POLICY ANALYSIS FOR FEDERAL BIOMEDICAL RESEARCH

PREPARED FOR THE PRESIDENT'S BIOMEDICAL RESEARCH PANEL

A. P. WILLIAMS, G. M. CARTER,
A. J. HARMAN, E. B. KEELER,
W. G. MANNING, JR., C. R. NEU,
M. L. PEARCE, R. A. RETTIG

R-1945-PBRP/RC
MARCH 1976

Rand
SANTA MONICA, CA. 90406
The work upon which this publication is based was performed pursuant to Contract No. N01-PP-6-2113 with the President's Biomedical Research Panel, Department of Health, Education, and Welfare. This research was also supported in part by The Rand Corporation as part of its program of public service. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.
PREFACE

This report was prepared for the President's Biomedical Research Panel. The work was supported jointly by a contract from the Panel (N01-PP-6-2113) and by The Rand Corporation as part of its program of public service.

The purpose of this report is to examine the state of the art of analysis that is or could become relevant to federal biomedical research policy decisionmaking and to describe an agenda of analysis that could improve policy decisions in the future. It is by nature an essay on the prospects for policy research, rather than a report of research findings.

It should be of interest to those concerned with federal biomedical research policy in particular, and science policy in general, as well as to those involved in policy research and evaluation.

Dr. Morton Lee Pearce is Professor of Medicine at the University of California at Los Angeles and a consultant to The Rand Corporation.
SUMMARY

Although federal involvement in biomedical research can be traced back to the late 19th century, the rapid growth of federal funding for such research began after World War II. Through the mid-1960s, the executive branch, the Congress, and the nation’s biomedical scientists jointly decided how much to spend for what in biomedical research. Government policymakers implicitly decided that the need for good biomedical research was greater than the supply and appropriated about as much money as was needed to fund almost all proposals that were judged to be sound by NIH’s peer groups. The scientists who submitted the proposals of merit implicitly decided “for what.”

By the mid-1960s, the nation’s biomedical research community had grown to the point that it could generate more scientifically sound proposals than the government was willing to fund. Efforts were made to evaluate research programs systematically and to equate program costs with expected benefits. However, although the broad objectives of biomedical research are clear enough, it is difficult to link particular expenditures with health outcomes. Any attempt to do so faces four fundamental problems:

- Even at the level of assigning probabilities to outcomes, it is difficult to predict scientific progress, with its complex and uncertain feedbacks from discipline to discipline and between various stages of research and development.
- The effects of federal and private program expenditure on productive scientific activity are poorly understood.
- Once the outcomes are known, there are problems of transforming scientific progress into medical practice and then into improvements in health.
- Finally, policy analysis in general faces the difficult task of assigning social value to improved health and longevity, even when the preceding problems are solved.

There is little likelihood that any one—much less all—of these problems can be solved. Thus, in the foreseeable future, we rule out the possibility of applying cost-benefit analysis, in the strictest sense of the term, to biomedical research programs. Although it will almost certainly fall short of providing answers to questions about what benefits particular research expenditures will yield, policy analysis can yield information that would lead to more sound biomedical research policy decisionmaking.

By definition, scientific inquiry deals with unknowns, but decisionmakers must have available predictions of the potential for success and the usefulness of results if they are to allocate scarce resources to such inquiry. Different types of scientific inquiry call for different criteria of usefulness to society and for different methods of anticipating the potential for success. The interactions between science and technology are complex and only crudely understood. Better models of these interactions are needed to enhance capabilities to predict the results of scientific programs and their likely benefits to society.
We have defined three categories of scientific activity—"knowledge generation," "advanced development," and "refinement for application"—that seem particularly useful for biomedical science. The first of these can be predicted, if at all, only on the basis of technical feasibility and the importance of gaps in (current) scientific knowledge. The other two categories can make greater use of society’s applications-oriented goals, such as improvements in health status.

Under current practices, NIH deals mainly with knowledge generating activities and uses the implicit predictions from a combination of evaluations by expert panels and Advisory Council reviews. These predictions are based largely on judgments about what is good science—to some degree, what is relevant to particular health problems. Decisionmakers could have greater confidence in the current practices and a better basis for initiating improvements if they knew the reliability of the priority scoring process for proposals, the assignment of proposals to IRGs, and the effectiveness of the categorical structure for funding Institutes.

Further consideration of the types of review activities that are undertaken for (and by) the biomedical research community may be desirable. Appropriate areas for such expanded support without pre-empting the private sector are another important subject for policy analysis. Evaluation of possible changes in management contexts and development strategies—using some of the lessons learned from scientific inquiry in other substantive areas—may also be called for as society’s objectives evolve and as pressures grow for the effective use of federally funded scientific inquiry. That proposition is valid as a general principle. However, to be policy relevant, it is necessary to move from this level to more concrete understanding of the individual programs that compete for scarce resources.

The first logical step in this direction is for NIH to improve its understanding of how its own research funding mechanisms affect the scientific activity of those who receive the money. The effects of length of funding and the choice of funding instruments (e.g., contract or grant) are straightforward problems for analysis. Experiments could be designed to examine the effects of alternative funding approaches for first time investigators and low-level maintenance of effort funding for continuation grants just below the "pay line."

The federal government cannot evaluate the role of its funding in the support of biomedical science in the aggregate until there is a classification system that describes the scientific characteristics of research from all major sources. The system used in NIH’s own IMPAC file is probably closer in content to what is needed than any of the several other systems in use, but its scope is limited. The scope of the Smithsonian Institution’s system is much broader, but it is not well suited to analysis that considers multiple characteristics of research. A comprehensive, flexible classification system would combine the desirable features of both systems.

Once such a classification system is in use, the federal government can begin to make a systematic examination of the effects of its programs on the nation’s health research activities. The first candidates for such an examination should probably be the large, "targeted" research programs of the National Cancer Institute. Data from the classification system would enable NIH to determine whether major increases in its funding in a field attract more funds or replace funds from other sources. The system would also permit an analysis of the effects of discontinuities in federal funding on the activities in a scientific field.

Every scientist and federal program administrator would agree that research
activity in his field of interest is relevant whether it is funded by a public or a private agency. Moreover, most federal policymakers would probably concede that there is no reason to spend federal funds for research unless the expenditures increase useful scientific inquiry in the aggregate. Yet there is no place in the federal resource allocation process where research activities funded by nonfederal agencies are considered systematically. Even if there were, policymakers would not find any convenient source of information for such an examination. The feasibility of creating the required comprehensive scientific information system is clear, because a combination of the attributes of existing systems would be sufficient to meet these needs.

There is widespread agreement that the purpose of biomedical research is to generate scientific knowledge in the life sciences in order to improve health. It would be useful to policymakers to have a greater understanding of how medical innovation occurs. First, an empirical picture of the process would provide some common ground for policymakers and biomedical scientists to identify and resolve differences about appropriate R&D investment strategies. Second, greater understanding of how medical science flows into medical practice would provide more information on the possibilities and limitations of government intervention in the medical innovation process. Third, research on medical innovation could bring together several diverse bodies of knowledge in a way that illuminates important policy questions.

The study of medical innovation needs to concern itself with the process of minor medical innovations as well as that of major medical "breakthroughs." The strategy proposed here moves away from an analytical focus on the single innovation and suggests that related innovations, whether major or minor, be studied within the context of (a) the management over time of patients with a given disease problem, (b) the changes over time in the practice of various medical specialties, and (c) the innovations flowing over time from particular knowledge "streams"—specific fields of science or technology.

An improved understanding of the process of medical innovation would provide the basis for determining the federal government's role in innovation and diffusion relative to the private sector. Insofar as it is appropriate for the federal government to sponsor or perform developmental activities, analysis can shed some light on the appropriate management environment for these activities. Although NIH has taken on some developmental activities, the large cost of these programs compared with costs of most research projects makes it difficult to strike the proper balance between the two, and the possibility of deriving economies of scale by combining activities related to more than one categorical area suggests that other organizational forms should be considered. Analyses of research and development management in other areas may contribute ideas to this study of organizational forms.

In the absence of some difficult and quite specific social and political choices as to the relative importance of various individuals, there can be no meaningful measure of the value of improved health for any group. If decisions are to be based on such valuations, a fairly complex methodology such as that based on individual preferences may be required. Simpler methodologies obscure the distinctions between the analytically tractable and the politically determinable and are likely to lead to major contradictions. There will never be a single number for the value of life and health. The best we can hope for is a rough comparison—subject to different political or value judgments—of how relief of some kinds of ill health would be
valued relative to others, and an even rougher guess of whether the money being spent is justified. Although this information would not supply simple answers to questions of policy relevance, they would illuminate to some degree a process of valuation that now must be carried on in almost total ignorance.
ACKNOWLEDGMENTS

The idea for this essay emerged from our discussion with Charles Lowe, Executive Secretary of the President's Biomedical Research Panel. We benefited from his comments and those of George Eaves and William Doak of the Panel staff. The comments and questions of Panel members during our briefing to them helped to sharpen the focus of our final draft.

Louisa Kasdon provided valuable assistance in this project by reviewing and summarizing a wide variety of literature that was relevant to the questions we addressed. Arthur Alexander and Walter Baer reviewed an earlier draft of this report and offered insightful and constructive criticism.

The editing and coordination efforts of Helen Turin have been invaluable. Marilyn La Prell maintained an up-to-date computer-based version of the report through dozens of separate revisions by the various authors.

Notwithstanding our debts to those who influenced or helped in preparing this essay, we bear full responsibility for any errors of logic or interpretation that may remain.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>Section</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. COST BENEFIT ANALYSIS—THE CONCEPT</td>
<td>3</td>
</tr>
<tr>
<td>III. THE STRUCTURE AND PREDICTABILITY OF SCIENTIFIC PROGRESS</td>
<td>6</td>
</tr>
<tr>
<td>Models of Scientific Progress</td>
<td>6</td>
</tr>
<tr>
<td>More Complex Models of Science and Technology</td>
<td>7</td>
</tr>
<tr>
<td>Distinguishing Activities Within Science and Technology</td>
<td>8</td>
</tr>
<tr>
<td>Patterns of Interactions Within Science and Technology</td>
<td>10</td>
</tr>
<tr>
<td>Interactions Between Science and Technology</td>
<td>11</td>
</tr>
<tr>
<td>Interactions Among Basic Medical Science, Clinical Medical Science, and Clinical Medical Practice</td>
<td>14</td>
</tr>
<tr>
<td>Current Practice of Anticipating and Influencing Scientific Progress</td>
<td>15</td>
</tr>
<tr>
<td>Analyses to Assess Reliability of Current Practices</td>
<td>16</td>
</tr>
<tr>
<td>Analysis For Improving Predictions</td>
<td>18</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td>IV. THE EFFECTS OF FEDERAL PROGRAM EXPENDITURES ON PRODUCTIVE SCIENTIFIC ACTIVITY</td>
<td>22</td>
</tr>
<tr>
<td>The Effects of Program Changes on Federally Sponsored Research</td>
<td>22</td>
</tr>
<tr>
<td>Classification of all Scientific Activity</td>
<td>25</td>
</tr>
<tr>
<td>Relationship of Scientific Activity to Federal Funding</td>
<td>26</td>
</tr>
<tr>
<td>Summary</td>
<td>27</td>
</tr>
<tr>
<td>V. FROM BIOMEDICAL R&amp;D TO INNOVATION IN MEDICAL PRACTICE</td>
<td>29</td>
</tr>
<tr>
<td>Strategies for Studying Medical Innovation</td>
<td>30</td>
</tr>
<tr>
<td>Disease State Management</td>
<td>31</td>
</tr>
<tr>
<td>Major Medical Specialties</td>
<td>32</td>
</tr>
<tr>
<td>Major Streams of Knowledge</td>
<td>32</td>
</tr>
<tr>
<td>Elements of a Research Design</td>
<td>33</td>
</tr>
<tr>
<td>A Comment About the State of Our Knowledge</td>
<td>34</td>
</tr>
<tr>
<td>The Federal Role in Innovation and Diffusion</td>
<td>35</td>
</tr>
<tr>
<td>Summary</td>
<td>38</td>
</tr>
<tr>
<td>VI. THE VALUATION OF LIFE AND HEALTH</td>
<td>39</td>
</tr>
<tr>
<td>Some Traditional Approaches</td>
<td>39</td>
</tr>
</tbody>
</table>
A Recently Formulated Approach—The Use of Individual
Preferences .......................................................... 42
Summary ................................................................. 46

VII. CONCLUSIONS ......................................................... 47
The Predictability of Scientific Progress ......................... 48
The Effects of Federal Program Expenditures on Productive
Scientific Activity .................................................. 48
From Biomedical R&D to Innovation in Medical Practice .... 49
The Valuation of Life and Health ..................................... 49
All diseases may by sure means be prevented or cured, not excepting that of old age.

—Letter from Benjamin Franklin to Joseph Priestly.

I. INTRODUCTION

In modern society, there is a near universal acceptance of the importance of science in the improvement of health care and disease prevention. Such names as Pasteur and Lister are associated not merely with improvements in health but with changes in the course of civilization.

The precedents for federal expenditures for health science are well established in American history. The first involvement in biomedical research can be traced back to the Federal Quarantine Act of 1878. Under the Chamberlain-Kahn Act of 1918, the Public Health Service made grants to 25 institutions to study venereal disease, setting the precedent for what became a multi-billion dollar National Institutes of Health (NIH) extramural research program. In 1922, a Special Cancer Investigations Laboratory was established at the Harvard Medical School by the Public Health Service, and in 1930, the National Institute of Health was created.

After World War II, the National Institutes of Health were the organizational conduit for rapidly expanding federal research funding, both to government scientists and to scientists at universities and medical centers. During the 1950s and early 1960s, Congress strongly supported the NIH programs, and the executive branch did not seriously challenge their justification.

From the end of World War II until the mid-1960s, the executive branch, the Congress and the nation’s biomedical scientists jointly decided how much to spend for what in biomedical research. In effect, the executive branch and Congress implicitly decided that the need for good biomedical research was greater than the supply and appropriated about as much money as was needed to fund almost all proposals that were judged by NIH’s peer review groups of scientists to be sound. The scientists who submitted the proposals of merit implicitly decided “for what.”

By the mid-1960s the biomedical research community had grown to the point where it could generate more scientifically sound proposals than the government was willing to fund. During the mid-1960s, the federal government also greatly expanded its involvement in social programs, including medical care for the poor and aged. As the federal functions increased in scope, budgetary pressures increased. The federal government sought ways to examine all programs—domestic as well as national security—more systematically.

