public expenditures for recreation against alternative claims. The principle that human beings need and should have recreation for its own sake means that its benefits and therefore its cost effectiveness cannot be compared with the benefits of obtaining some other objective such as physical or mental health, or the amelioration of social ills. Similarly, the enormous diversity of recreational interests among different individuals makes comparisons of the value of various kinds of recreational facilities not only invidious, but nearly impossible, except in the crudest fashion. An attempt should be made to meet the needs of large numbers of people before paying too much attention to small minorities, but the latter should not be totally neglected.

Diversity of interests and values among human beings is one of the most telling justifications for private philanthropy in our society, particularly where it can be combined with pioneering experimentation or with service to minorities. The farsighted Rockefellers have demonstrated the possibilities for private provision of public recreation by creating with their own funds a national park in the Virgin Islands. In the future, as recreation increases in public virtue, we can expect both foundations and generous individuals to contribute to it as they now give to higher education.

In a previous section, I have suggested the use of fees to even out visitor loads and to reduce the pressures of crowds in the central areas of national parks. A more fundamental question is, How much of the total cost of public outdoor recreation should be borne by the users, and how much by the taxpayers? An implicit partial answer to this question is given by the principle that all citizens have a right to outdoor recreation, just as they have a right to health, education, and welfare. The taxpayers must bear enough of the cost to insure that this right can be exercised by all who wish to do so. Fees should not be imposed when a charge might destroy or damage the social value of an outdoor recreation facility. To uplift the quality of life in an urban ghetto it may be highly desirable to provide attractive free outdoor recreation, including free transportation to and fro.

One principle of value seems clear: Irreplaceable and unique resources should be preserved, even at considerable cost, for they are beyond price. The California redwoods are a prime example. These ancient trees have outlived civilizations and occur only in one relatively small area of the earth. They could not be replaced within any meaningful timespan. It would be the height of arrogance and selfishness for our generation to deny them to all future human beings.

REFERENCES

[1] *Natural Beauty Message to the Congress.* President Lyndon B. Johnson. February 1965.

[2] *Restoring the Quality of our Environment.* Report of the Environmental Pollution Panel of the President's Science Advisory Committee. The White House. November 1965, pp. I–XIII and 1–317.

[3] *New Resources in an Urban Age.* Harvey S. Perloff. Paper presented at the ACTION Symposium on the Future American City, January 1966.

[4] *The Adequacy of Technology for Pollution Abatement.* Report of the Research Management Advisory Panel through the Subcommittee on Science, Research, and Development to the Committee on Science and Astronautics, U.S. House of Representatives. U.S. Government Printing Office, Washington, D.C., 1966, pp. I–X and 1–17.

[5] Ibid.
[6] Alan Kneese, *The Economics of Regional Water Quality Management.* 1964. Published for Resources for the Future. Johns Hopkins Press.
[7] *Summary of Total Air Pollution Data for Los Angeles County.* A Report of the Engineering Division, Air Pollution Control District, County of Los Angeles, January 1965, pp. i–ii and 1–41.
[8] *A Study of Pollution—Water,* A Staff Report of the Committee on Public Works, United States Senate, U.S. Government Printing Office, June 1963.
[9] *Waste Management and Control.* A Report to the Federal Council for Science and Technology by the Committee on Pollution of the National Academy of Sciences, National Research Council. Publication 1400. National Academy of Sciences, National Research Council. Washington, D.C., 1966, pp. i–vii and 1–257.
[10] Lynn White, *The Historical Roots of Our Ecological Crisis, in Science,* vol. 155 (March 1957) p. 1203.
[11] Walt Whitman, *Leaves of Grass* (1860 edition).
[12] Quoted by Stewart Udall in *The Quiet Crisis* (New York, 1965) p. 112.
[13] William Bartman, *Travels* (1791) quoted by Stewart Udall in *The Quiet Crisis,* p. 41.

ENVIRONMENTAL QUALITY AN INVESTMENT IN MAN'S WELL-BEING AND PRODUCTIVE CAPACITY

BY RICHARD D. GRUNDY*

In reviewing environmental pollution today, we are painfully aware of the interrelationship between technology and the human environment. For centuries and even up to recent decades, man has continued to go along modifying his natural environment without any harmful or even noticeable implications. Man has refined, processed, and changed the forms of enumerable natural materials and has used the air, water, and land to dispose of the resultant waste materials.

In the process man has relied upon what has been termed the "natural assimilative capacity" of the environment to reprocess most of his wastes without concern over its capability to do this. However, during recent times man has become acutely aware that we have exceeded the environment's capability to reprocess these wastes. We are reminded immediately of the words of the noted John Galsworthy when he said, "If you do not think about the future, you cannot have one." Although written some time ago, his words bear a message not only for the individual but for all civilization.

The belief that science and technology has provided the means for man to conquer nature and achieve a better life for us has been challenged by nature herself. To appreciate this we need but look at recent environmental experience. The majority of our Nation's rivers run black, unable to support the biological life we relied on to reprocess our wastes. Recent experience with foaming streams and rivers pointed to our lack of planning in the development of a biodegradable detergent.

Occasionally we may yearn for the "good old days" but we are quickly reminded of the benefits that have accompanied the problems at hand. Occasionally we tend to accept the benefit, discounting the accompanying environmental pollution and impairment of man's physical and mental well-being. Environmentally speaking, there is a price to pay for a livable environment whether in terms of impairment of man's productive capacity or in terms of cost of control and abatement. As each new problem looms before us, we are goaded into adjusting and revising our technology so that it assists nature in supplying and resupplying our life system needs.

What this Nation needs to do is learn as a society to take the long view rather than the crisis-to-crisis approach to environmental pollution. This requires that we emphasize human goals and values which provide for action before the fact, rather than reacting after the fact.

In the past, in an effort to avoid inhibiting economic development on the national or regional scene, control authorities have tended to lean over backward to make pollution control as inexpensive as pos-

*Professional Staff Member, Subcommittee on Air and Water Pollution, U.S. Senate Public Works Committee.

sible. It has become common practice to delay pollution control until the damaging effects on public health and the environment have been proven beyond all shadow of doubt. And when proof is forthcoming, pollution control is evaded on the justification that, relative to some other cause (i.e., traffic deaths, floods), the effect is insignificant.

Technologically we have the means to keep our rivers clean and our air free of pollution. The future is optimistic because there are few problems we face today which, strictly from a scientific and technological standpoint, could not be solved within a reasonable period of time using today's knowledge or the knowledge we hope to get from current investigations.

As a first step we must define long-term goals that are directed at restoring and maintaining the quality of our environment and man's productive capacity without disrupting the culture and economy. This is an ambitious undertaking and the price each citizen must pay for environmental pollution control and abatement has not been fully realized. Clearly what is needed is knowledge of the effects of man's activities on his environment and indirectly on himself. This must be accompanied by development of the means and know-how to measure and preserve environmental qualities which promote and enhance man's well-being and productive capacity. The success of environmental quality centers on man's ability to segregate pollutants from reusable resource.

Air Pollution

A positive and coordinated approach is necessary in air pollution control. An adequate approach is needed to both preserve the atmosphere in a condition that will protect the public health and welfare, as well as to insure enhancement of environmental quality. However, the successful improvement of man's atmospheric resources requires the identification of the specific air contaminants and the nature of their effects on health and welfare.

Unlike water quality which conceivably could be achieved by treating the water phase of the environment, restoration and maintenance of atmospheric quality can be achieved only through the minimization of wastes at their source.

To date our national approach to air pollution control has been a defensive, piecemeal attack on each newly identified atmospheric crisis. This is exemplified in the defensive response to a steadily increasing number of specific problems represented by the need to reduce automobile emissions in Los Angeles, Calif., to control of phosphate emissions in Montana, and to restrict the sulfur content of fuels used in the New York metropolitan area.

These crises and the growing number of more or less spectacular "air pollution episodes" are in some respects a dramatic short-term manifestation of the general problem of atmospheric pollution. The lower levels of atmospheric pollution have more persistent and insidious effects on public health, on vegetation, and on materials of all kinds and, in the long run, undoubtedly represent a greater damage and loss to individuals and to the Nation than does the occasional "air pollution episode."

However, there can be no doubt that the effects of air pollution on human health are unesirable, whether measured positively in relation to growth, well-being, and joy of living, or negatively in terms of

death, disease, and the economic loss which goes with the incapacity to work. Previous studies of the Joint Economic Committee have classified the economic benefits of environmental pollution control and abatement into two categories. First, an increase in "gross national product" as a result of the increased productivity of workers whose early death is averted or whose sickness is averted, shortened, or made less severe. Second, a decrease in the cost of the health services which are needed to provide treatment for those individuals whose health is impaired as a result of environmental pollution.

Experience over the last two decades has provided ample evidence of "air pollution episodes" which are characterized by high levels of pollution over relatively short periods of time resulting in death or acute illness. The most widely known of these episodes have occurred in the heavily industrialized Meuse Valley of Belgium in 1938; in Donora, Pa., in 1948; in New York City in 1953, 1958, 1963, and 1966; and in London in 1952, 1962, and 1963.

A typical picture of the health effects of an "air pollution episode" is an immediate rise in mortality which occurs within 24 hours of the onset of the episode. The age groups particularly affected are the very elderly, and to a lesser extent the later middle ages, and infants under 1 year of age.

Possibly the most extensively studied episode was the 1952 London episode in which 4,000 more deaths occurred in that city than would normally have happened during a similar period of time. A detailed analysis of the excess deaths during the fourth week of the episode showed a three-fold increase in babies (4 to 52 weeks) and in adults (over 55 years of age), and an approximately two-fold increase in deaths for all other ages. A detailed investigation which included 1,280 postmortem reports of persons who had died suddenly either before, during, or immediately after the episode revealed no case of sudden death which could not be explained by previous respiratory or cardiovascular illness.

A rough estimate of the loss in productive capacity during the 1952 London episode is not possible at this time; however, the loss of babies' lives and the accompanying long-term loss of their productive capacity could easily justify control without consideration being given to other age groups. Because of the lack of information on disease morbidity during the episode, it is not possible to estimate the economic loss caused by incapacity to work, but it is reasonable to assume that it was comparable to the loss due to death.

Similar "air pollution episodes" have been observed in New York City in 1953, 1958, 1963, and 1966, with an estimated 200 to 400 excess deaths in the 1963 episode alone. Comparable results are not available for other major U.S. metropolitan areas either due to the lack of suitable mortality or air pollution data.

The individuals dying in these episodes have had a previous history of cardiac or respiratory disease, but the fact remains that they become bodies in the morgue. This atrocity is compounded further by a growing amount of evidence that air pollution is a causative factor in disease etiology.

Naturally one assumes that the chronic effects of air pollution are manifested mainly in the respiratory system, and there is ample evidence that air pollution is an etologic agent in the occurence or aggra-

vation of acute nonspecific upper respiratory disease, chronic bronchitis, emphysema, bronchial asthma, and lung cancer.

Quantitative investigations into disease causation of air pollution is both time consuming and complex because not all individuals react to the same concentrations of irritants. At one extreme are the so-called normal individuals to which may be ascribed normal variations caused by social, ethnic, and cultural backgrounds in addition to the effects of smoking. At the other extreme are susceptible groups of individuals such as the bronchitic patient whose disease is in part an attempt by his body to protect his lungs against further irritation by secreting more mucus than is normal. At some stages in the disease etiology the individual's disease might make him able to cope with more irritants than his counterpart, while in the advanced stages of the disease a strain may be put on his heart and lungs by air pollution to the extent that death occurs. Regardless of the route taken, the end result is identical and dictates air pollution control and abatement.

While the investigator considering "causation" must inquire how and by what mechanism the effects are produced, the requirement to take action invariably will not allow control officials to remain inactive until the whole chain has been uraveled. A few links may have to suffice as a basis for action, depending on the circumstances.

While air pollution contributes to the etiology of respiratory disease, there is ample evidence that five other factors are involved also. These other factors have been defined in studies into the pathology and evolution of bronchitis. The first factor is "demographic" and defines the health status of a community in terms of disease mortality and morbidity, according to age, sex, etc. The second factor is the "statistical significance" between indices of health status of the community (i.e. mortality and morbidity) and the various indices of air pollution (i.e. particulate and sulfur dioxide levels in the atmosphere). Third are the "smoking habits" of the individuals exhibiting health effects. The fourth factor that must be evaluated is the effect of social class which is in part an environmental effect but unrelated to occupational exposures since a comparable effect is observed in both males and females. Fifth are the climatic factors such as temperature, humidity, rainfall, sun hours, barometric pressure, and season of the year which have been observed to correlate with adult admissions of respiratory patients to hospitals.

In the epidemiological investigation of air pollution health effects, these factors must be taken into consideration although it is difficult, if not impossible, to separate the contribution of each factor. However, the protection of "public health" is paramount and requires action based upon the best available evidence. While strong evidence of a cause-effect is desirable this does not imply the need to cross every "t" before acting. The fact that all scientific work is subject to revision in light of advancing knowledge does not confer upon us a freedom to ignore the knowledge we already have, or to postpone action when knowledge appears to demand it at a given time.

Air pollution control technology is available which can bring a large part of the problem under control. The key is a willingness to do so and an ability to make the investment. Much of this technology is expensive and it is necessary to balance the need for action against the cost; however, protection of the public's health is of first priority.

Research effort is required to develop less-expensive techniques. Considerable research was provided under the Clean Air Act into new and improved methods for the prevention and control of air pollution resulting from the combustion of fuels. However, the research to date was inadequate to meet the need in both funds and effort. Recognizing this, Congress passed the research sections of the Air Quality Act of 1967. Congress was convinced that no matter how successful the development of solutions to the remaining technical problems of air pollution control through the combined efforts of industry and government, it was unrealistic to consider that progress would be rapid enough to cope with the steadily worsening air pollution problem in the absence of the new authority contained in the Air Quality Act of 1967.

The proposed research makes provision for enlarging this Nation's research effort by the specific authority to enter into demonstration contracts with industry. Such contracts are to prove the technological and economic feasibility of the control of pollutants from the combustion of fossil fuel and to shorten the timelag between research, development, and full-scale practice.

It was recognized that current technology is inadequate to control all the sources of atmospheric pollution. Only through the continued, combined efforts of government and industry will alternate solutions become a reality. These solutions will not be completed overnight and the associated costs will require their application over many years so that this Nation's economy will not be severely disrupted. But responsibile industries will contribute significantly to the ultimate achievement of environmental quality.

Water Pollution

The total fresh water supply of this Nation and the world is relatively fixed, although populations and industrial needs are steadily increasing. The requirements for clean, fresh water are increasing so rapidly that our consumption will rise in 20 years from 370 billion gallons per day to 600 billion gallons per day. This increased demand for water will be felt by all segments of this Nation's economy. Concurrently, this Nation's waterborne wastes are increasing at a rate such that they will, by 1980, have the potential to consume all the oxygen in all the 22 river systems of the United States.

Accompanying this requirement for clean, fresh water is a growing concern over the ever-increasing introduction into water of chemicals and radioactive materials with carcinogenic, toxic, and physiological effects on man and his environment. However, this concern should not overshadow the need for adequately treated water from a bacteriological or enterovirus point of view. The World Health Organization has estimated that water-carried enterovirus rates in children under 15 years of age average 10 percent in the United States and can be expected to be higher in less developed countries. It would be very interesting to know the economic losses incurred as a result of recurring enterovirus epidemics in the United States alone.

In addition to chemical and biological degradation of ground and surface waters, consideration must also be given to physical forms of pollution, among which heat and radioactivity are the most important. Thermal pollution is critical because of resultant high temperatures of

surface waters and loss of dissolved oxygen due to de-aeration with a marked effect on the water ecology.

Considering the chemical contamination of surface and ground waters, we have as an example the history of pollution by detergents and their residues. Because of their chemical structure synthetic detergents were particularly resistant to decomposition by bacterial action. In this respect they were very different from soaps which undergo rapid decomposition. The resultant residues discharged to surface waters after sewage treatment caused a serious deterioration in water quality that spurred the industry, through public pressure, to develop a substitute more susceptible to bacterial attack. This brief example is presented as background to the more recent problem posed by the increasing use of synthetic organic compounds.

The presence of persistent synthetic organic chemicals in sewage only present a health hazard when they occur in domestic or municipal water supplies which has been the case. The presence of at least some of these contaminants in domestic water supplies is certainly undesirable.

Pesticides are a case in point. First, laboratory studies show that the direct toxicity of some pesticides to fish is extremely high, occuring for some substances in the range of parts per billion. Since these levels are often below environmentally detectable levels, protection is required before the fact. Second, it has been well established that pesticides present at low levels in water can be concentrated manyfold by aquatic organisms including algae. This concentration mechanism increases the degree of contamination at each step in the food chain of man. Thus, although fish may not themselves be killed at lower levels of pollution, their bodies may contain sufficient pesticide residues to harm birds and conceivably those members of society who consume a high proportion of fish in their diet.

The introduction of a waterborne contaminant into the food chain of man was exemplified in Japan in a case involving mercury where human disease and death resulted from eating fish taken from waters polluted by industrial wastes which contained mercury. Although the mercury concentrations in the water were not lethal to the fish, the fish had absorbed sufficient mercury from the water to be a toxic foodstuff. This example emphasizes the need to consider the routes by which man may be exposed to waterborne contaminants in addition to drinking water.

While this discussion focuses on water contaminants, food should not be overlooked as it may be a more important source of ingested contaminants than water and cannot be excluded in any consideration of health effects.

Water pollution by mineral contaminants is of particular interest although exposures are not high enough to cause clearly demonstrable toxic effects. However, it is possible to identify the presence not only of substances vital to the body functions such as copper, cobalt, molybdenum, manganese, and zinc, but also the presence of potentially toxic substances such as cadmium, chromium, vanadium, nickel, and lead.

In recent years it has been suggested that certain metals in concentrations normally present in the body may contribute to the development of chronic disease. There are a number of hypotheses on the

mechanisms involved, and considerable work has been done of the health effects of iron, cobalt, cadmium, mercury, and zinc.

This discussion has cited a few examples of the potential effects of water pollution on man, effects which may decrease man's productive capacity. With each subsequent reuse of our water resources the quality of the water is modified, generally to the disadvantage of subsequent users and of the aquatic life. There is clearly a limit to the waste products that a stream or lake can assimilate without serious effects on man's physical, mental, and social well-being.

Land Pollution

Solid waste matrials are quickly becoming a major disposal problem. Generated by domestic, industrial, and agricultural sources alike, the domestic solid wastes alone amount to 3.2 pounds per person per day on a dry weight basis. In contrast the total solid wastes carried away as water-borne wastes amount to less than 0.5 pound per capita per day, or approximately one-third the amount of wastes disposed of in a solid form.

Disposal of solid wastes may take the form of incineration, composting, dumping or burying on the land, or barging to sea. Regardless of the manner of disposal, each disposal procedure has accompanying public health problems associated with the chemical and physical composition of the waste materials.

The burning and composting of domestic wastes accounts for the organic components of the waste materials but leaves an inorganic portion which must be disposed of by other means. Air pollution by contaminants such as nitrogen oxides accompanies even the most efficient incineration. Of particular concern are the contaminants produced by the incineration of plastics. The combustion of certain types of plastics (i.e., teflon, fluorinated plastics, and vinyl plastics) produces chemical contaminants whose physiological effects may be similar to those of phosgene gas, a severe respiratory irritant used as a war gas in World War I. While the practice of burning or incinerating plastics along with other refuse may pose no immediate hazard at this time, it is one contaminant that requires action before the increased use of plastic creates a widespread and very serious health hazard. However, reports indicate that no efficient or economical means are known to reclaim plastics being used in ever increasing amounts for food containers and other household goods.

Composting and open dumping of domestic wastes, on the other hand, may increase the population of flies, rats, and other disease vectors. Buried organic wastes are subject to anaerobic decomposition with the accompanying production of methane and carbon dioxide which may result in pollution of ground waters.

The major problems associated with the disposal of industrial solid wastes are economic, while those instances which cause a direct health hazard are the exception rather than the general rule. The major health hazard arises from the combustion of these wastes either when they are burned or ignite spontaneously. However, when disposed of on land, subsequent leaching by rain can result in pollution of surface or ground waters. Of particular importance is the need to develop techniques to dispose of radioactive wastes from mining and refinery operations, nuclear reactors, and nuclear research facilities.

Solid agricultural waste problems center on canning plants, packaging plants, produce yards, etc., where huge quantities or organic wastes may accumulate. These wastes may subsequently decompose as well as attract disease vectors which result in pollution of the environment.

One of the greatest hazards in solid waste disposal is possible chemical pollution of the soil and the subsequent exposure of man through ground water, surface or drainage water, and the plants used as food or forage for domestic animals. An example of this is the contamination of soil by impurities in fertilizers used to fortify soil.

Ideally, only those chemicals that have been proven biodegradable should be used. This is reflected in the present trend to manufacture agricultural pesticides that are short-lived, minimizing the persistence of pesticide residues and their degradation products on food and forage crops. Whatever and wherever their source, most solid and liquid wastes are disposed of on the land. While gaseous wastes are generally discharged into the atmosphere and liquid wastes are discharged into surface waters, solid wastes are placed on or in the land. Some solid wastes are dumped into surface waters with resultant water pollution and some are burned or incinerated with consequent air pollution. Regardless of their means of disposal, all wastes contribute to environmental pollution and the subsequent exposure of man to additional biological stresses which may impair his health and well-being.

Discussion

If our quest for environmental quality is to be successful, every industry, every business, and every individual living in America must do his part. Then, and only then, will America's efforts to preserve our environment be successful.

Our current dilemma results from shortsightedness in the past. First, we were not aware of the changes we were causing, or of nature's ability to handle them. Second, it was economically cheaper and easier to disregard the waste as long as we got the desired product. We have found ourselves in a series of technological traps because we have relied on the crisis to crisis approach to handle environmental pollution, reacting to pollution in the same manner that we act to floods, drought, traffic jams, and riots. But time and our rapid national growth have exacted their toll and in order to insure environmental quality under future advances in our standard of living and increased urban-industrialization, we must practice environmental management.

Broad thinking and long term planning for action in environmental pollution control are among the necessities of our time. We have taken only the first step toward what many hope will be a more liveable world. But if we continue to limit our efforts to the equating of benefits with risks and are unwilling to make any sacrifices, if we continue to show a resistance to change and a dependency on others to act first, we will continue to find ourselves bogged down in the mire of our own shortsightedness.

Let us invent the future so that the environment is the kind that we and our children will want to live in and can live in. An environment that promotes and enhances man's productive capacity and well-being.

