11634

110548

REPORT BY THE

Comptroller General

OF THE UNITED STATES

Science Indicators: Improvements Needed In Design, Construction, And Interpretation

The National Science Board's Science Indicators reports are collections of measures which attempt to portray significant trends in the condition and direction of U S science and technology Development of these indicators is a very complex and difficult task, and the art is still in an early stage of evolution

The National Science Board and the National Science Foundation staff should continue to experiment in the Science Indicator series by developing and testing new indicators They should emphasize a more conceptual approach which first identifies what will be measured, and then generates the appropriate data Attempts should be made to develop indicators of the process and substance of research and to better differentiate between science and technology More interpretation of the meaning of indicators should be included in future reports

UNITED STATES GENERAL ACCOUNTING OFFICE

PAD-79-35

SEPTEMBER 25, 1979

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON D C 20548

B-135547

The Honorable Lloyd M. Bentsen, Jr.
Chairman, Joint Economic Committee
Congress of the United States

TNT00100

Dear Mr. Chairman:

As part of the Special Study on Economic Change, the past Chairman of the Joint Economic Committee, Representative Richard Bolling, asked GAO to review the measures used to indicate the state of U.S. science and technology. This report examines the measures presented in the biennial National Science Board's Science Indicators (SI) reports, with emphasis on SI76, the most recent edition at the time of this study. GAO reviewed the indicators in SI76 to determine their validity, their limitations, and possible improvements in their selection, design, and interpretation.

The subject studied in this report concerns many legislators. Accordingly, we are sending this report to several other interested congressional committees.

Copies are also being sent to the Director, Office of Management and Budget; the Director, Office of Science and Technology Policy; the Director of the National Science Foundation and the Chairman of the National Science Board; and the Departments of Commerce, Defense, Energy, and State.

Sincerely yours,

Comptroller General
of the United States

COMPTROLLER GENERAL'S
REPORT TO THE
JOINT ECONOMIC COMMITTEE

SCIENCE INDICATORS: IMPROVEMENTS NEEDED IN DESIGN, CONSTRUCTION, AND INTERPRETATION

D I G E S T

BACKGROUND

The Joint Economic Committee requested, as part of their Special Study of Economic Change, that GAO examine the measures used to indicate the state of U.S. science and technology. To do this, GAO reviewed the National Science Board's biennial Science Indicators reports (particularly Science Indicators 1976--the latest edition) because they contain measures which attempt to portray significant changes in the state of science and technology. These measures are potentially a valuable resource in Federal decisionmaking. Indicators in the 1976 report were examined to determine their validity, their limitations, and possible improvements in their selection, design, and interpretation.

The development of science and technology indicators is extremely difficult for many reasons, including: the complex nature of science and technology, the diverse and pervasive way both interact with society, and primitive understanding of the processes and linkages involved. These factors greatly impede selection of important concepts or kinds of information and measures. Also, much of the terminology generally used to develop measures (e.g., "health" and "vitality" of science) is vague and evaluative. (pp. 14-17).

In addition to these overall problems, the Science Indicators series is confused by disparate statements of purpose that reveal disagreement about whether (1) the reports should include evaluation of the state of science, and (2) whether technology should receive as much coverage as science. From the different statements (which to some extent indicate changes in perspective with

Tear Sheet Upon removal the report cover date should be noted hereon time and experience), GAO distilled a common-denominator purpose: the reports should quantify as many of the dimensions of both science and technology as is feasible, with a view towards identifying significant trends and developments, interpreting possible causal relations, and analyzing their possible meaning for the condition and direction of science and technology in the United States.

GENERAL PROBLEMS WITH 1976 REPORT

The implicit model of science and technology inferred from the selection of indicators in the 1976 report is too narrowly based on the notion that science and technology are directly related by cause and effect and are to serve societal needs exclusively. This simple input-output approach does not sufficiently differentiate science from technology and views these activities only in terms of economic resources and tangible products. It leaves out both the process of scientific work and the substance of scientific knowledge (pp. 18-23).

A major limitation of the 1976 report is lack of interpretation of the meaning of the indicators. This is at odds with the purpose of the report and limits its capability to describe the state of science and technology.

DIFFICULTIES WITH PARTICULAR INDICATORS

Some indicators in the 1976 report are improperly conceived and need to be reworked (p. 23). There are measures without proper conceptual development, some of which have been poorly constructed by uncritical adoption of economic indicators (e.g., research and development divided by the gross national product (p. 25). Indicators based on patent data, i.e., declining rates of U.S. industrial patenting versus that of foreign corporations in the United States, have little value due to incorrect assumptions about industry's incentives towards patenting. The Science and Technology Annual Report provides more realistic indicators using similar data (pp 27-28)

Other appropriate and important indicators were underused in the reports, or their limitations were not spelled out sufficiently. Indicators on technological innovation seem well-made and reveal significant information about the innovation process, but they have underlying limitations which need to be made explicit (p. 30). GAO believes the 1976 report was too conservative in its use of bibliometric data which involves counting both the number of publications in a given field and the number of citations to certain publications. Bibliometric data seem to have much potential for measuring both the process and the substance of science although they also have important limitations (p. 32).

The public attitudes survey done for the 1976 report suffers from (1) considering the public as a single, homogenous mass; (2) not separately surveying any of the scientific "public"; and (3) making all questions evaluative (i.e., based on good/bad distinctions)(p. 34).

RECOMMENDATIONS

GAO recommends that the National Science Board direct the Science Indicators staff to:

--Use different models of science and technology to present a spectrum of important concepts which need to be measured. Particular attention should be given to developing indicators of the process and substance of research.

--Emphasize a more conceptual approach in designing indicators which first identifies what will be measured, and then generates the appropriate data.

--Include overall interpretation of the meaning of the indicators without emphasizing short-term topical policy issues.

--Consider alternative indicators suggested by GAO.

--Continue to experiment in the series by developing and testing new indicators, and by reevaluating and improving old ones.

--Attempt to more clearly differentiate science from technology and develop distinctive indicators for each.

--Consider whether sufficient resources are available to perform essential research and experimental development of new and improved indicators.

AGENCY COMMENTS

The National Science Foundation reviewed a draft of this report and provided lengthy comments. The full text of the Foundation's comments and GAO's responses are in appendix II.

The Foundation noted that GAO's report is based on an analysis of the 1976 report and that many changes, consistent with GAO's recommendations, have been made in the 1978 volume, due to be released soon. Other key comments were:

--It is very difficult, and possibly not currently feasible, to develop models of science and technology on which to base the selection of indicators. Furthermore, indicators which are independent of the choice of a model may be the most useful ones.

--Science indicators used are generally based on definite concepts. Frequently multiple concepts are reflected by a single indicator or a combination of indicators.

--Every attempt has been made to clearly differentiate science from technology, wherever this is possible or appropriate.

GAO believes it is important that indicators be selected according to a variety of explicit, experimental models in order to gain a broad perspective of science and technology. Additionally, a conceptual approach to indicator development should be emphasized. Some of the existing concepts used by the Foundation are too broad and should be redefined. Finally, GAO does not agree that the Foundation has adequately separated science from technology, particularly in the chapter on public attitudes in the 1976 report.

Contents

ABBREVIATIONS

AAAS	American Association for the Advancement of Science
GAO	General Accounting Office
NSB	National Science Board
NSF	National Science Foundation
OECD	Organization for Economic Cooperation and Development
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
NASA	National Aeronautics and Space Administration
SI	Science Indicators (the two digits following SI indicates the year of the report, e.g., SI76 is Science Indicators 1976)
SSRC	Social Science Research Council
STAR	Science and Technology Annual Report

CHAPTER 1

INTRODUCTION

For its "Special Study of Economic Change," the Joint Economic Committee requested the General Accounting Office to

> "* * * (1) assess the validity of the data bases and interpretation of the surrogate measures used to measure the quality and vitality of U.S. science and technology; and (2) in particular critique Science Indicators and related reports dealing with the matters mentioned above."

The main objectives of this study were to evaluate the measures of science and technology now used by Federal decisionmakers and to suggest criteria for evaluating future measures.

SCOPE OF GAO REVIEW

Science Indicators, 1976 (SI76) is the main focus of this report. SI76 (published in February 1978--several months late) is the third and most recent of the biennial science indicator reports published by the National Science Board (NSB). (In response to congressional mandate, NSB has published an annual report on issues in science and technology since 1969. The Science Indicator series began with the publication of SI72 in 1973 and has been published as NSB's annual report for each odd-numbered year, alternating with reports on more topical issues.) Some attention is also given to the first Science and Technology Annual Report (STAR), published in September 1978. It was written primarily by the staff of the National Science Foundation (NSF) and includes an overview by the Office of Science and Technology Policy (OSTP). A third closely related report, also legislated by Congress, is the Five-Year Outlook report which is scheduled to appear in January 1980. These reports (SI, STAR, and the Five-Year Outlook) presumably will give Congress a thorough description of the current state of U.S. science and technology.

Our critique of the SI series, particularly SI76, is based on comments and suggested improvements found in relevant literature and interviews with officials from NSF, OSTP, congressional staffs, and other organizations that use Science Indicators.

TERMINOLOGY

The terms defined below are offered both as a glossary for this study and as suggestions for future common use.

Measure	- A numerical surrogate or substitute for, or a partial description of, some empirical aspect of a phenomenon.
Indicator	- A measure selected for elucidating a particular concept. For a full discussion of this term, see page 7.
Concept	- An idea or notion of some intangible social phenomenon or activity; such as poverty, intelligence, inventiveness.
Model	- A description of reality that removes the most important parts from it and reassembles them for easier understanding, analysis, and manipulation. Theories are considered highly complex and explicit models which attempt to define interrelationships and causation.
Interpretation	- Explanation or descriptions of various trends or interrelationships derived from data.
Evaluation	- Judgment concerning the significance, degree, condition, or validity of data and trends and their interpretations.

BACKGROUND

The development and use of science and technology indicators are part of a movement to quantify and thus more clearly define, understand, and measure social conditions and trends. Beginning with economic indicators, quantification now includes social indicators, which have been used in a variety of ways, including social and environmental impact assessment.

The impetus for creating economic indicators began with the desire to measure the extent of various economic problems during the Great Depression. Initial studies were made, biases and conceptual problems were slowly uncovered and resolved, and more reliable measures were developed. This process took a long time (at least 19 years for reliable unemployment measures, for example), and much conceptual work was necessary before reliable indicators were devised to satisfy a perceived need.

The idea for charting social trends began with a presidential committee that issued a report in 1933 entitled Recent Social Trends. This report, and similar efforts in the following years, had little affect on policymaking. Most social scientists attribute the recent social indicator "movement" to a National Aeronautics and Space Administration-sponsored study by the American Academy of Arts and Sciences completed in 1967. Social indicators are still relatively new, although their use and importance continue to increase. As with economic indicators, experimentation over time has been elemental to the success achieved thus far.

SCIENCE AND TECHNOLOGY INDICATORS

Present indicators and other measures of science and technology began with NSF's collection and publication in 1953 of periodic surveys on "the funding and performance" of research and development (R&D) in the United States. In 1968, the Congress mandated an annual authorization for the NSF budget, rather than the continuing authorization which NSF had been under since its formation in 1950. The legislation (P.L. 90-407) also required an annual report from NSB, the "corporate board" of the Foundation. The annual report was to be

> "* * * on the status and health of science * * * [and] shall include an assessment of such matters as national scientific resources and trained manpower * * *."

The first reports were topical, with such titles as: Toward a Public Policy for Graduate Education in the Sciences (first report, NSB-69-1), The Physical Sciences (second report, NSB-70-1), and Environmental Science: Challenge for the 70's (third report, NSB-71-1).

The Board, mainly on the initiative of one of its members, Dr. Roger Heyns, decided that assessing science would be well-served by including some form of measures. Therefore, for its fifth annual report, NSB published Science Indicators 1972, the first in a series of experimental biennial data books which have been published in alternate years with the other topical Board reports. Each year the amount of data has increased as the NSF staff has experimented with new indicators. The fourth in this series, SI78, is scheduled to be published in the fall of 1979.

The Congress has also expressed an interest in an overall assessment of science and technology by establishing OSTP and by mandating two reports--a Science and Technology Annual Report (STAR) and a Five-Year Outlook (P.L. 94-282, May 11, 1976). These reports were to appraise U.S. science and technology in relation to national goals. This appraisal was to include two aspects related to the use of science and technology: (1) how science and technology could contribute to the resolution of present societal problems; and (2) how problems specific to the scientific enterprise could hinder the future contribution of science and technology.

The use of indicators

Quantitative measures and indicators in science and technology policy have a variety of uses, mostly as "background information" for setting policy and as evidence for evaluating perceived problems and issues. According to a broad survey by NSF, 1/ its science and technology statistics are used to varying degrees in setting budgets, drafting legislation, planning and administering R&D programs, and monitoring trends of particular interest. The NSF survey listed the following as decisions which were influenced by science and technology statistics:

--reversal of the decline in Federal basic research funding,

--development of Federal science programs aimed at women and/or minorities,

--stimulation of industrial R&D funding and attempts to shift the distribution of R&D funding among academic and industrial institutions, and

--increase in Federal allocations of funds for energy R&D.

The foregoing list suggests the scope and significance of the public policy issues that science and technology indicators help clarify. Because these indicators are tools in the Federal decisionmaking process, it is important to determine their validity. Validity is the most important aspect of indicator evaluation, and it is a function of how the indicators are constructed. Chapter 2 examines the theory and construction of indicators and how to determine their validity.

1/"Report on the Inquiry on Uses of NSF Data and Studies," National Science Foundation, Division of Science Resources Studies, May 1978.

CHAPTER 2

GENERAL THEORY OF INDICATORS

General considerations of the measurement process are used to construct a working definition of an indicator as a particular type of measure. For this report, the term "indicator" refers to a measure or set of measures aimed at elucidating any type of social phenomena. Therefore, science indicators are considered a subset of social indicators.

LIMITATIONS OF THE PROCESS OF MEASUREMENT

Empirical measurement has long been regarded as inherently objective, yielding factual information that can help mankind understand the nature of the physical universe. Keplers' statement that "* * * the mind comprehends a thing the more correctly the closer the thing approaches toward pure quantity as its origin," 1/ and Lord Kelvins' well-known belief that if something cannot be quantitatively measured it cannot be understood, are but two examples of this traditional faith.

Social scientists have recently emphasized that a measure is a numerical surrogate or substitute for some empirical aspect of a phenomenon such as dimension, mass, population size, or expenditures. Quantitative measures alone offer only a a partial definition of phenomena. Quantification is a crucial limitation in the realm of social measurement because the complexity of most social interaction and change can rarely be understood purely in terms of quantifiable parameters. Yet, because quantitative measures are expressed in abstract symbols which can be easily manipulated and configured in a model, they hold considerable power for a variety of analytic purposes.

The measurement process is also inherently limited by the inevitable human selection of both the phenomena to be measured and the type of data considered relevant to the purpose of measurement effort. <u>Selection, then, indicates the degree to which a model (or theory) is explicit or implicit--and the</u>

1/G. Holton, "Can Science Be Measured," in <u>Toward A Metric of Science: The Advent of Science Indicators</u>, R. K. Merton, et al, eds., (1978), pp. 39-69.

relative importance of certain phenomena and the appropriate ways of measuring them. 1/ When the purpose of the measurement effort is vague or poorly understood, the utility of the model(s) used may decrease.