One approach to systematic examination was the Planning, Programming, and

---

2 The first all-out war on cancer was proposed by Senator Matthew Neely of West Virginia in 1928. The Senate passed the Neely bill but the House of Representatives failed to act on it. S. P. Strickland, Politics, Science and Dread Disease, Harvard University Press, Cambridge, Massachusetts, 1972, pp. 1-7.
Budgeting System (PPBS) instituted in 1965. It called for all programs to be associated specifically with desired outputs to justify budget expenditures. The biomedical research and training programs of NIH did not escape evaluation attempts. Although the broad objectives of biomedical research were clear enough, and expenditures could be categorized, there was no straightforward way to link particular expenditures to health outcomes.

Any attempt at policy analysis of biomedical research faces four fundamental problems:

- Even at the level of assigning probabilities to outcomes, it is difficult to predict scientific progress, with its complex and uncertain feedbacks from discipline to discipline and between various stages of research and development.
- The effects of federal and private program expenditure on productive scientific activity are poorly understood.
- Once the outcomes are known, there are implementation problems of transforming scientific progress into medical practice and then into improvements in health.
- Finally, policy analysis in general faces the difficult task of assigning social value to improved health and longevity, even when the preceding problems are solved.

These problems are the central concern of this essay. In Section II, we consider them in the context of a particular method of evaluation, cost benefit analysis. We discuss the underlying concepts of cost benefit analysis in order to make clear to the reader that it is inappropriate for complex problems such as those encountered in biomedical research decisionmaking.

Sections III through VI will address the four problems in sequence and consider ways to deal with them more effectively with policy-oriented research. However, even if the results of the proposed research were extraordinarily successful, they would still fall short of what is needed to provide analytic answers to questions about how much, where, and in what manner the federal government should commit funds to biomedical research.

The objective of this essay is not to establish the case for or against any particular, limited method of analysis. Rather it is to describe a variety of analyses that together can lead to better informed decisions in allocating resources for biomedical research.

The approach was developed to evaluate water resource projects, and that sort of application is most useful. See J. Hirshleifer, J. C. DeHaven, and J. W. Millman, Water Supply: Economics, Technology, and Policy, University of Chicago Press, Chicago, 1960.
II. COST BENEFIT ANALYSIS—THE CONCEPT

Before considering how the federal government should make decisions regarding funding of biomedical research or any other federal program, we should consider why any particular program is the concern of the federal government in the first place. One reason often cited for government intervention is the inability of the private sector for one reason or another to undertake an adequate level of some activity. Research is a good example of such an activity. The commodity provided by research is knowledge, and in most cases the researcher is powerless to control its dissemination. Each of us is likely to benefit from most advances in medical technology whether or not we have paid for the research that led to their development. Since each of us will reap the benefits of research regardless of whether we have paid for them, none of us will have much of an incentive to pay for research privately. If research is an activity that all of us wish to have carried on but that none of us feels individually compelled to support, we may find ourselves doing without something we all want. Similarly, because the benefits of research will inevitably pass to individuals who have not paid for them, it will not be profitable for an entrepreneur to fund research activity with the intention of ultimately charging those who benefit. It is not a coincidence that the largest area of biomedical research carried on by the private sector is pharmaceutical research, which has as its principal product not knowledge but medication, a commodity that can easily be distributed to those who pay for it and withheld from those who do not.

In this sense research is what is often called a public or collective good. Because it is difficult or impossible to deny a public good to those who have not paid for it, such goods are consumed publicly and are thus unlikely to be supplied privately. A solution to such a dilemma is for us to agree to supply such goods collectively through charities, cooperatives, or, as is usually the case, through government involvement. All of this is an economic rationale for federal funding of some activities, and it suggests that decisions regarding the amount and nature of this funding should be based at least in part on economic analysis.

There are other reasons for federal involvement in certain areas. One of the most important is the political requirement for a degree of equity beyond what the marketplace may provide. In the medical risks they face, men differ from women; blacks differ from whites; the young differ from the aged; indeed, those unlucky enough to become ill differ from those who remain healthy. It is considered a legitimate role of the government to transfer resources from some of these groups to provide better health for other groups. Decisions regarding the extent of such redistributions are distinctly not economic matters, but political ones. As a result, any attempt to deal with questions involving significant redistributions must be based on more than just economic analysis.

Many techniques have been developed in recent years to aid in sorting out the complexities of public decision making. Most of these techniques can be grouped under the general heading of cost benefit analysis. In a sense, almost any analysis of the attractiveness of a proposed action (either public or private) may be considered a cost benefit analysis since a decision must hinge on a comparison of what the action will cost and what is to be gained by it. Because of the multiplicity of similarly
named concepts and applications, there is no single, clearly defined procedure that represents cost benefit analysis. At the most general level, the name implies nothing more than a systematic attempt to consider as many as possible of the costs and benefits associated with some action. To the extent that relevant costs and benefits (nonmonetary as well as monetary, nonquantifiable as well as quantifiable) can be articulated, such an analysis will be of obvious usefulness.

It is frequently the case, however, that cost benefit analyses are quite limited in scope. In the standard practice, costs, measured as best they can be in monetary terms, are added up and arrayed against similarly computed benefits. Some summary measure—usually the difference between costs and benefits or their ratio—is computed. The final result of the analysis is a number that supposedly reduces the complexities of the problem at hand to a single decision variable.

A simple example will illustrate the logic of an application of cost benefit analysis. Consider a proposal to build a dam to provide hydroelectric power, flood control, and recreational opportunities. In theory, all those affected by the construction of this dam could gather to decide whether this proposal should be adopted. Each person who stood to gain from the services provided by the dam could pay some amount to see it built and still come out a net beneficiary. If the total of such amounts were to cover the costs of construction and to compensate those people whose farms would be inundated by the new lake, it would be possible for all to benefit from the building of the dam. Some functions of the dam (flood control, for example) are public goods in that each person would benefit from its construction even if that person had not contributed to its financing, and it might be impossible for such a project to be undertaken if the decision were left to private interests. If, instead, some governmental body can devise a scheme for taxation to finance the dam's construction by which each beneficiary is charged no more than the value of the dam to him, everyone will have gained. Each taxpayer will have gained the use of new recreation areas and protection from floods at a cost of no more than these are worth to him; and suppliers of the materials, labor, and land required for construction will be compensated for their sacrifices. (The electricity would probably be sold privately.) If alternative proposals for the location or size of the dam were available, the "best" choice would be the one for which benefits exceeded costs by the widest margin.

Of course, the presumed caucus of affected interests would never actually take place. Instead, analysts would infer the total amounts that people would pay. They would know, for example, the going rate for electricity in the area and thus the value of the power produced by the dam. Similarly, the value of flood control could be estimated by the resulting drop in insurance premiums, and the price of the farms to be flooded by the new lake could be approximated by the price for similar farms nearby. The worth of the recreational services or the loss of some particular scenic rapids would be harder to value. If these considerations were thought to be minor, they might simply be ignored. If not, some ad hoc procedure would be needed to assign them values.

It is not surprising that such analyses place a heavy emphasis on the quantifiable at the expense of nonquantifiable considerations, which may in fact be more important. To the extent that important aspects of the problem are subordinated to the need to derive numbers, such an analysis will be misleading. Still, in many
applications this narrow form of cost benefit analysis is quite appropriate, and it is attractive (at least superficially) on theoretical grounds.\footnote{The theoretical attractiveness of cost benefit analysis is based on the so-called compensation principle. This asserts that if the total benefits of an action exceed the total costs, those who benefit could, in theory, compensate those who bear the costs and still have some excess benefit left over. Each individual would then be better off (or at least no worse off) than he was before the action was taken. In practice, of course, such compensation is seldom made; and the fact that one group receives benefits greater than the costs borne by a different group is of doubtful relevance for policy formation.}

Certain characteristics of biomedical research make it unlikely that such a limited approach will be useful. First, it is extremely difficult to know in advance what is actually being bought for the costs of the program. Specifically, there is no reliable way to know in which areas scientific progress can be made, how federal funding will affect the rate of this progress, or to what extent this progress will be transformed into better health. Similarly, it has proved nearly impossible to predict the costs associated with the exploitation of new medical knowledge. Retrospective analysis makes the problem no simpler, since we cannot say with any confidence what advances might have been made in the absence of federal support. Even if we could, the rapid rate of change of biomedical research severely limits the value of the past as a prediction of the future.

A second difficulty in performing a simple cost benefit analysis of biomedical research is the nature of its ultimate products, improved health and increased longevity. It is very difficult to assign a dollar value to commodities as personal as life and health. Such current practices as valuing lives at the amount of earnings associated with them are clearly inadequate. At best, we might derive rough measures of the relative value of various changes in the prospects for life and health. These will fall far short of the numbers needed for a traditional cost benefit analysis.

A third problem lies in the importance of distributional or equity questions. In a traditional analysis, a dollar of benefit is thought to balance a dollar of cost no matter who reaps the benefits or who pays the costs. In the field of health, serious issues of equity are involved in any federal intervention. In few other areas are there such heated debates about the rights of citizens to services beyond their ability to pay. No analysis can do more than simply describe the distributional aspects of a particular program since the ultimate valuation of these distributional aspects is a political matter. To ignore their importance is to base policy on seriously incomplete consideration.

Finally, the nature of scientific progress itself prevents the traditional approaches from providing adequate guidance. There simply is no direct cause and effect relationship between the costs of research and the advance of science. Progress is by accretion, and it would be arbitrary to attribute so much of each step forward to any particular input to the process. Since the whole idea of traditional cost benefit analysis is to associate specific costs with specific benefits, it is likely to fail in the case of research.
III. THE STRUCTURE AND PREDICTABILITY OF SCIENTIFIC PROGRESS

In deciding to invest in biomedical research, the federal government is predicting that the results will lead to improved health care. Predictions about the progress of science are commonplace, ranging from the very general to the very specific. Some predictions are only implicit—for example, decisions to award funds to research proposals that promise useful results and to deny funds to others. Other predictions are explicit—for example, those periodically obtained from groups of experts such as the interdisciplinary clusters convened by the President’s Biomedical Research Panel to assess the state of biomedical science in 11 fields.

Although predictions about scientific progress are frequently sought and obtained, their makers and users are generally uncomfortable with the results, because a fundamental characteristic of all scientific inquiry is the uncertainty of its outcome. Good science is the careful, imaginative, and systematic investigation of the unknown. Professional co-workers in a field can judge careful and systematic investigation: They base their judgments largely on research design and on what is known in that particular area of science. Although it is difficult to evaluate the “imaginative” quality of the investigation in advance, it can usually be recognized after the fact. Using a working definition of “good science,” scientists predict where they and their peers will make progress.

Progress in science is not easy to describe. For example, scientific activity does not have to lead to positive results to be valuable. The demonstration that experiments do not produce expected results can point to errors in existing theory. Also research and development (R&D) demonstrating that seemingly promising avenues of investigation are not fruitful can often save society from committing resources to a technological innovation that will not succeed.

We can be certain that scientists will continue to surprise themselves as well as the public; and in this sense, science will continue to elude prediction. However, a better understanding of the overall structure of progress in various areas of biomedical science and, at a more modest level, a better understanding of the reliability of predictions that are a routine part of research program administration would provide a better basis for policymaking than is now available.

In this section, we begin by discussing current models of scientific progress and some promising refinements of these models. We then consider modest but practical steps that policy analysis can take to improve the prediction component of research policy decisionmaking in the short term.

MODELS OF SCIENTIFIC PROGRESS

Accurate and understandable models of scientific progress and its translation into medical practice would permit better communication between policymakers and the scientific community. Such models may make possible greater agreement on how to invest and manage the federal government’s biomedical R&D dollar.
Consideration of two models of the relationship of R&D to medical innovation will indicate the limitations of current thinking. The first—what we might call the "NIH" model—appears in the "Knowledge Development" portion of the FY 1975 "Forward Plan" of the Department of Health, Education, and Welfare. It portrays a sequence of activities—results flowing from basic research to applied research and development to clinical investigation, clinical trials, and demonstration programs—in parallel with control programs and professional and public education programs. This sequential model has some general validity and captures the popular understanding of the value of research. Unfortunately, it lacks the detail about interconnections of the activities or motives within the activities that would make it useful in specific decisions.

A second model, identified with Lewis Thomas, makes a distinction between "high technology" and "halfway technology." According to Thomas, high technology "comes as the result of a genuine understanding of disease mechanisms; and when it becomes available, it is inexpensive, simple, and easy to deliver." Examples are methods for immunization against diphtheria, pertussis, and various virus diseases and the contemporary use of antibiotics and chemotherapy for bacterial infections. The halfway technology of medicine consists of procedures done to compensate for the incapacitating effects of certain diseases we cannot prevent or cure. Organ transplantation and artificial organs are examples. While the media describe these as breakthroughs and therapeutic triumphs, they are actually enormously expensive makeshift procedures. A final category is "supportive therapy," which is administered when there is no knowledge of how to intervene to alter the course of a disease or its effects.

The paradigm of successful applied science in the Thomas model is poliomyelitis. Scientific progress changed the prognosis from death to life in an iron lung (halfway technology) to prevention by the Sabin vaccine (high technology). The model does suggest that investment in high technology may ultimately be more efficient than investment in a halfway technology, but again the theory needs much more elaboration before it can be used analytically.

The above discussion is not meant to suggest that the two approaches are simple-minded or wrong, but to show that these models are inadequate to the needs of policymakers dealing with real world complexities.

More Complex Models of Science and Technology

No single analytical model adequately captures the processes by which biomedical R&D results flow into medical practice. The search for such a comprehensive model is futile. Rather, we should try to develop a family of models of intermediate generality, able to handle a wide variety of cases with adequate respect for their detail.

Gruber and Marquis developed a general model that might serve as a basis. The model makes the point that scientific development, technological development, and

---

2 Lewis Thomas, "Guessing and Knowing: Reflections on the Science and Technology of Medicine," Saturday Review, January 1973, p. 54. All subsequent quotes of Thomas are from this article.
current practice are concurrent and not sequential processes. It emphasizes the several pathways by which knowledge flows from science to technology to use, and the various sources of innovation in patterns of use, whether previous practices, technological developments, or scientific advances.

Models of this type permit a more detailed set of relationships among the "streams" of science, technology, and their practical use. Each of these streams may be examined for illustrative purposes by reference to Fig. 1. For example, within the general course of technological development are physical engineering and bioengineering. The latter may be further subdivided into prosthetics, instrumentation, artificial organs, and nonthrombogenic materials. General developments within medical practice have been disaggregated in terms of medical practice specialties and families of disease entities.

Interactions occur across the entire map of scientific, technical, and application developments. Chemistry contributes to biochemistry (physical science to basic medical science), physics contributes to radiology (physical science to clinical medical science and clinical practice), microbiology to pathology (basic medical science to basic medical science), instrumentation to neurology (bioengineering to clinical medical science), hematology and materials to nonthrombogenic materials (clinical medical science and physical engineering to bioengineering), etc.

We can observe several things about the pattern of interactions within and among these streams.