88-744—68—vol. 2——20

ENVIRONMENTAL POLLUTION

BY ROBERT U. AYRES and ALLEN V. KNEESE*

CONTENTS

I. INTRODUCTION

Environmental pollution has usually been observed and studied in rather rigidly separate classes based on the environmental medium into which waste residuals are discharged. This tendency has also been quite evident with respect to governmental institutions devised to control these problems. Thus we have at the Federal, State, and often at the municipal level independent units of government to deal with water-borne wastes, wastes discharged to the atmosphere, and solid wastes. Such a division of responsibilities can perhaps be justified on the basis of specialized skills needed to deal with different problems and the differing areal extent of the geographical "problems sheds" involved. It is also possible, however, that this segmentation results in significant measure from—and in turn contributes to—a failure to recognize and to develop analytical tools for proper consideration of the strong interdependencies which exist between the various waste streams and the control measures appropriate thereto.[1]

*The authors are indebted to many persons for help in the preparation of this paper. Among them, our associates at Resources for the Future, Blair Bower and Richard Frankel, merit special mention. We are also indebted to George Löf, Richard McKenna, and Erna Belton for research assistance.

[1] In the last few years there have been several studies which took a more comprehensive view of the waste residuals problem. Among them are *Waste Management and Control*, National Academy of Sciences-National Research Council, Washington, D.C., 1966; *Restoring the Quality of Our Environment*, report of the Pollution Panel of the President's Science Advisory Committee, the White House, 1965; and *Environmental Pollution, A Challenge to Science and Technology*, report of the Subcommittee on Science and Astronautics, House of Representatives, 89th Congress, second session, Washington, D.C., U.S. Government Printing Office, 1966. An additional study incorporating more empirical information and with a regional focus (New York Metropolitan region) is cited in the final section.

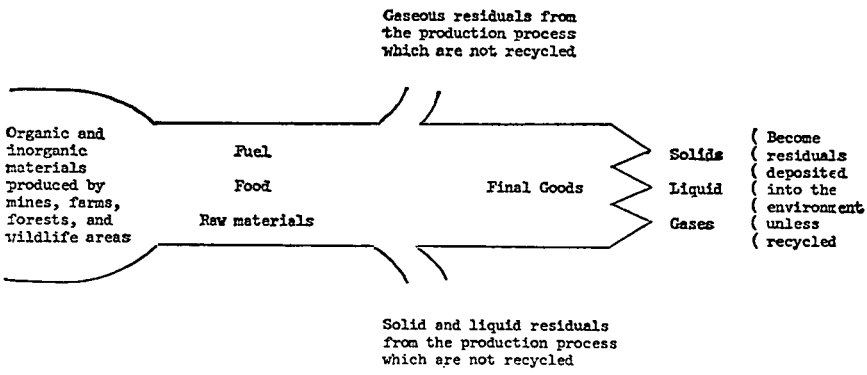
In this paper we do not pursue the subject of environmental pollution in terms of the usual air-water-solids categorization. Rather, we organize our discussion in terms of the main sectors of the economy which present the Nation—particularly, in its urban areas—with a waste disposal problem and attempt to make some progress toward viewing that problem in its entirety.

A MATERIAL BALANCE APPROACH

In this connection we find it useful initially to view environmental pollution and its control as a materials balance problem for the entire economy.[2] We start with a highly simplified schematic of how the goods and residuals production process works, as indicated in chart I.

CHART I

Schematic of the Goods-Residuals Production Process



Final goods in the schematic can be either consumer goods like food and autos, or intermediates in the productive process. The inputs to the system are fuels, foods, and raw materials which are partly converted into final goods and partly become waste residuals. Except for increases in inventory, final goods also ultimately enter the waste stream. For the sake of simplicity and also because it isolates the problems of greatest significance for urban areas, we will ordinarily begin our consideration of the materials flow after the basic fuel, food, and raw materials have been produced by mines, farms, forests, and wildlife areas, but before they enter into processing. It must be noted, however, that this does neglect some significant discharges of wastes into the environment—mine tailings, acid mine drainage, and salt water

[2] As far as we know, the idea of applying materials balance concepts to waste disposal problems was first expressed by Frank Smith in *The Economic Theory of Industrial Waste Production and Disposal*, draft of a doctoral dissertation, Department of Economics, Northwestern University. We also benefited from an unpublished paper by Charles Headley in which a pollution "matrix" is suggested. One of the authors has previously used a similar approach in ecological studies of nutrient interchange among plants and animals; see Robert U. Ayres, "Stability of Biosystems in Sea Water," Technical Report No. 142, Hudson Laboratories, Columbia University, August 1967.

pumped up with crude oil, for example. However, the streams included are the most relevant ones for urban environments.

In an economy which is closed (no imports or exports) and where there is no net accumulation of stocks (plant, equipment, inventories, consumer durables, or residential buildings), the amount of residuals which is inserted into the natural environment must be approximately equal to the weight of basic fuels, food, and raw materials entering the processing and production system, plus oxygen taken from the atmosphere.[3] This result, while obvious upon reflection, leads to the rather surprising and even shocking corollary that residuals disposal—in terms of sheer tonnage—is an even larger operation than basic materials production.

Chart II shows a materials flow of the type we have in mind in greater detail and relates it to the various sectors of the economy in terms of which our further discussion proceeds. In an open economy (such as any city, state, or nation engaged in trade across its boundaries) it would be necessary to add flows representing imports and exports. Similarly, in an economy undergoing stock or capital accumulation the production of residuals in any given year would be less by that amount than the basic inputs. In the United States, accumulation would account for about 10 to 15 percent of basic annual inputs, and there is some net importation of raw and partially processed materials amounting to 4 or 5 percent of domestic production. Table 1 shows the weight of raw material produced in the United States in several recent years, plus net imports of raw and partially processed materials.

Of the "active" inputs,[4] perhaps three-quarters of the overall weight is eventually discharged to the atmosphere as carbon (combined with atmospheric oxygen in the form of CO or CO_2) and hydrogen (combined with atmospheric oxygen as H_2O) under current conditions. This results from combustion of fossil fuels, and from animal respiration. Discharge of carbon dioxide can be considered harmless in the short run. There are large "sinks" (in the form of vegetation and large water bodies, mainly the oceans) which reabsorb this gas although there is some evidence of net accumulation of CO_2 in the atmosphere. Some experts believe that the latter is likely to increase greatly—as much as 50 percent—by the end of the century possibly giving rise to significant—and probably, on balance, adverse—weather changes.

[3] To simplify our language, we will not repeat this essential qualification at each opportunity, but assume it applies throughout the following discussion. In addition, we must include residuals such as NO and NO_2 arising from reactions between components of the air itself, but occurring as combustion byproducts.

[4] Excluding stone, sand, gravel, and other minerals used for structural purposes, ballast, fillers, insulation, etc. We also disregard gangue and mine tailings in this tally.

Chart II.--Materials Flow

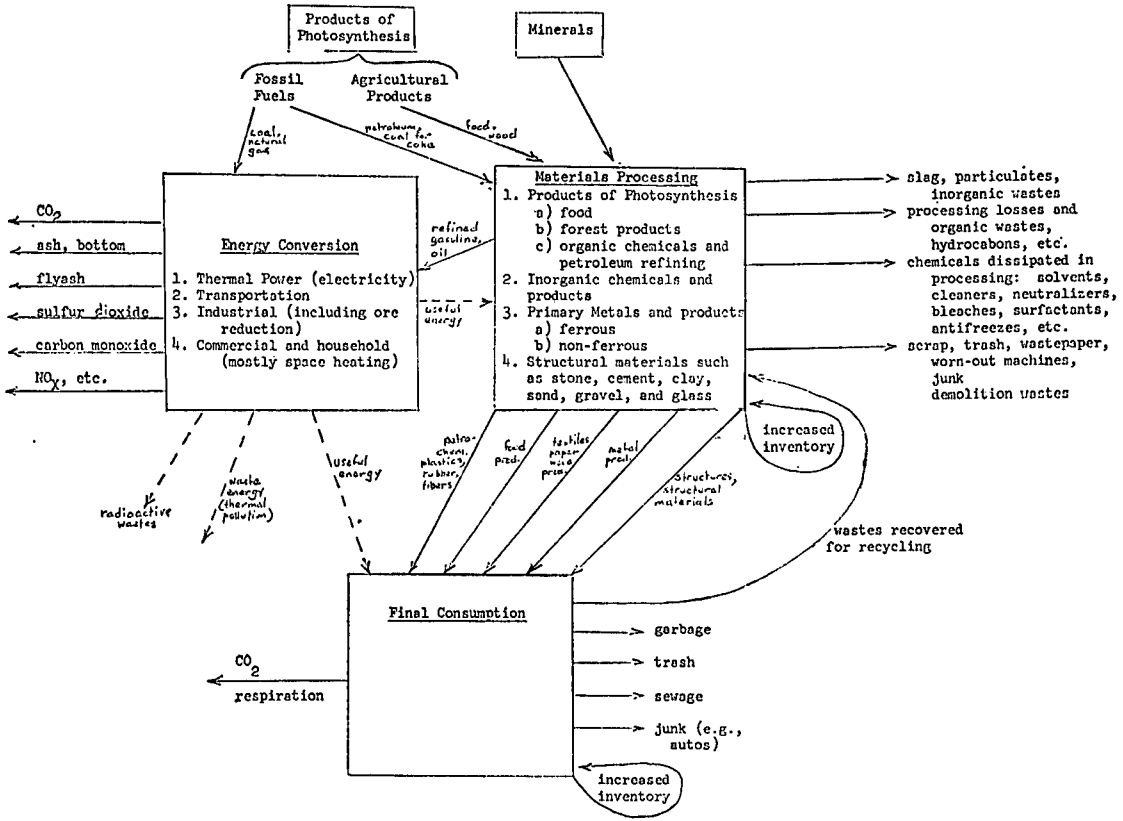


TABLE 1.—*Weight of basic materials production in the United States plus net imports, 1963*

[10^6 tons]

	1963	1964	1965
Agricultural (including fishery and wildlife and forest products):			
Food:			
Crops (excluding livestock feed)	125	128	130
Livestock	100	103	102
Other products	5	6	6
Fishery	3	3	3
Forestry products: (assumed 15% moisture content:)			
Saw logs	53	55	56
Pulpwood	107	116	120
Other	41	41	42
Total	434	452	459
Mineral fuels	1,337	1,399	1,448
Other minerals:			
Iron and ferroalloys	204	237	245
Other metals	161	171	191
Construction materials	1,579	1,668	1,763
Other nonmetals	125	133	149
Total	2,069	2,209	2,348
Grand total	3,840	4,060	4,255
Total active materials	2,261	2,392	2,492

The remaining residuals are either gases (like carbon monoxide, nitrogen dioxide, and sulfur dioxide—all potentially harmful even in the short run); dry solids like rubbish and scrap; wet solids like garbage, sewage, and industrial wastes suspended or dissolved in water. In a sense, the solids are the irreducible limiting form of waste. By the application of appropriate equipment and energy, all undesirable substances can be removed from water and air streams [5]—but what is left must obviously be solid. Looking at the matter in this way, clearly reveals a primary interdependence between the various waste streams.

But solid residuals, or for that matter those that remain in a liquid or gaseous state, do not necessarily have to be discharged to the environment. In many instances, it is possible to recycle them economically back into the productive system. We shall note some outstanding examples later in the discussion, along with a number of unexercised opportunities of this kind. The materials balance view underlines the fact that the total materials throughout necessary to maintain a given level of production and consumption decreases as the efficiency of utilization (that is, recycling) increases. Similarly, the useful lifetime of goods is closely related to the net throughput of the system. The longer cars, buildings, machinery, and other durables last, the fewer new materials are required to compensate for depreciation or sustain a given rate of capital accumulation.

Finally, the more efficient fuel combustion processes can be made (in the strict energy conversion sense), the fewer waste products there will be for the environment to receive, for a given total energy production. Perfect utilization of carbonaceous fossil fuels would leave only water and carbon dioxide as residuals, while nuclear energy conversion need leave no chemical residuals at all (although thermal pollution and radiation hazards cannot be dismissed).

[5] Except CO_2, which may be harmful in the long run, as noted.

One may reasonably protest that this is a very mechanistic and simplistic view of the problem: after all, improving combustion is not costless; increasing the durability of goods requires more expensive inputs; it may be cheaper to buy new raw materials than to recycle used ones, and it may cost less to buy additional gasoline than to achieve further increase in the efficiency of combustion. All these things are of course true. If the functioning of the economy gave rise to incentives (like prices), which fully reflected the cost of disposing of residuals including any costs to the overall society associated with their discharge into the environment, they would of course also be very much in point. But it is clear that, whatever other normative properties the functioning of a market economy may have, *it does not reflect these costs adequately.* Market economies are effective instruments for organizing production and allocating resources, insofar as the utility functions associated with two-party transactions affect only the parties. But in connection with waste disposal, the automatic market exchange process fails, because the utility functions involve third parties. The materials balance view gives us some insight into how large a failure this may be. However, before expounding the materials-balance principle further to elucidate the points made briefly above, some further comments on the basic economic issues involved seem to be justified.

THE ESTABLISHMENT OF VALUES IN A MARKET

To deal meaningful with the policy issues presented by discharge of residuals to the environment, it is important to have a concept of the functioning of the economy. While the American economy is a mixed one with heavy doses of public production and regulation, the best simple model for understanding its basic functioning and for providing a benchmark against which to measure its performance is the "competitive market." This is really an abstract, theoretical model of how choices concerning resources use are made in a decentralized decisionmaking system, where markets are competitive, and the individual decisionmakers (industries and individuals) act rationally to maximize their private benefit from whatever decisions they take. If we accept two basic assumptions, that is, that the overall distribution of income is justifiable on ethical grounds and that individual preference should be satisfied to the maximum extent possible given the distribution of income, a normative or "standard-setting" value is attributable to the outcome of the market process. Such a decentralized decisionmaking system will, by definition (subject to the foregoing assumptions), produce maximum welfare: the optimum mix of goods and services which can be procured within the specified income constraints. This process has been analyzed in great detail by economists but Adam Smith, who in the 18th century was perhaps the first to perceive clearly the possibility of an economic system functioning in this way, graphically described the process as the market's "invisible hand." The importance of this perception remains undiminished, even today, for it serves to isolate for us and permit us to understand the basic functioning of a decentralized economy—the organization of production so as to produce what each consumer wants within the limit of his income.

Like all models, the one we have just discussed is an abstraction from reality. But it helps us to identify the ideal functioning of a decen-

tralized system and to isolate and forecast the effects of departures from such an ideal system.

MARKET FAILURES

That things do not actually work out in the way specified by the idealized model will surprise no one who is exposed to the day-by-day operations of the economy. Many reasons for less than perfect operation of the economy are discernible. The particular aspect of "market failure" that concerns us in this paper arises from the fact that what the consumer buys and uses (we carefully avoid saying "consumes"), or what a business firm uses in the way of inputs, is only one aspect of the overall disposition of resources. Large residuals remain at each stage of processing and those that are not recycled become wastes, as do the "final" goods themselves when they become worn, broken, or obsolete. We have already seen that the total weight of these is greater [6] than the weight of fuels and raw materials initially introduced into the productive process. The environment has a certain limited capability to absorb these wastes without harmful effects, but, once the ambient residuals density rises above a threshold level, they become involuntary and unwanted inputs to other production processes or to final consumers.

At this point the automaticity of the market as an allocator of resources breaks down, in the sense that supply and demand—as expressed through a market mechanism—is not the only vehicle for determining what is received by consumers and what is used as inputs in business firms. There are flows of some goods and services that come to the consumer or business whether he wants them or not and without his paying for them or being able to avoid them by making a payment. The activities of an economic unit thus generate "real"—as contrasted with price or monetary—effects that are external to it. The economist refers to these as external effects or "externalities."

Among the clearest and most important sources of such externalities are the waste residuals resulting from production processes and consumption activities. It is apparent that the downstream water user, be he a recreationist or a manufacturer, can usually not control the quality of water he receives as return flow from upstream users, and the quality of this water is often degraded by residuals discharged into it. The breather of the air is usually not in a position as an individual to have any substantial impact on the quality of the ambient air by exercising his demand in a market. Similarly to the extent that chemical residuals are introduced into the environment after use, affected parties will find them in their food, air, and water, whether they want them or not, and they cannot avoid them by exercising a demand in the market.

The distorting effects of such externalities on resources allocation and use have been analyzed in considerable technical detail by economists. The basic problem which results can be expressed rather simply, however. A society which relies completely on a decentralized decisionmaking system and in which significant externalities (environmental overloadings) occur, as they do in any society which contains

---

[6] Because oxygen from the atmosphere enters into compounds in some of the residuals produced and combustion processes produce compounds from the air itself.

significant concentrations of population and industrial activity, will find that certain resources are not used optimally.

The capacity of the natural environment to assimilate residual waste is extremely valuable. To eliminate completely all residuals—as in a spacecraft—would be an immensely costly process; indeed, as the materials balance view shows, it necessarily implies complete recycling of *all* waste materials, which cannot be seriously contemplated.

But on the other hand, if no price or other use restriction is put upon the assimilative capacity of the environment, it will be used too much—and thereby abused. For example, when the assimilative capacity of a stream is exceeded, significant external costs appear: fish which have value for recreation and commerce will die, industrial plants and municipalities will have to treat their intake water more elaborately, etc. These are real costs which are external to the waste discharger and, since he does not bear them, he does not consider them in his decision-making. He will not design his manufacturing processes to take account of them; he will use too much of certain inputs, which have particularly large external costs; he will not recycle except where an internal economic benefit is apparent;[7] he will not control waste in the production process closely enough; he will not treat his effluents, even though doing these things might involve a smaller cost than is subsequently imposed on third parties. Similarly, the consumer will purchase insufficiently durable goods and—like the industrial producer—practice insufficient recycling; for example, he will not purchase returnable or reusable containers to an optimum extent. Under these conditions, not only will the environment be called upon to receive an excessive amount of residuals but the demand for natural resources inputs will exceed the optimum.

CONTROLS AND ENVIRONMENTAL MANAGEMENT

If we wish to realize the very great advantages of decentralized decisionmaking mechanisms—in other words the market—we must find appropriate ways to control these external environmental effects. That means that in some fashion the external costs imposed by residuals discharged to the environment—be they liquids, solids, or gases—must be weighed against, and balanced with, the costs of controlling the amount of these residuals including any external costs which may result from increased levels of residuals in one environmental medium as a result of control exercised in another.

There are two potentially workable ways whereby decisions on how much to produce and how to produce it—and similarly for consumption—can be induced to take better account of all the costs and benefits—private and external or social—flowing from the economic activity in question.

The first is to "internalize" the problem so that a single economic unit will take account of all of the costs and benefits associated with the external effects. One way to do this is to enlarge the size of the economic unit. Where external effects are limited in scope, there tends to be some incentive for merger,[8] because—as we implied earlier—

[7] In deciding whether to recycle, the industrialist should be forced to compare the value of the recovered material or byproduct *plus* the external cost avoided with the cost of recovery. Failure of the decisionmaker to bear external cost will clearly lead to too little recovery, although frequently (as we will see further) a certain amount of recovery is internally profitable.

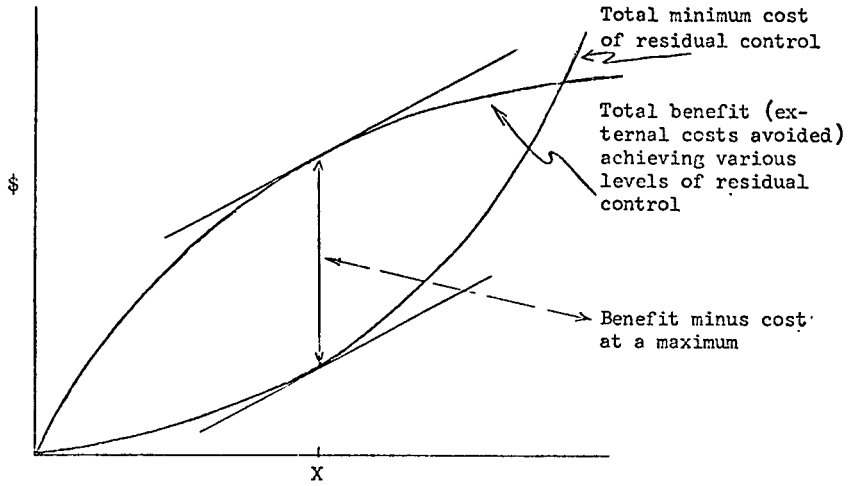
[8] An extreme form of this approach would be nationalization of all economic activity.

overall costs can be reduced, and overall net gains increased, when the external costs are properly taken into account. For example, combining a rolling mill with a blast furnace in an integrated plant facilitates the recycling of "home" scrap and saves on fuel and combustion residuals. But waste products are often so widely propagated in nature and affect so many diverse interests that the merger route is not feasible. Also the decisions of individuals and households bear upon the extent to which external costs occur and these decision units are not subject to merger in the usual business sense.

The second approach is closely related to the first but it involves the formation of a cooperative agency or governmental authority to induce (or require) systematic "internal" consideration of all costs and benefits by changing the framework within which the individual or firm would otherwise be operating. This is sometimes achieved by mutually negotiated restraints, but more frequently by external regulations enforceable at law. For example, certain emission standards may be required for automobile exhausts. There are numerous examples of public agencies placing regulations on economic activities such as production, purchase, and sale of property or goods to limit the external effects involved. Another means which is usually preferred by economists, but little used in practice, is to put an actual price (or tax) on the externality-causing activity. Thus, a tax on sewage effluents which is related to the quality and quantity of the discharge, or a surcharge on the price of fuels with high sulfur content which is meant to take account of the broader costs to society external to the fuel-using enterprise, are examples of this approach. This procedure uses economic incentives to allocate the resource (the waste assimilative capacity of the environment) similar to those generated where market mechanisms can balance costs and returns.

Internalization of external costs by one means or another presents complex problems, but even the best techniques for doing this cannot guarantee that environmental quality will be managed in the most economical way; that is, in such a fashion that the combined cost of controlling or alleviating the ill effects of residuals, plus the external damages imposed by uncontrolled final discharges, is minimized. In chart III this optimum level is shown at point X, which is defined mathematically as the point where the slopes (the derivatives) of the cost and benefit functions are equal. But for this to be the true economic optimum, the costs of residual control must be the lowest possible for a given level of control. In this connection, it is important to recognize that in most, if not all, fields of environmental quality management there are potentially effective and efficient measures which cannot be achieved either by setting standards or by imposing taxes upon the activities of individual decisionmakers that cause externalities (discharge residuals). In other words, there are potentially efficient environmental quality control measures unavailable to the individual firm, or even the individual municipality, not to mention the individual person.