For measuring an area as little understood as the science and technology enterprise, multiple models are needed to insure that as wide a spectrum of phenomena as possible is included. Gerald Holton expresses this when he says:

> "There are * * * reasons for welcoming the admission of a diversity of models and of corresponding indicators. One is that in the absence of conscious pluralism, one theory is likely to establish itself or, at least, discourage the others. * * * The absence of any explicit theory to guide the making and use of indicators may not be good; but the adoption of a single one is likely to be worse." 2/

Different views or models of science emphasize and attempt to measure different aspects of science. An economic approach concentrates on immediately quantifiable resources and results; a philosophical view emphasizes the progress and evolution of scientific knowledge; a sociological perspective might place scientists at the center of concern and examine their activity and their general cognition of the world; a scientist's view might focus on the contemporary substance of science and the state of scientific knowledge, including its advances and frontiers. Each of these perspectives adds to the variety of measures needed for assembling an accurate, comprehensive portrait of science and technology.

1/See for example A. Kaplan, The Conduct of Inquiry (San Francisco, 1964); T. S. Kunn, The Structure of Scientific Revolutions (Chicago, 1962); and S. Toulmin, Human Understanding The Collective Use and Evolution of Concepts (1972).

2/Holton, op. cit., pp. 56-57. See also relevant discussion in F. M. Andrews and S. B. Withey, Social Indicators of Well-Being (New York, 1976), chapter I.

INDICATORS: A WORKING DEFINITION

In the literature on social measurement, there is substantial variation in the definition of a social indicator. 1/ This report does not attempt to provide a final definition, but rather it attempts to develop a notion of indicators which is useful for understanding their construction and evaluation. It is worthwhile to examine indicators not in terms of what they *are*, but of what they *do*. The following are two examples of how measurements become indicators.

1. The prices of consumer goods are measures. They become indicators when certain of them are selected, given estimated numerical weights, added, and viewed over time according to a concept of "average price change" (consumer price index).

2. The number of patent applications in a particular field of industry over time is a measure. If seen in light of the concept of "inventiveness," the number becomes an indicator.

In each case, the empirical measure becomes an indicator only when seen in reference to a concept of some social phenomenon or activity. Indicators do not define the concepts of the phenomena, but act as surrogates pointing to or reflecting some aspect of them. *An indicator is a special kind of measure, one that has a special meaning due to its reference to a specified concept*. Common examples of such concepts are "poverty," "unemployment," "birthrate," and "gross national product." These concepts can be measured in many ways, but the validity of the measure (and consequently the data used in the measurement effort) depends on how well the measure fits the concept.

1/Andrews and Withey, *op. cit*. See also R. A. Bauer, ed., *Social Indicators*, (1966), pp. 1-68; B. Cazes, "The Development of Social Indicators: A Survey" in A. Schonfield and S. Shaw, eds., *Social Indicators and Social Policy*, (London, 1972), pp. 9-23; *Quality of Life Indicators*, Environmental Studies Division, Office of Research and Monitoring, Environmental Protection Agency, (1972), pp. 1-15; and M. Burge, "What is a Quality of Life Indicator?", *Social Indicator Research*, vol. 2, no. 1, June 1975.

APPROACHES TO INDICATOR CONSTRUCTION

The general construction of indicators consists of two basic approaches: the "theoretical" and the "empirical." 1/ The former starts with an explicit model or theory as the basis for choosing concepts, which are in turn used for selecting measures. The latter attempts to find data for indicators without reference to theory or concepts in the belief that these will be derived from the chosen measures. Such indicators are called "operational." Note that when using either approach, the model(s) that initially guide the collection of the data will be transferred to the indicators using that data.

The benefits of the theoretical approach appear to strongly outweigh the benefits of operationalism when it comes to evaluating and thus validating the indicators. Our discussion treats these two approaches separately. In practice, both are used in a single, broad effort to develop indicators.

Assigning a particular measure to a concept of some social trend or activity necessarily involves assumptions. An example might be to use low income as an indicator of "poverty." Such an indicator is based on a particular definition of poverty that either sees it strictly in monetary terms, or assumes that all other factors (e.g., physical and mental well-being, education, etc.) depend on the level of family income.

Clearly, it is essential to be aware of such assumptions when the validity of an indicator like income is being evaluated. One might note that many nonurban, low income families are healthy and quite content with their way of life. Thus, for nonurban families, income might have low validity as an indicator of "poverty." However, in urban areas, this indicator might have high validity. The assumptions underlying a concept determine the possible measures and thus must be the focus of any attempt to make the indicator comprehensive.

1/Much of the following discussion is adapted from de Neufville's treatise concerning indicator construction and design, and the relationships between models, concepts, and measures. See J. I. de Neufville, Social Indicators and Public Policy: Interactive Processes of Design and Application, (New York, 1975).

As demonstrated by the development of economic indicators, the attempt to create valid social indicators demands an evolution which proceeds through evaluation and eventually results in the improvement of both measures and concepts. When the assumptions that underlie the concepts are revealed and defined (made explicit), criticism will have a constructive focus. The operational approach to indicator construction tries to derive a concept directly from the data. This renders the assumption-making process implicit, and thus limits the usefulness of the indicators. Evaluation of operational indicators is therefore more difficult because the assumptions are implicit (hidden) and must be uncovered.

There is also a tendency to believe that since the concept is derived from data, it has high validity and that its assumptions are not significantly limiting. Operational indicators often use the most reliable data available or that can be collected. Nevertheless, such data do not insure the validity of the indicator. To reiterate: It is how well the data fit and measure a relevant concept that determines the indicator's validity.

A good example of an operational measure is the intelligence quotient (IQ). Until recently, performance on IQ tests was often used as a valid measure of "intelligence." The concept of intelligence was never explicitly modeled or defined for the original tests, and it remains vague and undefined today. The recent controversy over the racial and cultural biases of IQ tests has exposed the previously unrecognized assumptions behind the concept of intelligence--and hence the IQ indicator. This controversy shows that the central problem with operationalism is that basic assumptions remain implicit, and when this is the case, the validity of the indicators cannot be determined.

There are three main strategies for measuring a single concept: (1) single indicators, (2) combinations of several measures into single "indexes," and (3) multiple indicators. Single indicators are apparently best used where the concept is narrowly defined, e.g., "population size", or "birthrate." Broad or complicated concepts, such as "inventiveness" or "quality of life," often require a variety of measures and may involve both indexes and multiple indicators. There is thus a definite need to consciously match the strategy of indicator design to the scope of the concept being measured.

PROBLEMS AND POSSIBILITIES IN OVERALL INTERPRETATION OF INDICATORS

One of the long-term goals for developing social indicators is to use them as a framework for interpreting the condition and direction of society. Without the advantages of well-established theories or models, however, general interpretations about society are difficult. When many diverse models are used to formulate theories about social phenomena, they will yield a variety of data interpretations. Diversity and variety have values because they reveal agreement and conflict as to indicator meanings. Where consensus is shown to exist, the interpretations are useful for making policy decisions. But when conflict is apparent, the interpretations of the data should be applied with caution until greater understanding is achieved.

On the other hand, if data are collected for only one model, the scope of the social concept being modeled is likely to be too limited. Even if the indicators are valid (because the data fit the concept), they will probably be too narrow to provide interpretations that are meaningful for policy formation.

For general interpretation, an explicit model (or models) has several advantages:

1. It spells out the limitations of both the initial selection of concepts and the interpretations of the indicators.

2. It helps improve and develop more comprehensive theories by revealing where current models are strong and where they are weak. 1/

3. It provides a framework for "weighting" certain indicators, i.e., judging their relative significance.

1/de Neufville, *op. cit.*, p 65. For relevant discussion on the importance of first defining concepts see also *The Quality of Life Concept A Potential New Tool for Decision-makers*, Environmental Studies Division, Environmental Protection Agency, (1972), pp. 1-35; *Social Indicators 1973, A Review Symposium*, R. A. Von Dusen, ed., Social Science Research Council Publication, (1974), and especially Kaplan, *op. cit.*, chapters II and V.

4. It helps reveal those indicators that have cross-model "weight." Such indicators are essential to any comprehensive model and are suitable candidates for general acceptance and institutionalization.

Basically, then, the central point is that models, whether implicit or explicit, guide the selection of both measures and concepts. An effort to derive a model or theory operationally is an attempt to avoid or circumvent this fact. 1/

Ideally, the design and construction of indicators begins with an explicit recognition of both the purpose of the indicators (why, how, and where they are needed) and the choice of model or models. Following recognition and choice, initial concepts are chosen and loosely defined before the actual measures are sought. Finally, after data are collected and formed into indicators, the concepts are reviewed and their definitions are refined along with their underlying assumptions. Over time, measures, concepts, and models are continually reevaluated to improve the indicators.

This theoretical process of constructing indicators appears preferable. However, in reality, the large and continually growing reservoir of usable data that has been collected for purposes other than for the design of specific indicators will probably be used when indicators are formulated. Indeed, it would be wasteful to recollect any of it. Consequently, there will be an initial tendency towards using these data--i.e., towards operational indicators--"because it's there." This type of data collection will incorporate an unplanned, implicit model(s). If this is not recognized along with the underlying limitations (i.e., what the model does not cover), the perspective on what future types of measures might be

---

1/As Dr. deNeufville has stated:

> "It is only self-deception to think that by not making the choices explicitly we are somehow being objective. We are simply replacing our judgment with chance or someone else's judgment. The circularity is ultimately unavoidable in all analysis. It need not be a closed system, however, where we do not let reality and new facts alter our choice of methods and models for the next iteration. The circular process can be an evolutionary one as we test measures and models against experience and try to alter each with the aid of the other." Op. cit., pp. 125-26.

added can become narrowed. Basing the attempt to construct indicators solely on already collected data will likely lead to an overreliance on operationalism.

The disadvantages of operational indicators can be overcome, however, by recognizing the implicitness of the model and by finding and evaluating the assumptions underlying the indicators. In selecting data, it is often necessary to compromise between data that would make the most valid and significant indicators and the data that are accessible.

EVALUATION OF INDICATORS

There are several main criteria for qualitatively determining the validity of indicators:

1. In view of what is known about the phenomenon being measured, how limited is the definition of it?

    a. Is there a definition at all, even a loose one, or is the reader left to define the phenomenon only in terms of the measures presented? If there is no definition, the reader is being asked to accept the indicator on faith.

    b. Are the assumptions of the definition too narrow or arbitrary?

2. Does the measure behave similarly to the phenomenon it is intended to explain? If the phenomenon changes (e.g., increases or decreases), how well does the measure reflect this? Some measures which look at only gross changes will miss detailed shifts, some of which may be highly significant.

3. Does the measure behave in the same way as other indicators of the same or closely related phenomena?

    a. If a phenomenon is well-measured, different measures which use independent types of data will behave in the same way. When only interdependent measures are used, they will all naturally show the same or similar trends, and thus be no better than a single measure.

    b. If several measures are combined into a single index, they should be well-related to each other (and thus sensitive to changes in the phenomenon).

Each of these criteria need to be addressed in any evaluation of indicators. As a general note of caution, when the concept is very broad (e.g., "level of activity" of scientific research), it might be more profitable to consider the concept as a part of overall interpretation. Instead of being measured directly, and often incompletely, such concepts could be better covered by a number of more specific indicators with narrower concepts.

OTHER RELEVANT BACKGROUND READING

- Etzioni, A., and E. W. Lehman, "Some Dangers in 'Valid' Social Measurement," in Gross, B., ed., Social Intelligence for America's Future: Explorations in Societal Problems, Allyn and Bacon, Boston, 1969.

- Tropman, J. E., "The Social Meaning of Social Indicators," Social Indicators Research, vol. 3, no. 3 & 4, December 1976.

- de Neufville, p. 125. See also discussion in op. cit. Kaplan, Chapter II; op. cit. Kuhn, Chapters 2 and 3; and Blalock, H. M., Jr., and A. B. Blalock, eds., Methodology in Social Research, McGraw Hill, New York, 1968.

- Blalock, H. M., Jr., ed., Measurement in the Social Sciences: Theories and Strategies, Aldine, Chicago, 1974. (Chapter II is especially relevant.)

- AICPA, Social Measurement: Points of View, 1972.

- Gehrman, F., "'Valid' Empirical Measurement," Social Indicators Research, vol. 5, no. 1, January 1978.

- Fox, K., Social Indicators and Social Theory, Wiley, New York, 1974.

CHAPTER 3

CRITIQUE OF SCIENCE INDICATORS 1976

NSF has sought to improve the quality of its science indicators by supporting two Social Science Research Council (SSRC) conferences (1974 and 1978) 1/ that specifically addressed SI72 and SI76. These conferences have stimulated the academic community to conduct research in this area. NSF has also conducted a valuable internal review (completed in October 1977), which both evaluated SI76 by chapter and suggested new indicators. In addition, the SI staff has contracted for studies to develop indicators in selected areas. We were told that more money is now available for these contracts.

Our analysis of SI76 is based in part on comments, suggested improvements, and proposed alternative indicators found in relevant literature, interviews, and the efforts mentioned above. We also draw from the criteria discussed in chapter 2 and the statements of purpose for the SI series. Different versions of this purpose are contained in relevant legislation and written and oral statements by Board members and other NSF personnel (see appendix I).

SPECIFIC DIFFICULTIES IN MEASURING SCIENCE AND TECHNOLOGY

Developing a series of comprehensive, successful, and acceptable indicators of science and technology is a formidable task. The scientific enterprise is very complex. It is decentralized and, in general, poorly understood. This creates numerous difficulties that hamper decisions about how to select the most important aspects to measure.

Scientific work and advance are not separate from society, but are part of it and interact with other social forces in complicated and diffuse ways. The effects of this interaction are economic, military, political, and influence such things as education, communication, transportation, health, and even the individual's view of both the world and himself. By the same token, nonscientific developments strongly influence the

---

1/Papers from the 1974 conference have been collected and recently published in the book Towards A Metric of Science: The Advent of Science Indicators, R.K. Merton, et al, eds., (New York, 1978). At the time of this writing, no such book yet represents the papers presented at the later meeting. A number of these will be listed specifically.

state of scientific work and progress. It is thus difficult to find where the borders of the scientific enterprise end and other social forces begin. It is no less difficult to decide what is meant by the "state" of science and technology, since science and technology are tied in such a complex way to non-scientific influences.

The effects of science and technology are both short-term, as in the case of tangible products or daily processes dependent on technology (e.g., communication), and long-term, as expressed by changing views about the world. Short-term trends are generally easier to measure. They are often expressed in familiar terms such as expenditures, number of publications, and the like. It is the long-range trends, however, that are usually the least understood--particularly those dealing with "strengths and weaknesses." In effect, the lack of knowledge about the social context of scientific work affects what type of indicators are chosen. It is not surprising that many indicators of science and technology use economic data (i.e., resources and tangible results) because there is a more established framework for understanding changes in these data.

The long-term perspective is central to both science and technology, particularly for research. For technological innovation, there are natural lag times between the time that resources are put into research and the time that any innovations may result. These lag times vary considerably, even within a single scientific field. Very often, specific measurements of "return on investment" in research are simply not possible. It is the ever present potential of research to reveal applicable knowledge that is crucial. No simple cause and effect connection exists between research and technological advance. Consequently, desires for short-term measures of a researcher's ability to "produce" are often characterized by frustration or misleading simplification.