- Any basic medical science has its own complex and dynamic set of activities.
- The basic medical sciences interact in complicated ways.
- Progress within basic medical science relates to larger developments in the life and physical sciences.
- Results from an area of basic medical science may flow into several areas of the clinical medical sciences.
- Questions arising in an area of clinical medical science may become important research questions for several areas of basic medical science.
- Progress in the "technology" of medicine may lead to new developments in the basic medical sciences.
- Similarly, progress in or problems arising from medical practice can stimulate new lines of biomedical science or technology.

**Distinguishing Activities Within Science and Technology**

It is useful to disaggregate activities involved in both scientific and technological advancement into three broad categories—"knowledge generation," "advanced development," and "refinement for application." The first of these is often referred to as research (sometimes as basic research or general science as used above); the key characteristic is that we can neither measure the value of the results directly in terms of application-oriented goals nor specify them in advance with any precision. Not only are the values unknown, but often the outcome of the activity will fall totally outside the set of outcomes envisioned a priori.

The interdisciplinary clusters that reported on the state of science for the President's Biomedical Research Panel repeatedly found such unexpected outcomes. In the neurosciences cluster an attempt to isolate a postulated hypothalamic peptide
Fig. 1—Streams of interacting activities—science, technology, and medical practice
that releases somatropin (growth hormone) from the pituitary resulted instead in
the discovery of a factor (somastatin) that inhibited the release of somatropin. It is
now used in the treatment of juvenile diabetes. The microbiology and immunology
cluster showed that work on parasitic infections had implications for many other
problems including kidney and cardiac transplants, tuberculosis, sarcoid, regional
enteritis, and allergic disorders. In "knowledge generation" the uncertainty of
results and the distance between the acquisition of new knowledge and its applica-
tion to relief of suffering are so great that goal-oriented planning or budgeting is
likely to be ineffective. The major considerations in knowledge generating activities
are the technical feasibility of undertaking a particular activity and its importance
in relationship to the perceived gaps in scientific knowledge.

Technical feasibility is also an important consideration in the second category
of activities—"advanced development." By this characterization we mean the initial
use of new knowledge to explore potential applications. At this stage the links to
goals of direct interest to society in terms of improved health are less tenuous, and
the remaining uncertainties in the enterprise are reduced. Enders' successful re-
search on viruses in tissue culture made it clear that vaccines for virus diseases
could be developed. A large infusion of money for advanced development (in this
case, from the National Foundation for Infantile Paralysis) led to the early de-
velopment of two types of polio vaccine.

The final activity—"refinement for application"—involves the design, ex-
perimentation, and testing of practical devices, treatments, or diagnostic procedures
that, if successful, can lead to widespread use and be of direct benefit to society. At
this point the technical uncertainties are considerably reduced and the goals of the
activity are clearly defined. But different sorts of uncertainties appear. There are,
for example, the complexities of transferring useful laboratory results into success-
ful clinical trials, of controlling the experimental design in different institutional
environments, of ethical problems related to the possibility of harming instead of
helping the subjects of such trials, and even of appropriate educational functions so
that practitioners can effectively adopt the newly discovered technology.

Patterns of Interactions Within Science and Technology

The state of knowledge of the interactions within and among the streams of
science and technology is poor. However, we can suggest fruitful lines of research
on two problems in this area—how interaction occurs within and among the basic
medical sciences and how the basic medical sciences interact with the clinical medi-
cal sciences.

Small and Griffith have developed a promising technique that uses citations of
the scientific literature to map scientific specialties and their interactions. The
basic unit of science, they argue, is the individual scientist or scientific work group.

4 Sidney Goldring et al., Report of Neurosciences Interdisciplinary Cluster, draft report to the Presi-
dent's Biomedical Research Panel, October 1, 1975, p. 4 ff.
5 Herman N. Eisen et al., Final Report of the Immunology and Microbiology Interdisciplinary Cluster,
draft report to the President's Biomedical Research Panel, October 8, 1975, p. 61 ff.
6 Citation analysis of scientific literature has led to the development of a number of techniques for
analyzing the underlying patterns of behavior within the scientific community. Citation analysis is based
upon the assumption that the scientific literature represents the revealed nature of scientific develop-
188, 2 May 1975, pp. 429-432 for a general description.
The next highest level of organization of scientific research is the specialty group within science, which they call a "cluster." They isolate frequently cited papers, determine the frequency of co-citation among papers, and identify a cluster on the strength of co-citation links.

Small and Griffith assume that scientific specialty groups, rather than disciplines, constitute the true organizational pattern of the scientific community. Figure 2 represents the specialty mappings for medicine for 1972 and 1973. This mapping technique permits us to observe dynamic characteristics of scientific development. For example, two 1972 clusters—reverse transcriptase and chromosomes—converged in 1973 to form a viral genetics supercluster. In addition, a shift occurred from 1972 to 1973 that resulted in stronger relations between immunology and cyclic AMP. Also, the 1972 cluster microtubule protein emerged in 1973 as an important new specialty: muscle myosin and cytochalasin-B.

This technique suggests the possibility of reconstructing the specialty groupings retrospectively within science, and specifically within medicine, for successive years. On the basis of these successive "samplings" of the specialty groups, it will be possible to analyze the process of cluster formation, growth, merger, differentiation, and dissolution on a longitudinal basis. We will return to this approach in Section V.

A better understanding of the underlying developments within science would be of obvious value in predicting its future course. It would also provide a basis for examining the relationship between federal program emphasis and scientific developments, which is the subject of Section IV.

Although science and technology are often intertwined in contributions to biomedical progress, communication patterns exhibited by the development of technology are different from those within science. First, though there is an abundant literature in most technical fields, technical knowledge cumulates in the elusive "state of the art," which is only partly contained in the literature. "The published paper," Price writes, "is not in general the end product of a work in a technological subject." There is a special literature of technology consisting of engineering handbooks, manufacturers' catalogues, and the advertising pages of engineering magazines. Catalogues are typically bound in loose-leaf fashion, so that old pages can be removed and new pages substituted from time to time. Thus we cannot understand technology from the analysis of technical literature in the way we can understand science.

Interactions Between Science and Technology

According to the sequential model, the interaction of science and technology

---


2 These specialty groups are concerned with particular scientific phenomena and should not be confused with medical specialty groups. See Steve Aaronson, "The Footnotes," Monocle, March/April 1973, pp. 22-27, for the article from which these maps are taken.


Note: Map of biomedicine constructed by Henry Small and Belver Griffith from the 10,000 documents cited 15 times or more in the 1972 Science Citation Index. Application of the clustering program produced several groups of articles, including a 1000-strong group with obvious relation to biomedicine. The map shows this group displayed in its interlinked clusters according to the threshold criterion that, to be a member of a cluster, an article must be co-cited with one or more other members at least 11 times. The numbers on the lines show the sum of co-citations between documents in connected clusters.

Fig. 2a—1972 biomedical clusters
Fig. 2b—1973 biomedical clusters
consists of a flow of information from science to technology. Price has advanced four propositions that contrast with this view. One proposition holds that the typical pattern is one of weak interactions. "Only very rarely," Price writes, "does a new piece of science give rise directly and quickly to technological repercussions."11

A second proposition is that science is "squeezed out of a technical advance." In other words, technology flows into and undergirds science. Radiology is a good example of a clinical science that owes its existence to technological developments having their main locus outside of medicine. The improvements in radiotherapy through increased beam resolution and reduced beam side scatter indicate the debt owed to advances in nuclear engineering and accelerator physics.

A third proposition is that science affects technology "only slightly and with great difficulty through the literature."12 Scientific information must be distilled for the engineer into a more manageable and comprehensible form than found in the scientific literature. Furthermore, the engineer must assimilate it so that he can use that information.

A final proposition is that the interaction between science and technology occurs mainly in formal education when engineers learn their science, and scientists learn their technology. Thus, science and technology are available to each other at some distance from their research frontiers, the distance being "equal to about one generation of students."13

Interactions Among Basic Medical Science, Clinical Medical Science, and Clinical Medical Practice

The above model refinements and remarks on science and technology in general may be useful to understanding medical science, but care must be taken with their application. Each of the models illuminates different aspects of the real world of medical science, but no single model is sufficient to present a comprehensive picture.

Two studies of citation patterns in basic science and applied science suggest some interesting possible linkages between basic science and applied science. One study analyzed the literature of applied chemistry, and concluded "that applied chemistry depends most heavily on other applied chemistry and that it is only weakly coupled to basic chemistry."14 The material cited by applied chemistry was older and contained a much higher proportion of book or monograph material than was typical within basic chemistry. Applied chemistry appears to move independently of basic science, interacting weakly through literature that is clearly not at the research frontier of science. Applied chemistry, therefore, displays many of the attributes Price imputes to technology.

A study of citation patterns involving cancer journals and articles reveals a different pattern. Cancer research could be regarded as an applied medical science, and we might expect it to move independently of basic medical science. But the primary finding of this study is:

11 Ibid.
12 Price, Technology and Culture, p. 564.
13 Ibid.
Articles on cancer research cite basic-research and other non-cancer journals more frequently than they cite cancer journals. In other words, cancer-oriented research today seems to be learning much from basic research and is heavily dependent on non-cancer research.\textsuperscript{18}

What is the meaning of the difference between the interaction of applied chemistry with basic chemistry research and the interaction of cancer research with basic medical science? Are other medical specialties closer to the applied chemistry pattern? Should that pattern be a goal of applied medical research management?

We have no answers to these questions. However, we observe that, in considering various theoretical frameworks of scientific progress, we have reached a point where empirical tests of theory treating questions like those posed above would be useful to federal policymakers. To see why, and to identify the issues for analysis, we now consider scientific predictions explicitly in the federal policy context.

CURRENT PRACTICE OF ANTICIPATING AND INFLUENCING SCIENTIFIC PROGRESS

The National Institutes of Health support mainly knowledge generating activities. Planning for these activities is at two different levels. First are the broad resource allocation plans and decisions on levels of funding across Institutes or major areas within the Institutes. These allocation decisions are not strictly the province of NIH; questions often rise to the Cabinet level and surface during deliberations by the Congress. At a lower level are decisions on project funding. This subject is delegated to the individual Institutes with the advice of the peer review groups and Advisory Councils. The activities underlying program resource allocation deliberations are quite different from those focused on questions of project funding.

Both programmatic resource allocation and project funding decisions implicitly predict the potential for success of scientific activities. A prediction of success is explicit in a peer review judgment that prospective scientific activity, as set out in a proposal, is "meritorious"—will produce valuable results. It is thus important for policy consideration to understand how reliable are predictions of scientific success at the project level.

Assignment of research proposals and grants to Institutes implies a prediction that the results of such proposals will be relevant to the diseases for which an Institute is responsible. The assignment procedure has two basic problems: (1) that worthy proposed research may not be funded because it is only relevant to other Institutes than the one to which it was assigned; and (2) that the funding and successful completion of these proposals and grants will yield information that requires further development and refinement before it is practically useful. Inaccurate predictions of disease relevance may undermine attempts of the government to control the production of useful science by its control of Institutes' budgets, and the budgetary deliberations may be largely self-deceptive.

Programmatic decisions also involve predictions. At the highest level, the declaration of the war on cancer implied a prediction that the war could be won, or at

\textsuperscript{18} Eugene Garfield, "Journal Citation Studies. XV. Cancer Journals and Articles," \textit{Current Comments}, October 16, 1974, p. 5.
least that some ground could be gained against specific cancers. The Advisory Councils often denote areas of particular promise as “high priority areas.” Behind such designation is a belief that more funds will provide even more effective progress in these areas. But at any given time, some discoveries may be “ripe” while solutions to other problems may be impossible (because of basic knowledge gaps). Additional funding in these areas might merely result in wasteful proliferation of unsuccessful approaches. In Section IV, we propose analysis that would examine this problem.

The President’s Biomedical Research Panel has commissioned interdisciplinary cluster reports in a number of areas to assess the state of their field and the adequacy of available funds and manpower. Most of the panels have pointed to problems that scientists are on the verge of solving and extrapolate to the likely implications of the solution for health care. For example, in “Areas of Greatest Promise in Nutritional Science,” predictions run from the specific—“tissue culture systems using human cells are proving increasingly valuable in working out the causes of human lipid disorders”—to the general—detecting biochemical differences in individuals “offers unusual possibilities for improving the health of the nation.”

Such reports are useful since society must turn to the scientific community itself for judgment of feasibility and importance of different scientific activities. However, since this same scientific community is likely to be the beneficiary of any reallocation of resources toward its area of scientific competence, these predictions may appear self-serving. There may be a need for better control and direction of such “advice.”

Government agencies are not alone in predicting the potential for scientific success in resource allocation decisions. Private firms, such as drug companies, invest in scientific activity with the expectation of generating profits; for large firms this activity often runs the gamut from knowledge generation to refinement for application. Foundations and private volunteer agencies also make at least implicit predictions about the potential value of the scientific projects they choose to support. Predictions about the effects of incremental changes in federal resources allocated to programs should include the likely effect of these changes on decisions in the private sector.

**Analyses to Assess Reliability of Current Practices**

The NIH peer review process deals principally with the evaluation of knowledge generating activities. Since the priority scores of this process determine many project funding decisions, the reliability of the predictions of project success is an important question. Analysis can let decisionmakers know how much variability there is in the priority scores suggested by the members of an Initial Review Group (IRG). An experiment across groups—providing a series of proposals to more than

---

16 Evidence of the greater returns from smaller, targeted development efforts over large ones has been revealed, for example, by a comparison of computer hardware manufacturers; see A. J. Harman, *The International Computer Industry: Innovation and Comparative Advantage*, Harvard University Press, Cambridge, Mass., 1971, esp. pp. 69-93.

17 William Shive et al., Final Report of the Nutrition Interdisciplinary Cluster, draft report to the President’s Biomedical Research Panel, October 20, 1975, pp. 21, 27.

18 In fact, few members thoroughly review any one proposal, and the IRG has a proclivity for consensus as a result of the discussion. In such circumstances, an analysis would probably require obtaining individual priority ratings prior to the discussion (with some indication of panel members’ familiarity with the proposal) as well as a final rating.
one IRG—might identify attributes of the proposal (or the scientific field) that lead
to greater or lesser reliability of the priority scores. An NIH Grants Peer Review
Study Team is currently looking at the entire peer review process, but their report
is not as yet available.

There are often ambiguities about which IRG should review a particular propos-
al. The extent of such ambiguities could be examined by observing the correlation
between assignments made by more than one individual. Another possibility would
be to allow the study sections that are assigned proposals to transfer responsibility
for them to alternative IRGs.