CHART III



Example of such measures are reservoir regulation of the low flow of rivers to improve their residuals assimilation capacity, land-use restrictions to achieve a similar objective as regards the atmosphere (or "airshed") over an area, or coordinated solid waste disposal for a megalopolitan region as a whole. Economic efficiency requires that such measures be executed on a geographic scale which would normally encompass a number of political subdivisions, and our society is institutionally illequipped to accomplish this. Moreover, existing institutions for environmental management have no machinery or authority to allocate the external cost associated with residuals (or other external effects) to the originating activities, or—indeed—even to identify them unambiguously. This is true even when air, water, and solid wastes control are each considered in isolated categories and very much more so when the integral interdependencies between them are recognized.

We return to these points at the end of the paper but first we turn to a discussion of the types and sources of residuals, technologies for control, and costs of control measures associated with the several main sectors of economic activity having impacts on the quality of the urban environment. Throughout we place emphasis upon accounting for the materials flowing through the sector rather than focusing independently on air, water, or land as media of disposal. Our discussion in the next section often relates to national aggregates because data exist at that level to permit us to illustrate our basic methodology and explore a number of the implications which arise from the "materials balance" approach. However, more confined geographical regions are normally more relevant units for policy and management analysis. We return to this point in the final section.

It should also be noted that since our approach is relatively new and experimental and since suitable data for implementing it are often unavailable, numerous shortcuts and assumptions entered into the cal-

culations below. Accordingly, the specific numbers should always be regarded as illustrative rather than final and precise.

As illustrated in chart II, we find it convenient to order our discussion in terms of three major "sectors" of the economy; namely, (*a*) energy conversion, (*b*) materials processing and manufacturing (i.e., industry), (*c*) final consumption (i.e., households).

A formal materials input-output model is described in appendix A. A natural extension of the materials balance concept to include waste energy (heat) is discussed in appendix B.

II. Residuals Associated With Energy Conversion

The energy conversion sector, as a whole, obtained 4 percent of its output from hydroelectric generators and 0.1 percent from nuclear fuel in 1965 (although the latter contribution is increasing rapidly). The remainder was derived from fossil fuels as shown in table 2.

Table 2.—*Fuel consumed in energy production* [1]

Fuel source	Percent contribution to total primary energy	1965 consumption as fuel$\times 10^6$ short tons
Coal	23	465
Petroleum and natural gas liquids	43	503
Natural gas, dry [2]	30	337

[1] Figures from W. A. Vogely and W. E. Morrison, "Pattern of Energy Consumption in the United States, 1947–65 and 1980 Projected," World Power Conference, October 1966.

[2] Assuming an average molecular weight of 16 (methane, CH_4), whence 16 gm. (1 mole) occupies 22.4 liters; thus, 337×10^6 tons of natural gas is equivalent to 15.2×10^{12} cu. ft. (15,200,000,000,000).

In 1965, electric utilities consumed 20 percent of all primary energy, 24 percent went to transportation (mainly as gasoline), 32 percent was used in industry and 21 percent in households and commercial establishments. The overall breakdown of energy consumption, by source, of the various sectors is shown in table 3.[9] Coal dominates the electric power generating field and plays an important role in industry, especially in smelting of ferrous metals; petroleum even more heavily dominates the transportation area. These two fuels also cause the most serious residuals problems, as will be seen later.

Table 3.—*Sources of energy used by various sectors*

[In present]

	Utility	Transportation	Industry	Households and commercial
Coal	55	----	29.4	0.4
Petroleum	6	99.85	22.8	38.8
Natural gas	22	----	39.3	42.8
Utility electricity	(1)	0.15	8.5	14.4
Other	17	----	----	3.6
Total	100	100	100.0	100.0

[1] Not applicable.

[9] Derived from Vogely and Morrison, *op. cit.*

To discuss the pattern of utilization of mineral fuels in the U.S. economy, it is necessary to analyze them in terms of their primary constituents—carbon, hydrogen, and sulfur—as indicated in table 4. We can then follow these elements through a sequence of chemical transmutations without being unduly concerned with the exact physical form or chemical combination at each stage.

TABLE 4.—*Weight breakdown of various fuels used for energy production*

Fuels used	Percent Carbon [1] (ex. CO_2)	Percent Hydrogen [1]	Percent Sulfur [2]	Percent other (Ash, H_2O, CO_2, etc.)
Anthracite coal, 16.7 percent ash	76.0	2.6	0.6-0.8	~20.0
Bituminous coal, 7.5 percent ash (after washing at mine)	75.0	5.0	2.0	18.0
Electirc utilities coal 10 percent ash (after washing at mine)	75.0	4.5	2.5	18.0
Natural gas, unprocessed, 93 percent CH_4+4 percent C_2H_6	75.0	24.0	0	1.0
Gasoline (motor fuel), assumed C_8H_{18}	84.0	16.0	0	<.5
Distillate fuel oil (grades 1 to 4)	[3] 85.0	[3] 14.5	----------	.5
Residual fuel oil (grades 5 to 6)	85.7	11.5	2.0	.8
Residual for utilities	85.4	11.4	2.5	.7

[1] Figures on anthracite and bituminous coal from "Combustion Engineering," table 13-2, 1966 edition Figures on utility coal, residual oil, and natural gas from "Costs of Large Fossil Fuel Fired Power Plants,' Jackson and Moreland, J. & M. 636 (Apr. 30, 1966).
[2] These figures from "Sulfur in U.S. Coals," *Coal Age*, December 1955, pp. 78-79 and from "Coal Utilization and Atomospheric Pollution," by E. A. Fohrman, J. H. Ludwig, and B. J. Steigerwald, *Coal*, April. 1965, pp. 5, 6.
[3] Estimated.

THERMAL POWER [10]

As noted above, over half the utility electric power produced in 1965 used coal as the primary energy source. In 1965 about 251×10^6 tons out of a total U.S. domestic supply of 465×10^6 tons were used for this purpose.[11] Most of the remaining electric energy was produced in the West and Southwest with hydropower and natural gas. Combustion of natural gas produces comparatively small amounts of potentially harmful residuals.[12]

Some 16.5×10^6 tons of (high sulfur) residual oil were also burned by electric powerplants, mostly on the east coast. A variety of residuals which can result in external costs are associated with the use of coal. Among them are acid and turbid waste waters from coal cleaning at the mine, losses as soot or dust during transport, flyash (fine inorganic particulates in flue gases), gaseous stack emissions (primarily CO_2, sulfur oxides, and oxides of nitrogen or NO_x)—and, if flue gases are scrubbed, liquid and solid residuals result from this process. Indeed coal residual oil burning thermal plants contribute major fractions of the total amounts of sulfur dioxide (~50 percent), NO_x (~53 percent) and particulates (~25 percent) emitted to the atmosphere in the United States. In the case of particulates this remains true despite the fact that major control efforts have already been introduced. A summary of the (nonnuclear) residuals from thermal power production is shown in table 5.

[10] Much of the information for this section was obtained from an unpublished research report by Richard J. Frankel of RFF.
[11] About 100×10^6 tons each were used by the steel industry and other industries.
[12] Use of natural gas does cause some oxides of nitrogen to be produced as secondary products. During combustion, oxygen and nitrogen gas combine to form nitric oxide (NO), which will react with more oxygen to form nitrogen dioxide (NO_2) and other nitrogen oxides (generally written NOx) which contribute to Los Angeles type smog. It has recently been discovered, apparently by accident, that a two-stage combustion process can greatly reduce NOx from natural gas and oil fired plants but this technology is not applicable to coal fired plants.

TABLE 5.—*Thermal power combustion residuals, 1965*

Fuels used	Weight breakdown [1]				
	Total weight × 10^6 ton	Carbon × 10^6 ton	Hydrogen × 10^6 ton	Sulfur × 10^6 ton	Ash × 10^6 ton
Anthracite coal	2.0				
Bituminous coal	249.0				
Total, coal	251.0	168.0	11.3	6.4	25
Natural gas	51.0	38.0	12.5		
Residual oil [2]	16.5	14.2	1.9	.4	
Total	318.5	220.2	25.7	6.8	25

[1] See table 3.

[2] 16.5×10^6 ton = 101×10^6 bbl.

Note.—Residuals ($\times 10^6$ ton): CO_2, 807;
H_2O, 231;
SO_2, 13.7 [a];
Ash (total), 25;
Fly ash not collected, 2.4 [b];
NO_X, 3.7 [c].

[a] The Public Health Service estimates 12×10^6 tons of SO_2 emitted into the atmosphere by electric utilities, which might suggest that 1.7×10^6 tons is being removed from stack gases at present; however, it appears that SO_2 removal at present is actually much less than this.

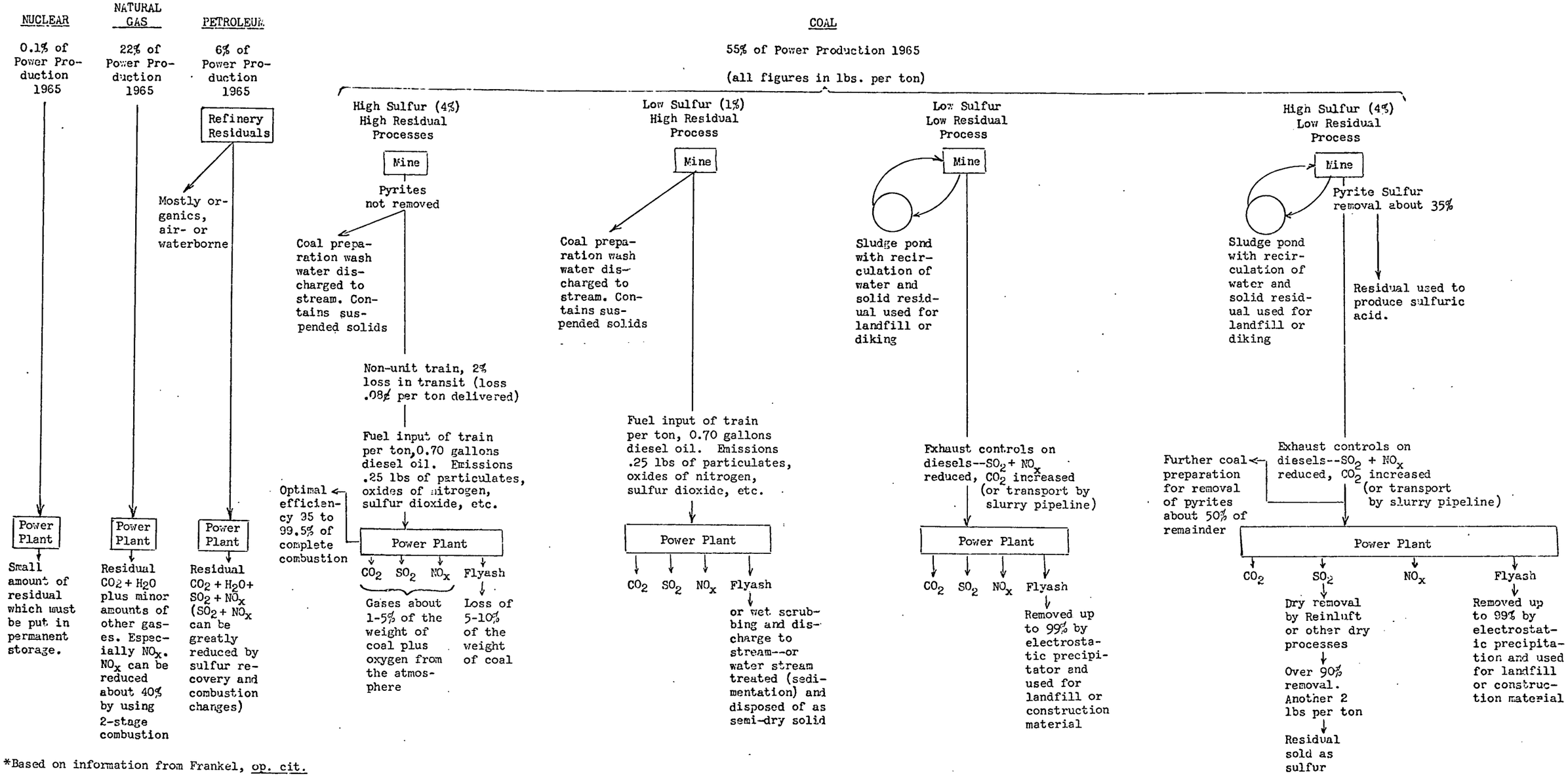
[b] Data from Frankel. Based on an industrywide collection efficiency of 86.5 percent for fly ash in the stack, and allowing about 20 percent for bottom ash.

[c] Assuming 28.4 lb. NO_X produced per ton of oil burned (Los Angeles). If carbon is the controlling factor (as seems to be the case), then we would expect 33.3 lb. NO_X per ton of contained carbon or 3.7×10^6 tons NO_X on a nationwide basis.

Chart IV shows several alternative process sequences giving rise to varying amounts of residuals. It focuses primarily on coal because of its quantitative importance in thermal power production and because of the availability of information. A high-sulfur coal burning operation with little or no control or recycling of residual materials back into the productive system can give rise to high levels of residual at various stages and potentially large external costs. On the other hand, we also give a schematic representation of sequence which uses high sulfur coal but produces very little final waste. Other process sequences might be sketched but the one shown seems a reasonable one for achieving low residuals, based on presently available or projected near-term technology. No control and reuse system as complete as this is, however, presently employed at any U.S. thermal powerplant.

In this process sequence water used for coal preparation at the mine is ponded and the resulting sludge is used for landfill. Next, pyrites (metallic sulfide crystals) are removed from the coal by mechanical means at the mine. (The pyrite can be used to produce commercial sulfuric acid on an economic basis.) The coal is then transported by unit train or slurry pipeline, thus reducing losses in transportation. After combustion, the CO_2 is discharged harmlessly (at least in the short run) into the ambient air. Up to 90 percent of the sulfur dioxide is removed from the stack gases by a dry removal process and the recovered sulfur is also salable. Flyash is removed up to 99 percent by electrostatic precipitators and used for landfill or construction materials (light aggregate). The overall amount of nonrecycled waste from such a high level residuals control process (even giving special weight to the potentially most harmful ones) is perhaps less than 10 percent of that resulting from the less strictly controlled process. Moreover,

CHART IV. Residuals from the Thermal Electric Industry*



*Based on information from Frankel, op. cit.

88-744 O - 68 - vol. 2 (Face p. 638)

since virtually all flyash and sulfur are productively reused, potential losses in other production processes are short circuited (for example, less elementary sulfur and pumice need be mined).

However, the dry SO_2 removal process does not remove NO_x from the gas stream nor (as was mentioned in footnote 12, p. 637) is two-stage combustion applicable to convention coal-fired furnaces. Wet scrubbing will remove some NO_x as well as SO_2 but the result is a waterborne waste stream which is difficult to treat and from which it is difficult to recover materials economically for recycle. The interdependencies and tradeoffs between various waste streams which must be considered thus are clearly illustrated.

It is not contended that the external costs avoided by high-level residuals control would always outweigh the internal costs of exercising the control. However, for a new 300 megawatt or larger plant (which almost all new plants are), additional capital costs would not seem to be prohibitive. It does not appear that net costs per kilowatt-hour would be increased more than a few percent, assuming a continued active demand for sulfur.

An obvious substitute for emissions control for "dirty" fuels is to use a "clean" fuel like natural gas or to convert to atomic energy. A growing trend toward using atomic energy is now clearly evident. The amount of residuals produced by nuclear fuels is quantitatively very small compared to fossil fuels since energy released per unit of material throughput is vastly larger. There are no significant discharges of residuals to the atmosphere from nuclear plants but some of the liquid residuals (small amounts) are so radioactive that they must be put in permanent storage. The full ramifications of this apparently have not been fully explored.

Despite the rapid growth of atomic energy, the use of fossil fuels for power generation will continue to increase until at least 1975 or 1980. During this period the use of natural gas is also likely to increase. Pipeline capacity could be augmented with a view to shifting gas gradually to household and commercial users as those demands expand and electric power generation shifts to atomic energy.[13]

So far we have focused on overall emissions reduction but it may also be possible to use systems for making better use of natural assimilative capacity. For example, high stacks (very cheap) help disperse gases and could in large systems perhaps be combined with systematic load shifting from plant to plant to take account of atmospheric circumstances.[14] These kinds of alternatives are discussed further in the final section where we address the matter of regional environmental management systems.

TRANSPORTATION

In 1965 the transportation sector accounted for 24 percent of all primary energy produced in the United States, and all but an insignificant percentage was accounted for by petroleum products: in 1965 the sector consumed 308×10^6 tons as fuel,[15] plus substantial quantities of additives, notably 2×10^5 tons of lead.

---

[13] It is also possible that shifts could be made to low-sulfur coal but because of its limited supply and high value to the steel industry it is doubtful that much of this will be available to the eastern utilities which now use high-sulfur coal.

[14] The possibility of load dispatching to take account of differential atmospheric conditions is discussed in Jerome K. Delson "Choices for Electrical Utilities in the Control of Air Pollution" mimeo July 18, 1967.

[15] Vogely and Morrison, *op. cit.*

In the United States transportation is overwhelmingly automotive. In 1965, 82 percent of all workers used private automobiles *en route* to work and a majority of the rest traveled on buses. Intercity travel is even more heavily dominated by motor vehicles: 89.4 percent of intercity passenger miles were accounted for by private cars and a further 2.5 percent by motor coaches. The movement of freight is still largely by railroads, but trucks were the second most important mode, with 24.4 percent of all freight ton-miles in 1965—and a virtual monopoly of local distribution of goods. Domestic airlines accounted for just under 6 percent of passenger miles and 0.1 percent of freight ton-miles.[16]

The more than 300 million tons of petroleum fuels and lubricants being consumed annually for transportation purposes is mainly gasoline burned by highway vehicles. In 1954 about 97 percent of all fuel used in transportation purposes (except on ships) was gasoline and the remainder was mainly diesel oil, split between large trucks and buses and railroads. In the same year aviation accounted for about 2.5 percent of all gasoline (in the prejet period). Even allowing for a factor of three relative increase [17] in the proportion of fuel allocated to air transportation, and assuming a roughly constant percentage of diesel, it is evident that the vast bulk (90 percent) of all petroleum fuels are gasoline consumed by automobiles and trucks. Hence, it seems reasonable to focus on the emissions picture for internal combustion engine vehicles.

Unfortunately, the data available are almost entirely presented in volumetric terms,[18] either per hour or per mile of operation, in a specified mode (i.e., acceleration, deceleration, idle, or cruise). Theoretically, from this one can deduce the total emissions produced by an average vehicle over any given driving cycle, which is a specified mix of the four modes. In practice, the modes are so vaguely defined that a synthetic driving cycle such as the well-known "LA–4 cycle" [19] need not very accurately reflect the total residuals actually produced per pound of fuel burned over a real driving cycle. Apparently this crucial coefficient has never been measured directly. As far as we can determine, national estimates were obtained by estimating emissions per mile, using a synthetic cycle, then multiplying by vehicle-miles traveled nationwide. Instead, we have inferred emissions per pound of fuel burned, from total emissions estimated for Los Angeles County [20] (because data for Los Angeles are far more complete than for any other place). The results shown in table 6 are substantially confirmed by similar survey data for New York.[21]

---

[16] From "Automobile Facts and Figures," 1966 ed., published by the Automobile Manufacturers Association.

[17] In 1955 airlines accounted for 3.42 percent of domestic intercity passenger-miles. By 1965 this had risen to 5.96 percent.

[18] Perhaps because emission standards have hitherto been expressed by volume, i.e., in percent or parts per million (ppm) of the exhaust. It is our belief that standards should be revised and stated in terms of total emissions per pound of fuel consumed.

[19] Sweeney and Pattison, SAE paper 660546, August 8–11, 1966.

[20] R. G. Lunche *et al.*, "Air Pollution Engineering in Los Angeles County," hearings before the Subcommittee on Science, Research and Development, of the Committee on Science and Astronautics, House of Representatives, July 1966.

[21] "Air Pollution in N.Y.C.," Council of the City of New York, M–970. An interim technical report of the Special Committee To Investigate Air Pollution, Jan. 22, 1965.

TABLE 6.—*Major emissions per pound of fuel consumed*

[Pounds/pound]

	CO	HC	NO_x	SO_2	Lead [1]
Otto engine:					
Los Angeles data	0.500	0.093	0.024	0.00014	0.00075
New York data	.472	.104	.018	.00014	
Diesel engine:					
Los Angeles data	.0045	.018	.018	.0045	.0045
New York data	.0067	.020	.025	.0045	.0125
Gas turbine [2] (Rover 25-140)	.02		.0003	.002	
	(340 p.p.m.)	([3])	(5.5 p.p.m)	(60 p.p.m.)	([3])
External combustion: [4]					
Steam (Williams)	.0075	.00006	.001	.0045	
	(500 p.p.m.)	(20 p.p.m.)	(70 p.p.m.)		([3])
Sterling (Philips)	.0015	([4])	.002	.0045	
	(100 p.p.m.)	(1.5 p.p.m.)	(150 p.p.m.)		([3])

[1] Figures on particulates are clearly greater than figures on lead, since most tetraethyl lead is emitted in the form of particulates.

[2] Calculated assuming 63 lbs of air/pound of fuel, volumetric emissions in parts per million as shown in parentheses and molecular weights as follows: CO=28 (=air), NO_x=36 (=1.3×air), and SO_2=64 (=2×air).

[3] Negligible.

[4] Calculated by assuming volumetric emissions shown in parentheses in parts per million, 15 lb of air/pound of fuel and molecular weights as follows: CO=28 (≅air), HC=140 (≅5×air), and NO_x=36 (≅1.3×air). Sulfur emissions assumed to be the same as Diesel.

Using the calculated coefficients in table 6 and the total gasoline consumption in the transportation sector we obtain the results shown in table 7. (No allowance is made for aircraft or diesel engines.)

TABLE 7.—*Residuals from automotive transportation, 1965*

[$\times 10^3$ tons]

	This study	PHS
Carbon monoxide (CO)	135.0	66
Unburned hydrocarbons (HC)	27.0	12
Oxides of nitrogen (NO_x)	5.7	6
Lead compounds	.2	
Oxides of sulfur	.4	1
Total	168.3	85

It will be noted that our estimates of carbon monoxide production and unburned hydrocarbons are roughly double the official PHS estimates. On the other hand, our estimates of NO_x agree closely with the PHS. Apart from what has been said previously about the derivation of the numbers, we are not able to explain the discrepancies.

Potential control techniques have been discussed at great length in recent years. Basically these techniques consist of:

(1) Recycling the blow-by gases which formerly escaped from the cylinders past the rings and were formerly released from a vent under the engine (this change has been made since 1963 on all U.S. cars; it eliminates about 20 percent of the unburned hydrocarbons and carbon monoxide).