The decentralized nature of science and technology in the United States makes the development of broad comparative measures difficult and tenuous. There is a multiplicity of both sponsors and performers of science and technology whose goals and general viewpoints vary considerably. Furthermore, the sponsorship and performance are largely autonomous within each of the three major sectors. Government, industry, and academia. Single disciplines in academic research are characterized by individual professional societies and organizations, journals, and jargon. A consequence of this plurality is conflict about what the most important characteristics of science and technology are and thus about measurement of them.

Another problem is the difficulty in distinguishing between science and technology. These two activities are different in scope, methodology, source of support, and professional societies. However, the border between them has grown increasingly fuzzy in recent decades as engineering investigations have involved a greater component of basic science. Consequently, while certain measures are appropriate for both science and technology (especially in the area of resources), the need for different types of measures remains strong.

The current state of science and technology in the United States (and in much of the western world) is characterized by significant controversy. Many traditional conceptions about science and technology, their goals, benefits, and context of operation have been recently challenged, and even rejected. 1/ Previous notions, which saw science as an autonomous, self-accountable activity whose advances lead naturally to benefits for society, are being criticized by certain interest groups, academicians, Government officials, and even scientists themselves. Recent academic work has introduced and supported ideas that conflict with those held by many, perhaps most, scientists. 2/ Furthermore, the strong Federal support that existed during the early and mid-1960s gave way rather suddenly to one of limited (even declining--until recently) resources and a concomitant increased emphasis on accountability. As a consequence, agreement is lacking about the nature, purpose, activity, and social context of science and technology. This lack of agreement has encouraged a diversity of partially overlapping but often conflicting viewpoints. In this light, both the purpose for measuring science and technology, and the model or models behind the selection of measures, can vary considerably. Such variety is displayed in the range of opinion offered by the participants in the recent conference on Science Indicators, 1976, sponsored by SSRC (May 1978).

A final difficulty in developing successful measures is the consistent vagueness and evaluative nature of many of the terms used to discuss measurement. A few of these terms, most

---

1/See, for example, D. Nelkin, "Changing Images of Science," Newsletter on Public Conceptions of Science, Harvard, #14, January 1976, pp. 21-31.

2/See discussion by G. Holton in "Limits of Scientific Inquiry," Daedalus, Spring 1978, pp. 227-234.

notably "health" (of science and technology) have enjoyed almost unquestioned use. Good "health" is a metaphorical standard, one which must be interpreted. 1/ While the health of a human being is generally well-defined and understood, for science and technology there is no standard or accepted definition of a healthy state, or even agreement if such a state exists. Our analogy supposes there is enough knowledge and information on the physiology of science and technology to make a modern diagnosis. In fact, the level of knowledge about the scientific enterprise is more analogous to that of a 19th century country doctor than to a modern physician's. Subjective terms may be useful in gleaning initial insights about what type of evaluation might be desirable, but, if used continually, they obscure specific, useful information. This kind of terminology also encourages numerous, often opposing, interpretations.

Other frequently heard terms such as "viability," "vitality," and "capacity" also suffer from overgenerality and implicit evaluation. When are science and technology "viable," or of high "capacity" and "vitality"? Answers to these questions are inescapably fraught with the respondent's judgment, and hence add to the confusion about science and technology. It would be at least of some benefit if nonevaluative terms such as the "state" or "condition and direction" (of science and technology) were used.

Despite the major difficulties facing the SI effort, there is a considerable amount of knowledge about science and technology, as well as a large body of data on some of the significant aspects of indicators. This discussion is meant to provide an overall "state-of-the-art" back-drop, which emphasizes both the problems in developing science and technology indicators, and the need for continued experimentation in this development.

STATED PURPOSES OF SCIENCE INDICATORS

Certain conflicts were noted among the different statements of purpose for SI. (See appendix I for list of these and full discussion.) Some of these are indicative of changes in perspective with time and experience. However, we have

---

1/D. Hornig, The Health of the Scientific and Technical Enterprise: An Advisory Panel Report to the Office of Technology Assessment, 1978.

found several basic objectives common to the different statements. We have distilled these common objectives into a single purpose: SI is to quantify as many of the dimensions of science and technology as is feasible, with a view towards identifying significant trends and developments, interpreting possible causal relationships between them, and analyzing their possible meaning for the condition and direction of science and technology in the United States.

PROBLEMS WITH THE MODELS OF SCIENCE AND TECHNOLOGY INFERRED FROM SI76

In evaluating a group of measures, it is important to understand the characteristics, assumptions, and thus limitations, of the model(s) (expressed or implied) which have inadvertently directed the selection of data. The most serious general problem with SI appears to be its de facto view or model of science and technology.

In examining a list of original "indicators" proposed by NSB in 1971, one can discern the model that is implicitly emphasized in the SI reports. The measures were first conceived in a long list, and later partitioned into categories such as "Scientific Output," "Activity," "Science Education," and "Manpower." The following is a sample of the original measures:

1. Number of papers in top quality, refereed journals.
2. Ratio of basic research funds to total R&D.
3. Federal support of total research by field of science.
4. Major new frontiers of science opened up during a specific year.
5. Utility of knowledge.
6. Ratio of applied research funds to total R&D.
7. Ratio of development funds to total R&D.
8. Federal basic research dollars by field.
9. Distribution of new baccalaureates, masters, and doctorates by field.
10. Number of science and engineering degrees as a percent of total degrees.

11. Relationship of U.S. R&D/GNP to GNP/capita among various nations.

12. R&D scientists and engineers per 10,000 population in different countries.

At the time these measures were selected, most of the data already existed "in hand" for NSB, i.e., within the Federal Government, particularly in NSF's Division of Science Resources Studies. With few exceptions (the fourth and fifth), these are measures without concepts and represent an operational attempt at constructing indicators.

Though this de facto model oversimplifies science and technology, it is the support-service view that Government naturally takes. It can be very broad, as in the past, when it was assumed that "what is good for science is good for the country."

The problem with the version of SI76 is that it is too constricted by an input-output framework. In this approach, science and technology are seen as resources which go into, and tangible results which come out of, a "black box." Inherent in this model are the following assumptions:

1. Science and technology are primarily to serve social (i.e., national) goals. 1/

2. The states of science and technology can be described primarily in terms of what is being added each year to the overall base of expertise, without sufficient reference to the accumulated base.

3. Science and technology exist in a social context and have minor self-regulating aspects. 2/

4. Science and technology can be measured in terms of resources (such as funding and "stocks" of personnel) and a few tangible results such as innovations and published papers. The process involved in doing science and technology are not important.

1/M. Kochen, "Models of Scientific Output," in Towards A Metric of Science, pp. 97-139. See also R. McGinnis, Science Indicators 1976: A Critique, in press (Social Indicators Research).

2/G. Holton, "Can Science Be Measured?", in Towards A Metric of Science, p. 56.

5. Money drives the system; more money, more "output."1/

6. Science and technology advance incrementally with no "sudden" changes.

7. Science and technology are related by cause and effects. 2/

8. University research is the mainstay of U.S. science. 3/

9. Physics, chemistry, biology, and math are the most important sciences, and hence deserve the most complete coverage.

10. International competition is of central importance to U.S. science and technology, more significant than cooperation. 3/

Basically, this support service model, with its input-output stress, is a technological view of science--one that ignores the internal sociology and conceptual progress of science and and technology and emphasizes its economic aspects. Scientists engaged in basic research are motivated primarily by a search for knowledge, whether or not it has any perceived relevance to social needs. Furthermore, the black-box view (into which a resource is put, and from which a result emerges) does not take into account the activity and process of investigation. This point is important, since research investigations are the core of science and crucial strengths and weaknesses are likely to have their roots in research. The inadequacy of the input-output approach is well demonstrated in SI76: input indicators of research activity in SI reveal a substantial decline since 1968 (funds for basic research decreased after 1968), while output indicators of research activity display an unbroken increase (number of overall publications has grown). Examination of the process of research is necessary to accurately identify blocks that would hinder the advancement of science.

1/McGinnis, op. cit.

2/J. D. Holmfeld, "Science Indicators and Other Indicators Some User Observations." Paper presented at the annual meeting of the Society for Social Study of Science held at Indiana University, Bloomington, Indiana, November 5, 1978.

3/Holmfeld, op. cit.

Another important part excluded by the input-output model is the sociology of the scientific process. Though science and technology are subsets of the larger social context, they also operate independently of it. The cohort and competitive aspect to research, the transfer of information, and the workings of professional societies, are important parts of how scientific work is organized and conducted. Another central aspect of this sociology is the "flow" of personnel, i.e., shifts between fields, in and out of scientific work, and cross-nationally. It is more important to know where and how scientists and engineers are employed, rather than the fact that they are. These aspects of the scientific enterprise are covered partially by the support service model, but only as

> "* * * pale reflections of the social structures, such as authority and reward, and processes, such as mobility and aging...that sociologists of science insist are fundamental to any real understanding of how the system of science operates." 1/

In general, the input-output model views those who perform scientific work as resources, like dollar allocations, and gives minor treatment to the internal relationships that characterize the process of this work. For identification of "strengths and weaknesses," this model is consequently inadequate.

By its narrow focus, the input-output framework used in SI leaves out both the knowledge base and the major advances in science and technology. Science is a major branch of knowledge; any view which would describe its state without examining its knowledge base and advances is deficient. Such information is needed if one is to know where new knowledge might best be applied to social needs. Identification of frontiers in science was originally on the NSB's list of proposed indicators, but it has not been developed, possibly due to a difficulty in collecting data. However, data on this area do exist in the form of citation analysis, a recent quantitative technique to monitor contemporary science, but SI76 made only small use of such data.

1/McGinnis, op. cit. See also the following papers presented at the May 1978 Science Indicators Conference sponsored by the Social Science Research Council in Washington, D.C.: C. Kuh, "Indicators of Scientific Manpower and Science Indicators"; J. Ben-David, "U.S. Science in International Perspectives"; and D. Tufte, "Statistical Issues in Science Indicators."

Education is another major aspect of science and technology that receives limited treatment. The input-output model is concerned almost entirely with how many personnel work in science and technology. It views education as a producer of degrees that leads to renewed resources for science and technology. As such, only the educated are examined, and the focus is again on employment. The process of how science is taught, what is taught, and what is happening in preuniversity and university training in the sciences is not covered. Thus, if problems are developing in the training of new scientists, the underlying framework of SI would not reveal them.

SI76's concentration on resources, especially expenditures, emphasizes the "hard" sciences. Since these fields are the most capital-intensive, they receive the most complete coverage and appear to be the most important of the sciences. Measures of the social sciences, environmental sciences, and engineering sciences are rarely field specific. The sciences are thus viewed in terms of their economic, not scientific, importance. This points up a critical way in which the model undercuts itself: by limiting indicators to largely nonscientific information, the model prevents perception of where sciences can best benefit society.

In general, it is the input-output framework that concentrates the model on economic resource data. To confine the "policy purpose" of indicators to measuring only "the resources allocated to the scientific enterprise or the fulfillment of its goals" as NSB does, is useful for some institutional aspects of science and technology, particularly those which are industrial and mission oriented. However, for science and much of technology, results are not material and cannot be treated in the same way as economic outputs. As R.B. Freeman states: "It is far from clear what we mean by output in this context." 1/ These terms (input and output) are insufficient when applied to research, for it is the process of research that must be examined to reveal problems, strengths, and needs.

1/R. B. Freeman, "The Impact of Science (R&D) on the Economy," paper presented at the mentioned May 1978 Conference on Science Indicators. A. Biderman has also criticized the input-output framework for social indicators as the "* * * domination of social data by the assumptions of liberal economics * * *"; and that, in general, "* * * the distinction between input and output is rather artificial * * *." In Social Indicators 1973. A Review Symposium, R A. von Dusen, ed., Social Science Research Council, 1974.

The terms "input" and "output" confuse the development of indicators by acting as substitutes for concepts of social phenomena. At the present early stage in the development of science indicators, all outputs are considered to be significant. This is an expression of the operational approach--i.e., data are collected on the basis of their presumed importance. However, we believe that certain important concepts can be identified and measured. A few of these have been derived from the collection of data on identified research outputs, but the search for the "outputs" of science and technology can draw attention away from finding the important concepts (e.g., frontiers in science and technology, mobility of personnel). The NSF internal review made an important point in this connection: the acknowledged enthusiasm for output indicators should not lead to their use beyond the state-of-the-art.

A final major problem with the model in _SI_ is that it develops, to a large extent, the same indicators for science as for technology. This combining of science and technology leads to overaggregation of the data and has prevented the perception of _where_ there might be important trends. Because of this combining of science and technology, the reader is given only an overall notion as to whether the entire scientific system is strong or weak.

These are the major limitations in the implicit model which has been central in the _SI_ effort. After three iterations, there has been only slight overall broadening of perspective and only little apparent change planned for the next. In order to interpret causal relations and analyze their possible meanings, more effort should be placed on relating the indicators to underlying theories. Statements in the introduction to _SI76_ acknowledge the need for "* * * an explicit model of the research enterprise, both in itself and in relation to the rest of society * * *" but note that existing models are considered inadequate to meet the need.

In our view, there is a definite need for incorporating new models of science and technology into SI, for if the input-output model remains dominant, the information which SI offers will be too restricted to comprehensively assess science and technology.

GENERAL LIMITATIONS AND PROBLEMS WITH INDICATORS

From _SI72_ to _SI76_, progress has been made in both the addition of topics covered and in the number and types of indicators used. The authors of _SI76_ are clearly aware of the limitations in the general interpretations which can be derived from the data presented. Throughout the text, they present important caveats that apply both to the data and to the use to which the data can be put. However, they do not address the conceptual limitations of the indicators.

The most evident of these limitations is that very broad, undefined concepts such as "level of activity" or "extent of effort" have been applied to a narrow range of measures. To adequately measure such concepts, one must have data on nearly every major aspect of science and technology--research, education, invention and innovation, and personnel. SI76 uses expenditures in basic research and numbers of publications as indicators of basic research "activity." The assumptions underlying each measure equate spending with work (i.e., 55 percent of the basic research work done in universities means 55 percent of the money for basic research is spent by universities 1/), and publications with the amount of research. Increases in both of these would thus indicate a more active research community, while decreases would mean less work is being done. However, SI76 shows that while expenditures have decreased, the number of publications has continually increased. Thus, the measures are too narrow to cover the concept of basic research activity.

On the other hand, the report gives a more specific concept to expenditures: when examined on the level of sectors over time, expenditures are said to measure Federal or industrial "priorities for research." Here, the measure has more meaning because the concept is defined more precisely.

A related problem is that several concepts of differing scope are applied to a single type of measure. Federal expenditures for general and specific areas in science and technology are said to indicate "extent of effort," "level of activity," "health of U.S. R&D effort," and "national priorities in the area of science and technology." Thus, one measure is assigned four concepts, the first three of which are vague and highly complex, while the last one is specific and meaningful. This is also the case where SI uses the number of scientists and engineers as indicators of "depth and direction of a country's R&D effort," the "magnitude" of such effort, and of "employment" and "unemployment." Only the last concept is specific enough to have any meaning. Assigning multiple concepts to a single measure confuses the meaning of that measure. SI could benefit from using, to the extent possible, single, specific concepts for each separate measure.