It might also be desirable to compare IRG ratings with other peer review systems
(e.g., "mail review" of the sort used by NSF). Such a comparison could be in terms
of the ratings given to a proposal (as a basis for funding) compared with the quality
of the final research as indicated, for example, by citations.\(^ {19} \) It should also consider
the costs to the government and to the scientific community of the ranking pro-
cesses; proposal development and review consume many manhours of effort of some
of the most productive members of the scientific community.\(^ {20} \)

NIH could perform a number of experiments using members of the IRGs to
predict scientific advances. During each IRG session members might devote some
time to consideration of the expected progress of science in the IRG's field of competence
and feed this back to the program planning activities of the NIH managers.
Session-to-session feedback could be institutionalized so that members of each group
could evaluate the quality of their predictions of the progress of science in their field,
as well as the merits of funded projects. After enough experience had been gained
with such predictions, it would be useful to study the causes of success and failure
of the prediction process to determine whether they are purely random and due to
the intrinsic uncertainty of the research activities, or whether there is some discern-
ible pattern. For example, an evaluation of why some high scoring projects had not
produced as expected, or of why some outstanding projects had received scores only
at the margin of the funding level, might reveal consistent and correctable errors
in the review process.

Policy studies of the assignment of research grants to Institutes would focus on
the predictability that breakthroughs in knowledge generation activities will actual-
ly help in the prevention or cure of diseases for which particular Institutes are
responsible. This topic, although formally a prediction problem, is part of our under-
standing of how research progress is translated into improved health and will be
covered in Section V.

A study could also be made of the accuracy of the predictions of the interdiscipli-
nary cluster panel reports. Predictions of scientific breakthroughs and their applica-
bility could be refined by polling the cluster panelists to estimate the likelihood that
a breakthrough would occur within a specified time period and the confidence the

\(^ {19} \) See, for example, G.M. Carter, Peer Review, Citations and Biomedical Research Policy. NIH Grants
to Medical School Faculty, The Rand Corporation, R-1583-HEW, December 1974, in which the NIH peer
review process by itself is analyzed in such a context. A study is currently underway by the National
Academy of Sciences Committee on Science and Public Policy to look at the process of distribution of
government funds for research in all areas and may provide insight into different mechanisms for peer
review and research outcome.

\(^ {20} \) As we articulate a "menu" of possible study activities involving various kinds of policy analysis for
improved policy judgments in biomedical research fields, we are not insensitive to the fact that our ideas
for studies also have costs associated with them. However, it is beyond the scope of this paper to detail
the work statements and probable price tags.
panelists have in these predictions (see the discussion of "controlled feedback" techniques later in this section for an elaboration of how this might be performed). The confirmation or lack of confirmation of these predicted dates may affect the way future resource allocation decisions will be reached.

Analysis For Improving Predictions

How "scientific" are the various approaches to prediction? Let us briefly review a few examples of these approaches.

The General Electric Company's corporate laboratory plans their knowledge generating and advanced development activities by a progression of evaluations of a broad spectrum of "candidate" R&D projects (that in their total exceed funding possibilities). The end result is the year's fundable "operating plan." Initially, GE reviews a program in terms of the corporate objectives, the demand for the (potential) end results, and the resources of the scientific community available to the company. From this review of all of the candidates, the program is divided into advanced development activities and knowledge generating activities. For the advanced development activities, a scoring algorithm has been devised involving eight variables, which cover the probability of the successful completion of a particular project and its success in the marketplace. This algorithm encourages the necessary information exchange between the researchers and others in the corporation and produces a priority ranking that lets top management focus on the 10 or 20 percent of projects near the budget-cutoff region.21

Another technique for eliciting subjective predictions is the Delphi procedure.22 Although this method has been used extensively, little evaluation has been done of its validity and reliability. A recent study has indicated that many applications of Delphi have been seriously deficient in design, execution, and analysis.23 The Delphi technique involves the (usually quantitative) assessment of a subject by panelists who remain anonymous to one another. Through one or more rounds of feedback of the group's responses (usually summarized by a median and some indication of the range of responses) the panelists provide a revised assessment (and give a reason for "extreme" positions with respect to the rest of the panel). Consensus is often obtained, but little has been done to establish the degree of confidence one can place in such a consensus.

Research on a general class of "controlled feedback" subjective methods for predicting scientific advances has been pursued recently at Rand.24 These methods, like Delphi, involve a panel, the members of which are anonymous to each other (possibly at remote locations); and protocols for eliciting judgments and for providing feedback to the panel members. They differ from Delphi in the devising and validat-

21 For information on corporate and long range R&D planning, we are particularly indebted to Rand colleagues G. K. Smith and E. S. Ojana for sharing the insights of ongoing research.


ing of instruments for eliciting judgments, in the type of information "fed back" to the panel, and in the statistical techniques for analyzing the results.

Such methods may be useful in obtaining cost-effective judgments of the prospect for scientific progress if expert panels can be found and protocols carefully developed. The expertise of the panel is crucial. They must have a greater degree of knowledge, intuitive understanding, and ability to predict in a particular subject area than the rest of us. It has been widely accepted that such expertise exists and that it can be impaneled (as the use of IRGs by NIH indicates), especially for the knowledge generating types of scientific activities. However, one should distinguish between questions of technological feasibility and scientific importance and questions of application. The latter require expertise concerning the potential usefulness to society of the end results of the activity.

Less subjective evaluation methods may be feasible on advanced development or refinement-for-application activities, even though separate consideration may still be required on the ways scientific advancement may benefit the users of new, technologically based systems. Several methods have been developed for the assessment of progress in technologies leading to new hardware. For example, one technique relates changes in the attributes of the new hardware over time to the point at which the attributes become available. In the provision of such a service as the prevention or treatment of heart disease or lung cancer, it may be useful to think of various scientific advances as affecting one or more of the attributes of the disease—e.g., the probability of being affected, the effectiveness of the treatment, the flexibility within which the treatment is provided, and the cost of the treatment.

For all methods of anticipating scientific advances, one must raise the issue of how foreseeable scientific progress really is. It is important whether the progress is "within paradigm" or "new paradigm," to use the Kuhn characterizations. The selection of the members of the panel may determine the success of the latter kinds of predictions. Within a particular scientific "paradigm," the problem-solving effort is based on assumptions accepted by a wide segment of the research community.

25 There have been several other approaches within the broad subject of making use of the attributes of a price of hardware (or a service) undergoing change, through the application of scientific advancement. There are also a number of recent books and articles reviewing the state of the art of assessing potential technological advancement by analytical methods. For further details, see A. Alexander and J. R. Nelson, "Measuring Technological Change," Technological Forecasting and Social Change, Vol. 5, pp. 189-203, October 1973; J. R. Nelson and F. S. Timson, Relating Technology to Acquisition Costs: Aircraft Turbine Engines, The Rand Corporation, R-1288-PR, March 1974; and J. R. Nelson, Performance/Schedule/Cost Tradeoffs and Risk Analysis for the Acquisition of Aircraft Turbine Engines: Applications of R-1288-PR Methodology, The Rand Corporation, R-1781-PR, June 1975. For the project funding perspective, the benefits if the research is successful are usually an integral part of the proposal. Even in this case, however, a number of proposals dealing with the same program area may be evaluated in the common terms of the range of benefits to potential users in society.


While changing to a new "paradigm," the underlying assumptions are subject to challenge and dispute.

An instance of the difficulty of predicting paradigm change is found in poliomyelitis research. The work of Simon Flexner at the Rockefeller Institute from 1910 to 1913 resulted in a number of major advances in the understanding of polio. But progress was fairly slow for three decades thereafter. On two issues—the portal of entry of the polio virus and the nature of the causal agent of polio—Flexner's views were later shown to be incorrect. Flexner's dominance in polio research may have caused many other researchers to pursue scientifically unrewarding pathways. However, the problem may instead have been ignorance of the fundamental differences between bacteria and viruses. Until the 1930s, viruses were known only by their pathologic effects. Not until the development of the ultra centrifuge and electron microscope in the late 1930s was the analysis of the structure and composition of viruses possible. Thus, the advance of polio research was dependent upon a number of broader scientific developments—in some sense a change in the scientific paradigm. The difficulties of prediction in these circumstances are readily apparent.

Assessments (e.g., by members of an IRG or other panel) of the potential for success of a new line of inquiry—perhaps involving a new paradigm—may differ radically. One advantage of a controlled feedback process of eliciting judgments from a panel is the possibility of obtaining reliable measures of the degree of consensus of the group about a particular question. These measures may make the degree of uncertainty of various program directions more explicit. In advanced development and refinement-for-application activities, techniques for program planning and resource allocation and for project funding decisions may differ from those currently used within NIH. Cancer planning has been undertaken for more than a decade; the time may be ripe for an analysis of its effectiveness. If federal funding of the more applied stages of biomedical research increases it will be appropriate to tailor the assessment techniques not only to elicit judgment of the possibility for scientific progress but also to devise an appropriate management context to use such information.

SUMMARY

By definition, scientific inquiry deals with unknowns, but decisionmakers must have available predictions of the potential for success and the usefulness of results if they are to allocate scarce resources to such inquiry. Different types of scientific inquiry call for different criteria of usefulness to society and for different methods of anticipating the potential for success. The interactions between science and technology are complex and only crudely understood. Better models of these interactions are needed to enhance capabilities to predict the results of scientific programs and their likely benefits to society.

---

20 Virology was a nascent science during this time: The first course in virology was given in 1922; the first textbook on virology appeared in 1928; and that textbook differentiated bacteria and viruses sharply for the first time.
We have defined three categories of scientific activity—"knowledge generation," "advanced development," and "refinement for application"—that seem particularly useful for biomedical science. The first of these can be predicted, if at all, only on the basis of technical feasibility and the importance of gaps in (current) scientific knowledge. The other two categories can make greater use of society's applications-oriented goals, such as improvements in health status.

Under current practices, NIH deals mainly with knowledge generating activities and uses the implicit predictions from a combination of evaluations by expert panels and Advisory Council reviews. These predictions are based largely on judgments about what is good science—to some degree, what is relevant to particular health problems. Decisionmakers could have greater confidence in the current practices and a better basis for initiating improvements if they knew the reliability of the priority scoring process for proposals, the assignment of proposals to IRGs, and the effectiveness of the categorical structure for funding Institutes.

Further consideration of the types of review activities that are undertaken for (and by) the biomedical research community may be desirable. Appropriate areas for such expanded support without pre-empting the private sector are another important subject for policy analysis. Evaluation of possible changes in management contexts and development strategies—using some of the lessons learned from scientific inquiry in other substantive areas—may also be called for as society's objectives evolve and as pressures grow for the effective use of federally funded scientific inquiry.
IV. THE EFFECTS OF FEDERAL PROGRAM EXPENDITURES ON PRODUCTIVE SCIENTIFIC ACTIVITY

In reviewing proposals for research and evaluating research performed in the past, every funding agency, whether public or private, attempts some assessment of the quality of the scientific activity it sponsors. Section III described ways that this assessment might be improved. The improvement is generally in the nature of better understanding of the qualitative dimensions of scientific activity, not the creation of a single quality measure. To complement the improvement in qualitative assessment, the analysis proposed in this section focuses on improved understanding of how the federal government influences the content of activity, independent of its quality.

The objectives of the federal funding programs are to provide general support for scientific research and to influence its direction. However, federal funds are just one of many influences on the direction and amount of scientific inquiry. A scientist’s choice of work is also guided by private and market demands, scientific esthetics, and his own invested training and research “momentum.” Those who shape federal biomedical research programs are undoubtedly aware that these other factors affect scientific activity. Nevertheless, there is little evidence that decisions on federal programs are considered in the larger context of factors that influence biomedical science. To ignore this larger context is to risk erroneous evaluations.

Considering federal programs in the larger context correctly implies the need for a more comprehensive understanding of scientific activity, but it is nonetheless possible to gain useful information by some analysis that focuses on federal programs alone. In this section, we discuss these analyses first because they are more conventional forms of program evaluation. Then we describe the characteristics of a broad classification system for scientific activity that is prerequisite to comprehensive analysis of the role of federally funded research in the nation’s scientific activity. Finally, we describe several uses of the proposed classification system that appear to be of high priority because of their relevance to federal policy.

THE EFFECTS OF PROGRAM CHANGES ON FEDERALLY SPONSORED RESEARCH

Allocation decisions are made at several different levels. The highly aggregated decisions—how much for defense as opposed to health, how much for health services as opposed to research, how much for cancer research as opposed to heart research—are usually made at cabinet level and within the Congress. Most decisions allocating funds to programs, and within programs to individual projects, are made at the Institute level within NIH.

It is fairly easy to decide which of two projects studying the same problem is better. With some confidence, groups of scientists evaluate and rank research proposals for coherence and ingenuity of approach as discussed in Section III. After the fact, the quality of findings and the conformity of the research to scientific methods serve as criteria for continued funding.
Decisions at the program level are more difficult—for example, whether program expenditures on heart function should be larger than those on artificial blood. Even if it were possible to decide on the costs imposed by various knowledge gaps, and even if we could decide that potential heart program payoffs were bigger than blood program payoffs, we would still be faced with a problem of quality. Should a good project in a high priority field be funded ahead of an even better project in a lower priority program?

Separate studies might address at least two issues using citations data and peer reviews of renewal funding requests. Analysis of the payoffs from projects that were given different priority scores could indicate whether the higher chance of success of an excellent project in a low priority field outweighed its limited scope. The other issue deals with the general progress of science in a given area, independent of project quality. It has been hypothesized that at any given time some discoveries are ripe, but a solution to other problems is impossible because of basic knowledge gaps; and additional funding in these areas merely results in wasteful duplication of unsuccessful approaches. Analysts could look at various crash research projects to determine how much the extra money bought and to find out what distinguished the successful crash project from the failures. The research on limits to the rate of progress in research and development mentioned in Section III would also be relevant here. All of these proposed studies address the question of how much returns to research diminish as more money is put into an area. A knowledge of this question should be useful in deciding how much discretion administrators at different levels should have in determining the allocations to different programs.

Because of federal decisions lowering the priority of certain programs and year to year changes in aggregate federal funding in broad areas, there is growing concern within the scientific community over the potentially adverse effects of discontinuities in federal funding. These effects occur at different levels: (1) the person whose ongoing research is interrupted because of a cutback, (2) the institution whose human and physical research capital is used inefficiently because of erratic funding patterns, and (3) the scientific field whose progress is impeded by the discontinuity, quite apart from the total level of funding.

Distinctions among the different levels at which discontinuities occur are often lost in discussions of adverse effects, as are the distinctions between discontinuities in and terminations of research funding. The effects of discontinuities at the first two levels can be analyzed by case studies focused on federal funding recipients alone. However, at the level of the scientific field they require investment in a more comprehensive system for tracking research activity, which we shall discuss below.

Scientific activity may be affected not only by how much money the federal government provides for research in an area, but also by the form in which the money is provided. In the variety of federal programs for supporting research one can distinguish several broad categories of mechanisms: intramural, contract, large grants such as center or program projects, and the traditional research project grant. Case studies of research performed under each mechanism, along with analy-

---

1 The distinction between discontinuity and termination is important. Termination of federal sponsorship may adversely affect the individual, the institution, and the scientific field, but such decisions are presumably based on scientific evaluations of the long term promise of the particular research area. Discontinuities may occur because of short term factors such as impoundments or year to year target switching.
uses of surrogates for research quality, such as citation data, should illuminate the
characteristics of the research for which each mechanism is best suited.