(2) Improving the efficiency of the combustion process by improved carburetion (or fuel injection). This is a matter of more accurately matching the fuel and air intake to the instantaneous demands of the engine, and particularly preventing excessively rich mixtures. Evaporation losses from the carburetor can also be considerably reduced by simple redesign.

(3) Unburned components in exhaust gases can be more completely consumed either in the manifold or the tailpipe, by some type of afterburner. The simplest method may be to introduce excess air into the hot manifold, where unburned gases will have an additional opportunity to burn completely. Catalytic mufflers and other add-on devices can also reduce exhaust emissions, although catalysts tend to be poisoned by the lead in the combustion products. Hence the latter approach is not much in favor at present.

(4) Evaporation losses from the gas tank can presumably be reduced by various means, such as an absorbent charcoal buffer.

It seems likely that by a combination of these means about 90 percent of the emissions from a typical gasoline internal combustion engine may ultimately be eliminated at an acceptable cost (less than $100 or so). Engines may even achieve somewhat better results than this, while they are comparatively new, but the level of control tends to be degraded with age and wear. The problem of keeping emissions low throughout the life of the car, under conditions of little or poor maintenance, remains a very difficult one. Meanwhile, if emissions control is effective primarily during the first half of a vehicle's life, the net gain would ultimately be something like 50 percent. However, an expected rapid increase in the number of cars on the roads would nullify most of this gain. Indeed, present approaches are unlikely to achieve much, if any, reduction in present air pollution levels. To "roll back the clock" more than a few years, more radical approaches would seem to be required, such as:

(5) Greater emphasis on mass transportation and rail—rather than highway—transportation of goods. As an example of what can be achieved by switching automotive traffic to rails, each 100 million passenger miles on electrified mass transit results in saving 16,000 tons of gasoline, which means a net decrease of 8,000 tons of carbon monoxide, 1,600 tons of vaporized hydrocarbons, and 320 tons of oxides of nitrogen. Assuming the power is produced in thermal plants, there will be a compensating increase of 350 to 500 tons in oxides of sulfur (assuming fuel with a 2.5-percent sulfur content). This shift might be accomplished in part by public subsidy of conventional or unconventional mass transit systems, tax or other disincentives to drive or park private cars in areas where alternatives exist, and construction of convenient (automated?) parking facilities near transit termini. Automated comprehensive underground goods distribution systems utilizing electric minirails, pneumatic tubes and/or moving belts may find a place in densely populated areas.

(6) A (partial?) switch to alternative power sources for vehicles, including gas turbines or turboelectric hybrids for large buses and articulated trucks, and external combustion (e.g., steam) engines or battery fuel cell electric propulsion for automobiles and taxis. Remote (i.e., wired or radiated) external power sources in a dual mode (e.g., road-rail) configuration are also conceivable possibilities for private vehicles in the longer time frame.[22]

---

[22] See a forthcoming RFF sponsored study by Robert U. Ayres on technological alternatives for urban transportation. Also the report of the Panel on Electric Cars of the Commerce Advisory Committee.

In general these latter technological alternatives can render transportation services with less throughput of materials and fuel than conventional motor vehicles and/or involve combustion of fuels in central power plants where it is more efficient and residuals can be more efficiently treated and disposed of. They can be encouraged by public (e.g., military) sponsorship of research, public investment, tax incentives such as faster writeoffs, effluent taxes on pollutants, enforcement of air quality standards, and by directed purchase of appropriate vehicles for the use of Government agencies. Thus the Post Office alone would provide a sizable market for an electric truck, and city, State and Federal agencies could provide an initial market for a steam engine or battery fuel cell (electric) vehicle. Once the economies offered by such vehicles are proven fleet owners such as rent-a-car concerns, taxi companies and finally private individuals might be induced to switch.

The externalities caused by electric power production and by automotive transportation make an interesting contrast. Electric power producers are acknowledged to be *utilities* subject to control and supervision by public agencies; hence the regulation route mentioned earlier is quite likely to be successful in the long run as a means of enforcing due consideration of the external costs of operation.

On the other hand, although railroads, airlines, and interstate trucking companies are regulated as utilities, the private automobile, which accounts for the great bulk of all transportation, is treated not as a utility but as a convenience. Neither the manufacturers nor the owners have, until very recently, been subject to any systematic regulation representing the public interest. Scattered State and Federal legislation now exists covering such topics as auto safety, compulsory insurance, and air pollution; but these functions are not unified under a single agency, nor are they likely to be in the immediate future. This fragmentation of governmental authority, coupled with the enormous economic power of the automotive industry, make a successful attack on the problem of automobile-caused environmental pollution more difficult.

INDUSTRY AND HOUSEHOLDS

As noted previously, coal is an important source of energy in industry, particularly in the metallurgical field and in the manufacture of lime and portland cement. Thus in 1965 the industrial sector consumed 188×10^6 tons of coal, of which 96×10^6 tons were first carbonized to yield 77×10^6 tons of coke plus coal-gas and about 4×10^6 tons of coal tar derivatives such as benzene, toluene, xyline, naphthalene, and creosote as byproducts mostly used as raw materials for organic chemicals. The best quality low-sulfur ($\sim$1 percent) bituminous coals normally are used for coking.[23] Thus the average sulfur content of all industrial coal (2 percent) is less than the average for utility coal (2.5 percent), as shown in table 4. This implies a total SO_2 emission from coal burned in industry of 7.4×10^6 tons in 1965.[24] Total particulates (flyash) produced by industrial and coking coal are

[23] These low-sulfur coals, mainly from western Pennsylvania and West Virginia, sell for much higher prices than utility coal, and supplies are largely preempted by long-term contracts.

[24] The Public Health Service estimated 9×10^6 tons of SO_2, which presumably includes SO_2 from other sources such as sulfide ore treatment.

14×10^6 tons, based on the average 7.5 percent ash content shown in table 4. Assuming an average collection efficiency of 62 percent implies residual particulate emissions of the order of 5.3×10^6 tons.[25] The Public Health Service estimates a total of 6×10^6 tons of particulate matter from all industrial operations.[26] Although we have no quantitative basis of verification, the same source also estimates that 2×10^6 tons of NO_x and 2×10^6 tons of carbon monoxide are produced annually by industry in the United States. Further discussion of energy conversion residuals production and controls in materials processing and manufacturing operations are best reserved for the next section where the industrial sector is analyzed explicity in more detail.

Space heating in industry and households is the other major demand for energy. The fuels used in this application are mainly distillate oils ($\sim 177 \times 10^6$ tons) and natural gas (286×10^6 tons), both of which are comparatively clean. About 26×10^6 tons of coal are still used in the household sector, although this is decreasing rapidly. The latter contributes on the order of 1×10^6 tons of SO_2 and—because of the virtually complete lack of flyash collection in small heating plants—probably 1.7×10^6 tons of particulates. The ash content (0.5 percent) of fuel oil probably contributes a further 1×10^6 tons. Fuel oil burned in industry and for space heating also contains a small amount of sulfur—perhaps 0.25 percent—which would result in about 1×10^6 tons of SO_2. (This may be somewhat underestimated.)

A figure for NO_x can be extrapolated from the coefficient used in table 5; namely, 33.3 pounds per ton of contained carbon, or 1.37 percent by weight. The industry and household category altogether consumed about 522×10^6 tons of contained carbon in fuel (1965), which would imply a production of 7×10^6 tons of NO_x. Information on carbon monoxides and unburned hydrocarbons during combustion is hard to find, but one may probably assume coefficients for distillate oil in the industry-household sector similar to those for external combustion engines (table 6), viz, 0.75 percent for CO and 0.006 percent for HC, by weight. Applying these coefficients indiscriminately to the total quantities of fuel burned for which no other data is available gives the numbers shown in table 8, although this is a dangerous extrapolation; in particular, it may be too high for CO and too low for HC, at least where natural gas is the fuel.

TABLE 8.—*Summary of residuals from energy conversion*

[$\times 10^6$tons]

	Carbon monoxide (CO)	Hydrocarbons (HC)	Sulfur dioxide (SO_2)	Oxides of nitrogen (NOx)	Particulates
Utility power	[1] 2.5	[2]	13.6	3.7	2.4
Transportation	135.0	27	.4	5.7	>.2
Industry and households	5.0	[2]	8.4	7.0	7.0
Total	142.5	27	22.4	16.4	9.6

[1] This is probably too high; the PHS estimates 1, which is more reasonable.
[2] Negligible.

[25] Frankel, RFF, unpublished research report.
[26] This is certainly too low if additional contributions from mineral ore beneficiation (particularly copper) and phosphatic fertilizer plants are considered.

III. Residuals From Materials Processing and Industrial Production

Describing materials flow and residuals production in manufacturing industries is particularly difficult. There are several reasons. First there is a notable lack of direct information about quantities and qualities of industrial inputs, processes, and waste residuals which is both comprehensive and dependable.[27] Second, industry generates an immense variety of residual products including most of those ultimately discarded by households, plus numerous others. Finally, while household, thermal power, and transportation activities are reasonably comparable across the country, the industrial mix of metropolitan areas varies drastically from place to place. For these reasons it is difficult to generalize meaningfully about industrial residuals (although some broadly applicable statements are possible). Thus at this point in time the following discussion is necessarily more illustrative than definitive. In many ways, materials balance concepts are more clearly applicable to industrial processes than to those in other economic sectors. But their comprehensive application must, like much other relevant analysis, await greatly improved data. At the end of this chapter we show, as an illustration, a detailed computed materials balance for one type of industrial plant under various conditions. This analysis yields some rather interesting conclusions.

For purposes of discussion it is convenient to subdivide the large and cumbersome "industrial" sector into several smaller subsectors, based on natural groupings among the materials involved in processing.

(1) One natural subcategory belongs to the products of photosynthesis, which comprises food processing, forest products, and organic chemicals (including petrochemicals and coal tar derivatives). The major classes of compounds involved are carbohydrates, fats and oils, proteins and hydrocarbons, which are in turn made up primarily of carbon, hydrogen, oxygen, and nitrogen.

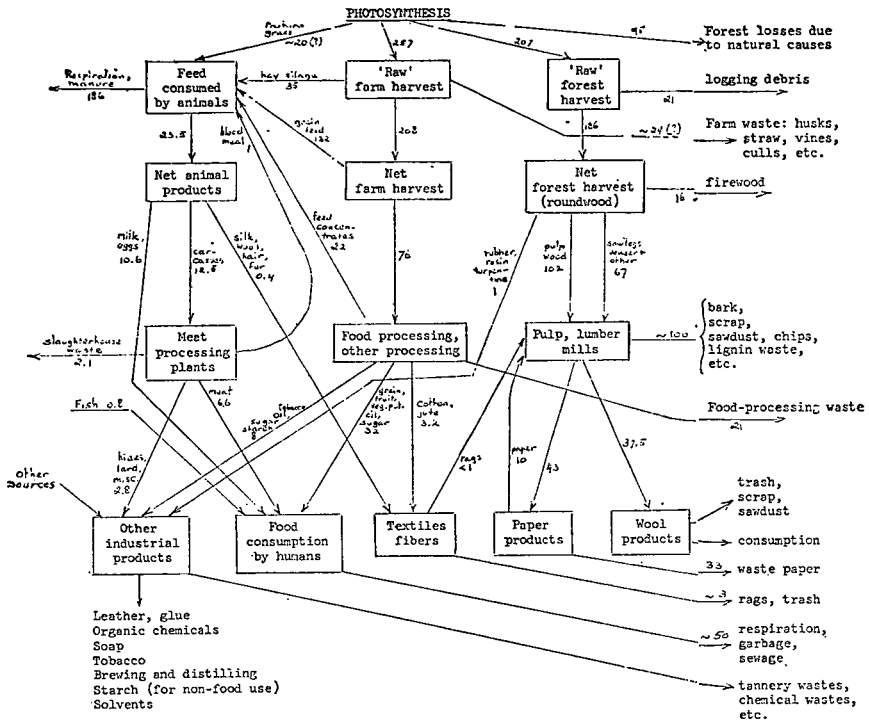
(2) A second large and slightly overlapping group comprises the inorganic chemicals, which are based mainly on alkali metals (sodium, potassium, calcium, barium), halogens (fluorine, chlorine, bromine), boron, sulfur, phosphorous, plus the four basic constituents of organic chemicals.

(3) A third group comprises the ferrous metals (iron, nickel, and their alloys), aluminum, magnesium, copper, zinc, and lead.

(4) The last group includes the relatively inert minerals such as sand (quartz), gravel, stone, pumice, feldspar, clay, gypsum, and so forth. The latter are used primarily in building materials, insulation, abrasives, cement, refractories, ceramics, and glass.

[27] For instance, only now is the first inventory of the content of industrial waste water discharges being assembled on a national basis (FWPCA study). Existing estimates are both old and somewhat ambiguous as to meaning. See, for example, William Rudolfs (ed.), *Industrial Wastes—Their Disposal and Treatment*, L.E.C., Inc., 1953.

Chart V.--Production and Disposal of Products of Photosynthesis[1/]



1/ Disregarding fossil fuels, which are treated separately.

Note: All figures refer to dry organic matter x 10^6 tons.

The materials balance approach can be applied with particular effectiveness in the analysis of organic wastes from the processing of food and forest products. A simplified "flow chart" for these activities is shown as chart V. From the detailed statistics compiled by the U.S. Department of Agriculture it can be determined that the total fresh weight of crops harvested for human consumption in the United States (excluding exports and industrial uses) is roughly 125×10^6 tons. This corresponds to about 59×10^6 tons of dry organic matter. An additional 210×10^6 tons of vegetable matter (dry weight) is harvested by and for livestock.[28] Wood and forest products harvested for consumption in the United States account for a further 207×10^6 tons (dry O.M.).

From the 210×10^6 tons (dry weight) of animal feed consumed annually in the United States, the net useful product consists of the items shown in table 9, as follows:

[28] Including an estimated 20×10^6 tons of pasturage and 35×10^6 tons of hay and silage. These figures are necessarily very rough.

TABLE 9.— *Materials balance for livestock*

	Dry weight ($\times 10^6$ *tons*)
Inputs feed and forage	210.0
Output:	
Milk	9.6
Eggs	1.0
Meat and poultry (edible)[1]	6.6
Hides, lard, and other animal products	2.8
Wool, hair, and fur	0.4
Animal feed (recycled)	1.0
Bone meal and slaughterhouse waste	2.1
Manure and respiration losses (approximate)	186.5

[1] "Edible portions" include some bone, skin, and fat which is later discarded in home preparation and appears as garbage.

Prior to the concentration of animal feeding activities in commercial feedlots and poultry farms near major cities, animal wastes (manure) were automatically recycled, thereby satisfying a major part of the nitrogen and phosphorous requirements of agriculture. Today, however, large quantities of manure are produced and accumulated where there is no market for the material, and as a consequence disposal has become an increasingly serious problem. Indeed the economies, which have hitherto justified feedlots, might be considerably diminished if the cost structure were broadened to include the disposal of the wastes by returning them to the soil in some form, or (alternatively) to take into account the depreciation of the soil which occurs when the manure is not returned.

Apart from the major (tonnage) residuals arising essentially from inefficiencies in the conversion of plant to animal calories, there are substantial residuals at the slaughtering stage. Although large meatpackers have stated that they utilize every part of the animal "but the squeal of the pig," this is an exaggeration. The edible portion of the animal is about 63 percent for hogs, 59 percent for beef cattle and 48 percent for sheep, or about 60 percent overall. An additional 15 percent consists of other useful or utilizable products such as hides, tallow, glue, bloodmeal, etc. The remainder, about 25 percent of the total, is essentially waste. Major meatpackers have achieved higher rates of economic recovery by incorporating some or most of this material in animal feeds and fertilizers; smaller local slaughterhouses cannot attain this degree of efficiency. Statistics in this area are difficult to find, but on a national basis probably at least 10 percent of the live weight is not recovered and is discharged to the environment. The utilization of hides to manufacture leather products is also a major source of unrecycled residuals. Tannery wastes can be particularly obnoxious sources of water pollution. The technology exists to reduce tannery wastes now discharged in water courses by using organic solvents and enzymes to remove unwanted material such as hair and fat from the hides, thus producing a much more compact residue than is now the case. This is not done in general because the high capital investment required does not seem justified in a static or declining industry. Moreover it would to some extent trade a water pollution problem for an air pollution problem.

The processing of food from plant origins for direct human consumption involves a smaller absolute quantity of organic waste. Table 10 shows the figures for harvested crops, excluding fibers, tobacco, crops destined for animals and exports; it also shows the quantities

of processed food available "for consumption" (prior to home preparation and cooking). The differences represent statistical inaccuracies, plus various processing losses including spoilage due to vermin. In some cases, such as sugar beets, most of the weight of residuals is recovered (for cattle feed, for instance) but the residuals discharged to the environment may be capable of exerting a highly degrading effect. Residual wastes from the processing of food products tend to be high in organics which reduce dissolved oxygen in rivers, for instance.

TABLE 10.—*Materials balance for processing foods of vegetable origin* [1]

[$\times 10^6$ tons]

	Input		Output			
	Net crops harvested for human consumption (in the United States)		Food available for consumption		Processing losses [2] and waste	
	Fresh weight	Dry weight	Fresh weight	Dry weight	Fresh weight	Dry weight
Grain	21.4	18.9	13.4	11.8	8.0	7.1
Potatoes and root crops	12.1	2.6	10.4	2.5	1.7	.1
Beans and peanuts	2.2	2.0	1.6	1.4	.6	.6
Green vegetables	21.0	2.9	19.2	1.5	1.8	1.4
Fruits and nuts	16.7	2.4	14.1	1.8	2.6	.6
Sugar (cane and beet)	49.3	27.7	10.2	10.2	39.1	17.5
Oil [3]	2.8	2.8	2.8	2.8		
Total	125.5	59.3	71.7	32.0	53.8	27.3

[1] *Agricultural Statistics*, 1966.
[2] A major part of these are recovered as byproducts.
[3] Vegetable oil is derived mainly from soybeans, cottonseed, and flaxseed; the pulp and meal is used as animal feed concentrates.

Although the demand for forest products, particularly pulp and paper, is rapidly increasing, the various processing stages are still comparatively inefficient on the average, resulting in enormous tonnages of unrecovered organic wastes. Harvesting efficiency is only about 50 percent, to begin with, whence half of the available organic matter is left behind in the forests as stumps, branches, and brush. Much of this waste material is ultimately burned.

From "roundwood" to final wood and paper products, there are also substantial wastages *en route*. In papermaking, for example, which absorded 102×10^6 tons of pulpwood in 1965 (dry O.M.), about 32 percent of the input organic material is utilized as pulp; the remainder is waste. In a large modern integrated pulp and paper mill some of this can be recovered and used as a raw material for organic chemicals—notably surfactants (detergents)—or as a growth medium for yeasts or bacteria which can, in turn, be used for animal feed. However, this degree of recovery is still relatively rare in the industry, and only about 15 percent of the lignin waste liquors are recovered.[29] Accompanying the 1965 production of 33.3×10^6 tons of paper pulp (dry weight) was a residual bark, cellulose, and lignin liquor waste of 69×10^6 tons (dry O.M.), plus a large number of other chemicals used in the processing. This clearly puts a heavy demand on the assimilative capacity of the rivers and streams which are the major means of disposal at present.

[29] See C. F. Gurnham, *Industrial Waste Water Control*, Academic Press, New York, 1965.

The reuse of wastepaper recommends itself as an obvious means of resource conservation and, of course, a method of reducing the residuals problem arising from production of new pulp. Indeed, the production of paper products in 1965 was 43.7×10^6 tons which means that about 25 percent of the total production is recycled. However as a means of reducing discharges to the environment, this approach has limitations since much of the potential supply of wastepaper (e.g., newsprint) must be reprocessed to remove fillers, coatings, and ink. Whereas most original pulps produce an effluent load (excluding bark) of 20 to 200 pounds of suspended solids per ton of product—depending on the amount of bleaching required—deinked wastepaper pulp yields 500 to 800 pounds of suspended solids per ton (25 to 40 percent).[30] Thus increased recycling of finished paper might not be very desirable, from the water pollution standpoint, unless it were preceded by a substantial improvement in the technology for recovering waste lignin byproducts. It would, however, contribute substantially to the alleviation of the solid waste problem.

Sawmills, plywood and veneering plants do not contribute large amounts of processing wastes but there are substantial residuals in the form of bark, sawdust, chips, and scrap. Again, in large, efficient integrated plants most of this material can be utilized either in "pressed wood" or as chemical raw materials. In the smaller local sawmills it is mostly disposed of as agricultural mulch, or simply burned. At every stage in the utilization of wood products—notably in furniture manufacturing and construction—there is a substantial wastage, the amount tending to increase inversely as the size of the operation. Thus local residential home and commercial construction is a major source of waste; subsequent demolition also contributes substantial quantities. In areas of low population density the burning of scrap lumber from construction poses no particular problem, but in dense conurbations the disposal of bulky demolition wastes (a major constituent of the "trash" category of refuse) can be quite awkward.

The chemical industry—both inorganic and organic—subsumes a large number of products which are used, or used up, in the preparation of other goods. A rather useful distinction can be found here between chemicals whose actual substance will ultimately become part of a final product, in some form, and chemicals serving other intermediate functions which do not appear physically in the final product.

The public is now well enough aware of the fact that agricultural chemicals such as pesticides, fungicides, and herbicides only serve their purpose (as presently used) when they are degraded and/or dispersed. Surface active agents such as detergents have also attracted a good deal of attention. It should be obvious, although it seldom seems to be pointed out, that the same is true of some dyes and pigments, virtually all industrial solvents and "carriers," softeners, fluxes, flotation agents, bleaches, cleaning agents, antifreezes, lubricants, explosives, and so forth. Tables 11 and 12 list some of the major tonnage organic and inorganic chemicals which are dissipated or dispersed, either immediately or gradually (as in the case of paints) into air, water, or soil in the course of *normal* use.

Apart from the existence of important dissipative uses in the chemical sector, there are also inevitable losses in production and refining.

[30] Gurnham, *op. cit.*

(See table 13.) This applies also, of course, to food processing and metallic ore reduction. However, as regards chemicals it is noteworthy that there are often a number of intermediate stages *at each one of which there may be significant losses.* For instance, consider the chain leading to nylon, illustrated in chart VI.