1/See Science Indicators, 1976, p. 75.

Another general problem with the indicators has already been alluded to· the predominance of two interrelated types of data--expenditures and personnel. Since these two basic trends show an overall decline since the late 1960s, the general indication of science and technology is continually gloomy. Furthermore, this predominance shows that, where SI concentrates on the economics of science and technology, data in this area are limited. Indices of prices or costs of such things as new scientific equipment are not attempted.

From the indicators, one notes that spending for defense and space has decreased significantly since the late 1960s, while civilian R&D spending has increased in "real" dollars. The salient decline in SI76, repeated by so many of the charts, appears due to this--particularly since the decline in numbers of personnel is shown to be closely related to that in expenditures. (See figures 1 and 2.) However, it would be a clear oversimplification to describe the general state of science and technology as so deeply dependent upon spending for defense and space. Yet this is an unavoidable conclusion, given that the data are presumed to reflect the levels of "activity" and "effort" in U.S. science and technology.

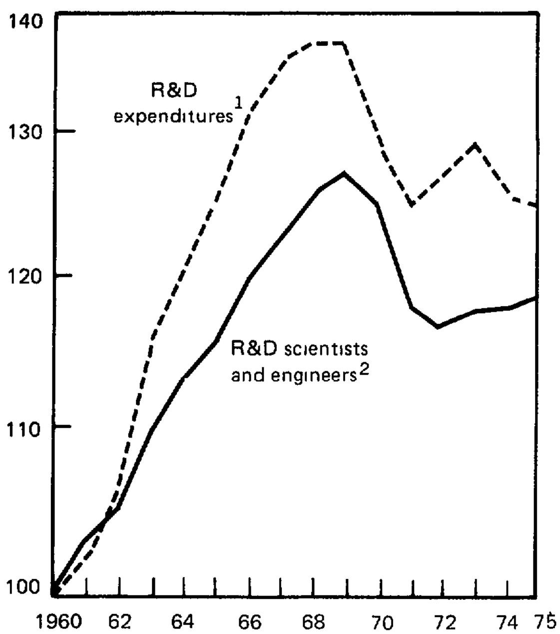
LIMITATIONS AND PROBLEMS WITH SPECIFIC INDICATORS

R&D/GNP

One widely used measure, which is misleading, is R&D spending as a fraction or percent of the gross national product (GNP). GNP is a well-known, well-established concept, used to help portray the condition and performance of the U.S. economy. The R&D/GNP measure, however, is a nonconceptual adaptation of it for purposes other than those for which the GNP concept was designed and developed. A significant amount of research and development money, which varies from year to year, is spent for Government purposes only. Each of these purposes will have different effects on future economic growth and productivity gains, depending on the nature of the research and its affect on commercial products and services. The R&D/GNP index thus combines measures that are poorly related: overall Government R&D spending can increase or decrease significantly but future GNP might be little affected. Furthermore, this ratio is even more questionable for international comparisons. The mix of support for R&D within governments as well as between governments and private enterprise varies greatly among nations.

Figure 1

Scientists and engineers engaged in industrial R&D, compared with constant dollar expenditures for industrial R&D, 1960 75



[1] GNP implicit price deflators used to convert current dollars to constant 1972 dollars

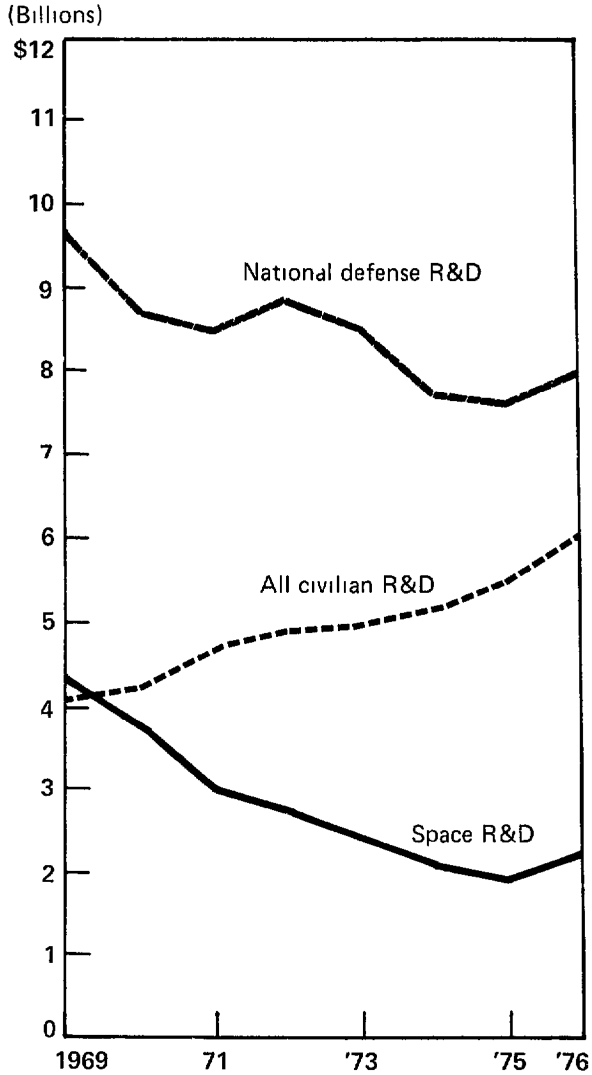
[2] Full time equivalent basis averaged for each year

NOTE Estimates are shown for 1975

Figure 2

Federal obligations for R&D by major function, 1969 76

Constant 1972 dollars[1]



[1] GNP implicit price deflators used to convert current dollars to constant 1972 dollars

NOTE Estimates are shown for 1975 and 1976

SOURCE Science Indicators 1976

A more thoughtful structuring of this data has been done in the recently released Science and Technology Annual Report (STAR), 1/ which examined industry-financed R&D as a fraction of GNP in manufacturing (over 90 percent of R&D in industry is funded by manufacturers). Even though this is an index without a specific concept, the individual measures have a more direct, known connection which makes the index more significant and usable. A concept such as "gross industrial R&D product" might apply to this measure, since the notion of a "gross product" can be transferred from GNP without misleading distortion.

Patents

SI76 includes a section of "inventiveness" indicators which have low validity because of the patent data base upon which they depend. According to Deborah Shapley, 2/ certain industries (especially electronics) have stopped patenting many of their inventions because (1) during the 2-year wait to complete the patent-granting procedure many inventions can become obsolete, and (2) making inventions public can give competing firms an "edge," since only minor modifications of some inventions can lead to marketable improvements. As a consequence, the simple tracking of patents granted as a measure of "inventive output" has little validity for at least one of the most inventive industries. The degree to which this is true for other major industries is uncertain. If this indicator is continued, it must be used with considerable caution. As a measure of general industrial "inventiveness," it is not valid.

For international comparisons, SI76 uses a "patent balance" idea which appears to be too specific to be used as a major indicator because it only looks at a nation's patenting in foreign nations. The balance is defined as "* * * the number of patents granted to U.S. nationals by foreign countries minus the number of patents granted to foreign nationals by the U.S." By this definition, foreign countries are patenting increasingly more in the United States than the United

---

1/Science and Technology: Annual Report to the Congress, National Science Foundation, with assistance from the Office of Science and Technology Policy, August 1978. See table 5-2, p. 73.

2/D. Shapley, "Electronics Industry Takes to 'Potting' Its Products for Market," Science, November 24, 1978, pp. 848-49.

States is abroad. If, instead, one examines the change in domestic patent applications of each country, as in the STAR, one finds that only Japan has continually increased its patenting from the late 1960s to 1975, and that the United States is not falling behind other western industrial nations. Examination of the percent decline in patenting for each country relative to its peak year (which varies among nations) shows the following: U.S. decline, 15.42 percent; U.K. decline, 21.9 percent; West Germany decline, 20.84 percent; France decline, 21.05 percent. Furthermore, when seen in light of the increasing disincentives for patenting, this decline in "relative inventiveness" is even more dubious. The "patent balance" concept thus appears insufficient for international comparisons, since it is centered completely on the United States. Examining domestic patenting for each nation has a broader base, each country being viewed on its own terms.

International comparison of
Government R&D expenditures

One particularly misleading indicator used in SI76 is the "Estimated Distribution of Government R&D Expenditures Among Selected Areas by Country." The data displayed in figure 3 shows that the U.S. Government spends the least in the area of "Advancement of Knowledge," while Japan whose total R&D emphasis is well known to be on development, spends the most. 1/ The crucial assumption underlying the structuring of the data is that the categories for expenditures (energy, health, space, defense, etc.) apply equally to all the countries examined. The Organization for Economic Cooperation and Development (OECD) which collected the data accepts the definitions and stated purposes of the cognizant governments without much further analysis, i.e., if money is budgeted for R&D by a military agency, that money is spent for military R&D only.

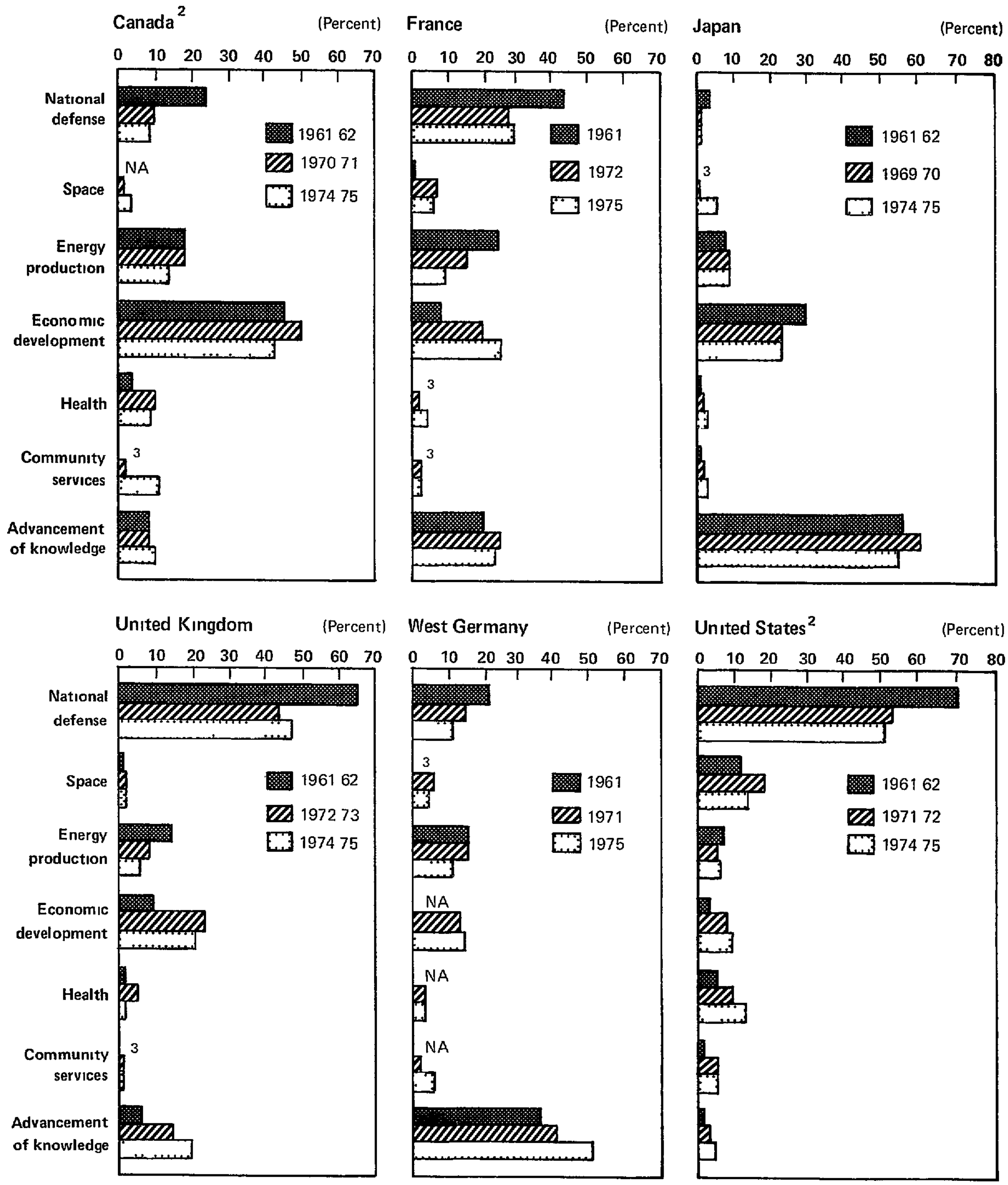
Since the U.S. Government allocates a large part of its research expenditures to these mission-type categories, much

---

1/See, for example, publications of the Japanese government: Outline of the White Paper on Science and Technology - Aimed at Making Technological Innovations in Social Development, Science and Technology Agency of Japan, Foreign Press Center, January 1977; Science and Technology Agency: An Outline for 1977-1978, Science and Technology Agency, Prime Minister's Office, Tokyo, Japan; and 1976 Report on the Survey of R&D in Japan, Bureau of Statistics, Office of the Prime Minister, Japan.

Figure 3

Estimated distribution of Government R&D expenditures among selected areas[1] by country 1961 75



[1] Function categories are not the same as those of Appendix Table 2 11 e g Advancement of knowledge does not equal Science and technology base

[2] Advancement of knowledge does not include general university funds

[3] Less than 0 5 percent

NA=not available

Source Science Indicators 1976

of the money that is spent on research by mission agencies does not show up under "Advancement of Knowledge." Moreover, where a large part of U.S. research and development is sponsored by the Government, Japan's industries fund and perform the great majority of this work. Consequently, the indicator is misleading. We believe that future SI reports would benefit by either discarding this indicator or recategorizing the expenditures in a way that more precisely reveals the different priorities characteristic of different nations.

Innovation

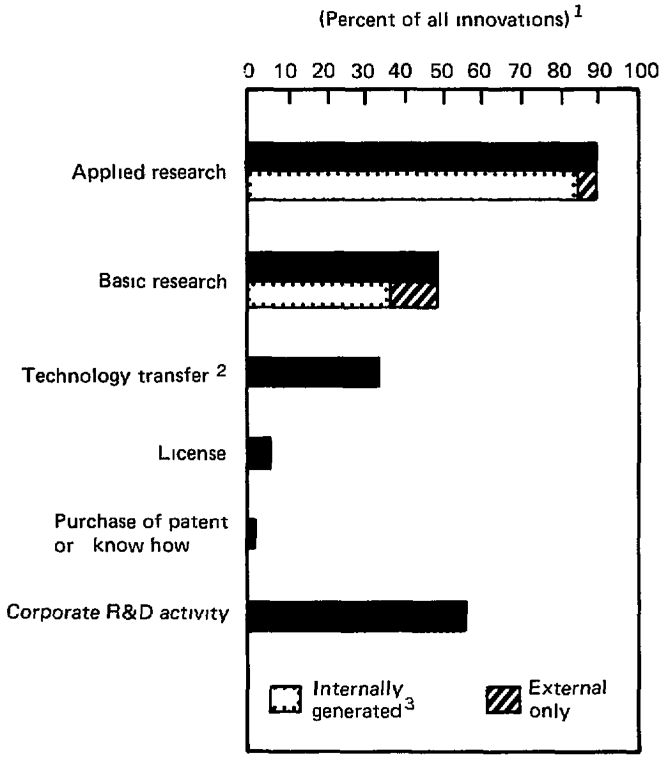
SI76's indicators of technological innovation reveal some important information about the processes involved. The data for these indicators are based on a 1976 report done by Gellman Research Associates (Indicators of International Trends in Technological Innovation) for NSF in which a panel selected 500 innovations for study. A number of underlying assumptions and limitations need to be mentioned. For instance, the original selection of the innovations was based on perceptions of an innovation's "* * * technological consequences as well as its primary and secondary impacts in socioeconomic and political terms." The major problem with these criteria is that recent innovations are more likely to be chosen on their technological merits, since their social effects are as yet unknown, unclear, or nonexistent. Older innovations have had time to reveal their social effects and will probably be judged more on the stated criteria.