In interpreting the results of such studies, analysts need to consider the role that
each mechanism plays in the national program for the support of biomedical re-
search and recognize the complementarities among different ones. For example, it
might be true that the traditional research project grant has an advantage for most
knowledge generating research performed within a single discipline but that some
of this type of research must also be performed within the intramural program and
center programs if these institutions are to fulfill their roles in interdisciplinary
research and training.

Many aspects of a research effort might be affected by differences in funding
levels. For example, we would expect an NIH research contract with a short term
to attract fewer Ph.D. students to a scientific area than a research grant with a long
term and a high probability of renewal. It is also possible that different animals
might be used in similar experiments on projects of long and short duration, even
if funded at the same annual rate. Institutions might react differently in treating
principal investigator salaries as allowable costs and in providing facilities for re-
search sponsored by agencies that do and do not pay indirect costs. The NIH in-
tramural and extramural programs do not seem to compete with one another for
funds in any direct way, which would be appropriate if they were strongly comple-
mentary. A systematic examination of differences and similarities would test the
validity of the complementarity assumption.

An important characteristic of the funding process is its efficiency. One measure
of efficiency is the relationship between the resources devoted to program admin-
istration—preparing proposals, making funding decisions, submitting program re-
ports, etc.—and productive scientific activity. In theory, time spent on nonscientific
aspects of a program should be justified in terms of the aggregate quality of the
scientific output, but it is uncommon for decisionmaking regarding program ad-
ministrative burdens to consider this criterion systematically.

It is possible to design studies of the research funding process to deal with some
important questions in this area. In particular, a combination of surveys, small
experiments, and ex post analyses of different funding procedures could shed light
on a number of efficiency questions. Examples include:

- **Length of project funding.** The more often a project comes up for renewal,
  the more time is consumed in preparing and evaluating renewal proposals.
  A simple experiment might assign a set of projects two, three, four, and five
  year funding periods to determine the effects of these differences, among
  other factors, on the production of publications and citations and the priority
  score of the renewal proposal.

- **Content of project proposal.** Members of IRGs are deluged regularly with
  thick proposals to review. These proposals are abstracted by a member of
  the IRG, but it is unclear how different parts of the proposals are used in
  the decision process. An experiment could determine the sensitivity of the
  review process to material in different parts of a proposal.\(^5\)

\(^5\) This experiment could be combined with one dealing with IRG reliability proposed in the previous
section. One concern is that when IRGs are forced to rank proposals that are very close in scientific merit,
they may exhibit "nitpicking" behavior in which they focus on elements that are irrelevant to overall
quality.
Alternative funding for initial proposals. A persistent concern with the NIH peer review process has been its treatment of first time investigators. Although such analysis of the process as has been performed suggests that they are treated fairly in relation to established investigators, it is possible that a more efficient mechanism could be developed for first time investigators. An experiment might be designed to award grants to institutions that could then be awarded to individual first time investigators for a short period. The results of these experimental "seed money" grants could be compared with those of other first time grants.\footnote{A study of this sort is being performed by The Rand Corporation under contract to the Office of the Director, NIH.}

Alternative treatment of approved but unfunded renewal grants. As budget constraints impose greater funding stringency, there is increasing concern over the losses of efficiency due to funding hiatus in research of substantial scientific merit. There may be cost to the investigator, the government, and the institution if an investigator is forced to dispose of such research "working capital" as animal colonies to cut costs while he is resubmitting a proposal to NIH to rectify a minor deficiency. If an examination of the behavior of unfunded investigators showed a high incidence of return to productive research but a high cost of hiatus, it might be useful to experiment with a system of short term, low-level maintenance of effort funding for investigators with priority scores near the "pay line."\footnote{Although this proposal may sound much like the General Research Support (GRS) grant, the purpose is much more explicit. GRS grants were available to the institution to support a wide range of activities other than those of new investigators.}

As funds for biomedical research become scarce, the efficiency of the process becomes more important; the money must be made to count. Inefficiency can occur in at least two ways. First, inferior choices may be made among the research proposals offered by the scientific community. Second, more and more effort may be devoted to justifying the choices among competing proposals with the result that fewer real resources are available for productive research. An efficient awards process must seek a balance that will minimize likely losses of efficiency from the two sources. The literature on organizational behavior suggests that a bureaucracy is at least as prone to the latter inefficiency as to the former.

CLASSIFICATION OF ALL SCIENTIFIC ACTIVITY

An examination of the effects of federal funding on scientific activity cannot progress very far without a system for cataloging activities under federal and other sponsorship. An ideal system would describe activities at the broadest and narrowest policy levels, such as the charter of any one of the categorical institutes of NIH, each of the programs within the Institute, and implications for many disease problems by discipline and subdiscipline, and even within or between disciplines.

Although we do not now have an ideal cataloging system, there are several that we can use to answer some of the questions and that could conceivably be the basis for a more complete system.

The IMPAC file of NIH provides information about applications and awards to
NIH, the Alcohol, Drug Abuse and Mental Health Administration, (ADAMHA), the Federal Drug Administration, the Center for Disease Control, and the Health Resources Administration. It gives information on most of the federally sponsored health research by the awarding organizations and by the programs within most of the organizations. In addition, for many NIH records, the IMPAC file uses a coding system that describes the discipline and field, the body systems, and the research materials to be used and thus allows an analysis of the relationships between scientific activity and program. This system is described more fully in another study. The IMPAC file contains descriptions of applications as well as awards and has the priority score awarded to the application, which can be used as a measure of the scientific merit of the proposal. It therefore permits analysis of the supply of research opportunities by scientific categories and could be used to examine the responses of scientists to shifts in funding levels in different research areas and the response of NIH to shifting scientific opportunities.

The drawbacks of the IMPAC system are that it is restricted to research sponsored by DHEW and, since it uses a fixed coding system, the scientific classification of the research is not flexible enough to describe in detail what a project is about or to accommodate a description of large grants such as program projects and centers. The CRISP system, which uses key words to describe all research projects supported by the Public Health Service, can handle the description problem. Since CRISP is linked to IMPAC, all program and award information in the IMPAC file can be used in analysis of activity using the CRISP system. Descriptions of unfunded applications are not entered in CRISP.

The Smithsonian Scientific Information Exchange covers all federal agencies as well as many state and local governments and private agencies. However, even here the coverage of the private sector is incomplete, and the system does not provide the flexibility of analysis available from either IMPAC or CRISP.

The Foundation Center compiles a record of awards made by about 400 private foundations in many fields, including biomedical science. However, it does not cover some of the largest foundations and does not have a detailed classification of the scientific substance of the research.

Most of the elements of a useful catalog of scientific activity are currently available. Exploratory work should be done on the kind of cataloging system that would be most useful for policy research and an attempt made to include all biomedical research in a single comprehensive system, perhaps by expanding and modifying one of the existing systems.

RELATIONSHIP OF SCIENTIFIC ACTIVITY TO FEDERAL FUNDING

Two of the most important reasons for developing a comprehensive classification system for biomedical research are to determine how the funding source affects scientific activity and how federal funding affects funding from other sources. If a large proportion of NIH projects are indistinguishable from those funded by founda-

---

tions, federal dollars may be viewed the same as foundation dollars and aggregate dollars are what matter. However, if the scientific composition of federal research projects differs significantly from projects funded by private organizations, then federal science policy should concern itself more with questions of balance. Similarly, if federal dollars displace private dollars in a particular scientific field, a different strategy is called for than would be the case if federal dollars attracted private dollars.

The first logical use of the classification system discussed above would be to determine if scientific activity differed by funding source. A pilot effort could combine research funded by the National Cancer Institute intramural and extramural programs with research funded by the American Cancer Society and other private foundations involved with significant cancer-related research. Scientific characteristics from a uniform classification system with funding source identifiers (NCI, ACS, etc.) would then be used in a multivariate clustering routine to develop clusters of related scientific activity. If the scientific activities under different funding are indistinguishable, the funding source will not affect the cluster and vice versa. After the pilot study developed the clustering routine, analyses could use the approach to examine differences in research funded from different sources, including those across various Institutes of NIH.

One of the most important policy issues to be addressed using the proposed classification system is the effect of so-called research targeting by the federal government. Apart from any scientific considerations, the case for targeted research assumes that the action affects the overall composition of scientific activity. The extent of the effects depends not only on the federally sponsored portion but on all research. A well-designed classification system would permit an assessment of targeting on various aspects of research—for example, the disease orientation (cancer), the scientific field (radiation therapy or immunology), or the involvement of human subjects.

A special case of the targeting question is the federal multiplier on private research. It has been hypothesized that the infusion of federal funds in an area attracts private money. If such is the case, federal targeting in, say, cancer or sickle cell anemia will have a multiplicative effect on research in that area. If federal funds instead supplant private funds, the real effects of incremental federal funding would be in the area where the supplanted private funds are applied.

**SUMMARY**

Analysis of the effects of federal program expenditures on productive scientific activity is fundamental to the justification of federal involvement in biomedical research. That justification depends on the proposition that the federal government needs to support such research for the health of society as a whole and that without such support society would get less than it is willing to pay for.

We do not question this proposition. Indeed, we accepted it as a premise. Rather,

---

* Such a clustering routine has been developed for use with the descriptive scientific information in the IMPAC file. Ibid.

* The breadth of the field to be considered for the multiplier effect is obviously an important consideration in the analysis, as well as in the policy.
our objective is to outline analytic approaches that might enable decisionmakers to move from the level of general principles to policy-relevant understanding of the individual programs that compete for scarce resources.

The first logical step in this direction is for NIH to improve its understanding of how its own research funding mechanisms affect the scientific activity of those who receive the money. The effects of length of funding and the choice of funding instruments (e.g., contract or grant) are straightforward problems for analysis. Experiments could be designed to examine the effects of alternative funding approaches for first time investigators and low-level maintenance of effort funding for continuation grants just below the "pay line."

The federal government cannot evaluate the role of its funding in the support of biomedical science in the aggregate until there is a classification system that describes the scientific characteristics of research from all major sources. The system used in NIH’s own IMPAC file is probably closer in content to what is needed than any of the several other systems in use, but its scope is limited. The scope of the Smithsonian Institution’s system is much broader, but it is not well suited to analysis that considers multiple characteristics of research. A comprehensive, flexible classification system would combine the desirable features of both systems.

Once such a classification system is in use, the federal government can begin to make a systematic examination of the effects of its programs on the nation’s health research activities. The first candidates for such an examination should probably be the large, “targeted” research programs of the National Cancer Institute. Data from the classification system would enable NIH to determine whether major increases in its funding in a field attract more funds or replace funds from other sources. The system would also permit an analysis of the effects of discontinuities in federal funding on the activities in a scientific field.

Every scientist and federal program administrator would agree that research activity in his field of interest is relevant whether it is funded by a public or a private agency. Moreover, most federal policymakers would probably concede that there is no reason to spend federal funds for research unless the expenditures increase useful scientific inquiry in the aggregate. Yet there is no place in the federal resource allocation process where research activities funded by nonfederal agencies are considered systematically. Even if there were, policymakers would not find any convenient source of information for such an examination. The feasibility of creating the required comprehensive scientific information system is clear, because a combination of the attributes of existing systems would be sufficient to meet these needs.
V. FROM BIOMEDICAL R&D TO INNOVATION IN MEDICAL PRACTICE

There is widespread agreement that the objective of federal biomedical R&D support is to facilitate the advancement of scientific knowledge in the life sciences to improve health. Substantial differences arise over the appropriate methods by which this objective should be pursued.

Historically there have been two broad strategies for investing in biomedical R&D: heavy investment in basic research across the full range of the biomedical sciences, and the mobilization and coordination of resources to attack a single disease. Typically, an NIH research program reflects a mixture of both. The critical aspect is the degree of emphasis.

The basic research approach relies upon existing institutional mechanisms and incentives for the incorporation of research results into medical practice. The single disease approach more actively mobilizes research resources in an effort to speed the pace by which results are carried into application. Within the past five years, this active orientation has been strongly reinforced with the addition of programs for disease control, professional education, and public education to traditional NIH research responsibilities.

Selection of the appropriate research strategy can be informed and assisted by a clearer understanding of the actual patterns by which biomedical R&D results contribute to innovation in medical practice. Policy research can help with this problem to the extent that it can develop conceptual frameworks of the relationships between biomedical R&D and medical innovation.

As was discussed in Section III, there are many sources of medical innovation. It may stem from improvements introduced by medical practitioners as they engage in the direct provision of therapy to patients, or it may come from clinical discoveries that work in practice but whose mechanisms are not understood. In short, not all medical innovation necessarily proceeds from the base of scientific knowledge.

We tend to think of advances in medicine in terms of major “breakthroughs,” which overlooks the importance of minor innovations in medicine. Substantial evidence from other areas indicates that minor innovations over an extended time account for a great deal of progress.

Because there are no adequate models of the relations between biomedical R&D and medical innovation, we need new methods of analysis. One method is to work backward from identified medical innovations to their scientific and technical antecedents by sampling innovations occurring in a specific time period for “disease state management” activity or for medical specialty groups. Another method is to work forward from a given stream of scientific activity to trace the flow of results into innovations in medical practice. This approach would help to explain the lags between R&D and innovation and the reasons why some results do not affect practice. These approaches are discussed more fully below.

Once we have a better understanding of the relationship of biomedical R&D to medical innovation, we may analyze the federal government’s role in stimulating such innovation. Questions of where the government can intervene pertain to the
competence, resources, and organization of existing federal government agencies. Questions of where the government should intervene have to do with the government's appropriate role with respect to the private sector. These considerations will be discussed at the end of this section.

STRATEGIES FOR STUDYING MEDICAL INNOVATION

The relationships between biomedical research and development and innovation in medical practice are diverse and complex. The sampling strategies we discuss below are designed to bridge important gaps in our knowledge of the sources of medical innovation. They are: (1) innovations in "disease state" management, (2) innovations in medical specialties, and (3) innovations flowing from major streams of scientific-technological knowledge. Before we consider these, however, it is important to think about medical innovation more broadly.

Most popular images of major medical innovations come from such introductions as penicillin and the antibiotic drugs, the Salk and Sabin vaccines for poliomyelitis, the heart pacemaker, and oral contraceptives. These major "breakthroughs" have already been the subjects of considerable analysis.¹ The advantage in analyzing major medical innovations is that the literature is substantial. A prime example of studies of this type is the 1968 TRACES study.² It had the following goals:

- Identify innovations suitable for study and analysis,
- Delineate the research and development events that were key to successful innovations, and
- Analyze the unique or common factors that influence the transition from nonmission research to innovation.