Chart VI

Chemical Intermediates in Nylon Production

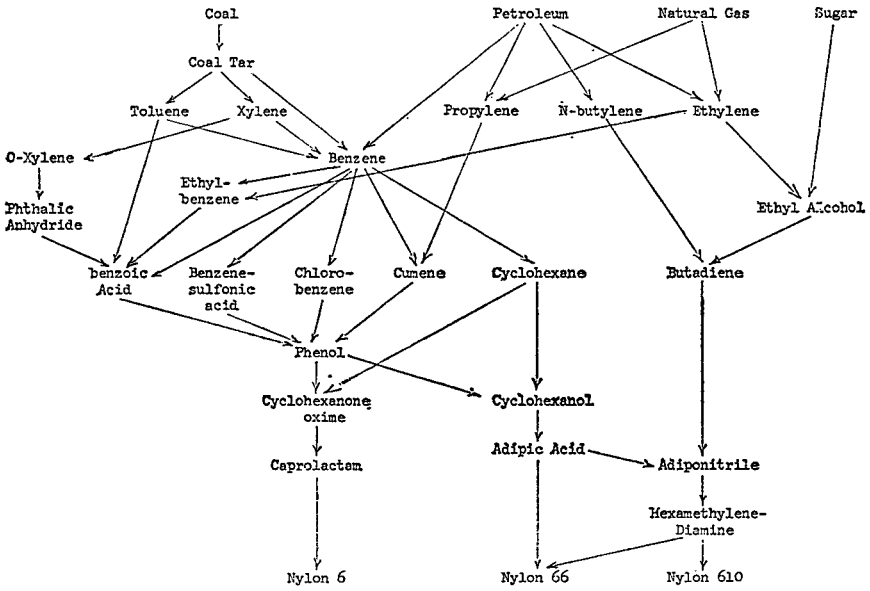


TABLE 11.—*Major dissipative uses of industrial organic chemicals, 1963*

	Approximate tonnage ($\times 10^6$)
Solvents and thinners[1] (excluding gasoline and naptha)	≤10
Antifreezes (ethylene glycol, alcohols, etc.)	~0.5
Pesticides and herbicides (DDT, 2–4–D, malathion, parathion, benzene hexachloride, etc.)	0.32
Surfactants[2]	1
Lubricating oil additives	0.2
Explosives (TNT, nitroglycerin, nitrocellulose, picric acid, ammonium nitrate, etc.)	?

[1] The tonnage of organic fluids used as solvents is very difficult to determine, since the major solvents (benzene, xylene, methyl-, ethyl-, or isopropyl alcohols, glycol ethers, acetone, methyl-ethyl-ketone, carbon disulfide, carbon tetrachloride, vinyl chloride, and various other chlorinated hydrocarbons) are also used for other purposes as well as intermediates. Total production of these chemicals including turpentine for *all* purposes in 1963 is about 10×10^6 tons. In addition, an unspecified amount of gasoline is used as a solvent (e.g., for dry cleaning).

[2] Surfactants are used extensively in the detergent and soap industries, along with fats and oils, sodium tripolyphosphate, sodium sulfate, and other inorganic chemical inputs. Most detergents and soaps are used in households, of course, rather than in industry. Total detergent production (1963) was just under 2×10^6 tons and soap production was 0.5×10^6 tons.

TABLE 12.—*Major dissipative uses of industrial inorganic chemicals*

	Approximate tonnage ($\times 10^6$)
Metal cleaning and pickling (H_2SO_4, HCl, etc.)	~1.25
Neutralization of excess acid, e.g., in paper manufacturing (CaO, $CaCO_3$, $NaCO_3$, NaOH)	≧1.5
Bleaches (H_2O_2, ClO_2, $CaCl_2ZnCl$)	~0.1
"Antichlors" (Sulfates, sulfites, NaOH)	>0.8
Water softening (CaO, etc.)	0.85
Deicing (roads and highways) (CaCl, NaCl)	~4.5
Paints and pigments (lead oxide, lead carbonate, titanium dioxide)	~0.8
Beet sugar refining (CaO)	0.6
Fluxing (fluorspar, cryolite, CaO)	~3
Insecticides and fungicides (lead arsenate, copper sulfate, etc.)	<0.1
Other (Fabric finishing, tanning, photography, aerosol propellants, etc.)	?

TABLE 13.—*Typical processing loss rates* [1]

Industry	Aerosols	Gases and vapors	Typical loss rates
Petroleum	Dust, mist	SO_2, H_2S, NH_3, CO; hydrocarbons, mercaptans.	0.25 to 1.5 percent by weight of material processed.
Chemical processes	Dust, mist, fume, and spray.	Process—dependent	0.5 to 2 percent by weight of material processed.
Pyrometallurgical and electrometallurgical processing.	Dust, fume	SO_2, CO; fluorides, organics.	0.5 to 2 percent by weight of material processed.
Mineral processing	----do	Process—dependent	1 to 3 percent by weight of material processed.
Food and feed processing	Dust, mist	Odorous materials	0.25 to 1 percent by weight of material processed.

[1] A. H. Rose, D. G. Stephan, and R. L. Stenburg, "Prevention and Control of Air Pollution by Process Changes or Equipment," in *Air Pollution* (New York: The Columbia University Press, 1961).

The cyclohexane route and the butadiene route involve the fewest intermediate stages (4) but a substantial fraction of all nylon is derived from phenol, which implies five or six intermediates. Nylon production in 1966 exceeded 0.5×10^6 tons.

A large number of other important chemical products are also derived via multistage processes, including synthetic rubbers, polyethylene, polypropylene, polyesters, PVC, polyurethane, insecticides, surfactants, etc. Needless to say, if 1 percent of the material processed is lost at each stage, the total wastage in producing a complex product may be 5 percent or so of the final weight. Of course the loss rate varies, depending on a number of factors, notably the mode of handling and the volatibility of the material. Thus, light petroleum and natural gas fractions are particularly subject to evaporative losses during refining and processing. If the overall loss rate to the atmosphere is only 1 percent in each case (which may be low), the total quantities of hydrocarbons dissipated would be of the order of 8×10^6 tons, although it is hard to see how this figure could be very closely checked.

In the inorganic sphere one of the largest tonnage operations is fertilizer production. We have not classed this as a dissipative use, although fertilizer consumption does involve dispersal, since presumably the active constituents constitute no hazard or pollution problem in the dispersed condition, except insofar as they are leached from the soil by surface water and thereby contribute to water pollution problems. Problems tend to arise, if at all, only during fertilizer processing, the most noteworthy example being the treatment of

phosphate rock to produce commercial "superphosphate." This involves breaking down an insoluble complex material with the empirical formula

$$Ca_5\ (PO_4)_3\ (Halogen, OH)$$

where the halogen is usually fluorine, by treating it with sulfuric acid. The result is a mixture of $CaSO_4$ and $CaH_4(PO_4)_2$, with a small residue ($\sim$3 percent) of fluorides or chlorides. If fluorides escape into the air they constitute a serious local air pollution problem, especially for commercial crops. This is currently a problem in Florida and Colorado.

Unlike the products of the chemical industries, the output of the mineral and metallurgical industries is almost entirely nondissipative. Residuals consist of bulky solids (e.g., slag), gaseous emissions associated with the energy conversion aspect of ore reduction—plus substantial quantities of particulates—and liquid wastes resulting from cleaning ("pickling") the metal during various phases of fabrication or treatment from ingot to final product, usually to remove oxide scales which form when the hot metal comes in contact with air. Sulfuric acid is the major pickling agent for steel, but hydrochloric, nitric, and hydrofluoric acid are also used.

Slag, which was once a serious disposal problem for steel mills, has become a valued byproduct used mainly for road ballast and as aggregate for concrete products. It consists of a mixture of impurities from the ore plus fluxing agents such as limestone (used mainly for smelting iron), cryolite, or fluorspar.

Substantial quantities of particulates are produced by foundries, both as "flue dust"—which is largely recovered by Cottrell precipitators and wet scrubbers, because of its high metal content—and as soot from coal burning.[31] The latter is a major cause of sulfur dioxide and particulate emissions. The Public Health Service estimates that 2×10^6 tons of carbon monoxide, 9×10^6 tons of sulfur dioxide, 9×10^6 tons of sulfur dioxide, 2×10^6 tons of NO_x and 6×10^6 tons of particulates are produced by industry as a whole; presumably the majority of these pollutants come from the energy-consuming ore reduction, refining, and other operations in the primary metals sector.

Waterborne residuals from the metallurgical industry are extremely difficult to recycle, because of the large bulk involved and the low price of the constituents. The spent pickle-liquor from steel mills consists primarily of ferrous sulfate ($FeSO_4$) plus some excess sulfuric acid. The latter can be neutralized with lime, but further treatment such as ponding or evaporation have not proved economically feasible to date. The quantity of such residuals can be deduced roughly from the amount of acid used for pickling. (See table 12.) About 0.75×10^6 tons of H_2SO_4 were used in the iron and steel industry in 1963; this would correspond to roughly 1.1×10^6 tons of $FeSO_4$, and a loss of about 0.35×10^6 tons of metallic iron. Similar processes are used in other metallurgical operations, particularly copper and brass mills. Other metallic sulfates discharged annually probably add an additional 0.75×10^6 tons, although the economics of recovery might be better in some cases. Apart from sulfate wastes, there are smaller but locally

[31] Blast furnaces usually use coke, but other metallurgical operations such as reheating furnaces and rolling mills normally utilize coal.

important quantities of other salts (chlorides, nitrates, and fluorides) and oily wastes from rolling operations.

Among the nonferrous metals (copper, lead, and zinc), ore benefaction causes additional problems; roasting of sulfide ores creates very concentrated sulfur dioxide fumes which are lethal to vegetation and have literally created sizeable deserts around copper smelters in Tennessee and Montana. However, it has recently become economically feasible to recover most of this sulfur in the form of sulfuric acid, and many of these blights have already been substantially reduced (although the ecological consequences have by no means disappeared as yet).

Although the fourth group of materials processing industries far overshadows the others in tonnages of annual throughput, it is not a major source of environmental pollution problems; such problems as do exist are mainly traceable to the energy conversion aspects of materials processing; for instance, 63×10^6 tons of portland cement were manufactured (1963) from some 102×10^9 tons of mineral ingredients—mainly limestone, dolomite, quartz, and alumina, fused together at very high temperatures in a kiln by burning coke. The difference in weight is primarily due to carbon dioxide driven off by the heat. Plaster is similarly made from raw gypsum (hydrated calcium sulfate) by heating it to drive off the excess water; approximately 11×10^6 tons of gypsum were thus converted in 1963. Bricks, tiles, refractories, ceramics, and glass also require intense heat for their preparation, and thereby contribute to the demand for fossil fuels.

POTENTIAL IMPROVEMENTS THROUGH TREATMENT AND PROCESS CHANGES

There is a wide range of treatment methods available for application to waterborne industrial effluents. Some common categories of treatment processes are as follows:

Screening*	Wet oxidation
Flocculation	Fermentation
Chemical coagulation*	Emulsion breaking
Flotation	Evaporation
Sedimentation*	Distillation
Centrifuging	Incineration*
Filtration	Biological filtration*
Stripping	Activated sludge*
Neutralization	Anaerobic digestion*
Chemical oxidation*	Stabilization lagoons*
Chemical reduction	Spray irrigation*

Processes to neutralize or remove gaseous or particulate emissions include:

Two-stage combustion	Wet scrubbing
Afterburners	Condensation
Catalytic oxidation	Absorption
Recycling	Adsorption
Electrostatic precipitation	Filtration

These treatment processes change the chemical composition of the residuals or convert them from one form to another. Clearly none of them *eliminates* the residual.

The processes marked with an asterisk are also applicable to, and frequently used on, waterborne household effluents. Degradable organic wastes are treated in the same way whether of industrial or household origin although the treatment of industrial wastes can often be improved by combining them with household wastes for "fertilization" purposes. Possible environmental problems associated with the residuals from these treatment processes are discussed in the section on household residuals.

Among the nondegradable industrial wastes—apart from combustion products—suspended and dissolved solids are the major ones. Suspended solids can be removed by sedimentation, with or without the aid of flocculants—such as polyelectrolytes—by filtration through various kinds of screens and filters, and by centrifuging. Dissolved solids can be removed, to any desired degree, by one or more of the processes of distillation, ion exchange, electrodialysis, and reverse osmosis. It is possible to obtain completely pure water from waste discharges. However, it should be remembered that although "pure water" can be produced, there is still some sort of residual waste to be disposed of in some manner, either a concentrated brine or a semisolid sludge. The overall waste management problem is not solved by the production of reusable water from these various processes. As we have frequently noted, ultimate disposal still remains a problem.

It should be emphasized that technological changes affecting residuals have not been, and are not now usually instituted because of water or other environmental problems. In fact, almost all changes in production technology have been stimulated by factors unrelated to environmental quality, and have been developed without consideration for their external costs. On the other hand, various stimuli to management—such as effluent charges—in the few instances where they have been used, have resulted in process modification greatly reducing waste loads.[32] As more systematic means are developed to bring to bear on industry the external costs associated with the discharge of waste materials into the environment, waste residuals generation, and control factors including recycling of recovered materials and production of useable byproducts will receive more prominent consideration in process design. Studies of several industries have made clear that process design changes leading to residuals recovery or their conversion to usable byproducts can, in some instances, profoundly affect wastes generation.

In the following discussion we lean heavily upon the beet sugar industry as an illustration. This is not because it is the industry with the most important waste problem (although in 1950 it was estimated to discharge about 15 percent of the organic wastes, measured in terms of "biological oxygen demand" (BOD) coming from all industries) but because its processes are comparatively simple, it has recently been intensively studied by RFF [33] and we have been able to estimate a complete materials balance for representative plants using different processes.

---

[32] For a detailed treatment of this experience see Allen V. Kneese and Blair T. Bower, *Managing Water Quality* (forthcoming).

[33] See George O. G. Löf and Allen V. Kneese, *The Economics of Water Utilization in the Beet Sugar Industry*, Resources for the Future, in press. The materials balance described in later pages was calculated by George Löf.

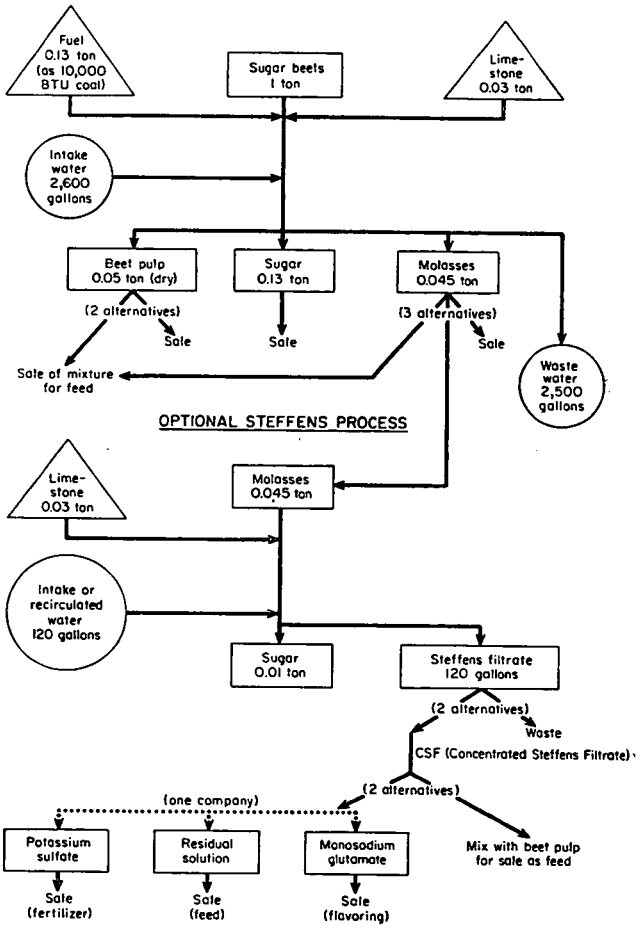
The waterborne waste load generated in pounds of BOD per ton of beets processed has been reduced greatly in the beet sugar industry as a whole in the last two decades by comparatively simple and economical alterations in processes. The main changes are the substitution of drying of beet pulp for storage of wet beet pulp in silos and the use of Steffens waste for the production of byproducts. (See explanation below.) These changes reduce BOD generation by about 60 percent. The other process change, i.e., a shift from cell type to continuous diffusers, is integrally related to recirculation of screen and press water. This further reduces the BOD generated by about 10 percent.

Chart VII indicates the main process and waste water residuals streams in representative beet sugar plants. Chart VIII shows residuals streams in a plant practicing no material or byproduct recovery and discarging all of its residual wastes to a watercourse. A few cases approaching this still exist. Chart IX shows a plant in which all water is fully recirculated and there is no external discharge of waterborne residuals. There is one plant in the United States which uses basically this system; the others fall in intermediate positions. Charts VIII and IX are not only helpful for understanding table 14 but also the materials balance for a beet sugar plant presented later.

It should be noted that the "closed" plant requires treatment (in the form of clarification) for its recirculating water stream, despite the fact that materials recovery and byproduct production have greatly reduced waterborne residuals. Even where opportunities to utilize process changes and increase recovery are favorable some waste residual usually remains. The stream containing this residual may be treated, thus producing a solid or gaseous residual or changing the chemical composition of the waterborne residual.

Chart VII

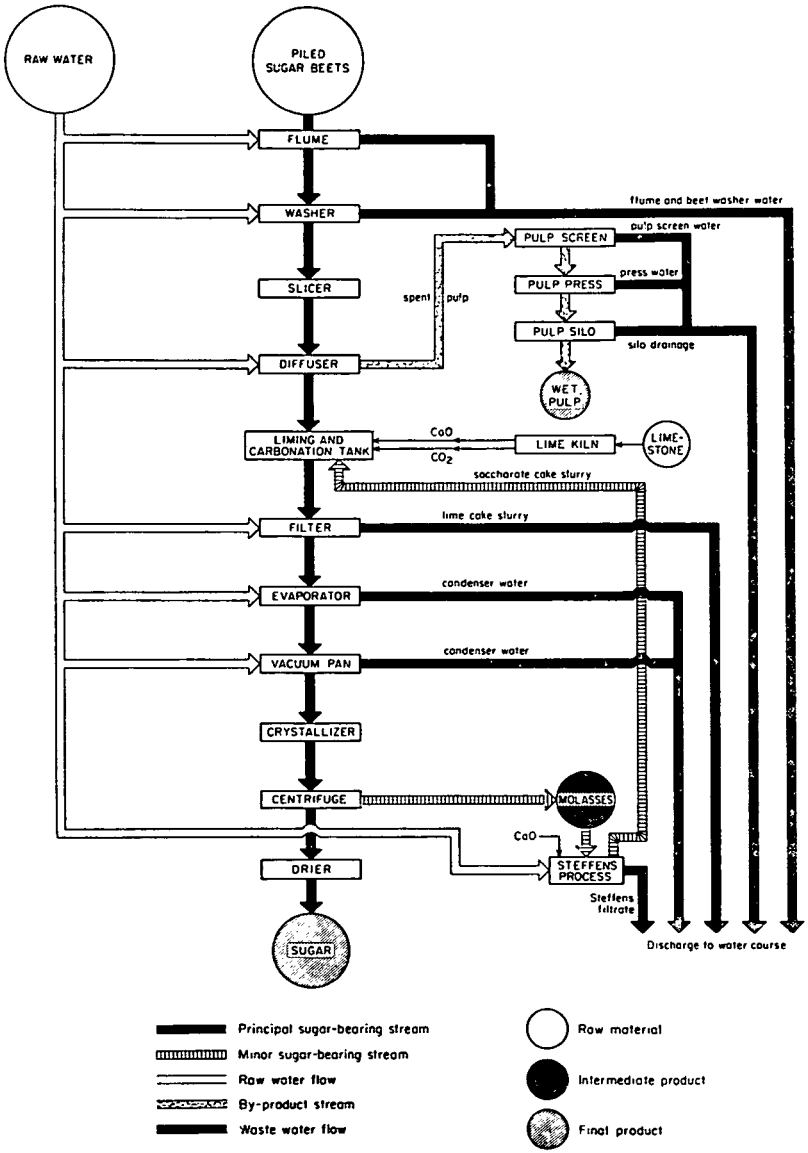
Main Processes in a Beet Sugar Plant



Source: Löf and Kneese, *op. cit.*

Chart VIII

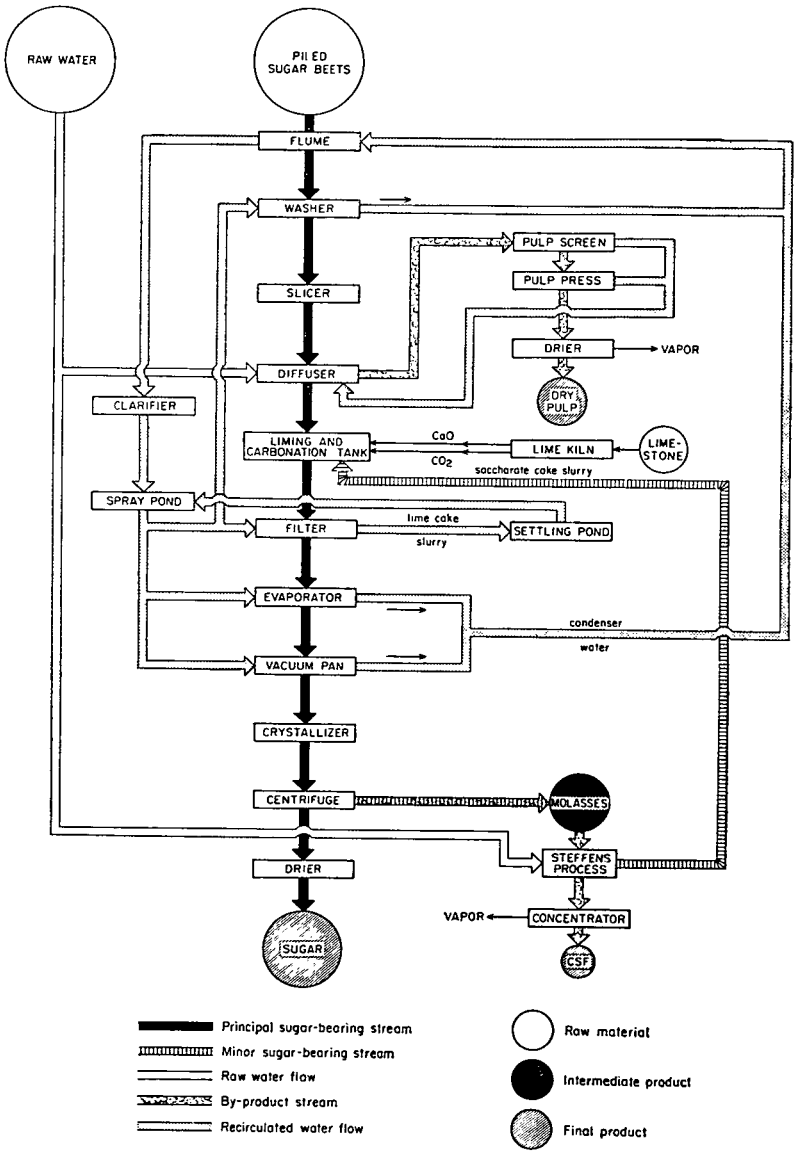
High Residual Beet Sugar Production Process



Source : Löf and Kneese, *op. cit.*

Chart IX

Low Residual Beet Sugar Production Process



Source : Löf and Kneese, *op. cit.*

TABLE 14.—*Estimated reduction of BOD in beet sugar processing, 1949 and 1962* [1]

All data in 1,000 pounds per day]

Type of waste	1949					1962				
	BOD generated [2]	BOD removed by process changes [3]	BOD removed by waste treatment	Total BOD removal	BOD discharged	BOD generated [2]	BOD removed by process changes [3]	BOD removed by waste treatment	Total BOD removed	BOD discharged
Flume and washer water	510	----------	----------	100	410	710	----------	----------	270	440
Cooling water and condensate	80	----------	----------	[4] 10	[4] 70	110	----------	----------	30	80
Pulp screen and press water	550	50	70	120	430	840	630	60	690	150
Silo drainage	[5] 1,390	[6] 660	140	800	590	[5] 1,940	[6] 1,920	10	1,930	10
Lime cake slurry	730	0	350	350	380	1,030	0	960	960	70
Steffens filtrate	610	[7] 160	80	240	370	770	[7] 560	160	720	50
Total BOD	3,870	----------	----------	1,620	2,250	5,400	----------	----------	4,600	800

[1] Based on 158,000 tons of beets per day processed by 58 plants operating in 1962. In 1949, 113,000 tons per day were processed. To enable direct comparison, the data for 1949 were extrapolated to production of 158,000 tons per day, assuming constant proportions.