Also, determining the significance of contemporary results from science and technology involves major assumptions. Though the importance of some innovations is known beforehand (e.g., the jet engine, the nuclear reactor), the longer-term significance of many cannot be determined (e.g., the photocopier). This caveat might also be applied to determinations of the "radicalness" of innovations, which the Gellman study does. The social significance and scientific radicalness of an innovation are important concepts, but ones which depend upon contemporary judgments. Over time, inventions that are first perceived as radical may later be seen as part of an evolution which produced even more important inventions.

The foregoing discussion relates directly to the indicator which attempts to measure the "distribution of major U.S. innovations by source of technology." (See figure 4.) The science underlying a particular innovation often has a long history behind it. Transistors, for example, resulted from the study of semiconducting materials, which in turn was based on an evolution of knowledge about solid-state physics. If traced, the history of most innovations eventually goes back

Figure 4

Distribution of major U S innovations by source of technology, 1953 73



[1] Multiple responses were accepted
[2] From an existing product of the same company
[3] Wholly or partly internally generated

SOURCE Science Indicators 1976

to "basic science." The particular immediate work leading to an invention can be of several kinds, but there still often exists a continuum that connects the invention to a series of previous inventions and basic advances. Thus, it is very important to know the assumptions involved in how the "source" is determined. Without these, the indicator cannot be evaluated and thus cannot be trusted.

Bibliometric data

Data which focuses on research publications appears to have a good chance for uncovering information about the process of scientific work. Due to its potential and underutilization in the SI reports released so far, a brief review of this controversial subject follows.

Bibliometrics counts the number of publications in different fields over time and the number of times an article is cited in other articles (citation analysis). Much controversy has centered on whether citation analysis can serve as a valid measure of both research quality and quality of professional performance.

Publication counts are used as a measure of "research activity." 1/ The assumption that more research produces more papers seems valid when applied to academic research, especially to individual disciplines and their subfields "Activity," however, is a vague and general term that needs a more strigent definition before the meaning of the measure is clear. The major limitation here is that an increase in publications does not necessarily mean a proportional increase in knowledge.

Knowledge coming from research, its utility, significance, and presumed quality, is treated by citation analysis. Underlying citation counts is the assumption tnat the numbers of citations strongly correspond to the utilization and significance of an article, which in turn strongly correlates with the quality of the scientific work. By examining where and how citations and co-citations (the citing of the same

---

1/F. Narin, and M. Carpenter, "Science Output Indicators Based on Bibliometric Techniques." Discussion notes and informal lecture delivered to Science Output Indicators Seminar in Washington, D.C., January 16, 1979. For general reference see also articles by: N. Wade, "Citation Analysis A New Tool for Science Administrators," Science (May 2, 1978), pp. 429-432, and S. Aaronson, "The Footnotes of Science," Mosaic, NSF, March/April 1975, vol. 6, no. 2.

pair of articles in other articles) cluster in separate fields, one can presumably measure how scientists relate to each other and the specific and overall progress of science.1/ The basis for these assumptions is the belief that scientists recognize both the immediate and long-term significance of their work at the time of, or soon after, its publication. A highly cited article is "recognized" as both significant and of high quality.

Excellent research, however, does not always yield results that are considered immediately significant and thus highly cited. Science's last 100 years abound with cases of contemporary scientists failing to appreciate work which later proved highly significant. The inattention to Mendel's work in genetics is only one well-known example. Moreover, areas of research become obsolete and papers once highly cited drop into obscurity.

Citation analysis is characterized by several problems that add bias to the counting. A few of these are--

- An article may be cited frequently because of its poor quality, concern with a controversial subject, heuristic value only, or other solely controversial or objectionable aspects.
- Heads of laboratories, research teams, or advisers of graduate students are known to place their names on articles which are the work of others.
- Authors may cite their own works or the works of personal friends, regardless of significance or quality.

Because of these problems, we do not think that citation counts should be used as an indicator of quality.

There are several other important caveats concerning the lack of quality of certain indicators.

- Apparently there is a greater incentive for foreign authors to publish in the U.S. journals than in their own. This makes international comparisons difficult.

1/E. Garfield, M. V. Malin, and H. Small, "Citation Data as Science Indicators," Towards A Metric of Science, pp. 179-209.

--In many foreign journals, editorial and space limitations allow little room for citations.

--Language barriers and lack of communication between communist and free-world countries lead to problems. For example, the Russian authors predominantly cite themselves.

- o Citation analysis mainly consists of journal articles and does not include research publications such as monographs, pamphlets, books, anthologies of papers presented at conferences, and the like.

Despite these caveats, this type of analysis gets at crucial aspects of science such as evaluation of frontiers and the ways scientists relate to each other. Actual "maps" of scientific change have emerged for particular areas. 1/ International publication patterns, including "field internationality," "journal internationality," and "country distribution among fields" (i.e., concentration) can be outlined and analyzed.

Our primary reservation with citation analysis is that some proponents outside of NSF presumed that it can be used to measure research quality. Besides the stated problems, the use of bibliometrics may draw attention away from measuring other important aspects of the research process.

Public attitudes survey

The chapter on public attitudes is an important section of *SI* because of the increasing emphasis on the social effects of science and technology. The general state of knowledge and opinions people have about science and technology is an essential part of any overall portrait, particularly in view of the growing public controversies about such subjects as nuclear power, laetrile, and artificial sweeteners. Also, since the Congress is the principal client of the *SI* reports, the knowledge and opinion of the electorate take on added importance.

SI76's series of survey questions, however, appear to have major limitations. The first is that only the general public was surveyed. A separate section, containing interviews with the scientific community, would have added valuable information. The views of scientists need to be solicited by the same kind of survey as that used for the general public. While the questions should not be the same for each, the type

---

1/E. Garfield, et al, *op. cit.*, see especially pp. 193-201.

of information should be. In this way, valuable comparisons can be made.

Second, some significant studies in the area of public attitudes have been done, 1/ but SI76 did not make adequate use of this available data. One social scientist working in this area, Dr. Todd LaPorte, criticizes SI76 for consistently "muddling" science and technology, treating them as indistinguishable (e.g., "benefits from science and technology"; "Do science and technology do more harm than good?"; "Do science and technology change things too fast, too slow, just right?") 2/ His own research, in fact, shows that the public appears to perceive the effects of science as radically different than those of technology. 3 / Those whom LaPorte has surveyed blame ill-effects less on science and more on technology, and they prefer "* * * near freedom for one and apparently increased regulation for the other." 4/

The third limitation, as LaPorte points out, is that for policymakers, treating the public as a "homogeneous mass" is too simplistic. He mentions that his research shows that the opinions revealed by a person's response to a questionnaire often display a certain degree of inconsistency. An example of this in SI76 is shown in table 6-13 (p. 176), where the majority of opinionated respondents (45 percent) blame science

---

1/A. Etzioni and C. Nunn, "The Public Appreciation of Science in Contemporary America," Daedalus, Summer 1974, pp. 190-212; T. LaPorte and D. Metlay, They Watch and Wonder: Public Attitudes Toward Advanced Technologies, Final Report to Ames Research Center, NASA, from Institute of Government Studies, University of California, 1975; T. LaPorte and D. Metlay, "Technology Observed: Attitudes of a Wary Public," Science (April 11, 1978), pp. 121-127; and J. Miller, "Selective Attentiveness: A Conceptual Framework for Understanding Public Attitudes Towards Organized Science," a paper presented to the 1978 Annual Meeting of the Society for Social Study of Science, Bloomington, Indiana, November 4, 1978.

2/These examples are taken from the survey questions in SI76.

3/LaPorte and Metlay, They Watch and Wonder, op. cit. See also T. LaPorte and D. Chisholm, "Indicators of Public Attitudes Toward Science and Technology: Science Indicators 1972, 1974, 1976. A Review and Prospective Reflection," a paper presented at the mentioned May 1978 SSRC Conference on Science Indicators,, pp. 21-33.

4/LaPorte and Chisholm, op. cit.

for "* * * air, water, environmental pollution * * *," but not for "* * * insecticides used the wrong way."

In examining what he calls the "potential public for technological politics" (those more likely to participate, by whatever method), LaPorte found that these respondents to his survey had far more

> "* * * organized, internally consistent perceptions of the effects of technology and * * * tended to have generally more favorable views of science." 1/

A survey of groups (like those identified by LaPorte) would add considerable information to the overall portrait of public attitudes. Furthermore, for policymakers, information on groups that are likely to be politically active would be especially relevant.

Overall, SI76's survey questions seem too concerned with how well science and technology are thought of by the general public. Nearly all the questions ask for evaluations on how good or poor science and technology appear. The result is that there is little information on what science and technology mean to the public, how important it is to them, or what their level of scientific knowledge is.

A general problem with evaluative questions is that people's feelings about social conditions are easily transferred to specific responses. For example, the response in SI76 which shows decline in the status of all professions, as well as of those who run major institutions, can be attributed to a general disenchantment caused by inflation, the Vietnam War, and political scandal. The response which blames Government decisionmakers for the problems resulting from science and technology is also a probable example of evaluation transfer. SI76's use of evaluative questions allows too much uncertainty to enter into the survey.

CONCLUSION

The foregoing are examples of some major specific difficulties with SI76. Basically, the majority result from a narrow approach to developing indicators, which is limited by a restricted view of science and seems not to recognize what an indicator is, or how it "indicates." One of the main points of our critique of SI76 is that the need to understand the central role that models play, to know indicators are chosen and constructed (either through formulation or acceptance of

1/LaPorte and Chisholm, op. cit., p. 18.

underlying assumptions that limit the relevance of the data), and finally to make explicit in some way the limitations both in the data (as has been done) and in the assumptions. Thus far, SI shows some significant advances in this direction, particularly the internal review of SI76, completed in the fall of 1977. This review recognizes the importance of models, and notes that selecting data solely on the basis of its present use is limiting. It also emphasizes that overall interpretation is needed. Furthermore, the review recommended that there be a more explicit rationale for presentation of indicators, and that implicit conclusions be questioned or explained. However, it only examined problems with the data itself, not with the assumptions and limitations underlying their use for indicators. The review also failed to examine many of the conceptual problems in the input-output model of SI. Thus, the review did not adequately address some of the fundamental weak points in the SI effort.

We believe the experimental nature of the SI attempt needs to be further stressed: it should be expected that developing science and technology indicators will require the same gradual development that was necessary for economic indiicators. SI needs to generate new data, as well as to make use of as much existing data as possible. In general, for science, measures of process and substance are needed, such as those which come from the analysis of research publications. For technology, economic measures and, to some extent, the input-output oriented view are far more appropriate, since material resources and results are of central importance.

As yet, SI has developed and used new data in each report, but has not made adequate use of data available from other sources. For example, data generated in work on social returns from R&D has been done by Dr. Nestor Terleckyj 1/ and Dr. Edwin Mansfield 2/ and might easily be more fully incorporated in SI reports. Other research in measuring science which SI could incorporate has been going on in academic circles (e.g., by Derek de Solla Price). NSF should not continue to exclude such relevant research in the SI series.

1/See, for example, N. E. Terleckyj, State of Science and Research: Some New Indicators, National Planning Association, (February 1976).

2/E. Mansfield, "Determinants of the Speed of Application of New Technology," in Science and Technology in Economic Growth, B. R. Williams, ed., (New York, 1973); and E. Mansfield, et al, "Social and Private Rates of Return from Industrial Innovation," Quarterly Journal of Economics, (May 1977), p. 221.

Furthermore, a central part of the *SI* experiment rests on interpretations that can be derived from the indicators. The *SI* staff has recognized that the absence of this has been one of *SI*'s greatest weaknesses, particularly in view of its own stated objective that *SI78* is designed to include considerable amounts of interpretation. Such interpretation needs to be improved over time, just as the selection and structuring of the indicators does. Due to the recent and experimental nature of *SI*, constant reevaluation and change are necessary. This is the only way to make sure that problems and limitations do not become codified in the series of reports.

CHAPTER 4

SUGGESTIONS FOR IMPROVING SI76 AND FOR DEVELOPING ALTERNATIVE INDICATORS

This chapter offers specific suggestions that might improve the *SI* reports. These suggestions pertain to the general aspects of the report and to individual indicators, but they are not meant as a comprehensive scheme for improvement. Numerous ideas for altering the report or for new types of indicators have been proposed by other commentators on *SI*, including NSF's internal review of *SI76*.

GENERAL SUGGESTIONS

The improvement most often suggested by other critics of *SI* is that more overall interpretation of the indicators be included.

We have been advised by the *SI* staff that *SI78* will contain a large amount of interpretation, some of which will center on selected policy issues. However, too much interpretation related to popular issues could result in an overemphasis on the political context of science and technology. Certain broad, long-standing issues need to be addressed before a general assessment of science and technology is possible, e.g., where there are particular strong or weak points, where science and technology are focused (i.e., which fields, which industries), and where they are less intensive.

The primary contribution of overall interpretation is to help find the general, long-term meanings of indicators, not those related to topical issues which may shift from year to year. General interpretation is an elemental part of the improvement of *SI* and, in our view, should not be subservient to short-term considerations of utility.

We see the improvement of *SI* as also tied closely to its experimental aspect. The NSF staff should continue to expand and develop indicators--discarding those that are not valid, and testing new ones to see if they are sound. In this manner, *SI* continually will improve its portrait of science and technology in the United States.

We suggest two additional changes relating to the presentation of the data. First, given *SI*'s extensive caveats concerning the reliability of data, there might be some estimation of the uncertainty in the graphs shown. This could take the form of quantitative measures of confidence. Second, it might be highly useful to include a summary chapter which actually juxtaposes many of the time-series graphs so that possible interrelationships could be seen better.

In light of the generally recognized difficulty involved with using the categories "basic research," "applied research," and "development" for expenditures in all sectors, we suggest a different category already in use by several Federal agencies: "science and technology base." This label might include research not uniquely related to any particular mission or product family, but which contributes to the general reservoir of scientific knowledge and emerging technology. In our view, this category is better suited for cross-sector comparisons because it is based on the recognition that scientific and engineering research are intertwined. It is less uncertain, for example, than "basic research" is for industry. Nearly all academic research and exploratory or generic industrial research would fall under "science and technology base."