The TRACES group confronted three problems: (1) How to choose innovations that will provide a representative cross-section of historical lineages. (2) How to make distinct and consistent differentiations among the three categories of "scientific advancement." (3) How to determine where to begin and end a historical lineage. Five "tracings" were analyzed: magnetic ferrites, the video tape recorder, the oral contraceptive pill, the electron microscope, and matrix isolation. Recently, a number of other studies have substantially increased our knowledge of the nature and extent of the processes of technological innovation. These include a 1973 study by Battelle Memorial Institute, which considered ten case studies, two of which (the heart pacemaker and oral contraceptives) were specifically medical.³


There are two major problems with studying the relationships between biomedical R&D and medical practice by major medical innovations. First, major innovations are not representative of all innovations. What does the comparison of the heart pacemaker with the oral contraceptive tell one about the similarities and differences between innovation in cardiology and therapeutic drug development? Medicine, after all, is extremely heterogeneous. Second, one learns nothing about the innovations that are overlooked, particularly minor or incremental ones. Although we know little about their place in medicine, the value of incremental innovation in other areas has been shown to be extremely important. For example, in a study of the sources of unit cost reduction in the production of viscose-rayon at several plants of the du Pont Company, 1929-60, Hollander concluded that "the cumulative effect of minor technical changes was in fact greater than that of major changes." Technical change in general was much more important to unit cost and reduction than improved inputs, labor, organization, and plant expansion.

Although they have not been studied systematically, there are few reasons to believe that incremental changes in medicine are unimportant. They are, in any case, more common, more predictable, and thus should be more susceptible to changes in federal R&D strategy than major breakthroughs. To emphasize the need to study incremental innovations is not to underrate the importance of major innovations; instead, it is to broaden our knowledge of scientific contributions to improved health care.

Disease State Management

The study of disease state management would sample medical innovations with respect to a particular disease. One could sample the cardiac patient, the diabetic patient, etc., but one would probably wish to disaggregate the sample to a much more specific medical problem. Diabetes mellitus might be narrowed to the brittle, juvenile-onset, acute-onset, or insulin-dependent patient on the one hand, or the adult, maturity-onset diabetic patient on the other. The first step in this approach, then, is the identification of a set of disease states for analysis.

The management of patients with a particular disease state has probably changed significantly over time. The period of time to be analyzed must therefore be sufficiently long to permit an identification and characterization of each of the major periods of patient management. This characterization would probably emphasize the dominant mode of therapy for each period and the factors precipitating the emergence of new modes. Typically it would range from the stage of nontechnology (or no treatment) through current practice to examination of prospective therapies.

Emphasis on management of patients with a given disease state throws into sharp relief the importance of diagnostic advances. The history of medical innovation is not simply of medical science reducing uncertainty about the state of nature,

---


2 The significant innovations "are the incremental innovations and adaptations of existing technology which contribute to the economic growth and survival of firms and to the technological changes that characterize our times." S. Myers and D. G. Marquis, Successful Industrial Innovations: A Study of Factors Underlying Innovation in Selected Firms, National Science Foundation, Washington, D.C., 1966, p. 19.

but of doing so in relation to particular disease states. The latter, in turn, are a matter of great importance in their own right. When was the disease first recognized? How was it classified? What were the diagnostic procedures used? How did the given disease state become a matter of scientific interest? How did changes in diagnostic techniques lead to changes in understanding of the particular disease? What were the scientific and technological antecedents of changing diagnostic capability? How did changing diagnostic knowledge affect the dominant method of therapy? These and related questions would cast light upon a particular disease and on the means by which it has come to be understood.

Focus on management of patients provides an economical means for aggregating the entire picture of medical advance for a given disease relative to diagnosis, therapy, and prevention. Given the major periods of patient management, how did the advancing knowledge of the disease state interact with advancing medical scientific knowledge in relation to diagnosis, treatment, and prevention? The study of disease state management is likely to aggregate the complex array of innovations in a manner that could never really be captured by a study of major innovations.

Major Medical Specialties

A different approach would be to sample innovations by major medical specialty or subspecialty. Representatives of each selected specialty could be asked to identify important medical advances in the past few decades. These would then be analyzed to establish a profile of the innovation process within that specialty.

The major difference between sampling of medical specialties and sampling disease state management is that medical specialists are likely to encounter a number of differing disease states. Thus one would be sampling not only within a major specialty but across a family of related disease states.7

Sampling medical specialties would show how innovation relates to the organizational patterns of medicine. Many different patterns have developed around the specialized medical practice groupings, and the approach suggested here would permit the analysis to take these into account more adequately than a strict focus on knowledge development would permit. Furthermore, the existence of professional specialty organizations and networks of professionals means that information sources—published and unpublished documents, as well as rich personal sources—are likely to be accessible.

Major Streams of Knowledge

In this approach one would look at the outputs from the R&D activity in selected major streams as they flowed into both disease state management and medical specialties, rather than analyze the scientific-technical antecedents of selected innovations.

One problem would be to identify the major knowledge streams. Some of those streams are not scientific but technological. Consequently, putting questions to

---

7 An excellent analysis of the contribution of medical research to clinical advances in cardiovascular and pulmonary diseases was published just as this report was going to press. See Julius H. Comroe and Robert D. Dripps, "Scientific Basis for the Support of Biomedical Science," Science, Vol. 192, 9 April 1976, pp. 105-111. This analysis represents something of a cross between disease state and medical specialty sampling of medical innovation.
physicians or medical scientists about the major streams of knowledge may generate responses that overlook the importance of physical technology in the medical innovation.

Bioengineering constitutes a major knowledge stream within biomedical R&D. Medical instrumentation is a source of continuing innovation and would fall within this category. Similarly, artificial organs, such as the heart pacemaker or the artificial kidney machine, are developments where physical technology is joined to biomedical knowledge to produce some of the more spectacular changes in medical practice of recent years. Though many of the innovations based on physical technology may fall in the "half-way technology" of the Thomas classification scheme, and thus be regarded as inferior by some, the analyst must still look at them. Indeed, since these innovations raise major problems for the practice of medicine, it is important for us to understand how they are generated and how those processes differ from innovations more directly rooted in biological or biochemical knowledge.

ELEMENTS OF A RESEARCH DESIGN

Because of the complexity of medical innovation, no one of the above sampling approaches can create a complete picture of the process. Multiple approaches are necessary.

The first step in any study is the choice of candidates for analysis. Whatever the basis for sampling, it is sensible to consult expert opinion to identify the important innovations, advances, or changes in medical practice. The use of experts can extend to selection of candidates or confirmations that selected candidates are representative.

The appropriate experts should indicate whether a given innovation is "major" or "incremental." There may be analytic benefits from the distinction. Minor innovations in medicine may have a substantial effect over time and may often be related easily to government investment in biomedical R&D.

Different candidate innovations would have different time frames. The disease state management approach would probably require a longer time than the medical specialty approach since one would wish to know how patient management has developed from the time when the given disease state was first identified and recognized. Studies of the process of scientific and technological innovation have indicated that it takes a long time from scientific discovery to the incorporation of that knowledge into a utilized innovation. This must be reflected in the research design.

The heart of the analysis is a retrospective reconstruction of each selected innovation. The basic element of that reconstruction is the tracing of the scientific-technical antecedents of the innovation. Which were significant, which decisive? Which were directly biomedical, which came from the larger body of scientific and technical knowledge? The methodology of the Battelle study cited above, or some modification of it, might serve as the basis for this portion of the analysis.

Another aspect of the reconstruction is to analyze the passage of the innovation from experiment to established therapy. Fox and Swazey discussed the difficulties in setting a clear boundary. They indicated that researchers, for instance, tend to

---

*See, for example, Everett M. Rogers and F. Floyd Shoemaker, *Communications of Innovations*, The Free Press, New York, 1971.*
see procedures as experimental far longer than clinicians with a strong orientation to patient care.⁹

To complete the picture, it would be worthwhile to reconstruct the process of adoption and then widespread utilization of a medical innovation. The analytical techniques for doing so are worked out in the literature on the diffusion of innovations in organizations. Some problems, however, are peculiar to the health care system. For example, common sense suggests that drug side effects and interactive effects with other drugs will be slow to emerge; these effects are rarer and often less noticeable than the main effects, and little work is done on them.¹⁰ Another problem results from the combination of “technological imperative” and third party payments; a study of overly expensive technologies might uncover desirable ways of influencing such technologies prior to development.

Did the given innovation allow the medical community to do something new, or did it improve established procedures? How did diffusion occur, and how might it be improved? Unfortunately, very little has been written on the spread and effects of new medical technology. One useful study develops a model of the forces that direct technical change in hospitals and the effects of this change on the average cost of care.¹¹

The discussion of the interactions among biomedical research and development and medical practice that lead to medical innovation emphasized the abstract aspects of knowledge development, transformation, and transmission rather than the role of prominent individuals, groups, or institutions. The importance of such actors and the need to address them in an analysis of medical innovation requires some justification for our approach.

In our judgment, it is analytically useful to separate the knowledge development analysis from the institutional analysis. The two are frequently fused in even informed discussion, often with prescriptive overtones. Basic research, for example, is often identified with university and medical school scientific departments, and arguments for continued or expanded support of the former are often proxies for claims for support for the latter. While this behavior may be understandable, it does not help in furthering our understanding of the process of medical innovation.

An understanding of the process of science will lead to better analysis of the institutions involved. This analysis should concentrate on the public agencies, since the purpose of analysis is not only to understand the process of medical innovation but to know where and how and in what circumstances public agencies can intervene intelligently in the innovation process.

---

A COMMENT ABOUT THE STATE OF OUR KNOWLEDGE

The findings of sociology, history, and the other social sciences are potentially appropriate to our analytic task, but they have many gaps, and they have not been


¹⁰ The Boston Collaborative Drug Surveillance Program surveys 24 Boston hospitals for information on side reactions of drugs, but more work is needed. Since these effects are rare, a large data base is necessary.

systematically related from an analytical standpoint. The sociology of science deals with norms and values, with incentives and rewards, with communication patterns, and with the development of scientific institutions. Unfortunately for our purposes, it has concentrated on mathematics and physics, rather than the medical sciences, and has emphasized academic science. There has been little analysis of interactions between the contemporary scientific community and the government agencies that support R&D and of the effects of such interactions upon science.

Medical sociology has neglected biomedical research. It has instead more broadly addressed the health care system without analyzing biomedical R&D as an integral component of that system. The sociology of science and the sociology of medicine have intersected on the ethical questions of human experimentation. Insights on this matter are of great social importance but have limited application to the problems of medical innovation.

The past two decades have seen growing interest in the intersection of the histories of medicine and science. The studies that have been done address both the “internal” developments of scientific-medical knowledge and the “external” factors in such developments. These latter have been studied through biographies and institutional and social histories. However, few of these studies go beyond the 1920s. The major limitations to doing more contemporary work are technical competence and the limited support for such work.

Some fragmented work on the relationship of biomedical R&D to innovation in medical practice has been done within political science and economics. Such research requires both enough historical perspective to comprehend the nature of the process of medical innovation and enough scientific training to grasp the content of the scientific and technical developments that underlie medical innovations. These demands on the background of the researcher prevent most social scientists from concerning themselves with the relation of biomedical science to the practice of medicine.

offsetting the fragmentary nature of current knowledge is the expectation that a modest amount of research support provided within the context of the conceptual framework outlined here could bring together a wide range of pertinent research skills. There are enough research approaches with potential applicability to the problem of medical innovation, and there are enough researchers with an active interest in parts of the problem, at least, for a well-designed program of research support to yield a detailed view of medical innovation within three to five years.

THE FEDERAL ROLE IN INNOVATION AND DIFFUSION

The federal role in medical technology does not end with the sponsoring of scientific research that makes new knowledge available. It may sponsor the complete development of new techniques, evaluate innovations, and encourage the diffusion of innovations throughout the practicing medical community. It may encourage the commitment of private sector resources for science—especially refinement-for-application activities—by the creation of a demand for the end product. The increased demand makes the profitability of such activities more likely. This form of encouragement often occurs when legislation is passed that some segment of the
society shall have increased access to certain forms of medical treatment (such as dialysis) not previously available or affordable.

If instead of increasing the demand for the end products of medical technology, federal funds are directly committed to develop, refine, or diffuse an innovation, these funds may simply substitute for resources that would have been committed by the private sector for the same general goals. Analysis could determine the extent to which this occurs and also identify areas where the market is not functioning effectively and federal funds are needed. Economic theory suggests that the market would provide enough incentives when it is possible to profit from developmental activities, as in the areas of biomedical equipment and drug development.

Even here the situation is not clear, as the development and introduction of the EMI scanner illustrate. In 1917, the Austrian mathematician J. Radon demonstrated that a two- or three-dimensional object can be reconstructed from the infinite set of all its projections. Although this concept recently has proved to have important implications in the development of cross-sectional diagnostic radiology, there was no available technology at the time.

Oldendorf, a neurologist working at UCLA, was dissatisfied with the pain and morbidity associated with the contrast studies necessary for the study of the brain and related soft structures. They have similar densities, so contrast was produced by replacing spinal fluid with air or injecting radio-opaque dyes into the arteries going into the brain. Although he was unaware of Radon’s theory, Oldendorf conceptualized a technique of making measurements of X-ray density with a collimated gamma emitter on one side of the head and a recorder on the other. A large number of projections could be made by rotating the emitter and receiver around the head. Since it is necessary to have a working model to obtain a patent in the United States, although Oldendorf showed the feasibility in 1937, he did not receive a patent until 1963. He built and demonstrated a model using his son’s electric train to rotate the gamma emitter and crystal detector around a skull into which he placed objects he wished to image. He presented this concept to several manufacturers who did not pursue it because they thought the cost of solving the engineering and computing problems involved would far exceed any potential profit.

Unaware of Oldendorf’s publication, which had appeared in a radiologic journal, Hounsfield, an engineer working on radar development at EMI Ltd. in England, started work on the same concept in 1967. Since a working model was not obligatory, it took him only two years to obtain a British patent in 1969. His company was also unwilling to assume the financial risks, but in this case, the British government granted funds for four prototypes, with the idea that there was a fair chance to break even from the sale of this equipment. They recouped their money: Currently several corporations are building and selling these noninvasive cross-sectional X-ray machines.

It may be necessary for the government to play a role in the evaluation and implementation of an innovation even after exploratory applications have proved successful. The Food and Drug Administration has traditionally enforced standards of safety and efficacy for drugs used in therapy. However, there may be segments of the health care industry where additional mechanisms of evaluation are needed.

---

such as verifying the effectiveness of new medical practices by private practitioners. A case in point is the introduction of new surgical procedures without adequately controlled clinical trials. For example, well after the internal mammary ligation for the relief of angina pectoris had been introduced into practice, a clinical trial indicated that a sham operation could give the same relief and the operation was abandoned.