[2] Based on BOD per ton of beets sliced in an "unimproved" plant, from Industrial Waste Guide to the Beet Industry, U.S. Public Health Service, December 1950.

[3] By process changes and recirculation.

[4] Based on estimated 10 percent reuse as diffuser makeup water.

[5] BOD which would be generated if all spent pulp were handled in silo; i.e., no pulp drying.

[6] BOD not generated because of use of pulp driers.

[7] By recycle-to-production process and CSF production.

Source: Löf and Kneese, *op. cit*

CHART X.--High Residual Process, No Recirculation

Intake 5250 gallons/ton sliced beets--regular
" 175 " " " " --Steffens additonal

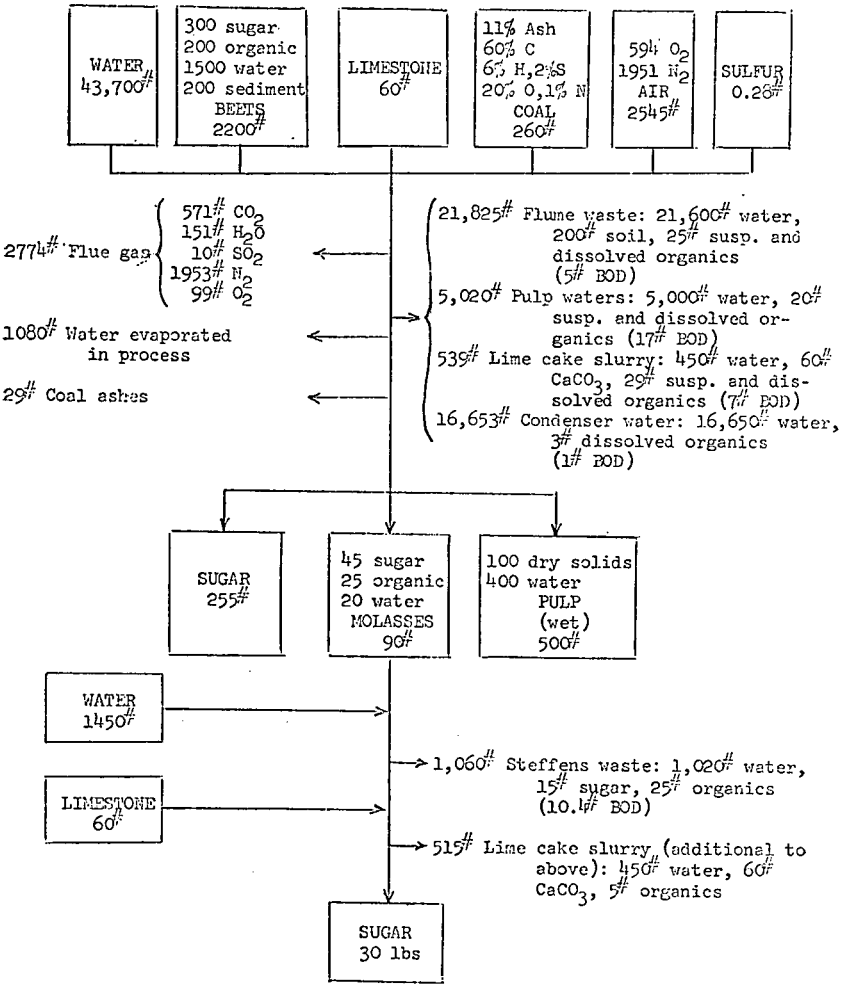
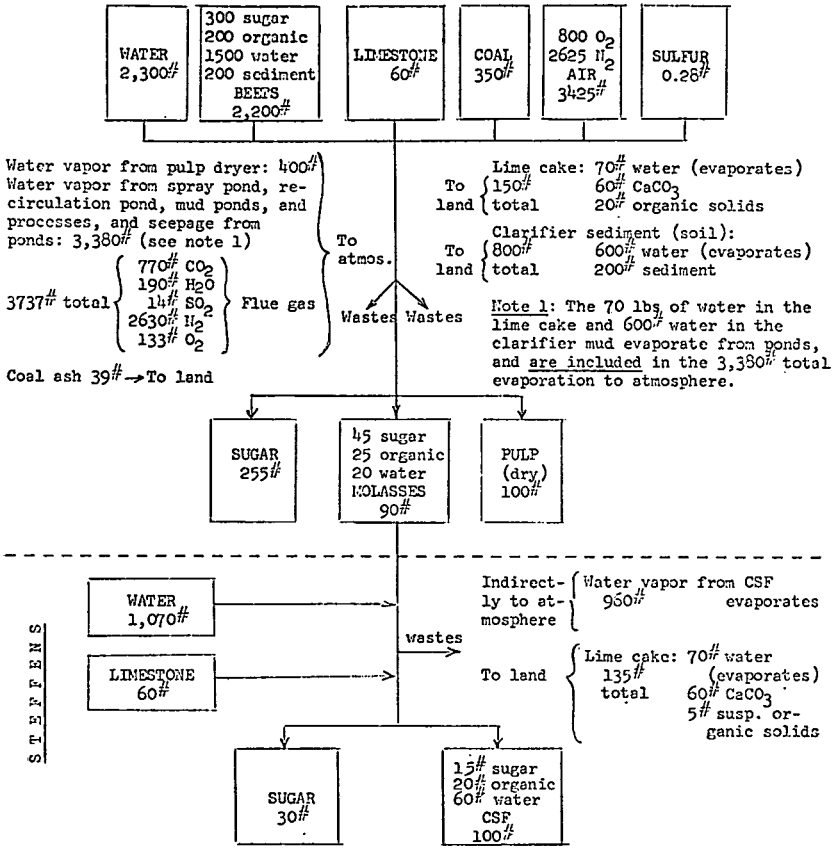


CHART XI.--Low Residual Process, with Extensive Recycle

Intake 270 gal/ton sliced beets--regular

Intake 128 gal/ton sliced beets--Steffens

Coal quantity based on 260# for simple plant + 60# for pulp drying + 25% for CSF production (evaporation).
Coal assumed 10,000 Btu/lb., 11% ash, 60% C, 6% H, 20% O, 2% S, 1% N.
Coal assumed to provide all the plant heat requirements, including pulp dryer.



Notes: Extra fuel for Steffens process included in original total quantity shown.
Lime cake from Steffens process actually combined with lime cake from straight process. Quantities shown separately.

To conclude this section we present a detailed materials balance calculation for two beet sugar production processes. One of these we term "high residual" and the other "low residual." Charts X and XI, showing the materials flow and residual materials, correspond to charts VIII and IX above which show the water circulation streams. In a wet-process industry like beet sugar the two are closely related. Table 15 summarizes a few salient figures from the materials balance.

It will be noted from the table that from the point of view of waste residuals a large reduction in organic residuals was purchased at the expense of a comparatively small increase in potentially harmful gas and inert solids. Again the interdependency between the residual waste streams is revealed. Considering the environment in which most beet sugar factories operate, away from large cities but near small streams with very limited capacity to assimilate organic wastes, the tradeoff shown is probably favorable.[34] A conventional study which looks only at air, water, or solids problems individually would never reveal such tradeoffs and permit their examination from the point of view of the full range of industrial costs—internal and external.

TABLE 15.—*Selected figures from materials balance for 2 beet sugar processes*

[All quantities per ton of beets processed]

	High residual inputs (pounds)	Low residual inputs (pounds)
Beets	2,200	2,200
Limestone	60	60
Coal	260	350
Sulfur	.28	.28
PRODUCTS OUTPUTS		
Sugar	285	285
Pulp	[1] 100	[2] 100
Concentrated Steffens filtrate (used for stock feed to recover monosodium glutamate and potassium sulfate)		100
Waste Residuals:		
SO_2	10	14
$CaCO_3$	120	120
Coal ashes	29	39
Organics	122	25
Soil	200	200

[1] Dry weight of wet pulp. [2] Dry.

IV. Residuals Associated With Final Consumption: Households

The household is where final consumer goods are utilized and give rise to the residuals remaining at the end of the materials flow through the economy to the consumer. We have already considered one important class of household activities giving rise to large amounts of residuals—the transportation of members of the household. We included this discussion in the energy conversion section because the interdependencies between the various types of residuals and possible methods for residuals control can be better analyzed in the context of overall urban transportation systems.

The other major residuals from household activities are gases resulting from space heating and home incineration or other combustion,

[34] Combustion of coal in beet sugar plants is conducted on too small a scale to make recovery of sulfur practical. (See section on thermal power.) However, fuel substitution might be feasible in certain instances.

sewage (solids in water suspension but containing also dissolved solids like chlorides and nitrates) and solid wastes including garbage and refuse as well as junk autos and the like.

Generally speaking, the space heating and incineration [35] activities of households do not appear to contribute a major portion of the gaseous residuals found in the atmosphere of cities, nor are gases generally a very significant portion of the weight of residual materials stemming directly from households. However, there is considerable variation among cities in this regard. In table 16 for Sacramento, which reflects conditions perhaps characteristic of areas having little heavy industry, using relatively clean fuels, and where home incineration is relatively uncommon, the preeminent position of autos in regard to gaseous emissions is clearly evident. Also it is apparent that residential incineration and fuel combustion contribute only marginally to total emissions. However, the type of fuel used in homes is very significant in this respect. Natural gas is a relatively "clean" fuel in the sense that combustion is rather complete. For example, the relative quantities of particulates emitted to the atmosphere per B.t.u. of fuel are 1 for coal, 0.6 for oil, and 0.15 for natural gas. Also natural gas is relatively free of sulfur compounds while certain coals and residual fuel oils contain up to 3 or 4 percent of sulfur, as we have already seen. Consequently, in cities where space heating with lower quality fuels is dominant, the contribution of households to overall emissions may be considerably greater. Table 17 shows, for example, estimates of emissions from various sources in New York City. In general it is difficult and expensive to control harmful emissions from households by treatment. Fuel substitution or centralized provision of heating services—via central steam plants or electricity—would usually be more practical.

TABLE 16.—*Atmospheric emissions in Sacramento County, 1964*

[Tons per day]

Source	Contaminant			
	Carbon monoxide (CO)	Hydrocarbons (HC)	Sulfur dioxide (SO_2)	Oxides of nitrogen (NO_x)
Stationary sources:				
Downtown business	2	18.0	1.0	4.0
Railroad and airport	1	3.0	1.0	3.0
Travis Air Force Base	7	5.0	<.5	1.0
Woodburning	3	<.5	<.5	<.5
Asphalt plants	2	4.0	1.0	4.0
Aerojet-General Corp	2	1.0	<.5	1.0
Residential and incinerators	2	3.0	.3	2.0
Municipal refuse	1	1.0	<.5	<.5
Motor vehicles	760	140.0	2.0	45.0
Total	780	175.0	6.0	60.0

Source: Aerojet-General Corp., *California Waste Management Study*, a Report to the State of California, Department of Public Health, Rept. No. 3056 (final), Azuza, Calif., August 1965.

Another point illustrated by table 17 is the importance of refuse incineration as a source of emissions—especially particulates and hydrocarbons. Some of these come from central incinerators but in New York and other large cities apartment house incinerators are also

[35] We do not include here the residual from incinerators *after collection* of solid wastes. These are discussed below.

a significant contributor. Combustion processes in these incinerators are usually inefficient and particulates are poorly controlled, if at all.

In the control of gases from household activities, there are important tradeoffs with other residual discharges both as to type and spatial distribution. We will return to this point later.

TABLE 17.—*Atmospheric emissions in New York City, 1964*

[Tons per day]

Source	Contaminant				
	Carbon monoxide (CO)	Hydrocarbons (HC)	Sulfur dioxide (SO_2)	Oxides of nitrogen (NOx)	Particulates
Stationary sources:					
Electric power	1	4	754	254	35
Industrial	2	2	140	54	26
Commercial, institutional, and large apartments	76	21	678	212	37
Small residential	4	4	67	78	24
Refuse combustion	291	120	6	7	75
Miscellaneous [1]	750	31	----------	705	7
Motor vehicles	3,784	695	14	162	22
Total	4,908	877	1,659	1,472	226

[1] Mainly gasoline marketing, commercial drycleaning, etc.

Source: "Electric Power and Fuel Consumption, 1965-85," July 1967, Tri-State Transportation Commission, New York.

One of the major components of household residuals is sewage. Most of the dry weight of sewage (about 0.55 lb. per person per day) from households not using garbage grinders is composed of human excrement. In addition there are various organic and inorganic substances which result from cooking and washing operations. (See table 18.) From households using garbage grinders, the total dry weight of sewage is about doubled. Actually the materials balance implies that the total dry weight of all organic wastes (sewage plus garbage) must be equal to the dry weight of food products entering the household, less the carbon content of CO_2 produced by respiration and minor corrections for annual accumulation (increase in population) and deaths. See table 19.

TABLE 18.—*Average per capita solids and BOD_5 in domestic sewage* [1]

[Grams per capita per day. 1 gram per capita equals 2.2 lb. per 1,000 population]

State of solids	Mineral	Organic [2]	Total	BOD_5
Suspended	25	65	90	42
(*a*) Settleable	15	39	54	19
(*b*) Nonsettleable	10	26	36	23
Dissolved	80	80	160	12
Total	105	145	250	54

[1] Source: Gordon M. Fair and John C. Geyer, *Water Supply and Waste Disposal*, New York: John Wiley & Sons, 1956, p. 563.

[2] Of the organic matter in average domestic sewage about 40 percent is composed of nitrogenous substances, 50 percent of carbohydrates, and 10 percent of fats. The detailed composition (carbon, hydrogen, oxygen, nitrogen) after primary treatment varies, depending on the process used: carbon content ranges from 52 to 67 percent, hydrogen from 7 to 9 percent, nitrogen from 3 to 4 percent, and oxygen from 21 to 38 percent. See R. Rickles, "Pollution Control 1965," Chemical Process Monograph No. 10, Noyes Development Corp., 1965.

TABLE 19.—*Materials balance for humans, 1963*

[Dry weight×10⁹ tons]

	Carbon	Oxygen	Hydrogen	Nitrogen	Sulfur	Total
Food inputs:						
Carbohydrate	24.1	26.8	3.25	----------	----------	54.15
Fat	12.04	1.81	1.89	----------	----------	15.74
Protein	6.28	2.73	.83	1.9	0.12	11.86
Total	42.42	31.34	5.97	1.9	.12	81.75
Outputs:						
Garbage	~9.8	~5.1	~1.3	~.15	----------	~16.5
Respiration [1]	~21	~16	~3	----------	----------	~40
Sewage (solids)[2]	~9.1	~8.4	~1.2	.7±.1	.1	19.4
Sewage (soluble)[2]	~1.3	~1.5	~.3	~.9		~4
Losses due to death [3]	.75	.2	.1	.1	----------	1.15
Added biomass [3] (population growth)	.45	.12	.06	.06	----------	.7

[1] Not including oxygen from the air; proportions based on combustion of sugar ($C_6H_{10}O_5$) yielding CO_2 and H_2O.

[2] Sewage solids estimated at 0.55 lb. per capita per day; nitrogen content of sewage sludge (3 to 4.3 percent) taken from O. E. Albertson "Low Cost Combustion of Sewage Sludges," Water Pollution Control Federation meeting (1963).

[3] Assuming the population increase is 1.2 percent (of the biomass) per year, and the death rate is 2 percent (of the biomass).

NOTE.—All figures prefaced by (~) are estimates based on plausible allocations of protein, carbohydrate and fat.

Of most direct concern in connection with household wastes are the degradable organic materials which impose a demand on the dissolved oxygen of waters to which they are discharged. This is usually measured in terms of the 5-day biochemical oxygen demand or BOD_5. Since the BOD_5 varies directly with temperature it is usually measured on a common temperature base of 20° centigrade. It is convenient to think of BOD_5 as a substance in the water—a measure of the degree of waste residual which can be added or removed. Of increasing significance too are the plant nutrients which are the final stage of the breakdown of degradable waste (added to this plant nutrient contribution from human excrement are the phosphate builders used in detergents).

Numerous detailed studies of the techniques used in treating domestic sewage are available so we will not discuss them extensively. The most basic processes, invented more than 50 years ago, use the principles of settling solids (sometimes with the assistance of chemical flocculents) and biological conversion to innocuous substances.[36] With these methods about 90 percent of the BOD_5 can be removed. Higher rates of removal can be achieved by chemical and filtration processes. Some of these (like adsorption with activated carbon) can achieve up to 98 or 99 percent removal of organic materials from domestic sewage. Costs vary approximately in proportion to $\frac{1}{1-E}$ where E is the removal efficiency in percentage.

A kind of ultimate treatment would involve distillation of the final effluent, producing pure H_2O as one final product. But even this does not mean that the residual materials are thereby eliminated: a solid or semisolid sludge remains which must either be used to produce reusable materials or must be dispersed into the environment. The most common procedure is to "digest" sludges resulting from sewage treat-

[36] For a detailed discussion of advanced treatment techniques together with cost and performance estimates, see Allen V. Kneese and Richard J. Frankel, *The Economics of Water Reclamation*, paper prepared for presentation at the symposium of the Institute of Water Pollution Control, London, England, Nov. 28–29, 1967.

ment in heated anaerobic tanks which produce CO_2 and methane gas (the latter being used for fuel in the plant or burned off), a supernatant liquor which contains most of the nitrogen and phosphorous originally contained in the sewage, and a stable organic material (mostly cellulose) which is referred to as the "digested sludge." The latter is most often used as landfill or simply dumped. The supernatant liquor is usually discharged to a convenient receiving watercourse where the additional nutrients may cause excessive fertilization and unpleasant or harmful algae blooms.

Another method for disposing of sludge involves its combustion (some techniques like the Zimmerman process do not require its prior digestion). Combustion—whether for power or not—results in the discharge of gases and still leaves a remaining, relatively small, residual amount of solids which must be disposed of. If the sludge has not been disgested prior to combustion the solid residual will be rich in plant nutrients,[37] and therefore potentially useful as fertilizer. Again we see the strong interdependencies between the control of liquid, gaseous, and solid waste streams and the important role which recycle may play in controlling external cost. For example, while it might seldom, if ever, be economically justified, distillation of municipal sewage with the solids resulting from perfect combustion reused as fertilizers would result in only CO_2 and H_2O residuals being discharged, as wastes, into the environment. The spectrum of sewage treatment methods currently employed in the United States is summarized in chart XII.

In addition to gaseous residuals and sewage households produce wastes which are in solid form, including garbage, rubbish and ashes.

(1) Garbage wastes from preparation, cooking, and serving of food.

(2) Rubbish:

Combustible: Paper, plastics, cartons, boxes, barrels, wood, excelsior, tree branches, yard trimmings, wood furniture, bedding, dunnage, etc.

Noncombustible: Aluminum foil, tin cans, metal furniture, dirt, glass, crockery, etc.

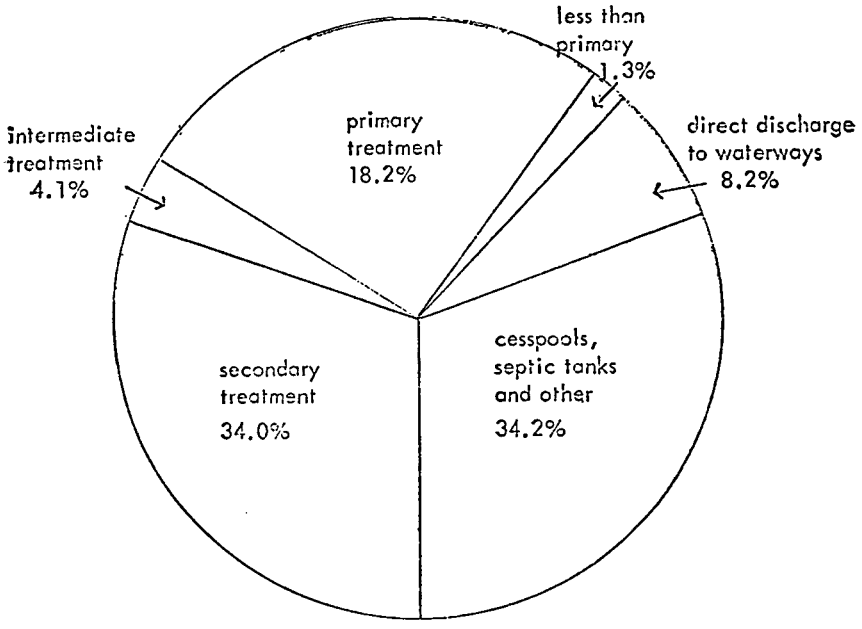
(3) Ashes: Incombustible residue from fires used for heating and on-site incineration.