The categories now used by SI appear most relevant to the academic sector. We suggest that SI explore the preferences of industry. A recent Industrial Research Institute study (Definitions of Research and Development, October 1978) indicates that industry prefers categories defined by business objectives. Research includes basic and applied research, as well as exploratory development of emerging technologies related to general or long-range business strategy. Product and process development are clearly identified with short-term manufacturing and market objectives. The Department of Defense has already established a "technology base" research category, and it may be feasible for other agencies to separate expenditures similarly.

Another improvement, mentioned in chapter 5, is that SI include more indicator work by others. Dr. Nestor Terleckyj, for example, has developed a series of indicators concerned mainly with the social returns to R&D, an area only lightly discussed in SI76. Academic researchers are also conducting relevant studies. For example, Dr. Derek de Solla Price's graph 1/ has compared different nations' scientific research intensity with development intensity by comparing the number of scientists per capita to kilowatt hours per capita. The validity of this particular indicator appears somewhat limited (kilowatt hours being a limited measure of development), but

---

1/D. de Solla Price and S. Gursey, "Some statistical results for the numbers of authors in the States of the U.S. and the Nations of the world," preface to Who is Publishing in Science, 1975 Annual, Philadelphia, Pennsylvania, 1975, pp. 26-34 (Figure 5).

it adds a new dimension to international comparisons. More recent research by Dr. Charlotte Kuh has shown that where there has been a decline in the hiring of mathematics Ph.D.'s between 1973 and 1977, the decline has been highly concentrated in 4-year colleges (which show a 50 percent decrease), while universities show only a slight change in their rate of hiring. 1/ We have also been advised that Dr. Stephen Cole is conducting research which examines the relationships between age and scientific "performance" or "creativity" by synthesizing data on personnel and bibliometrics. There are also researchers concerned with surveying public attitudes towards science and technology which SI could draw on. These are some examples of a considerable variety of academic research concerned with measuring important aspects of science and technology. Some of this work could be adopted for SI. Since it is an evolving, presumably experimental effort, SI need not be conservative in its use of relevant data. It appears that a significantly broader data base for indicators exists than has been utilized by SI. We believe that such data should be used, at least on a trial basis.

ALTERNATIVE INDICATORS

In addition to the general suggestions, there are many alternative indicators that might be relevant for future SI reports. We have listed and briefly discussed specific examples of these alternatives in this section. The suggestions described in chapters 1 through 3 are not included here.

1. SI should expand the use of bibliometric data. For example:

   a. Identification of frontiers, either by co-citation, or by identification of active fields and subfields through examination of publication activity. 2/

1/C. Kuh, "Indicators of Scientific Manpower and Science Indicators," a paper presented at the 1978 SSRC Conference on SI76.

2/F. Narin and M. Carpenter, "Science Output Indicators Based on Bibliometric Techniques," discussion notes and informal lecture delivered to Science Output Indicators Seminar in Washington, D.C., January 16, 1979. For general reference see also articles by N. Wade, "Citation Analysis: A New Tool for Science Administrators," Science, May 2, 1978, pp. 429-432; and S. Aaronson, "The Footnotes of Science," Mosaic, NSF, March/April 1975, vol. 6, no. 2.

b. Use of "activity indices" for various countries, as defined by Narin, et al. 1/ Activity Index =

$$\text{Activity Index} = \frac{\text{Percent of a country's share of world's publications in one field}}{\text{Percent that field shares in total world's publications.}}$$

This activity index might be used to approximate a "significance" index, which would indicate a country's research priorities.

c. Another "significance" index might be based on Narin's "Cross Country Citation Matrix," which compares citing country with cited country for a certain field. Here, significance would be defined differently than in the above example, being based on peer recognition of contribution, the assumption being that the greater the <u>recognized</u> significance of the work, the more international citing it will receive. The major limitation with this measure also involves the incentives behind international citation, i.e., that citing a certain country's work will add more support or prestige to a certain article. However, this still involves <u>recognized</u> significance. The diagonal of the matrix also gives some indication of the relative degree to which each country cites its own work, i.e., of its publication "isolation."

d. Growth in the number of journals within the past 30 years, both in the United States and other nations, measured by numbers of both field-specific journals and more general journals and "newsletters" aimed at a broad spectrum of the scientifically trained, e.g., <u>Scientific American</u>, <u>Science</u>, <u>Science News</u>, <u>Mosaic</u>, etc. These two main types might be disaggregated to separate the "need" for more field-specific information from the "incentive" to inform both scientists and laymen about general scientific developments.

2. Another possible concept to measure is the "growth" or "contraction" of university science departments in selected fields. This could be measured by trends in the numbers of faculty, graduate students, and non-faculty researchers such as research associates and

---

1/F. Narin, <u>op</u>. <u>cit</u>.

post-doctoral researchers. A potentially useful breakdown would involve separating faculty from graduate students, or constructing and tracking a "faculty accessibility ratio" measured by $\frac{\text{number of faculty}}{\text{number of students}}$ for each field. This ratio might help to further characterize certain differences in the education of new researchers.

3. "Capital intensity" of different fields might be measured by the percentage of expenditure for each field that goes towards equipment.

4. An indicator of "research scale" might be measured by a comparison in trends in grant funding for research teams (e.g., more than four researchers) with that for single investigators.

5. An indicator that would probably add considerable information to _SI_ reports would be an "employment demand index" for scientists and engineers. The widely used, long-standing Deutsch, Shea, and Evans Index is based upon square footage of employment advertising in both newspapers and professional journals such as _Chemical and Engineering News_. This index has existed since 1961, is seasonally adjusted for recognized fluctuations in labor demand, and is constructed through the weighting of its different major components. It is a gross monthly index, lacking field specificity, and does not distinguish scientists from engineers. It is thus most useful for industry, less so for Government, and of little value for academia. However, when combined with the information offered by surveys done by the College Placement Council, which collect "demand" information for bachelor's, master's, and doctoral degrees (how many offers, at what salary, for which degrees), an overall estimation of demand in both industry and Government can be derived. These surveys are done three times a year, and thus collect data which is timely.

   The basic assumption underlying the Deutsch, Shea, and Evans Index is that there is a strong correlation between the number of job openings and the amount of advertising. This correlation appears to be true only on the gross level, since many individual fields have different networks for announcing their openings. For example, annual meetings of professional societies are often used for recruitment purposes, and recruitment officers from numerous companies visit many campuses in search of new talent. Simple word-of-mouth is also a frequently used method of recruitment for several fields. Also, the growing number of professional recruitment agencies (or contractors) registered many job openings that otherwise may not be advertised. However, on the whole, the Deutsch, Shea, and Evans Index appears to us a reasonably valid gross measure because the amount of advertising will to some degree reflect the overall need for relevant personnel. One user of the index, Ms. Betty Vetter, Chairwoman of the AAAS Committee on Scientific Manpower, has stated that the indicator can be best trusted when it shows a trend spanning more than 2 or 3 months; i.e., if it turns down for a 6-month period, it is relatively safe to say that a decrease in industrial (and partially Government) demand for personnel has also taken place. Thus, if followed over a number of years, the indicator may reveal significant trends. Also, if combined with other data gathered on recruitment, this indicator would be considerably strengthened.

6. An indicator of professional movement of personnel in science- and technology-related areas would be a useful addition to SI's present data on manpower. Present data are often expressed in terms of stocks of particular types of people rather than flows of resources from one area to another.

   Such an indicator would view an individual's participation in scientific work as having a life cycle composed of a number of states: pre-high school, high school, college work in science and engineering (S&E), non-S&E college work, graduate training in S&E, etc. Given the appropriate statistics for each year, one can construct a matrix showing the flows of individuals between the different states of the system in a given time period (see figure 5). Each individual is classified by his or her state in the initial year one wishes to look at and by the status in the final year. The diagonal elements would show those people who remain in a particular state over the year, while the off-diagonal elements would show transitions between states. It is then possible to identify patterns of movement from one state to another after adjusting the matrix to account for overall population growth. 1/

Figure 5

Matrix of Personnel Flow

States a/	Years 1	2	3	4	5	6	7	8	9	10	11
1. Pre-college	.	.	.	.	.	.	.	.	.	.	.
2. S&E college work	.	.	.	.	.	.	.	.	.	.	.
3. Non-S&E college work	.	.	.	.	.	.	.	.	.	.	.
4. S&E graduate training	.	.	.	.	.	.	.	.	.	.	.
5. Non-S&E graduate training	.	.	.	.	.	.	.	.	.	.	.
6. Full-time R&D work in industry	.	.	.	.	.	.	.	.	.	.	.
7. Full-time R&D work in academia	.	.	.	.	.	.	.	.	.	.	.
8. Full-time R&D work in Government	.	.	.	.	.	.	.	.	.	.	.
9. S&E teaching	.	.	.	.	.	.	.	.	.	.	.
10. Non-S&E-related work	.	.	.	.	.	.	.	.	.	.	.
11. Other	.	.	.	.	.	.	.	.	.	.	.

a/These classifications listed above are for purposes of illustration only and are not meant to be definitive or exhaustive. One could disaggregate the academic R&D work by field (physics, chemistry, etc.) or possibly disaggregate the industry R&D workers by research and nonresearch or development, for example.

1/This suggested indicator has been adapted from Mervyn A. King, "Primary and Secondary Indicators of Education," Social Indicators and Social Policy, Andrew Shonfield and Stella Shaw, eds., Social Science Research Council, 1972.

This type of matrix, properly used, could serve to do the following:

a. Find the average time in years spent in a particular state; i.e., how many years might the average student spend in graduate education, or how long might the average researcher expect to be employed in full-time R&D work before moving to another job or retirement.

b. Find a number of interesting probabilities from the data:

--What is the probability when entering college of going into full-time R&D work; what is the probability of moving from one discipline to another or to interdisciplinary work; or what is the probability of doing non-S&E-related work, given graduate training?

This indicator of personnel flow might also improve SI's treatment of education in science and technology, since many of the possible states involve education (i.e., scientific and engineering graduate training).

7. SI might also use alternative price indices for R&D. Inflation is universally recognized by the science policy community as having had a large effect on the conduct of and funding for R&D. However, it is not so well agreed that deflating the figures for R&D spending by the GNP deflator (as is done in SI76) is the best way of taking the effect of inflation into account. For example, Langdon Crane is skeptical about the accuracy and desirability of deflating R&D budgets. 1/ William Carey, in an August 26, 1977, editorial in Science magazine, expresses the view that the research dollar may have been devalued by one-third in the last decade due to such things as equipment obsolescence, increased paperwork, and rising overhead rates in universities and colleges. In view of the experimental nature of SI, some attempts at developing an overall (or disaggregated) R&D price index would be valuable. To this end, we list here

---

1/T. Langdon Crane, "SI74 and Basic Research: A Partial Analysis," in measuring and evaluating the results of federally supported research and development, science output indicators - Part I, Special Oversight Hearings, Committee on Science and Technology, House of Representatives, 94th Congress, May 19 and 26, 1976, pp. 76-90.

some of the attempts to derive alternative price indices for R&D.

Lawrence Goldberg, in a paper entitled "Federal Policies Affecting Industrial Research and Development," developed R&D price indices for 14 manufacturing industries who fund 94 percent of the total R&D expenditures by industry. 1/ These indices show some differences when compared with the overall GNP deflator. For example, Goldberg's R&D price index for the chemical industry increased 15 percent between 1973 and 1975, compared to the GNP deflator which increased 19.4 percent over the same period. Battelle Laboratories also has developed an R&D price index. 2/ This index shows that the cost of doing R&D has increased much faster than the GNP deflator since 1960. Although there may not be agreement among the various papers in this field, SI should attempt to incorporate past work or to develop new indices in this area.

1/Lawrence Goldberg, "Federal Policies Affecting Industrial R&D." Unpublished paper presented at the Southern Economic Association meetings, Washington, D.C., November 1978.

2/Probable Levels of R&D Expenditures in 1977: Forecast and Analysis. Batelle, December 1976. Cited in Support of Basic Research by Industry.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The complexity of science and technology and the diverse manner in which the two are interrelated with almost every societal activity, permits for only a general description of the dimensions of the enterprise, and defies precise definition. However, quantitative measurement is an important part of the appraisal that seeks precise definition. Thus, as discussed in this report, the quantitative measurement of science and technology is an arduous but nonetheless vitally important activity.

In previous chapters, we have examined quantitative measures of science and technology in a recurring Government report. In our opinion, quantitative measurement, when properly constructed and presented, serves as a valuable resource for policymaking. Such measures, although an important tool, are only one of many resources used by policymakers.

It appears that SI, the STAR, and the Five-Year Outlook report (first biennial edition scheduled for January 1980), will be viewed in concert as the most comprehensive appraisal of U.S. science and technology. Although the first volume of the Five-Year Outlook is not yet completed, our observations on the role of this report are based on information from NSF. This trio of documents is augmented by publications on selected topics, such as those issued by the National Research Council, NSB reports issued on alternate (non-SI) years, the annual American Association for Advancement of Science report on Federal R&D policy, nonrecurring special topical reports (e.g., Smith and Karlesky, The State of Academic Research), and the large range of statistical publications by the Science Resources Studies Division of NSF.

PURPOSE OF THE REPORTS: SI, STAR, AND FIVE-YEAR OUTLOOK

There has been some confusion concerning the interpretation of the mandate for SI. Additionally, the initial purposes of the STAR and the Five-Year Outlook were apparently confused due to the transfer of the responsibility for writing the reports from OSTP to the Director of NSF (Reorganization Plan No. 1 of 1977 and Executive Order No. 12039, February 24, 1978). Therefore, the staff of NSF has been trying to separate and clarify the purposes of the three reports for which they are now responsible. We offer the following observations concerning these purposes.

First published in 1973 (SI72), SI is a biennial series resulting from NSB's interpretation of its vague 1968 mandate for an annual report. The essence of the mandate was that NSB should "assess science." NSB originally cast SI as an experiment that would attempt to provide policymakers with important statistics concerning the state of the science and technology enterprise. Alternate year reports by NSB would then be concerned with more topical issues.

In 1976, as part of legislation to centralize science policy responsibility in the Executive Office of the President, Congress mandated two additional reports--the STAR and the Five-Year Outlook--which affected the role of SI. In contrast to the vague mandate for the NSB's annual report, the executive branch was given more specific instructions as to the nature of the STAR and the Five-Year Outlook. This mandate requested reports which were to be substantially different from the existing SI in one fundamental way: whereas SI had been an attempt to quantitatively describe the state of science, the 1976 legislation called for the STAR and the Five-Year Outlook to appraise and evaluate how science and technology could contribute to society. Thus these two later reports were explicitly directed to discuss how problems within the scientific enterprise would inhibit the societal contribution of science and technology. This mandate reflected the prevalent rationale for extensive Government involvement in science and technology.

These three reports provide different kinds of information for policy, and thus they are complementary. SI has begun to design a broad statistical basis for understanding and assessing the science and technology enterprise. We believe that SI can provide a necessary background component for policy. Furthermore, it can also serve as a significant data base for both the STAR and the Five-Year Outlook to test, support, or identify perceptions, knowledgable opinions, and analyses on topical issues of concern to policymakers.

NSB seems to be faced with a dilemma here. On the one hand, it may wish in the future to develop indicators around topical policy issues so that SI can be of immediate use to policymakers. On the other hand, NSB could broaden its approach and orient SI primarily toward the assessment of the condition and direction of science and technology, while maintaining an objective detachment from popular short-term issues.