In other instances, the private sector has been slow to adopt a new technology, and the government has intervened in time of national emergency. A case in point is the discovery of penicillin in 1928. Fleming made the serendipitous discovery that a common mold inhibited the growth of certain pathogenic bacteria. Several bacteriologists were aware of this observation and used it as a laboratory tool. The pharmaceutical industry did not pick up the discovery because their typical development strategy was to synthesize organic compounds and then test them thoroughly for a variety of uses (the sulfonamides were discovered this way). In 1938 Florey produced enough penicillin for a small, successful therapeutic trial, but manufacture on a large scale proved to be technically overwhelming. With the demands for more effective treatment for infected wounds in World War II, the U.S. Office of Research and Development referred the problem to the U.S. Department of Agriculture. Using techniques derived from the brewing and distilling industry, they were able to produce penicillin in adequate amounts, first for the military and ultimately for civilian use.

More recently, NIH has taken on a role in development, refinement for application, and diffusion of results through large scale clinical trials, contract programs, and demonstration and control programs. Clinical trials are frequently difficult to design. They normally involve a number of investigators from several different institutions, and prolonged negotiation is often required to obtain agreement on the protocol for the trial. The constraints arising from ethical considerations of human experimentation may be significant. In addition, there are difficulties in maintaining the interest and participation of high quality investigators over the life of the trial.

These development and evaluation activities also raise important resource allocation issues because of the large amounts of money required. A single clinical trial may require most of an institute's uncommitted research funds for several years. It may be more than we can expect from any one agency to strike an appropriate balance between funding many research projects or a few clinical trials.

Because of the size of the expenditures involved, it is particularly important that these trials have careful experimental designs so as to extract as much information as possible for a wide spectrum of issues involving potential applications. For example, it may be possible to design some of the large clinical trials to gather information on more than one disease problem. If this is the case, then the categorical structure of NIH is not appropriate for these activities. The Institute to which the protocol has been assigned may not be aware of its application to other problems. In some instances, additional observations could be "piggy-backed" on the original protocol. For example, for only a modest additional cost, information about cancer or gastrointestinal disease might be obtained from a trial targeted at reducing coronary heart disease.

Further analysis is needed to determine if such large scale projects should be funded separately so as not to be in competition with the usual grant applications.
It may be worthwhile to consider a series of studies that would clarify the circumstances in which such trials should be conducted and the appropriate institutional arrangements within or outside of NIH for supporting them.

**SUMMARY**

There is widespread agreement that the purpose of biomedical research is to generate scientific knowledge in the life sciences in order to improve health. It would be useful to policymakers to have a greater understanding of how medical innovation occurs. First, an empirical picture of the process would provide some common ground for policymakers and biomedical scientists to identify and resolve differences about appropriate R&D investment strategies. Second, greater understanding of how medical science flows into medical practice would provide more information on the possibilities and limitations of government intervention in the medical innovation process. Third, research on medical innovation could bring together several diverse bodies of knowledge in a way that illuminates important policy questions.

The study of medical innovation needs to concern itself with the process of minor medical innovations as well as that of major medical "breakthroughs." The strategy proposed here moves away from an analytical focus on the single innovation and suggests that related innovations, whether major or minor, be studied within the context of (a) the management over time of patients with a given disease problem, (b) the changes over time in the practice of various medical specialties, and (c) the innovations flowing over time from particular knowledge "streams"—specific fields of science or technology.

An improved understanding of the process of medical innovation would provide the basis for determining the federal government’s role in innovation and diffusion relative to the private sector. Insofar as it is appropriate for the federal government to sponsor or perform developmental activities, analysis can shed some light on the appropriate management environment for these activities. Although NIH has taken on some developmental activities, the large cost of these programs compared with costs of most research projects makes it difficult to strike the proper balance between the two, and the possibility of deriving economies of scale by combining activities related to more than one categorical area suggests that other organizational forms should be considered. Analyses of research and development management in other areas may contribute ideas to this study of organizational forms.
VI. THE VALUATION OF LIFE AND HEALTH

Even if we knew in which areas of research scientific progress were most likely, and if we knew how federal funding would affect the rate of this progress, and if we knew how this progress would be reflected in improved health, we would still have to devise a way to attach values to the ultimate products of biomedical research: improved health and increased longevity. We would have to do so both to decide which particular research programs to fund and to determine the appropriate level of funding for biomedical research in general.

It might seem at first that life and health are of unlimited value. A moment's consideration, however, suggests that this is not the case. They are not of unlimited value to society since we often refuse to take some actions—widening a road, for example—that we are certain will save lives in following years. We refuse because we value the alternative uses for the required resources more highly than we do the lives that would be saved. Neither does each of us attach an unlimited value to his own life. We all risk our lives and health every day for minor rewards. We travel by automobile, we ski or climb mountains, we enjoy rich food, alcohol, and tobacco.

Few subjects involve such a tangle of legal, ethical, financial, and religious considerations as do life and health, and it will be impossible ever to capture the richness of attitudes toward their value on a scale that can easily be compared with the dollar costs of a health program. Some aspects of the valuation of life and health will defeat any analysis. Often these are manifested in paradoxes in the attitudes of society toward life saving. The picture of a small child suffering from some serious illness will elicit a flood of donations from a population that may have refused to appropriate money to hire school crossing guards who would save (statistically) several children per year. There are preferences in favor of treating diseases that are esthetically offensive to nonvictims or that have recently struck a prominent person, at the expense of treating other diseases that may be just as dangerous to the population at large. Such preferences for saving an identified rather than a statistical life or for focusing social efforts on the elimination of particular diseases are real. Whether these preferences are to be treated as valid for purposes of policy formation is not clear.

Many criteria have been advanced by which the value of improved health might be judged. Some of these are potentially useful; others can be seriously misleading. In this section we outline the advantages and disadvantages of the most frequently encountered traditional approaches and then introduce a more recently formulated approach based on preferences of individuals for health and life.

SOME TRADITIONAL APPROACHES

Perhaps the most frequently used methodology for valuing a program that reduces morbidity or mortality is the so-called human capital approach. In this view, humans are considered solely as productive entities—capital—and a program is valued in terms of the output that would have been lost in the absence of the program as some of these productive entities became incapacitated because of death.
or illness. The value of an individual's output is assumed to be approximated by his wage or salary, with future earnings discounted appropriately. In the human capital framework, the most valuable program is the one that increases total earnings by the largest amount. This concept is usually behind such statements as "Disease X cost the U.S. $100 million in lost output last year."

At first glance, the human capital approach may appear to provide a straightforward, "objective" measure of the value of health programs. There are, however, some serious flaws in the logical structure of this framework. Should the goal of national health policy be to maximize GNP by keeping the labor force large and productive? If our goal was simply to increase the size of the labor force, the most efficient program would be to outlaw contraceptive devices rather than to invest in health care or prevention of specific diseases. If a productive person dies, society is deprived of his output, but it also no longer has to afford him all those things he consumes. Thus, we should really care about a person's net productivity—what he produces minus what he consumes. This view allows consideration of the benefits of birth control (since an additional person might consume more than he produces), but it leads to other disturbing conclusions.

The human capital methodology obviously fails to apply in the case of the retired. By definition, this group consumes more than it produces. If we consider only gross productivity, there is no reason to save the lives of such people. If we choose to consider net productivity as the appropriate measure, it becomes a social benefit to let them die. Since diseases of the aged are a large fraction of the health problems we face as a society, the human capital methodology is inappropriate to many of our policy decisions.

Earnings are often a misleading measure of an individual's productivity. If it is true that women and blacks earn less for producing the same output as white males, programs aimed at eliminating breast cancer or sickle cell anemia will be undervalued relative to programs aimed at heart attacks, which most frequently strike white males. Further, to the extent that the production of certain groups (most notably housewives) lies outside of the market economy and is thus not valued fully by money earnings, human capital calculations will be in error.¹

Beyond these difficulties, human capital computations simply measure the wrong thing. The reasons for undertaking programs to save lives or prevent illness are much more fundamental than is implied by the human capital construct. We do not desire life and health only that we may work. Life and health are desirable for themselves, and there is no reason to think that the strength of these preferences is related solely to the effect of changed health status on our income.² Although the concept of human capital may be useful for understanding some aspects of individual behavior (the use for which it was originally intended), it is inappropriate for health policy planning.

It has been argued that the amount people do in fact pay for improvement in health should represent a lower bound on the value of these improvements and that


² It is probably true that there would be little loss in GNP if a white-collar worker were confined to a wheelchair. Human capital calculations would therefore not attach much importance to the prevention of such an occurrence. The affected worker, however, would much rather give up his livelihood than lose the ability to walk.
a health program should be assigned a value at least as large as the reduction in
total medical costs that would result if the program were adopted. The first and most
serious difficulty in adopting such an approach is that some of the most important
improvements in health cannot be bought by a private individual at any price. It is
clearly an important characteristic, for example, that a particular program will save
lives, but where can we observe what people are willing to pay to save their lives?

Another source of difficulty is that advances in medical technology often actually
increase direct medical expenditures. To be sure, the quality of care has risen as
a result of these advances, and few of us would prefer that these advances had not
been made; but such advances cannot be justified on the grounds that they save
medical costs. To provide an extreme example, if we were truly interested in doing
nothing but reducing total medical costs, it would be difficult to justify programs to
reduce drownings since these cause fairly cheap deaths of people who might other-
wise have died very expensively from, say, cancer.

There are also disturbing distributional questions involved in this sort of calcu-
lation. It is undoubtedly true that wealthy women spend more money for weight
reduction treatments than do welfare mothers for the prevention of lead poisoning
of their children, but it would be ludicrous to suggest on the basis of this evidence
that a weight reduction program for wives of corporate executives has more social
value than an inner city program to repaint cribs with lead-free paint.

Finally, we must question even that the amounts spent on medical care repre-
sent a lower bound on the value of health services. That an individual consumes a
certain amount of medical services does not necessarily imply he would be willing
to pay the costs of these services. Once the consumer has paid his insurance premi-
urn or become eligible for Medicare, health care is much cheaper to him although
the costs must still be borne by other policyholders or by taxpayers. It is possible that
in some cases the consumer would rather do without care than pay the full costs
himself.

The problems associated with the inability to buy some kinds of health improve-
ments and the distorting effects of health insurance have given rise to a more
sophisticated version of this measure of value. Although only some of us become sick
or die at any given time, we all pay the costs of this poor health in the form of
insurance premiums. If a federal program were to reduce morbidity or mortality,
each of us still might be stricken, but presumably the amount we would pay for
insurance would decrease as our risks fell. It has been suggested that a health
program be valued in terms of the reductions in insurance premiums that would
result if it were adopted. This approach has the advantage of recognizing that poor
health is a problem not only for those unlucky enough to become sick, but for all
of us.

The major shortcoming of this insurance-oriented approach is that insurance
buys not improved health itself but only relief from the financial burdens of ill
health. The fact that a person buys a certain amount of life insurance does not
indicate how he values reductions in his probability of death relative to other
commodities. Instead, it says something about how he values reductions in the
probability that his survivors will be left without support.

3 For an example of this approach, see R. Eisner and R. H. Strotz, "Flight Insurance and the Theory
Even when the benefits of an insurance policy accrue to the victim, we may have no indication of the value of reductions of risk. In many cases the kinds of insurance policies we would like to buy are not offered for sale. It is possible to buy insurance to cover the medical costs of paralysis, say, but it is impossible to insure that paralysis will cause only a minimal decrease in the victim’s quality of life. Such insurance would pay some sum above and beyond medical cost to help compensate the victim for his misfortune. Indemnity insurance of this sort is not unknown, but it is rare. If important types of policies are unavailable, then the amount people actually spend on insurance will not reflect the amounts they wish to spend, and our measure will be inaccurate.

A few other criteria seem to be implicit in some statements that argue for or against particular programs. They do not provide serious possibilities for analysis, but they are worth mentioning. The first revolves around the view that the federal government should be in some ways analogous to a profit-making firm: that any program that will reduce government expenditures (or increase tax revenues) without adversely affecting the health status of the population should be adopted. This sort of argument is most often encountered in arguing for programs of preventive medicine. It seems clear that any program satisfying this criterion should be adopted in the interests of governmental efficiency, but few programs fall into this category. The government is, of course, not a profit-making firm, and one of its primary functions is to absorb the costs of socially valuable activities that for one reason or another do not generate sufficient profit to be undertaken by the private sector.

Another idea for making health policy decisions is the suggestion that we adopt the values implicit in past health policy decisions. But the variance in the estimates of the value of human life that result from viewing past decisions is too large to be useful, and there is a disturbing degree of circularity in using policy decisions to make policy decisions. To its credit, this approach does recognize the ultimately political nature of health policy decisions and makes consistency in governmental decisions over time a desirable goal.

Another argument is that if the federal government finds it necessary to stimulate the economy through increased federal spending, such spending should be for "useful" activities, specifically for medical research, rather than for make-work projects or for simple tax cuts. Obviously, if such a strategy were adopted, medical research would have to be curtailed whenever it became necessary for the government to restrain the economy. The special burdens imposed on the research community by instability in funding make medical research a poor choice for countercyclical spending.

A RECENTLY FORMULATED APPROACH—THE USE OF INDIVIDUAL PREFERENCES

In recent years there has been an attempt to formulate a new approach to overcome some of the inadequacies of traditional measures of the value of programs for saving lives or improving health. For the most part this work has focused on the

4 For more on this subject, see Jack Carlson, "The Valuation of Life Saving," Ph.D. dissertation, Harvard University, 1963.
use of individual preferences for life and health. Clearly people vary with respect to the risks they are willing to take with their health; and to the extent that we believe in policy formation by means other than benevolent despotism, these varying individual preferences should be taken into account. The problem is to devise some way to articulate these preferences (about what are certainly difficult subjects) so that they are useful in planning public policy.

We can view a given health program in two ways. First, we may focus on the huge gains it provides for a few members of society. A disease detection program, for instance, may save a hundred lives each year. To value such a program we would need to value each of those hundred lives. This is more or less the view adopted in the human capital methodology. Alternatively, we may focus on the much smaller benefits the program provides for a very large number of people. The detection program is viewed as slightly reducing the chance of death this year for all of those served by the program, presumably many more than the one hundred who would actually have died in its absence. This is the view in the insurance-oriented approach. What is offered is a contingent commodity—the move from one set of probabilities of death and illness to another presumably more favorable set of probabilities. This is the conceptual orientation for the individual preference approach.

We prefer this second approach for two reasons. The first is that in the context of individual preferences there is simply no such thing as the "value of life." Most of us would be willing to give up everything we had to stay alive, perhaps having made some provision for the support of our children. If we could borrow two or three times that much, we would gladly pay it. The fact that most of us are willing to spare no expense to save our own lives does not imply that society feels the same. As suggested above, most of us are also willing to accept increased chances of death or illness for finite benefits in the the form of money, time, and convenience. Thus trying to value "life" or "health" per se is not likely to prove fruitful.