---

[37] In some areas where water pollution is reasonably effectively managed sludge disposal is deemed the major unsolved problem. The Ruhr area of Germany and the Metropolitan Sanitary District of Greater Chicago are examples. In Chicago, until 1961, much of the wet undigested sludge (about 900 tons a day—dry weight—produced at the West-Southwest sewage treatment plant) was deposited into huge sludge lagoons along the main canal where the discharge of gases caused a great odor nuisance. Some sludge is still air dried and continues to cause significant odor problems. Also about 500 tons a day are heat dried and sold as fertilizer at a net loss to the district of $38 per ton. The district is also experimenting with a 200 ton a day Zimmerman plant. The ash resulting from this process is presently dumped. See statement of Vinton W. Bacon, general superintendent, Metropolitan Sanitary District of Greater Chicago to Natural Resources and Power Subcommittee, on Government Operations, House of Representatives, Congress of the United States at Chicago, Ill., Sept. 6, 1963.

CHART XII

U.S. Sewage Treatment (as % of Population)

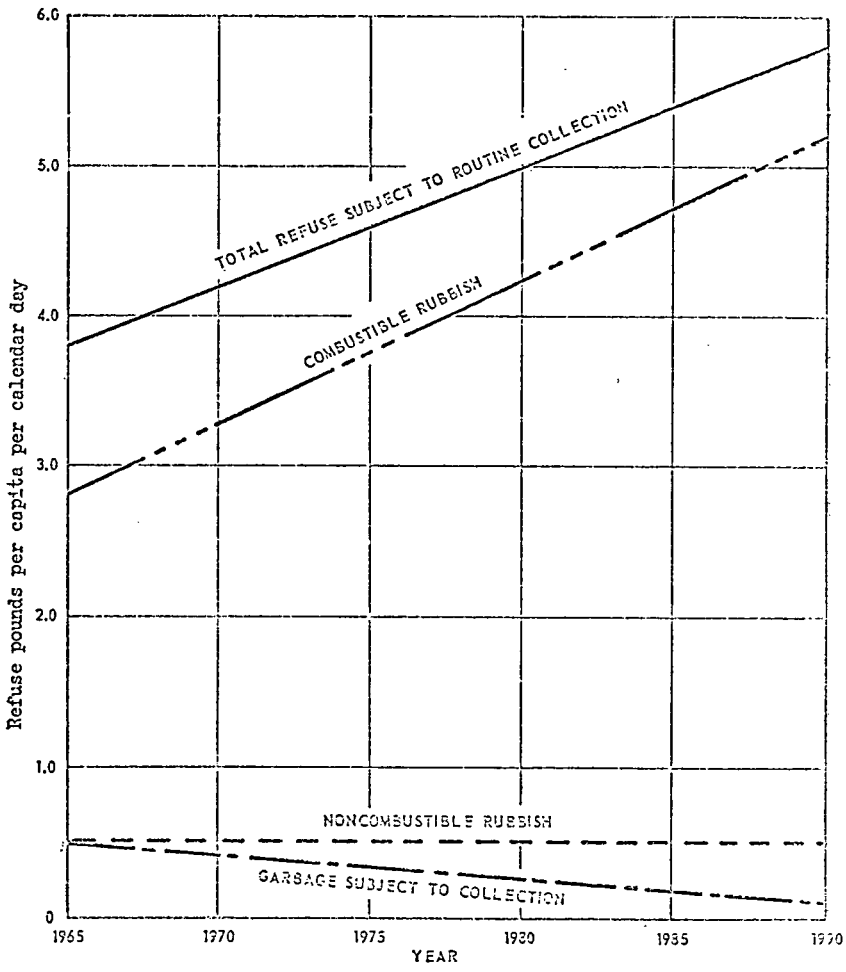


Source: USPHS estimates; and Chemical Week, May 30, 1964

Altogether the average member of a household will throw away somewhere between 3.5 and 4.5 pounds of solid wastes per day. (See table 20 and chart XIII.)

CHART XIII

Projection of Refuse Production Trends



Source: Aerojet-General Corporation, *California Waste Management Study*, a report to the State of California, Department of Public Health, Report No. 3056 (final), Azusa, Calif., August 1965.

TABLE 20.—*Total refuse collected in 5 U.S. cities, 1957–58*

Cities	*Pounds per capita per day*
Los Angeles, Calif.	4.61
Washington, D.C.	4.36
Seattle, Wash.	3.75
New York, N.Y.	3.63
San Francisco, Calif.	2.18

Source: Aerojet-General Corp., *California Waste Management Study,* a report to the State of California, Department of Public Health, Report No. 3056 (Final), Azusa, Calif., August 1965.

These wastes are disposed of in a variety of ways in our urban areas. Backyard burning and burial has greatly diminished and the pattern has shifted historically to neighborhood dumps, and then to central incinerators and countryside landfills. But the present practices still give rise to major external costs—indiscriminate dump sites still litter the countryside; odor, smoke, insects, and rodents accompany open dumps (which are often burning to reduce volume and help control insects and rodents); contributions to air pollution arise from improper incineration and uncontrolled burning; and water quality degradation results from drainge through dumps and direct disposal of solid wastes into waterways. As part of a very good study of the solid waste problem by the Maryland Department of Health a survey was made in July 1966 of solid waste disposal practice in all 23 Maryland counties and the city of Baltimore. Practices in the State could perhaps be considered typical of those in the megalopolitan east. At least 40 percent of the 155 disposal sites were clearly producing external costs.[38]

As presently practiced solid wastes collection and disposal costs are high. It is estimated that annual local government costs for collection and disposal are $1.5 billion. This is exceeded in local government budgets only by expenditures for schools and roads. In addition it has been estimated that the annual expenditures of the private sanitation industry are nearly as high.[39] Of the total cost of disposal about 80 percent is for collection. Collection is labor intensive and technologically primitive. One among the several significant external costs associated with dosposal of solid wastes is the high rate of injuries among sanitation workers. For example, in New York the injury frequency rate is reported at 70 injuries per million man-hours worked as contrasted with 12 injuries per million man-hours in manufacturing[40] as a whole.

Collection costs are also highly sensitive to certain variables. A cross section statistical analysis of St. Louis found that an increase from two weekly pickups to three increased collection costs by nearly one-third, while moving pickup location from the curb to rear of the house almost doubled the cost.[41]

Because collection is such an expensive and cumbersome process, it is attractive to think in terms of different and less labor intensive collection systems. One proposal that recurs frequently in the literature

[38] See *Collection and Disposal of Solid Wastes—a Maryland Program,* Maryland State Department of Health, Aug. 1, 1966.
[39] Wesley E. Gilbertson and Ralph Black, *Solid Wastes, National Commission on Community Health Services, 1964.*
[40] *Ibid.*
[41] Werner Z. Hirsch, "Cost Functions of an Urban Government Service: Refuse Collection," *The Review of Economics and Statistics,* February 1965.

is to grind solid wastes and deliver them to the sewers. Since sanitary sewerage is now over 99 percent water, the existing sewers could accommodate a considerable increase in solids without an expansion in capacity. Additional grinding could be done in the home or at grinding stations located relatively close to the points of pickup. In the latter cases various organic substances in addition to garbage could be ground up. This would reduce (but certainly not eliminate) collection costs, and it would require additional sorting.

This procedure would, of course, increase the waterborne wastes and put an additional loading on the sewage treatment plants and/or the stream. The interdependency of various waste streams and means of control is again evident. On the other hand, it would reduce the load on the transportation system by reducing the number of collection trucks and thus increasing the overall speed of surface traffic.[42] Savings realized at the collection end might possibly compensate for increased treatment costs.

As far as disposal is concerned, combustion looks favorable because, as shown in chart XIII, so large a portion (80 percent) of the total rubbish and garbage is combustible. Incineration reduces the total volume of the solid wastes to about one-fifth of that before combustion.[43] Consequently, from a narrow cost minimization point of view it would be desirable to conduct the incineration very close to the source (in apartment house incinerators and backyards). This, however, results in converting the major part of the residuals to gases and airborne particulates which, in highly developed areas, are likely to give rise to large external costs. If efficient techniques for controlling emissions from small incinerators could be worked out, burning close to the source would have much to commend it.

In recent years solid waste residuals from households have been substantially increased by a heavy trend toward nonreturnable containers of all types. For example, the use of one-way bottles almost doubled between 1965 and 1966.[44] In many instances, also, nonreturnable containers are being made of less degradable materials than previously—primarily plastic and aluminum. Overall, each year, 48 billion cans, 26 billion bottles and jars, and 26 billion metal and plastic caps are produced, virtually all of which are eventually discarded. Rates of reclamation and reuse are rather low. About 10 percent of the plastic produced is recovered and about 15 percent of the rubber products (mostly automobile and truck tires).[45] of the 43×10^6 tons of paper products produced each year about one-third is reclaimed.

One source has cited comparative costs for incineration and composting, including recovery of salvageable items as shown in table 21. These costs are very uncertain as they are affected by the markets for scrap and compost and by numerous other local factors.

---

[42] Experiments at the University of Pennsylvania indicate that ground up solid residuals (including metal and glass) can be successfully pumped as a liquid slurry through surprisingly small diameter pipes. A 2-inch pipe would suffice for a city of 10,000 or 15,000. This might provide an alternative to conventional collection and transportation but would present a problem of liquid disposal.

[43] At several locations in Europe, incineration is carried out in thermal electric plants. At Munich's Nord power station, for example, refuse is burned along with powdered coal. Electrostatic precipitators are used to control particulate emissions in the stack gases. The electric power system charges the refuse collection agency $1.50 per ton to burn the refuse delivered to the plant.

[44] *Chemical Week*, vol. 99, No. 19, 1966, p. 44.

[45] "Environmental Pollution: A Challenge to Science and Technology," report of the Subcommittee on Science Research and Development of the Committee on Science and Astronautics, U.S. House of Representatives, 89th Cong., second sess., U.S. Government Printing Office, 1966,, p. 44.

On the basis of this calculation it is clear that the compost operation can only make a profit if he can dispose of the product for fertilizer value and has a commercial outlet for his salvaged items. It should be noted, however, that the calculations do not take into account external costs which may be associated with gaseous and solid residuals from the incineration process. In some—perhaps many—instances these could be sufficiently large to swing the balance toward recovery and reuse.

TABLE 21.—*Comparative cost of incineration and composting*

[Basis: 100 tons per day, no cost for collection]

Inputs:	
Fuel oil	6 cents per gallon.
Electricity	1.5 cents per kilowatt hour.
Labor	$4 per hour (overhead, etc., 150 percent).
Amortization, taxes, interest [1]	15 percent
Value:	
Tin cans, rags, et al	$1.50 per ton feed.
Product:	
$50 per ton, nitrogen	Effective selling value after selling expense: marketing, shipping, etc.
$35 per ton, P_2O_5	
$20 per ton, K_2O	
$5 per ton, organic	

	Incineration	Composting
Raw materials	$4.00	$2.50
Labor	11.00	14.00
Capital charges	[1] 13.00	15.00
Total	28.00	31.50
Credit for scrap	----------	1.50
Subtotal	28.00	30.00
Credit for compose	----------	2.00
Total	28.00	28.00

[1] No taxes—municipally owned.

Source: R. Rickles, "Pollution Control," chemical process monograph No. 10, Noyes Development Corp., 1965.

A major source of solid wastes which poses quite special problems is wornout and discarded automobiles. At the present time there are about 90×10^6 automobiles and trucks operating, with annual production of about 9×10^6 vehicles, 6 to 6.5×10^6 vehicles scrapped each year, and about 2.7×10^5 vehicles added to inventory.[46] This accounts for about 30 percent of production.

Inputs to the "automotive transportation" subsector accounted for—

1.6×10^6 tons of new rubber of which about 1×10^6 tons were tires for new motor vehicles and the remainder were replacements or rubber recaps (60 percent of total U.S. rubber consumption).[47]

17×10^6 tons of steel [48] for automobiles and trucks (15.5 percent of U.S. consumption).

[46] Based on the anticipated growth in total vehicle population from 1960 to 1970. Estimates published by the FPC, February 1967, based on various projections by the Bureau of Public Roads, the Bureau of the Census, RFF, and the American Automobile Association.

[47] Data on rubber, glass, and other nonmetals was inferred from a variety of sources. Breakdown on weight composition of automobiles from G. A. Hoffman, *Automobiles, Today and Tomorrow,* the Rand Corp., RM–2922–FF, November 1962. Total rubber production is inferred from carbon black production, published by the Bureau of Mines, assuming rubber products are 33 percent carbon black by weight, and 60 percent of total rubber output goes to the automotive industry.

[48] From *Mineral Facts and Problems,* Bulletin 630, Bureau of Mines, 1965 edition.

1×10^6 tons of fabrics, carpets, paint, insulating materials, plastics, and other nonmetals.[49]

0.5×10^6 tons of glass.[49]

0.6×10^6 tons of lead for batteries (30 percent of U.S. consumption).[50]

0.15×10^6 tons of copper[50] (9 percent of consumption).

0.5×10^6 tons of aluminum, zinc, nickel, and other metals.

As noted above, about 30 percent of these inputs goes to increasing the inventory of vehicles.

The recycle picture is mixed. About 6.8×10^6 tons of steel were recovered in 1965 from scrap auto, bus, and truck bodies.[51] It is extremely difficult to say exactly what happened to the remainder. Presumably some was lost during manufacturing processes and, probably, most of this was returned to production as "new" scrap of unidentifiable origin. Essentially all of the lead in storage batteries is ultimately recovered, and the majority of tires apparently are recapped at least once. About 0.3×10^6 tons, or 15 percent of all rubber, is recycled—mostly old tires; however, a substantial tonnage of rubber—possibly 0.5×10^6 tons per year—is powdered. Some enters the air as organic dust which may be a health hazard. Ultimately it is mostly washed into rivers and streams; the remainder is probably burned or added to the growing inventory of old tires which are accumulating haphazardly in garages, on vacant lots, etc.

At present prices not all discarded automobile bodies can be utilized, as open-hearth or oxygen-process scrap, due to the presence of awkward contaminants, such as copper (~30 pounds per automobile), lead (30 to 40 pounds per car), aluminum (30 to 90 pounds per car), zinc (30 to 45 pounds per car), nickel, and chromium as well as nonmetallic impurities. Iron and steel constitute only about 79 percent of the curb weight of an average automobile, and, even with tires, fluids, and batteries removed, about 15 percent of the residual weight consists of materials which the potential scrap user does not want.[52] Thus, the annual supply of discarded motor vehicles (running at close to 6.5 million per year) is increasing faster than demand for this resource. The surplus piles up in automobile "graveyards" or as abandoned hulks scattered over the countryside, or when the concentration of eyesores reaches the threshold of toleration—the junk cars are sometimes incinerated and/or compressed and used in landfill operations. However, the main outlet in recent years has been to increase the inventory of junk cars in automobile graveyards.

Where a potential demand for scrap exists locally and the supply is large enough, new technology makes it possible to improve the quality of the scrap metal at an economic cost. The method utilizes giant shredders which automatically fragmentize the auto bodies (minus engines, radiators and gas tanks) and magnetically sort the ferrous

---

[49] See footnote 47 on p. 671.

[50] From *Mineral Facts and Problems*, Bulletin 630, Bureau of Mines, 1965 edition.

[51] Testimony by W. E. Gilbertson, Public Health Service, 1966 air pollution hearings. Mr. Gilbertson estimated that this accounted for 94 percent of the vehicles scrapped in that year, but this assumed an average weight of steel per vehicle (2,131 pounds) based on automobiles alone. In actual fact the average weight per vehicle must be somewhat higher since about 12 percent of the vehicles scrapped were trucks or buses, some of which were equivalent to several automobiles in weight. Hence it might be fairer to assume 90 percent of the vehicles were accounted for.

[52] This suggests the possible design ability of a disposed tax graduated to take account of the ease with which the materials contained in the automobile lend themselves to recycle.

from nonferrous materials. Combustible material is burned off, and the ferrous residue is heated and rolled to increase its density. The resultant product is equivalent in quality to desirable No. 1 heavy melting scrap (e.g., steel rails) in quality and degree of contamination. About 2.5×10^6 tons were processed in this way in 1965; unfortunately the capital investment required for such equipment is such that it can only be justified in densely populated areas.

The processing required to return the solid wastes (junk autos) to the production cycle results in discharge of residuals to the air as contrasted with compression and sanitary landfill *without* incineration or dumping in shallow coastal waters where the hulks can become convenient breeding grounds for marine fauna. At the present annual rate of motor vehicle discard about 2.5×10^6 tons of nonmetallic materials must be disposed of. Assuming this is 80 percent combustible complete combustion would yield about 0.5×10^6 tons of ash—much of it, however, injected into the atmosphere as particulates.

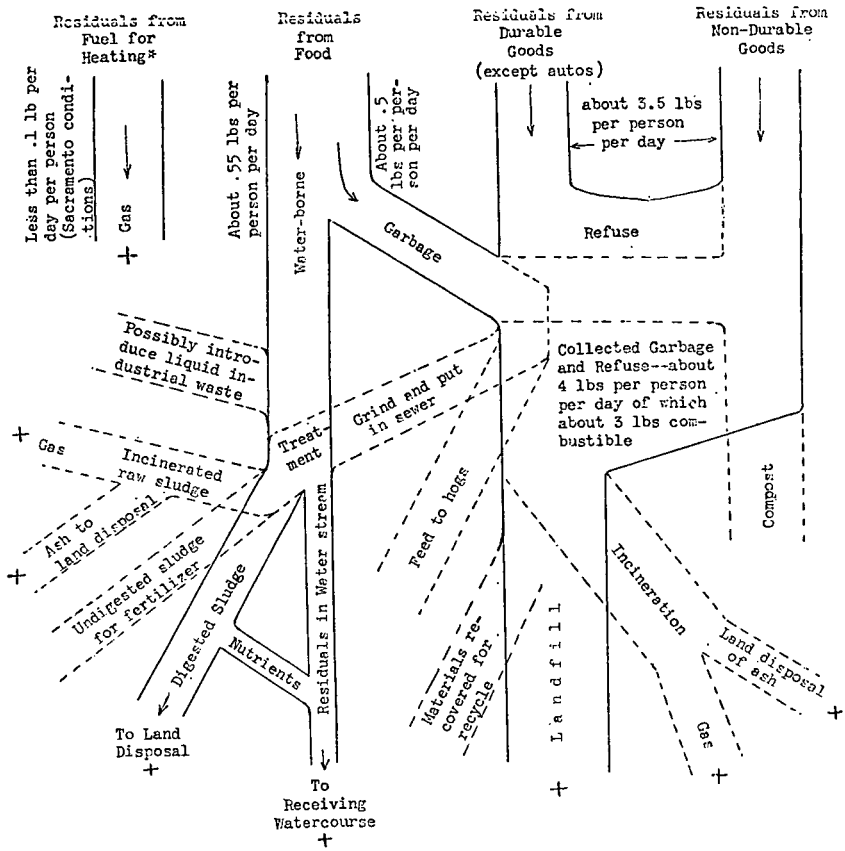
THE WHOLE SECTOR—INTERDEPENDENCIES

Chart XIV shows a schematic residual materials flow for the household sector, excluding automobiles. From it one can obtain a conception of the interdependencies of the various waste streams. For instance, if waste waters are given high-level treatment and if all sewage sludges and the maximum amount of garbage and refuse are incinerated, the majority of the final residual wasteload from households, which must be disposed of into the environment, is gaseous. On the other hand, if no incineration is conducted at all and sewage treatment is pressed to the point where all solids are removed from the waste water stream, all but a few percent (gases from space heating) of the weight of residuals flowing from households become a solid or semisolid waste. Finally, should garbage and refuse grinding and discharge to the sanitary sewers be practiced on a large scale and no waste water treatment to remove solids conducted, very little gaseous waste would flow from households and perhaps half the weight of residuals would be deposited into the aquatic environment. The remainder would be solids.

A general alternative to trying to "throw away" these residuals is to reincorporate them into the productive process thus reducing the overall throughput of the production-consumption system. At some cost, virtually all the household residuals could be recaptured for useful purposes. Thus, sewage solids and wet garbage could be composted and recovered for fertilizer. Metals and some other durable materials (including glass) can be, and sometimes are, recovered for reuse. Plastics pose perhaps the most difficult recycling problem. Still, application of known recovery technologies could reduce the net outflow of residuals from the household sector to a very low level.

We are *not* proposing that this is the appropriate solution. The cost of recovering the material may be, and often is, considerably higher than obtaining new materials from nature, even when external costs associated with disposing of residuals into the environment are taken into account. But there are reasons to think that the present system for organizing production (the market combined with rather ad hoc controls) does not produce the optimum degree of reuse possible with

CHART XIV.--Household Residual Materials Flow
(per capita)



* This stream is cut out if electricity is used. The waste stream then appears in the electric power sector.

+ Potential for external cost.

---- Alternative flows.

today's technology, nor does it stimulate an appropriate rate of improvement in that technology. Under current conditions few, if any, individuals and private or public agencies systematically bear the external costs of disposing of waste residuals into the environment. Water and air pollution control are in separate agencies and solid waste disposal is another still. Restrictions relating to water pollution usually say nothing about air pollution or disposal of solids. Often the water pollution authority is pushed to protect watercourses by producing gases or solids which it "throws away" into the environment. We will not here pursue the matter of how the matrix of controls and regulations might be altered to improve the situation but we do want to underline that arrangements which fully reflected all external costs would tend to lead to higher rates of reuse, less net throughput, and a more rapid rate of technological improvement in recovery systems.

V. Conclusion

In a necessarily partial and illustrative way we have applied an approach to the residuals problem which does not use as central categories the conventional division between liquid, gaseous, and solid residuals. Rather it attempts to account for residuals, regardless of form, resulting from the various materials and fuel-consuming economic processes and functions in national and regional economies.

As explained in the introductory section, this approach provides certain general insights concerning the quantity of residuals generated, the factors bearing on this quantity, and about the failure of market mechanism to deal appropriately with the entire resources allocation problem. The interdependencies between residuals streams, which become so clear in this approach, also suggest that the conventional organization of governmental regulatory and management authority along the lines of the environmental media to which residuals are discharged—air, water, and the solid earth—need reexamination.

To recapitulate, given the population, industrial production, and transport services in a region, it is possible to visualize combinations of social policy which could lead to quite different relative burdens placed on the various residuals-receiving environmental media; or, given the possibilities for recycle and less residual-generating production processes, the overall burden to be placed upon the environment as a whole. To take one extreme, a region which went in heavily for electric space heating and wet scrubbing or stack gases (from steamplants and industries), which ground up its garbage and delivered it to the sewers and then discharged the raw sewage to watercourses, would protect its air resources to an unusual extent. But this would come at the sacrifice of placing a heavy residuals load upon water resources. On the other hand, a region which treated municipal and industrial waste-water streams to a high level but relied heavily on the incineration of sludges and solid wastes, would protect its water and land resources but at the expense of discharging waste residuals predominantly to the air. Finally, a region which practiced high-level recovery and recycle of waste materials and fostered low residual production processes to a far-reaching extent in each of the economic sectors might, as we have seen, discharge very little residual waste to any of the environmental media.