We believe that NSB should adopt the latter approach for SI. To restrict indicator selection or interpretation to current issues would compromise the long-term benefit of SI in providing a continuing portrait of science and technology.

Analysis of topical issues could be offered separately, in other reports, perhaps based on some of the data developed for SI. This is precisely the mandate for the STAR and Five-Year Outlook--to analyze topical policy issues and programs, utilizing the resources available. These two reports, with their emphasis on the social utility of science and technology, could draw from the various models of science and technology developed by SI in order to understand present policy issues. Additionally, NSB may wish to offer judgment and evaluation in special or alternate-year reports.

RECOMMENDATIONS FOR IMPROVING SCIENCE INDICATORS

We believe that SI should be a broad effort designed to yield quantitative information on how science and technology are faring in the United States. In this manner, the SI series can be of continuing utility. We share NSB's view that SI is an experiment. Since the development of indicators is still in its infancy, the need for further research on science and technology indicator design and construction merits emphasis.

In order to yield as thorough a portrait as possible of the changing state of science and technology, we suggest that SI expand its view for examining the scientific enterprise. At present, this perspective is restricted by the implicit input-output model of science and technology. A major problem with this approach is that it focuses on resources and results, and leaves out both the process and substance of science. Science is seen mainly in terms of its economic, not scientific, importance, as leading directly to technology. This view has had the effect of excluding information about scientific advances and where they directly might benefit society. The input-output approach is consequently inadequate to cover many of the most important elements of U.S. science and technology. By viewing science and technology from a variety of models, SI could better examine the actual processes and operations involved and thus provide a more substantial quantitative basis for understanding science and technology. In seeking this improvement, it may be necessary to draw directly on both specific research and general knowledge concerning the sociology of science and the processes involved in technological innovation.

An attempt should be made to take more of a conceptual approach to the design and structuring of indicators. It was natural that the initial SI reports would be based largely on an operational approach, deriving indicators from the readily available data on the basis of suspected importance. This approach, however, incorporated a limited view of science and technology, and led to the construction of a number of indicators whose underlying assumptions are tenuous or invalid. The approach also increased the difficulties which are inherent in overall interpretations by confusing the meaning of some of the measures. The result has been that the assessment originally called for by legislation to date has not been achieved.

A more conceptual approach should be adopted whereby the important concepts to be measured (such as "research activity", "frontiers in science") are chosen and explicitly defined in the text along with their limitations and underlying assumptions. _SI_ reports would then be open to more focused and constructive criticism concerning the validity of their indicators. This would continuously advance the evolution of this kind of measurement of science and technology.

The extensive research still needed may require resources in addition to those presently available to NSF in the _SI_ effort. We believe that consideration should be given to whether sufficient resources are presently allotted to the _SI_ staff.

We recommend that NSB direct the _SI_ staff to:

--Use different models of science and technology to present a spectrum of important concepts which need to be measured. Particular attention should be given to developing indicators of the process and substance of research.

--Emphasize a more conceptual approach in designing indicators which first identifies what will be measured, and then generates the appropriate data.

--Include overall interpretation of the meaning of indicators without emphasizing short-term topical policy issues.

--Consider our suggested alternative indicators.

--Continue to experiment in the _SI_ series by developing and testing new indicators, and by reevaluating and improving old ones.

--Attempt to more clearly differentiate science from technology and develop distinctive indicators for each.

--Consider whether sufficient resources are available to the _SI_ effort to perform essential research and experimental development of new and improved indicators.

STATEMENTS OF PURPOSE FOR SCIENCE INDICATORS SERIES

The following are the different statements of purpose, including legislative mandates and the oral and written statements of NSF personnel closely connected with the SI effort.

1. Original legislation - Public Law 90-407, July 18, 1968, required the National Science Board to publish an annual report (which SI later came to satisfy every odd year).

   "* * * shall render an annual report to the President * * * on the status and health of science and its various disciplines. Such report shall include an assessment of such matters as national scientific resources and trained manpower, progress in selected areas of basic scientific research, and an indication of those aspects of such progress which might be applied to the needs of American society."

2. From testimony by Roger Heyns, first Chairman, Science Indicators Committee (NSB), Special Oversight Hearings on Science Indicators before the Subcommittee on Domestic and International Scientific Planning and Analysis of the House Committee on Science and Technology, May 19 and 26, 1976.

   "By late 1971, we [the Board] had identified seven major purposes or functions of the science indicators reports, most of which in retrospect still are appropriate:

   1. To detect and monitor significant developments and trends in the scientific enterprise, including international comparisons.

   2. To evaluate their implications for the present and future status and health of science.

   3. To provide the continuing and comprehensive appraisal of U.S. science.

   4. To establish one new mechanism for guiding the Nation's science policy.

5. To encourage quantification of the common dimensions of science policy, leading to improvements in R&D policy setting within Federal agencies and other organizations.

6. To stimulate social scientists' interest in the methodology of this type of report as well as their interest in this important area of public policy.

7. To provide a regular focus for the Board's annual reports."

3. NSF internal review (1977) committee based their critique on the following stated purposes or aims of _SI_, as stated by H. Averch, Assistant Director, NSF, in his original memo to the other _SI_ evaluation task force members.

"* * * whether _SI_ content or format are appropriately contributing to (1) the identification and illumination of policy issues related to science and technology; (2) the prediction or anticipation of future problems in science and technology; and (3) an integrated portrait of the state of science and technology.

"A list of limited objectives for the _Science Indicators_ reports might include the following:

1. to provide information that portrays and relates to the state of, and trends in, various aspects of science and technology;

2. to assess the trends as to their implications for the present and future state of science and technology;

3. to illuminate existing policy issues and options, especially to shed light on the existence or nonexistence of alleged problems;

4. to identify possible new policy issues."

4. In each of the _SI_ reports (1972, 1974, and 1976):

> "The purpose and function of science indicators is to follow changes in the scientific enterprise and its components over time, and thereby to reveal strengths and weaknesses as they begin to develop."

With slight wording changes this is to be found in each of the introductions to the three reports. Furthermore, in the latest of these, SI76, there are also other stated aims:

> "* * * continuing effort to develop indicators of the status of science and technology in the various sectors of the U.S. economy."

> "* * * should provide an early warning of events and trends which might reduce the capacity of science--and subsequently technology--to meet the needs of the nation."

> "* * * this effort to better understand the scientific enterprise."

5. Comments from interviews with members of NSF concerned with SI:

> "* * * science indicators is not supposed to be evaluative, but should concentrate only on measurement, on quantifying economic-type indicators for science."

6. Public Law 95-99, August 15, 1977. 1978 Authorization for NSF:

> "The Board shall render an annual report to the President, for submission to the Congress on or before March 31 in each year. Such report shall deal essentially, though not necessarily exclusively, with policy issues or matters which affect the Foundation or with which the Board in its official role as the policymaking body of the Foundation is concerned."

There is both agreement and discrepancy among these different statements of purpose. For example, what is stated in the introduction to each SI report agrees with the seven objectives in Dr. Heyns' testimony which interprets the original legislation. Both of these would direct SI at some evaluation or assessment of how the "scientific enterprise" is doing, both scientifically and in its ability to meet national needs. Strongly at odds with these, however, are the remarks by NSF personnel involved with SI which state that the reports were meant to emphasize quantitative data and not venture at all into evaluations or assessment. Quantification is certainly central to the effort, but it was not the single, underlying objective in these statements of purpose. In fact, there is even some discrepancy with respect to what should be quantified--"the common dimensions of science policy" (underlining added) or "the condition of science and research in the U.S."

A central difficulty we have noted is the degree to which the indicators are supposed to deal with both science and technology. Initially, the emphasis was clearly on science; technology was considered a result or "output" of science. However, as shown by later introductions to SI reports, though it was apparently recognized that this was an oversimplified view of technology, the degree to which technology indicators needed to be included remained unclear. Their inclusion has increased to the point where they occupy nearly half of SI76. Yet it cannot be discerned whether the original emphasis on science still exists, even in a subdued form. This area of confusion is not directly dealt with in the other statements of purpose. Also, it is unclear what the term "scientific enterprise" includes--whether all of science and all of technology, or only a subset of this. There needs to be clarification here since such confusion directly affects the basic scope of the SI reports. Our interpretation of SI's purpose mentions both science and technology, thereby keeping this scope as wide as possible.

Each of the stated objectives for Science Indicators, even the listed seven "major functions," are highly general, in places, vague. It appears that the initial hopes about the potential of SI to construct "a comprehensive appraisal of U.S. science" predominated over considerations of the inherent difficulties, some of which have been discussed. As a consequence, the achievable scope of SI has had to be reevaluated and altered as the project has evolved. An example of this is the initial focus on science, whereas the inclusion of technology has increased in the later two reports. The crucial point seems to be that confusion in the stated purposes has come from an increasing awareness of the inherent limitations in the overall effort.

In examining the collection of stated purposes, it is apparent that SI has been intended to do more than merely collect and present data with a few basic interpretations. Instead, it appears that SI's "assessment" should involve overall interpretation which discusses causal relationships to the degree possible. Furthermore, it appears that though SI was not meant to present by itself a comprehensive portrait of U.S. science, it was intended to develop quantitative information on as many major dimensions of science (and later, technology) as possible. The added requirement from Public Law 95-99, 1978 NSF Authorization Act, that "policy issues or matters" also be considered in the formulation of the report is relevant to future SI reports (i.e., post-SI76). In our opinion, this mandate is sufficiently broad to allow the earlier, more direct statements concerning the specific goals of SI to remain relevant. We then derive the following central purpose for the SI reports: SI is to quantify as many of the dimensions of science and technology as is feasible, with a view towards identifying significant trends and developments in these parameters, interpreting possible causal relationships between them and analyzing their possible influence on the condition and direction of science and technology in the United States. This stresses the need for overall analysis and interpretation, while maintaining that evaluations do not belong in SI.

NSF COMMENTS AND GAO RESPONSES

LETTER WITH ENCLOSURES

NATIONAL SCIENCE FOUNDATION
WASHINGTON D C 20550

August 21, 1979

Mr Harry S Havens
Director, Program Analysis Division
U S General Accounting Office
441 G Street, N W
Washington, D C 20548

Dear Mr Havens

We have recently received for review the final draft of a report from the General Accounting Office (GAO) titled "Science Indicators Improvements Needed in Design, Construction and Interpretation (Code 971350)" We appreciate the opportunity to provide you with some comments and the interaction of the authors of the report with NSF officials who have played a key role in the preparation of the science indicators series. Nevertheless, some misconceptions still seem to be reflected in the report, and we hope that our comments will help to clarify these

As Science Indicators--1978 is now at the printer, it was not possible for the GAO team to review this document and the GAO report is based primarily on a review of Science Indicators--1976 Since the development of science indicators is still a relatively young endeavor dealing with very complicated systems and concepts, it is not only natural but expected that continuous, mainly beneficial evolution is taking place Consequently, it is noteworthy that some of the recommendations of the GAO report have already been incorporated in Science Indicators--1978 Section A of the enclosure to this letter provides a detailed listing of these items It is reassuring that the GAO staff and the authors of the science indicators series agree on these improvements.

As may be expected, there is not agreement on all points The GAO report makes five major recommendations on which we would like to comment First, we too believe that more interpretation of the presented quantitative information will be useful, and this approach is already reflected in Science Indicators--1978 Second, we agree that continuous experimentation with new indicators is an absolute necessity and are actively engaged in such efforts

However, we do have difficulties with the recommendation that different "models" of science and technology could be used now in the development of science indicators This suggestion ignores the fact that the science and technology enterprise represents an extremely complex system affected by many internal and external factors which are far from understood This is reflected in the lack of agreement among expert analysts of the system as to whether valid, reliable total models are possible or whether existing partial models are correct or applicable Thus, while further studies of possible models should be continued, the application of models at the current state-of-the-art has to be very tentative As a matter of fact, indicators which are independent of the choice of a model may be the most useful ones

The science indicators used are generally based on definite concepts, whenever this is possible Frequently, multiple concepts are reflected by a single indicator or a combination of indicators However, caution has to be used in interpreting indicators only in terms of single concepts since this may be misleading when the suggested conceptual approach turns out to be faulty Again, Science Indicators--1978 in its greater interpretive mode tries to point out alternative interpretations when this is called for or possible

Finally, every attempt has been made to clearly differentiate science from technology, whenever this is possible or appropriate Thus, the report contains a whole chapter on Basic Research, which is clearly science, and another on Industrial R&D, which is primarily technology However, the report appropriately takes into consideration that science and technology represent a continuous spectrum of both activities which are naturally dependent and interactive

Besides the comments on major GAO recommendations, we have a number of more specific comments on statements made in the draft report Some of these represent factual errors, others are differences in point of view These comments are listed in Sections B and C of the enclosure

We hope that our comments and remarks will clarify points of agreement and disagreement and will provide you and the Congress with a better understanding of a complex subject matter

Sincerely yours,

Norman Hackerman
Chairman
National Science Board

Richard C Atkinson
Director

Enclosure

ENCLOSURE

A GAO Recommendations Incorporated in Science Indicators--1978

A number of the comments or recommendations of the GAO report reflect critiques of early Science Indicators versions and are no longer applicable to the current science indicator effort, and in particular to Science Indicators--1978

GAO Draft Report Page Numbers	
iii	SI--1976 includes relatively little interpretive narrative. Consequently, the alleged "gloomy" tone can only be inferred from its quantitative information. By 1978, the general support picture had improved, which will be reflected in SI--1978. Thus, the overall tone of each Science Indicators report is set primarily by the data it contains
32	A different and more comprehensive literature data base is presented in SI--1978 than the the one discussed by GAO
41-43	Many of the problems concerning patent data which GAO discusses have already been dealt with in SI--1978 For example, the "patent balance" concept was dropped and replaced by other approaches Domestic patenting trends in different countries have been added to SI--1978 and analyzed. Detailed description of the meaning of, and uses of, patent data have been introduced in SI--1978, drawing upon and citing some of the same sources cited in the GAO report In addition, four experts on patents and their use as indicators participated in preparing this section
43-45	GAO feels that international comparisons of Government R&D expenditures are misleading, particularly between the United States and Japan, because of country differences in priorities and funding patterns The differences between Japan and the United States in terms of public versus private sources of R&D funding are highlighted in SI--1978, thus making it even clearer that the indicator of Government expenditures by objectives reflects only Government R&D priorities, not the total national effort Three new indicators in SI--1978 provide information on support of R&D by other sectors