The second reason for choosing the probabilistic view of the benefits of the program is that we are seldom faced with public choices of the form, "Shall we as a society keep citizen X alive?" Instead we are constantly required to decide whether to supply some service that we may reasonably expect will save some lives but whose effect on any specific life is unknown.

Even though an individual cannot sensibly answer the question of how much his life is worth to him, he can say that he would pay some amount far less than his full wealth to reduce his chance of death or illness by a small amount. Conversely, he would be willing to incur small increases in his chances of death above those he normally bears if there were sufficient return to taking such extra risks. It is not claimed that it should be easy for him to decide what trades of this sort he would be willing to make or even that he makes these decisions consciously. Many of these decisions might more properly be classed as nondecisions since they result from habit, routine, or rules of thumb. The fact remains, though, that such trades are made and that they would continue to be made (although possibly in different patterns) if each of us carefully considered each decision.

---


6 If a person is willing to pay $50 to reduce his probability of death this year by one chance in one thousand, it does not mean that a year of life is worth $50,000 to him. It means nothing more or less than that he is willing to pay $50 to move from his present state to one that is identical except that his chance of death this year is reduced by 1/1000.
In theory, it might be possible to observe carefully what trades people make and infer from these the rough outlines of their preferences. In practice, however, this is unlikely to be successful because so many of these trades are made carelessly or in ignorance of their true implications. It is also unlikely that we could find situations in which the trades were sufficiently simple to allow us to identify what is in fact being gained in the trade. Changes in risks are often accompanied by such other changes as different working conditions, increased opportunities, greater mobility, and so on. The difficulty of valuing these other changes will confound the process of valuing changes in risk.

If direct observation is impossible, we are forced to consider posing hypothetical questions about such trades directly to individuals. Surveys are by their nature difficult and the results always subject to question, particularly when sensitive issues such as those involved here are treated. If we are to learn anything about individual preferences as they relate to life and health, however, there seems little alternative to a survey. Whether a survey instrument that yields interpretable responses could actually be constructed remains in doubt, although some preliminary attempts appear promising.

Ideally, we would ask two kinds of questions. The first would seek to determine, for example, what risk of death an individual might be willing to accept to avoid being bedridden for life. The answers to questions of this sort would allow the construction of a scale (formally known as a von Neumann-Morgenstern utility scale) along which could be placed various states of health ranging from perfect health to death, permitting comparison of the relative attractiveness of quite complex situations involving different probabilities of being in these states. It would be possible to determine on the basis of a few questions whether the respondent would prefer that his chances of finding himself in some state A be reduced by some amount X rather than that his chances of being in state B be reduced by an amount Y. This is the same as preferring that a program aimed at disease A be undertaken rather than one aimed at disease B.

The second kind of question would explore what risks the respondent would take for certain monetary rewards. These questions would allow us to determine how much he would pay for reductions in his chances of death or illness. It is likely that all of us would encounter much more difficulty in considering the sorts of tradeoffs posed in this second type of question than we would when we are required only to compare the (un)attractiveness of various states of health. Questions of this second type are necessary if we are to decide how much to spend on a particular program; they are not required if we need only to compare the benefits of two competing programs.

The entire structure of this approach can be symptom-oriented. That is, a respondent may be asked his feelings about being bedridden, restricted in movement, etc. It seems much more likely that a person will understand what is involved in suffering a particular symptom than in having a given disease, and it is the symptoms themselves rather than the name associated with them that are the concern.

of public policy. Further, a wide range of diseases could be characterized, at least roughly, by a small number of symptoms, and the number of specific preferences that would have to be probed could be kept to a minimum.

This methodology concentrates on the value of a person's health to himself. We could word questions to incorporate the preferences of the respondent's immediate household. Further, it is possible to adjust our questions and interpretation of their answers to recognize the financial interest of others (principally his insurers) in a person's health. We could, then, capture most of the relevant preferences, leaving out only the humanitarian instincts of unrelated people. It is likely that these are negligible compared with those preferences we have taken account of, and thus our conclusions should not be far from the mark.

It is important to stress that what is suggested here is not just another poll of the sort: What do you think is a more important health problem, cancer or heart disease? The proposed survey would be designed to elicit preferences in such fundamental situations as suffering pain to avoid risk of death or spending money to avoid a likelihood of pain. Modern economic theory allows these preferences to be incorporated directly into a framework that values changes in the probability of ill health.

The articulation of individual preferences is not all that is necessary in an analysis of the social value of a research program. Somehow these individual preferences must be aggregated to produce what might be called a societal preference. This will not be simple or straightforward because preferences differ among individuals. An athlete, for example, might be expected to value his ability to run more highly than would other people. Since any policy decision must inevitably conform more closely to some preferences than to others, some scheme will have to be devised to attach the appropriate relative weight or importance to each individual's preferences. We are constrained by law and custom to treat all individuals equally. Disease, of course, is not similarly bound; and a decision to fund research in any particular area has implicit in it some value judgment about the relative importance of preferences of different groups. The choice to do research into heart disease rather than into cancer implies a high weight for the preferences of men as opposed to those of women. To combat obesity rather than malnutrition may be to favor the rich over the poor.

It is not enough to say that we require equitable treatment for all groups. There are many specific definitions of equity. We could assign equal values to the deaths of all citizens; we could assign equal values to a particular kind of pain for all citizens. We could seek equal improvements for all, or we could equalize per capita expenditures for all. This list could go on and on. Each of these definitions of equity has some attractiveness, but each would lead to a different set of priorities for biomedical research.8

---

8 One particular interpersonal value judgment that arises with some frequency and does not seem appropriate is that implicit in the following use of individual preferences. A survey is designed to discover how much each individual would pay for a particular reduction in his chance of death. These amounts are then simply added up, and if the total exceeds the costs of the program providing such reductions, the program is deemed worthwhile. The implicit concept of equity in this case is that a dollar a poor man is willing to pay is worth exactly the same as a dollar a rich man is willing to pay. (A similar concept of equity lies behind most applications of human capital.) Because a rich man is willing to pay more for anything than a poor man, he is willing to pay more for improvements in his health; by this measure, research aimed at diseases of the affluent would have an advantage over research aimed at diseases of the poor. Because there is a strong social tendency to view the right to health as independent of income, such an application should be considered suspect.
There can be no objective way of choosing among these concepts of equity. This is a serious stumbling block if we are seeking a simple approach to the value of life and health. However, if we are seeking useful insights into this complex subject, the lack of a simple equity rule may be quite useful since it forces us to face the fact that any analysis of the value of improved health will contain an important political component. Indeed, it is perhaps the greatest attraction of the individual preference approach that it focuses attention on these ultimately political issues. Where other methodologies give a misleading impression of "objectivity," this approach makes clear the two-step nature of the process of attaching a social value to anything: First we must ascertain individual preferences, and then a political decision is required to indicate how to aggregate these preferences.

It may in fact be in the broader interests of society not to force such difficult questions as the relative value of different lives. Any attempt to do so would require the explicit resolution of such potentially divisive issues as whether the lives of the elderly were as valuable as those of the young or whether the health of highly skilled workers is to be insured before consideration is given to the health of the unskilled. If there is no explicit set of interpersonal value judgments, however, there can be no meaningful content to an analysis of the value of improved health. We may finally be forced to choose between having a firm analytical basis for policy planning and avoiding socially unpleasant considerations. In this sense some of the traditional approaches that gloss over some serious distributional issues may represent sensible compromises.

**SUMMARY**

In the absence of some difficult and quite specific social and political choices as to the relative importance of various individuals, there can be no meaningful measure of the value of improved health for any group. If decisions are to be based on such valuations, a fairly complex methodology such as that based on individual preferences may be required. Simpler methodologies obscure the distinctions between the analytically tractable and the politically determinable and are likely to lead to the sorts of contradictions outlined above. There will never be a single number for the value of life and health. The best we can hope for is a rough comparison—subject to different political or value judgments—of how relief of some kinds of ill health would be valued relative to others, and an even rougher guess of whether the money being spent is justified. Although this information would not supply simple answers to questions of policy relevance, they would illuminate to some degree a process of valuation that now must be carried on in almost total ignorance.
VII. CONCLUSIONS

This essay has addressed two questions that may appear similar on the surface: First, is it possible to perform analysis—cost benefit analysis—that will provide answers to the questions of how much and where the federal government should make its investment in biomedical research? Second, if outright answers are not possible, is it possible to perform analysis that will lead to better informed judgments as to how much, where, and in what manner the federal government should commit funds to biomedical research.

The answer to the first question is simply "No." Much of the information needed for cost benefit analysis is not known; some of it will never be known. We presented a nontechnical description of the underlying concepts of cost benefit analysis and illustrated its application in a simple example. The reasons for wanting to do cost benefit analysis make it appropriate for biomedical research, but requisite information for such analysis is almost totally lacking. The reason for devoting a section of the essay to cost benefit analysis to arrive at what may seem an obvious conclusion is to make it clear there are no simple answers to important policy questions in the biomedical research sphere.

The answer to the second question is "Yes." However, there is no single analytic approach that leads to confident biomedical research policy decisions. Instead, there are many questions that can be addressed by many forms of analysis. Since analysis itself is both costly and time consuming, judgments will have to be made about the highest priority topics of analysis. Policymakers ultimately have to make these judgments themselves, but they can benefit from the advice of experts, such as those on the President's Panel, and from the practitioners of policy analysis.

To provide a structure for considering the potential value of analysis of questions related to biomedical research policy, we examined four broad subject areas: (1) the predictability of scientific advances, (2) the links between federal program expenditures and scientific advances, (3) the relationship of scientific advances to improved health, and (4) assessments of the value society places on improved health and longevity. Analysis in these four areas would provide some but not all of the information necessary to link biomedical research program costs to their benefits. That is, in the first three areas we describe analyses that would provide information relevant to estimating how much it would cost in federal research dollars to bring about some desired change in health. In the fourth area we consider the problem of determining how much such a change would be worth to society.

In each area we have discussed a set of questions that are both relevant to policy and show promise of being analytically tractable. We do not expect definitive answers to all of the questions we have outlined, but we are confident that well-designed analysis can yield useful information in almost every case.

We have summarized important topics for policy analysis at the end of each of the four preceding sections. Rather than repeat the summaries here, we recapitulate only what we view as the central points in each section. Then we venture our own views on priorities for analysis in each area.
THE PREDICTABILITY OF SCIENTIFIC PROGRESS

Although predictions about scientific progress involve irreducible uncertainty, they are often sought, made, and valued. Policy analysis lacks a systematic examination of predictions—particularly implicit ones—to determine their reliability and the measures that can be taken to improve them. In general, the main ingredients for such an examination are at hand: a well-organized and generally well-documented peer review process, citations data, and periodic predictions of a comprehensive nature.

To use these ingredients effectively requires circumspection while making predictions and retrospection in evaluating past predictions. We need to make some implicit predictions more explicit and to record more about how all predictions are made. We need to assess the accuracy of past predictions and to discover potential sources of improvement.

In our view, analyses of the peer review process can make the most immediate contribution to more confident decisionmaking. The peer review process is central to resource allocation decisions at NIH and is viewed by many as the strength of the NIH system of research awards. Although no analysis to date has discredited peer review, there are many who question some aspects of its reliability and some who doubt its validity in general. Experiments dealing with reliability are straightforward and could be performed at modest cost. They would provide answers to legitimate questions and they might also lay some doubts to rest.

We accord lower priority to studies designed to improve the overall understanding of the progress of biomedical science because they deal less directly with the immediate concerns of policymakers. By their character, these studies involve knowledge generation and, consequently, their outcomes are uncertain. In the longer term, however, they may be of greater value in improving the quality of federal programs than the more narrowly based experiments we propose.

THE EFFECTS OF FEDERAL PROGRAM EXPENDITURES ON PRODUCTIVE SCIENTIFIC ACTIVITY

Although the purpose of its programs is to influence the rate of progress and overall direction of all biomedical science, the federal government makes resource allocation decisions without systematically considering the broader context of these decisions. Because of its preoccupation with effective evaluation of individual projects, the government may easily overlook the medium and long term consequences of its short term funding actions and the interrelationships among its own programs. More comprehensive analysis of the federal government's own programs can be performed with currently available data and methodology, but evaluation of its programs in the broader context will require the development of a new information system.

In our view, a simple ordering of priorities is inappropriate for analysis of federal effects on scientific activity. Instead, the government should simultaneously adopt a short term and a long term strategy to deal with these questions.

For the short term, we would attach high priority to analyses aimed at discovering ways to improve the efficiency of different federal funding mechanisms. Fairness of funding decisions should not be sacrificed to gain efficiency. However, if the
awards process can be made to operate as effectively using fewer resources, more resources will be available to support productive scientific inquiry.

For the long term, we would attach the highest priority to developing a comprehensive scientific classification system. The cost and time required for such development should be modest because some desirable characteristics are present in several existing scientific information systems, including NIH's own IMPAC file. Since the long term value will depend on its soundness, the development should not be undertaken as a crash effort. If a comprehensive classification system were in place now, there would be immediate demand for its use in biomedical policy research, and in the course of its development, more demands will be generated.

FROM BIOMEDICAL R&D TO INNOVATION IN MEDICAL PRACTICE

There is widespread agreement that the purpose of federal biomedical research expenditures is to discover ways to improve health, and there are numerous examples of the contributions of science to health that justify past federal expenditures. Although that justification must inevitably depend on the final linkage to health care, the government has not systematically devoted its attention to the final stage of applications. In part, this important gap may result from inadequate understanding of the process by which scientific progress is translated into health care.

There are no well-developed research methods to study the translation of science to health care, much less the federal role in that process. The most promising analytic approaches are case study methods like those used to trace the scientific lineage of major technological breakthroughs. Unlike most past studies of breakthroughs, the studies we propose extend beyond the development stage to implementation in medical practice.

We believe the highest priority in this area should be given studies that deal with incremental improvements in medical care. This is not to undervalue the importance of "breakthroughs" but rather to focus analysis where least is known and where the scientific contributions should be more common and susceptible to influence. Within these studies of incremental improvements, priority should be given understanding of the development and implementation phases where the federal and private roles are least understood.

THE VALUATION OF LIFE AND HEALTH

Federal research expenditures aimed at improving health and extending life imply that society attaches values to the expected outcomes. Indeed it does, but the valuation process defies description, much less meaningful measurement. The methodologies commonly used to assign them value shed some light on particular aspects of better health and longer life. However, these methodologies often emphasize easily quantifiable characteristics at the expense of the more truly relevant but obscure. Attempts to assign a single number value to life and health must be discarded in favor of analysis that acknowledges the complexity of the question and accepts the more modest goal of improved understanding of social preferences.
We recommend an approach that involves analysis of a carefully prepared survey to elicit preferences of individuals with respect to life and health and other things of value. It is experimental, and the results it will yield are uncertain. In no circumstances will it supply simple answers, but it may illuminate to some degree a process of valuation about which little is known.