In the transportation sector—which is probably the greatest single source of gaseous residuals and also a very significant source of solid wastes—a variety of waste and cost tradeoffs are also possible. For example, powering automobiles electrically and supplying power from conventional steamplants would tend to greatly reduce carbon monoxide and hydrocarbons emissions to the atmosphere but at the same time would increase SO_2 unless the waste gases from the powerplant were treated. If treatment took the form of wet scrubbing, discharges to the water environment would be increased. If dry sulfur recovery processes were used, overall discharges to the environment would be reduced but noxious emissions less so than the others. If nuclear power were used, residuals discharges would be practically eliminated but the small quantities of radioactive solid waste would require special and expensive disposal methods. Mass transit substituted for individual cars would greatly reduce emissions as would the substitution of steam

(external combustion) engines for conventional internal combustion engines. If electric vehicles were, as has been suggested, longer lived than conventional automobiles, another major residuals problem would be significantly reduced in intensity.

We feel that the approach sketched out in this paper, while clearly experimental at this point, could be developed and refined into a much more powerful technological-economic tool than has previously been available for projecting and analyzing environmental quality problems in regional and metropolitan areas.[53] A regional or metropolitan area focus is the appropriate one, for it is at this level that the analysis of residuals flows can be brought into a meaningful relationship with the geographical, meteorological, and hydrological systems into which the waste residuals flow and understanding of which permits the analysis of concentrations, durations, and probabilities. The latter must be obtained as a basis for gaging external costs and devising appropriate mechanisms and levels of control. Damages, or external costs, must also be evaluated on a regional basis. Mere calculations of residuals on a national level tell us little of specific value in devising efficient management schemes.

The kind of approach to planning for environmental management we have in mind is the following: [54] As a first step a more or less refined materials balance would be estimated for the area. This would involve some major complications since regional economies are very "open." In other words, all material (including goods at various stages through the manufacturing process) imports and exports would have to be accounted for. Getting this information as well as that needed to estimate materials balances for various industries would involve considerable primary data collection. But once completed the materials balance would provide a picture of the residuals flows in the area. Economic base input-output models [55] could then be used to project levels of activity and industry mixes into the future, as well as transportation requirements, population, and other parameters of interest in consrtucting a new materials balance. In the first instance present technology and low levels of materials recycle and byproduct recovery could be

---

[53] A significant first step toward the type of approach we have in mind is found in Walter M. Phillips, Blair T. Bower, Gordon P. Larson, and Abraham Michaels, *Wastes Generation and Wastes Management in the New York Region,* a report to the Regional Plan Association, September 1967.

[54] There are very difficult problems associated with selecting the appropriate region and we will not attempt to pursue them in any depth here. In general it will be necessary to conceive of at least two regional areas of interest. One may be termed the "source and control area" and the other the "impact" area. The source and control area will include the region where residuals emissions are concentrated and affect a natural environmental system or systems in a substantial way (usually a metropolitan area or groups of areas). If control of the natural system is possible to affect environmental quality, for example, release from storage reservoirs to improve the assimilative capacity of streams, the point where control can be exercised should also be included. This is why it is appropriate to encompass portions of a watershed which may not be affected adversely by residuals discharges but where reservoir sites exist that could be used to store water to augment low flows. The "impact" area is the region over which external costs occur resulting from emission in the "source and control area." Clearly the two may not correspond. For example, air pollution may (and does) damage crops outside a source metropolitan area and even outside the watershed in which the source area is located. Such damages should "source and control area." From the technical economic point of view of achieving optibe identified and evaluated wherever they occur and weighed against control cost in the "source and control area." From the technical economic point of view of achieving optimum level and techniques of environmental control, it is only necessary that the "source and control" area be under some kind of unified management. But as the political scientists remind us, this may not suffice to obtain a "constituency" for control. Effective and efficient environmental management requires "institutional design" and our ability to do this is very primitive.

[55] See John R. Meyer, "Regional Economics: A Survey," *The American Economic Review,* March 1963.

assumed. This procedure would differ from conventional approaches in that it would relate levels of residuals generation logically to the industrial and population base and account for all residuals in an internally consistent model. Conventionally levels of emissions to the air, liquid borne, and solid wastes are extrapolated separately.

As a further step it would permit analysis of the overall impact on all residuals of control measures instituted for one or more of them. As we have seen, for a given level of economic activity, a given efficiency of energy conversion and a given degree of recycle and byproduct recovery, reduction of one type of residual must come at the expense of creating another. Moreover, residuals should be classified and quantified in terms of their potential recoverability as a basis for economic analysis of this alternative.[56] Also, and quite important, the overall residuals implications of projected changes in waste treatment and recovery technologies could be tested.

Another major effort would have to go into mathematically simulating the natural system, especially the metorological and hydrological systems in such a way that concentrations and their probability and duration could be estimated, as well as secondary effects of residuals discharges. The latter might include photochemical reactions in the atmosphere and the reduction of dissolved oxygen in watercourses resulting from the discharge of organic wastes to them.[57] Where possible, concentrations of residuals and their secondary effects should be related to quantitative monetary measures of damage. Where this is not possible they may be compared with conventional standards of environmental quality. But such standards should be subjected to cost sensitivity analysis before a final working standard is adopted.[58]

A further step in the planning process would be to examine the potential of larger scale control systems in the region. Research on water quality management has shown clearly that such measures as reservior regulation of river flows, mechanical aeration of residuals receiving watercourses, and treatment of municipal and industrial wastes together in collective regional treatment plants can enter efficiently into regional water quality management systems. Changes in regional transport systems and design and operation of electric power systems might also enter efficiently into regional environmental quality management systems.

The final result of such a planning activity would be a coherent set of projections of residuals, estimates of optimal emissions control equipment and procedures (including recycle and byproduct recovery and treatment to change the form of wastes where appropriate to maximize net benefits or achieve standards), and large-scale measures which would fit integrally and efficiently into a regional management system.

Finally, it might be necessary to design an institution or institutions to perform management functions and to supply it with ap-

[56] A distinction should be drawn between recycle and byproduct recovery processes which market conditions will make internally profitable to the firm and those which might be stimulated by public policy, residuals taxes for example.

[57] Much progress has been made in recent years in devising mathematical models of atmospheric diffusion and the waste carriage and assimilative characteristics of watercourses.

[58] For an excellent example of such cost sensitivity analysis in a more limited context, both from the point of view of the level of the standard and the probability of its violation, see Robert K. Davis, *The Range of Choice in Water Resources Planning* (forthcoming).

propriate policy instruments. The latter would include authority to regulate emissions, levy charges or taxes on them, and implement, or encourage the implementation by other agencies, of large-scale regional measures. Management would also include continuing operating, coordinating, and planning responsibilities. How to design an institution which would perform these tasks effectively, efficiently, and equitably, and which would be politically viable, is perhaps the most difficult and important problem in the environmental management field.[59]

[59] There is already clear evidence that rethinking of the traditional government approach to these problems is going on. A case in point is Mayor Lindsay's proposal for an Environmental Protection Administration, now before the New York City Council for enactment.

APPENDIX A

The Formal Materials Flow Model

Since one of our purposes is to lay out the groundwork for future research, it seems worth while to devote a short space to discussing the characteristics of a formal theoretical model for analyzing materials flow, residuals formation and disposal in the national (or regional) economy.

The keys to a serviceable model are (1) division of the total economy into appropriately defined sectors and (2) the choice of (a set of) uniquely defined variables by means of which materials flow can be measured precisely. An obvious, but not necessarily ideal, sectorization system already exists, well known as the standard industrial classification of activities. While these groupings are not as salient for our purposes as they might be, the advantages of using a familiar system—which greatly simplifies the data acquisition problem, among other things—would seem to outweigh the disadvantages which will appear subsequently.

However, for purposes of illustration, we propose to introduce a very much abbreviated sectorization consisting of the following gross activity categories:

(0) The external environment, which may be a source of inputs (e.g., ores) or a "sink" for waste streams.

(1) Energy conversion (including transportation).

(2) Primary processing: beneficiation (of ore), grading, sorting, rough cutting, ginning, combing, etc.

(3) Secondary processing: refining, milling, smelting (of metal), pulping, carding, and yarnmaking.

(4) Tertiary processing: alloying, rolling, machining, diecasting, grinding, papermaking, weaving and spinning, canning and preserving.

(5) Final processing: fabrication, assembling, construction, paper products, printing, garment making, tailoring, cooking, packaging, finishing, etc.

(6) Final consumption (households, municipalities, other).

As for the choice of variables, it is clear that neither dollar values nor gross tonnages of materials are pertinent measures from the standpoint of the purposes of this paper. In many processes the gross tonnages are dominated by inert or harmless substances, such as water, carbon dioxide, silicon dioxide, nitrogen (as gas), or various complex mineral aggregates, such as soil, clay, gravel, and stone. Apart from esthetic, and possibly macrometeorological issues, these substances do not constitute a "pollution" problem in any sense. Moreover, during various stages of industrial materials processing, these substances are freely withdrawn from or discharged to the environment in large quantities, and keeping track of all these transactions would merely be tedious without being illuminating.

On the other hand, it is clear that substances such as hydrogen, carbon, nitrogen, oxygen, sulfur, phosphorus, the halogens, most metals (and the compounds of the above), tend to be biologically

active. Moreover, most organic compounds and many metallic salts are alien to the environment and, to a lesser or greater degree, disturbing to it. Hence, it is these elements (or tracers) which we would like to keep track of as they evolve from raw materials through various stages of processing to final goods. In short, we contemplate an input-output model of the economy in which, however, transactions between sectors (whether SIC or our own simplified categories) are described in terms of physical transfers of carbon, hydrogen, sulfur, iron, and so forth, rather than transfers of dollars.

It is convenient to define functions φ_{ij}^k to represent flows (in tons per year) of the k^{th} element (tracer) from the i^{th} to the j^{th} sectors, while $X_i^k(t)$ is the cumulative total of the k^{th} element in the i^{th} sector at any given moment; thus, $\frac{dX_i^k}{dt}$ would be the instantaneous rate of accumulation of the k^{th} element in the i^{th} sector and $\left\langle \frac{dX_i^k}{dt} \right\rangle$ would be the time-averaged rate of change (over 1 year). Referring to the simplified group of categories previously defined, it can be taken for granted that there is no significant net change in inventory from one year to the next within the various intermediate processing stages (1–5). Flows between sectors 1–6 correspond to economic transactions: the flow of materials is accompanied by a (usually reverse) flow of dollars. However, flows to and from sector 0 may represent nonmarket interactions for which there is no dollar transaction. In any case, we have the following generalized equations (for the k^{th} tracer element) summarizing flows into and out of sectors 1–5:

$$0=\sum_{j=0}^{6}(\varphi_{j1}^k-\varphi_{1j}^k)$$

$$0=\sum_{j=0}^{6}(\varphi_{j2}^k-\varphi_{2j}^k)$$

$$\vdots \qquad\qquad \vdots$$

$$0=\sum_{j=0}^{6}(\varphi_{j5}^k-\varphi_{5j}^k)$$

$$\left\langle \frac{dX_6^k}{dt} \right\rangle=\sum_{j=0}^{6}(\varphi_{j6}^k-\varphi_{6j}^k)$$

where the subscript $j=0$ refers to the external environment (whence φ_{0i}^k is a transfer from the environment to the j^{th} sector, and φ_{j0}^k is a transfer in the reverse direction).

Although the summations are formally carried out over all values of j, it is immediately obvious that φ_{ij}^k is always zero (for every element k) for certain combinations ij. In fact, putting it another way, the only nonzero transfer functions are the following:

φ_{02} mining, harvesting, logging, fishing, etc.

φ_{23}, φ_{34}, φ_{45}, φ_{56}, successive transfers from one processing stage to the next.[60]

[60] Skipping a processing stage is excluded by convention.

φ_{21} low-grade fuel (for consumption in thermal power plants).
φ_{31} refined fuel (for vehicles or metallurgical purposes).
φ_{61} recycled wastes burnt as fuel (garbage, sewage, trash).
φ_{43}, φ_{53}, φ_{63}, recycled scrap, from various stages.
φ_{10}, φ_{20}, φ_{30}, φ_{40}, φ_{50}, φ_{60}, waste streams from various stages.

The nonzero transfer functions constitute the following array:

Transfer matrix

$$\varphi_{ij}^k \left| \begin{array}{ccccccc} \cdot & \cdot & \varphi_{02}^k & \cdot & \cdot & \cdot & \cdot \\ \varphi_{10}^k & & & & & & \\ \varphi_{20}^k & \varphi_{21}^k & & \varphi_{23}^k & & & \\ \varphi_{30}^k & \varphi_{31}^k & & & \varphi_{34}^k & & \\ \varphi_{40}^k & & & \varphi_{43}^k & & \varphi_{45}^k & \\ \varphi_{50}^k & & & \varphi_{53}^k & & & \varphi_{56}^k \\ \varphi_{60}^k & \varphi_{61}^k & & \varphi_{63}^k & & & \end{array} \right. \quad \text{(for all } k\text{)}$$

Note that the diagonal elements are arbitrary, since they automatically cancel out of the equation. The equation written previously can be written as a single equation in matrix notation:

$$\frac{d\boldsymbol{X}^k}{dt} = (\boldsymbol{u}' \cdot \varphi^k)' - \varphi^k \cdot \boldsymbol{u}$$

where $\boldsymbol{X}^k$ is a column vector consisting of the elements X_0^k, X_1^k, . . ., and $\boldsymbol{u}$ is a unit column vector (i.e., a vector whose elements are all 1), and, of course, the transposed vector $\boldsymbol{u}'$ is arow vector.

As given, these relationships are exact and absolute but the equations are not soluble as such until the φ_{ij}^k are specified as functions of t or of the X_j. Many assumptions are possible, the simplest being that transfer rates from sector i to sector j will be strictly proportional to the accumulated amount of the k^{th} element already in the j^{th} sector, viz., $\varphi_{ij}^k = C_{ij}^k X_j^k$. The generalized transfer functions would thus be effectively replaced by input-output coefficients, to be determined empirically. Constraints can be introduced absolutely in the form of linear inequalities, of which a trivial example might be that $\varphi_{02}^k \leq M$, where M is the maximum limiting output of mines, oil wells, and other basic sources of materials. Finally, we could introduce an "objective function" (analogous to the role of GNP in a conventional input-output model) to be minimized subject to the various constraints by means of a linear-programing technique. In this case, the objective function would be an appropriately chosen function (possibly a simple weighted summation) of all the residuals returned to the environment. In the terminology of our simplified sectorization, for instance, the objective function might be the sum overall k and j:

$$R = \sum_k A_k \sum_{j=0}^{6} \varphi_{j0}^k = \sum_k A_k \sum_{j=0}^{6} C_{j0}^k X_j^k$$

where the A_k are coefficients reflecting the relative "polluting" importance of given quantities of each element.[61]

The coefficients C_{ij}^k (or the more general function ϕ_{ij}^k) reflect the technology involved in the materials transfer from the i^{th} to the j^{th} sector. Hence, the coefficients, or functions, tend to change as older methods are phased out and new ones are introduced. Thus the gradual shift from sulfite to sulfate pulp, the introduction of the basic oxygen process in steelmaking, and the decreasing relative importance of the Solvay ammonia-soda process are three examples. The effects of technological change would be far more noticeable in a model using the two-digit S.I.C. sectorization; needless to say, the more disaggregated the model, the more pronounced the impact of a specific technological change on a particular coefficient might be.

While linear models such as the above are convenient and lend themselves well to electronic data-processing techniques, they may not be satisfactory representations of real interactions between sectors. For instance, let us consider the elementary economic fact that if an element is scarce with respect to demand at a given price, the price tends to rise until the excess demand disappears. An obvious corollary is that, where demand is elastic, there often exists some potential for switching from one technology to another (or from one material to another). The inability of a linear input-output model to handle this problem adequately is well known. A transfer function capable of reflecting technological elasticities might have the mathematical property that the transfer rate of an element from the i^{th} to the j^{th} sector would be related to the relative magnitude of the current inventories (or outputs) of the two sectors. Thus, if the inventory of the k^{th} element in the i^{th} sector is depleted because overall demand is greater than new supply, then the j^{th} sector will find substitute input materials. Intermediate and final demands for k (for instance, copper) may very in time due to price fluctuations, resulting in complementary fluctuations in demand for other elements (such as aluminum). A two-parameter nonlinear function such as

$$\varphi_{ij}=C_{ij}X_j\left(\frac{X_i}{X_j}\right)^p$$

(where the superscript k is assumed throughout) would have the required elasticity, as can be seen by examining its asymptotic behavior in the two limiting cases: $X_i<<X_j$ and $X_i>>X_j$.

Evidently, the basic flow relation is modified by a term which decreases as the inventory in the i^{th} sector decreases; but on the other hand, if the inventory in the j^{th} sector drops, the rate of transfer from i to j tends to increase to make up the deficit.

Further discussion of the formal model will be reserved for other publications.

---

[61] It will be pointed out quite properly that there is no way to determine the A_k uniquely, since dangerous pollutants are normally chemical compound rather than elements. However, this is only an illustration of a possible approach.

APPENDIX B

Waste Heat [62]

In the United States in 1965, gross energy consumption from all fuels, including hydroelectric power, was very close to 4×10^{16} B.t.u. or 1×10^{16} kilocalories. Of this total it can be assumed roughly that all energy content in fuel used to generate electricity as well as virtually all electrical energy (except for a small portion used in electrolytic decomposition of aluminum, magnesium, sodium and phosphorous compounds) is eventually reduced to heat. Similarly, most of the energy value of fuel used in transportation (24 percent of the total) and for space heating is reduced to heat, except for unburned hydrocarbons which account for about 2–2.5 percent of the total. In the industrial category, the energy content of some petrochemicals is largely preserved in the final product, viz synthetic rubber, carbon black, artificial fibers, plastics, lubricants, asphalt, and road oils. This category accounts for 5 percent of the total energy content of all fuels. Moreover, a substantial part of the heat content of some fuels (mainly coke) is used to reduce metallic ores—mainly iron ore—from the oxide to the metal, or in the manufacture of lime and portland cement from limestone. The primary metals sector also accounts for 5 percent of total energy (3 percent iron and steel, 2 percent other metals) of which possibly half is wasted. Altogether perhaps 90 percent of the energy consumed in the United States is returned promptly to the environment as "waste" (i.e., irrecoverable) heat. In quantitative terms, this now amounts to 3.6×10^{16} B.t.u. per year, and is increasing at about 3.3 percent per year.

The great bulk of the waste heat is returned to the atmosphere. A comparatively small fraction ($\sim$15 percent), mostly from electric utilities and large industrial users, is dissipated into rivers and streams (but most of this is eventually returned to the atmosphere by convection or evaporation and condensation) where under some circumstances it can cause substantial external costs.[63] Since the biochemical demand for oxygen (BOD) increases in proportion to temperature, waste heat discharges tend to decrease the assimilative capacity of rivers and streams. Also, elevated temperatures often tend to increase the sensitivity of local fauna to certain toxic substances which may be present in the water.

By far the largest amount of waste heat discharge to watercourses comes from thermal powerplants. At present efficiency levels about 6,230 B.t.u.'s of heat must be dissipated by cooling water, per kilowatt-hour of electricity generated. This is the result of a fuel consumption of 10,700 B.t.u. of which some 10 percent goes up the stack or is lost in a number of individually minor ways, about 3,400 is converted into electrical energy and the remainder is transferred to the cooling water. Improvements in energy conversion efficiency (which have been occurring perisistently over the years) reduce the amount of waste heat which must be dissipated. Cooling towers can be and often are used to take the heat load off the stream (or reduce water intake) but only by transferring heat to the ambient atmosphere instead. The temperature

[62] Most data from Vogely and Morrison, *op. cit.*

[63] Perhaps partly offset by some external benefits. For example, a small percentage of the water delivered through municipal systems is heated and an increase in the temperature of intake water can save a certain amount of energy—the palatability of the water is reduced, however. Also winter sport fishing is sometimes improved below large heat discharges because the fish tend to congregate there—if they manage to survive the overheated summers.

of the water coming from the condensers is reduced by partial evaporation, the balance of the water being recirculated. Generally speaking recirculation is rather low in cost—usually less than 1 cent per thousand gallons.[64]

At first it would appear that the discharge of waste heat to the atmosphere is of little significance (except locally) but recent evidence indicates that this may not be so. Cities have measurable and even rather large effects on their own climates in terms of differences in temperature, humidity, precipitation and fog frequency, and wind speed between the city and its environs. Not all of these effects listed in table B–1 are necessarily negative (for example, longer periods free of frost may be favorable), but others are, such as persistent smog and dust cover. Part of the difference in temperature between the central city and its suburbs, which is related to the other phenomena, is due to the tremendous amounts of heat generated by various energy-consuming activities in the city.[65] For example, a blanket of warm air over the city combined with rapid cooling of surfaces such as rooftops and streets in the early morning hours can contribute to a temperature inversion and thereby increase the level of air pollution.

TABLE B–1.—*Differential climatic effects between a city and its environs*

[Environs assumed to be 100 percent]

	City/environs (percent)	difference (percent)
Solar radiation (insolation) on horizontal surfaces	85	−15
Ultraviolet radiation, summer	95	−5
Ultraviolet radiation, winter	70	−30
Annual mean relative humidity	94	−6
Annual mean wind speed	75	−25
Frequency of calms	115	+15
Frequency of cloud cover	110	+10
Frequency of fog, summer	130	+30
Frequency of fog, winter	200	+100
Total annual precipitation	110	+10
Days with less than 0.2 inch of precipitation	110	+10

Source: *Scientific American*, August 1967.

The exact role of heat discharges in influencing the climate of cities has not been defined but it appears to merit study. Limiting heat discharges would be difficult and costly. There is no technological means on the horizon which would permit generating electric power in large quantities without waste heat production (although continued modest increases in efficiency can be anticipated). The most plausible approach to limiting the external costs is to concentrate on minimizing the effects. Thus heated but otherwise clean water could be pumped out to sea or to the deeper waters of large lakes.[66] This type of program requires regional planning to determine the optimum locations of large power thermal plants not only with respect to local markets, but also from the standpoint of making maximum use of the thermal assimilation capacity of the environment.

[64] See Paul H. Cootner and George O. G. Löf, *Water Demand for Steam Electric Generation*, Resources for the Future, Inc., 1965.

[65] See William P. Lowry, "The Climate of Cities," *Scientific American*, August 1967, p. 15.

[66] In the Great Lakes this might have a beneficial effect by helping to break up a thermal stratification layer which has recently kept the warmer surface waters from mixing with the deeper waters and thus greatly reduced the waste assimilation capacity of the lakes.

O