- 2 -

GAO Draft Report Page Numbers	Comment
45-48	The GAO discussion regarding the innovation indicators from the Gellman Report is no longer applicable For some of the reasons repeated in the GAO report, the "Gellman Study" of innovations was not used in SI--1978
48-51 (and 33)	GAO feels that Science Indicators has made little use of bibliometric data such as citation analysis Science Indicators has been one of the main supporters and users of research in this area and has continued to improve and increase the bibliographic data series with each report Although GAO presumes that citation ratios are used to measure research quality, SI--1978 does not use citation ratios as an indicator of quality, nor did SI--1976 Rather, they serve as measures of the influence and utility of U S scientific literature Several of the GAO report's caveats regarding the use of S&T literature indicators were explicitly stated in SI--1976 and expanded in SI--1978
51-54	The Public Attitudes survey, which is not included in SI--1978, has been redesigned technically to include policy issues and to identify separately the subgroup of the public that actively follows science and technology The redesigned survey will be undertaken in the Fall of 1979 and published later
55	The GAO report agrees with the internal NSF recommendation that there be more explicit rationale for presentation of indicators The indicators in SI--1978 are introduced with the rationale for their inclusion There is also an increase in SI--1978 in the treatment of con-conceptual limitations of specific indicators
55-56 61-62 (iii)	Throughout the report, GAO cites the lack of interpretation as one of the "major limitations of SI--1976 " SI--1978 has considerable interpretation of both short- and long-term policy issues for the indicators

- 3 -

GAO Draft Report Page Numbers	
56 63-65 (v)	GAO has recommended the inclusion of information developed by academic researchers not directly connected with NSF SI--1978 contains, for example, much of the work by Terleckyj and by Mansfield (whose research is recommended by GAO) A number of the researchers cited by the GAO report were actually reviewers of draft SI--1978 chapters
62	Although SI--1978 does not have a summary chapter per se, as GAO recommends, it does have summary sections in each chapter and for the first time a detailed index

B General Comments

1 Several major conceptual issues are raised in the GAO report that should be put into a state-of-the-art perspective It certainly would be desirable to have Science Indicators based on an explicit, detailed model of the operation of the science-technology system However, very little recognition is given to the fact that this system is very complex and not understood sufficiently well to permit the development of overall models Some subelements are better understood and whenever this is the case, interpretation in terms of causal factors has been attempted in the as-yet-unpublished Science Indicators--1978

2 The GAO report claims that SI "is too constricted by an input-output framework " It is correct that SI includes many "input" measures and fewer corresponding "output" measures However, these measures are included because they are objective and quantitative, in line with the stated objectives of SI (pointed out on p 28 of the GAO report) and not because of an oversimplified plan to adopt the rigid notion of input-output The implicit input-output or "black box" model does not exist and it would be incorrect to use such a model if it did exist only by itself

3 While one section of the GAO report (p 28) correctly states the objective of SI, other parts seem to ignore the objective Thus, the criticism that the internal operations of science and advances in knowledge are not covered fails to take into consideration that these aspects of the science and technology system are not suitable for quantitative description or analysis Similarly, while Chapter 5 points out correctly the complementary nature of the Science and Technology Annual Report (STAR), SI, and the Five-Year Outlook Report, criticisms in the Summary, Digest, and earlier chapters

assume that SI itself is intended to provide an overview of all aspects of the scientific and technological enterprise As pointed out in SI, a report dealing primarily with quantitative aspects of the system cannot do this as many facets of the system do not lend themselves to this type of approach

4 While SI can certainly be improved, few of its positive aspects are mentioned

5 Many of the inherent assumptions (p 30) ascribed by the GAO to the alleged SI model are not self-evident For example

> "2 The state of science and technology can be described primarily in terms of what is being added each year to the overall base of expertise, without sufficient reference to the accumulated base "

As a counter-example to this alleged assumption, SI analyzes the activities of the total number of scientists and engineers in great detail To take another example of an alleged assumption (p 31)

> "9 Physics, chemistry, biology, and math are the most important sciences, i e , deserve the most complete coverage "

Coverage of the social sciences is in line with their share of overall scientific activities in all fields

Other alleged assumptions deal with areas not even discussed, covered, or even assumed in SI, e g , "science and technology exist in a social context and have minor self-regulating aspects" or "science and technology advance incrementally with no 'sudden' changes "

6 It is unfortunate that the GAO report misunderstood what NSF staff said about greater interpretation in SI--1978 SI--1978 does not narrow its interpretation to immediate popular issues, but also provides strategic information on broader, longer term issues for science decision makers Furthermore, the material related to policy issues was not selected on the basis of popularity or political impact

7 Differences in time are sometimes not adequately taken into consideration Thus, the alleged disparity of statements of SI purpose (GAO Report Appendix) do not reflect a state of ambiguity, as stated in the report, rather, they consist of statements made at various times during the last nine years and thus reflect evolution and sharpening of the concept accompanied by increased

- 5 -

awareness of the implicit limitations of the approach Similarly, the fact that STAR shows more optimistic trends than SI--1976 is due to the fact that STAR, which was published in 1978, covers two more years of data SI--1978 will show similar trends

8 While SI frequently discusses R&D *in toto* to provide an overview of the overall U S technical effort, science and technology are covered separately to the extent possible Thus, one out of six chapters is completely devoted to basic research which is completely in the science domain and the industry chapter separately treats industrial basic research from other R&D activities which are generally in the technology area

9 The critique of some specific indicators frequently reflects lack of appreciation of the multiple use of statistical data, i e , the fact that data of one type, such as expenditures, can be analyzed through various types of crosscuts (p 38) or complete knowledge of the nature of the data

10 While scientific personnel and expenditure measures represent a significant portion of the indicators, this information is not duplicative but rather complementary Both give indicators of activity, but not necessarily identical ones Thus, the data clearly show that changes in expenditures are frequently not accompanied by corresponding changes in personnel

- 6 -

C Specific Comments

GAO Draft Report Page Numbers	Comment
17	Science Indicators reports have never been based only "on data which is already collected " Examples are international literature, indicators on public opinion surveys, patent data by SIC groupings, Soviet R&D resource statistics and many others
30	The statement that nearly all the measures listed on page 29 of the GAO report ("original" indicators proposed) were available from within the Federal Government, particularly from SRS, is incorrect for the first, fourth, fifth, eleventh and twelfth items Thus, it is true for only 7 out of 12 Furthermore, the items listed represent only about one-fourth of the original items, many of which were even less available from existing Government sources
31-32	The GAO report questions the use of basic research expenditures and S&T literature because the data show a decline in basic research funds and "an unbroken increase in the number of overall publications through 1975 " This is a result of the lagged effect between the actual research expenditure and the actual date the publication is reflected in the abstract journal An important point was made in SI--1976 "Available data for 1975 showed steady growth in only 3 of 11 fields since 1973, five fields not changing substantially from the 1973 level, and three dropping below the 1973 level " (p 88) Thus, the data for 1975 may have just begun to reflect the lessened financial support
43-45	The GAO report states that "The Organization for Economic Co-operation and Development (OECD) which collected the data accept the definitions and stated purposes of the cognizant governments without further analysis, i e , if more money is budgeted for R&D by a military organization, it is for military research "

- 7 -

GAO Draft Report Page Numbers	Comment
43-45 (Cont'd)	On the contrary, the OECD nations themselves developed the classification scheme and definitions for R&D expenditures across functional areas, under the auspices of OECD The scheme was adopted by the various nations as the most effective way to show Government support for R&D in these areas. OECD staff analyze and verify the responses for adherence to definitions and classifications
45	The GAO report uses the Japanese emphasis on Development as an argument for the invalidity of these "functions" categories for comparing the U S with other countries A 1975 OECD report, however, showed that Japan and the U S have the same share of total R&D devoted to development 64 3 percent and 64 4 percent, respectively
52	The GAO report states that SI--1976 apparently made little use of available data from other studies in the Public Attitude Chapter In fact, six outside sources of data contributed to tables 6-3, 6-4, 6-5, 6-12, 6-13, and 6-19 and ten more were used in the text, including one recommended by GAO
52	The Public Attitude Chapter of SI--1976 does not "muddle" science and technology, as the GAO report asserts They are separate for many of the survey questions See tables 6-2, 6-3, 6-15, 6-17, 6-18, 6-19, and 6-21
53	The GAO report criticizes the Public Attitudes Chapter of SI--1976 as treating the public as a "homogeneous mass " On the contrary, many demographic subgroups are discussed in the text
56	Contrary to the GAO report's assertion, not all data in the SI reports have been Government-generated or Government-sponsored (e g., College Placement Council salaries of beginning S/E's, NAS information on international congress attendance, public attitude surveys, etc) Furthermore, much of the Mansfield and Terleckyj work cited by GAO as examples of non-Government-sponsored efforts has been, in fact, Government-sponsored, as is over half of university-based social

GAO Draft Report Page Numbers	Comment
56 (Cont'd)	science research Government sponsorship does not in itself negate the value of research and analysis
62	The GAO report urges attention to "where science and technology are focused" and "where they are least intensive " SI--1976 has a large number of indicators showing where R&D is performed, where S/E's are employed and fields of science for R&D and S/E's In addition, R&D-intensiveness is specifically treated in Figures 1-28, 1-29, 1-30, 4-16, 4-17, 4-32, 4-37, 4-39 and 4-41

GAO RESPONSES

The letter from NSF highlights comments, most of which are presented in more detail in sections B and C of the enclosure. Our responses are therefore directed as noted to each section of the enclosure.

Section A

We are pleased that SI78 will include the improvements stated by NSF. However, since NSF declined to make a draft of SI78 available to us, we were unable to assess these changes.

We have changed the text on p. 34, to make it clear that NSF was not presumed to used citation analysis as a measure of research quality.

Other comments in section A which indicate disagreement with statements in the GAO draft report are repeated or expanded in sections B and C. To avoid redundancy, we respond directly to sections B and C.

Section B

1. For the present state-of-the-art, it is very important that SI base indicator construction on a variety of experimental models--not any individual one. We agree that the complexity of the system of science and technology makes this difficult. PP. 14-17 of this report discuss these difficulties.

2. Our report discusses the input-output model that we believe is inherent in the selection of data and indicators used in SI. We do not impute any motive or intent to limit SI to a restricted model.

3. We agree that SI is meant to be quantitative, and that not all aspects of the science and tecnnology system are fully quantifiable. Indeed, on pp. 5-6, we discuss the limitations of quantification in general. We believe that some quantitative indicators for advances of knowledge and the internal operations of science can be developed. Examples of such indicators are proposed in our report (p. 41). SI should further experiment with new indicators and not simply rest with the assumption that some aspects of science and technology are not quantifiable.

4. Favorable comments are included throughout the report, as appropriate.

5. These are assumptions which we believe to be part of the model implied by the choice of data and indicators included in the SI series.

6. Since NSF declined to make draft copies of SI78 available to GAO, the statements in our report should not be construed as an evaluation of SI78. To be most useful in identifying significant trends and developments in science and technology, we believe that interpretation in the SI series should emphasize the long-term perspective, rather than current issues.

7. We do not agree that all of the confusion over the purpose of SI is due to its evolution. Variances also were evident in recent statements. We disagree that the differences in tone between the STAR and SI can be attributed to changes over time. In many instances, the STAR used different data largely from the same time period and analyses it differently to yield interpretations at odds with SI. Our report cites two examples of such analyses--for R&D/GNP and patents (pp. 25-28).

8. Although science and technology are treated separately to some extent, we do not believe that it is enough. Technology is sometimes separate, but sometimes treated as an "output of science." Additionally, science and technology are lumped together in most of the questions used in the survey of public attitudes section (pp. 34-36).

9-10. We agree that the same data can be used in different ways. But over reliance on just two data sets--funds and personnel, which are strongly interrelated--causes misleading results. We recommend more emphasis on a conceptual approach to measurement, in which ideas on what is to be measured are first generated, then data is found.

Section C--Note: The page references by NSF correspond to our draft report. The page numbers in parentheses refer to the present text.

Page 17 (12) - The test refers to a hypothetical instance, not necessarily to methods employed by the NSF staff in selecting data for SI.

Page 30 (19) - We believe that "most of the data" did exist in the Federal Government at that time.

Pages 31-32 (20-21) - We believe that this NSF comment further supports our view that the implicit input/output approach or model is inadequate by itself. We do not agree that the data are sufficient to support the NSF conclusion that the decline in basic research funds and the increase in number of publications "is a result of the lagged effect between actual research expenditure and the actual date the publication is reflected in the abstract journal." Conceivably, there could be other explanations such as an increase in the number of small groups of scientists working on a greater number of projects, thus yielding more papers. Increased use of computer-aided data processing could result in more timely and more frequent publications. Major scientific discoveries frequently lead to proliferation of related articles by many authors who recognize opportunities to expand on the first breakthrough. The simple input-output approach does not yield this kind of information. To establish causal relations, other types of indicators are needed which are related to concepts and models of the research process. Furthermore, although we acknowledge the NSF letter's quote of the text of p. 88 in SI76 with respect to publications in individual fields, we found two graphs in SI76 that appear to contradict the text. Graph 2-18 (p. 64, SI76) and the top graph of figure 3-20 (p. 89, SI76) clearly show "an unbroken increase in the number of overall publications through 1975."

Pages 43-45 (28-30) - OECD explains its approach to Government R&D data collection on pp. 77-81 in the report Changing Priorities for Government R&D. The following excerpted paragraph (pp. 79-80) states the OECD approach:

> "An international organization attempting a retrospective analysis of objectives is restricted by the type of data available for the countries chosen for study. In principle we have adopted the 'purpose' approach and have centred our analysis on the moment when the funds are committed, distributing expenditures on the basis of governments' intentions in supporting R&D programmes and institutes. This has, however, involved a certain amount of 'assessment' of government motives on our part but we have attempted to keep it to a minimum. * * *"

Page 45 (30) - This is precisely the point which we are making--due to categories which have different meanings for each country, the chart does not reveal the similar percent spent on development by Japan and the United States.

Page 52 (35) The text has been changed by substituting "not made adequate use" for "made little use."

Page 52 (35) - We disagree with NSF. In 17 of 21 tables, we believe that science and technology are incorrectly lumped together, including 3 tables (6-15, 6-17, and 6-18) for which NSF believes otherwise.

Page 53 (35-36) - We believe that the discussion of demographic groups in the text of this public attitude section of SI76 is inadequate.

Page 56 (37) - We agree with NSF that Government sponsorship does not negate the value of research and analysis. It is unfortunate that an earlier draft of the report implied this and we appreciate NSF pointing this out. This has been clarified in the text.

Page 62 (39) - We appreciate that SI76 includes some indicators dealing with where R&D is performed, where scientists and engineers are employed, and the fields of science for R&D. We believe that the SI series should further develop indicators with a broader perspective using concepts relating to the process of research and advances in knowledge.

(973150)

☆U S GOVERNMENT PRINTING OFFICE 1979 -620-167/327

Single copies of GAO reports are available free of charge Requests (except by Members of Congress) for additional quantities should be accompanied by payment of $1 00 per copy

Requests for single copies (without charge) should be sent to

U S General Accounting Office
Distribution Section, Room 1518
441 G Street, NW
Washington, DC 20548

Requests for multiple copies should be sent with checks or money orders to

U S General Accounting Office
Distribution Section
P O Box 1020
Washington, DC 20013

Checks or money orders should be made payable to the U S General Accounting Office NOTE Stamps or Superintendent of Documents coupons will not be accepted

PLEASE DO NOT SEND CASH

To expedite filling your order, use the report number and date in the lower right corner of the front cover

GAO reports are now available on microfiche If such copies will meet your needs, be sure to specify that you want microfiche copies

AN EQUAL OPPORTUNITY EMPLOYER

UNITED STATES
GENERAL ACCOUNTING OFFICE
WASHINGTON, D C 20548

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U S GENERAL ACCOUNTING OFFICE

US MAIL

THIRD CLASS