# DOCUMENT RESUME

ED 320 491 HE 023 574

| | |
|---|---|
| TITLE | Biomedical and Behavioral Research Scientists: Their Training and Supply. Volume III: Commissioned Papers. |
| INSTITUTION | National Academy of Sciences - National Research Council, Washington, DC. Office of Scientific and Engineering Personnel. |
| SPONS AGENCY | National Academy of Sciences, Washington, D.C.; National Institutes of Health (DHEW), Bethesda, Md. |
| PUB DATE | 89 |
| CONTRACT | N01-OD-2109 |
| NOTE | 186p.; For related documents, see HE 023 572-573. |
| AVAILABLE FROM | Studies Section, Office of Scientific and Engineering Personnel, National Research Council, 2101 Constitution Avenue, NW--Room GR 402, Washington, DC 20418. |
| PUB TYPE | Viewpoints (120) -- Information Analyses (070) |
| EDRS PRICE | MF01/PC08 Plus Postage. |
| DESCRIPTORS | *Behavioral Sciences; *Biomedicine; Employment Patterns; Evaluation Methods; Higher Education; Labor Market; Physicians; Productivity; Professional Training; *Research and Development; *Researchers; Scientists |

ABSTRACT

Volume Three of the Biomedical and Behavioral Research Scientists study contains five commissioned papers. The first paper, "Evaluating the National Research Service Award Program (NRSA): A Review and Recommendations for the Future," (Georgina M. Pion) reviews previous evaluation activities of the NRSA program and proposes an agenda describing the types of evaluation activities that should be carried out over the next 5 years. The second paper, "Productivity" (Helen Mofer Gee) discusses the complexities, issues, and problems in the use of specific productivity measurements. The third paper, "Health Services Research Personnel: Demand, Supply, and Adequacy of Training Resources" (Elizabeth McGlynn) describes the training of individuals engaged in health services research, the various pathways used to produce these researchers, the current levels of support for training, and future events that might affect labor demand. The fourth paper, "Training of Physician/Scientists" (Lloyd H. Smith, Jr.) addresses the role of physician/scientists in conducting the nations's biomedical research program and how national policy should be shaped to ensure that a sufficient number of physician/scientists are available in the future. The fifth paper, "Biomedical/Behavioral Cohort Model: A Technical Paper" (Joe G. Baker) provides a descriptive overview of a model used to project the future labor market for biomedical and behavioral scientists. (GLR)

ED320491

# BIOMEDICAL AND BEHAVIORAL RESEARCH SCIENTISTS: THEIR TRAINING AND SUPPLY

## Volume III: Commissioned Papers

Committee on Biomedical and Behavioral Research Personnel

Office of Scientific and Engineering Personnel
National Research Council

in collaboration with the
Institute of Medicine



NATIONAL ACADEMY PRESS
Washington, D.C. 1989



HE023574

**NOTICE:** The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

This study was supported by contract N01-OD-2109 between the National Academy of Sciences and the National Institutes of Health.

*Copies available from:*
Studies Section
Office of Scientific and Engineering Personnel
National Research Council
2101 Constitution Avenue, NW - Room GR 402
Washington, DC 20418

Printed in the United States of America

# BIOMEDICAL AND BEHAVIORAL RESEARCH SCIENTISTS: THEIR TRAINING AND SUPPLY

## Volume III: Commissioned Papers

# CONTENTS

# EVALUATING THE NATIONAL RESEARCH SERVICE AWARD PROGRAM: A REVIEW AND RECOMMENDATIONS FOR THE FUTURE

*Georgine M. Pion*[1]

The focus of this paper is (1) to review previous evaluation activities of the National Research Service Award (NRSA) program and (2) to propose an agenda describing the types of evaluation activities that should be carried out over the next 5 years. In line with this emphasis, a description of the major evaluation questions of interest to key program constituencies will be presented. Then, previous evaluation efforts will be discussed in terms of whether they addressed these questions and provided answers that could be viewed with a reasonable degree of confidence. The "match" between the questions of interest and the availability of sound evidence for answering these questions as gleaned from previous evaluations to date will serve as the basis for recommending future evaluation priorities.

Throughout this paper, two considerations should be kept in mind. One concerns the diversity of the NRSA program itself. At first glance the overarching goal of this program is relatively straightforward: to train individuals for health-related research and teaching careers. However, in achieving this mandate, several different components and activities are encompassed by the program. For example, NRSA awards support training in a heterogeneous group of disciplines, ranging from genetics to health services research, and activities are administered by a variety of federal agencies and institutes, each with varying levels of experience in supporting research training. The training sponsored by these agencies also differs in terms of academic level (undergraduate, predoctoral, or postdoctoral), target populations, e.g., M.D.s, Ph.D.s, or ethnic minorities), and strategies (e.g., short-term training versus formal degree programs or disciplinary versus multidisciplinary approaches). Further, distinct funding mechanisms (individual) fellowships versus institutional training grants) are used to

---

[1]I would like to thank David Cordray, Peter Rossi, Robert McGinnis, Grace Carter, and Robert Boruch for their critical and insightful comments on previous versions of this paper. Also, all the individuals interviewed, particularly Charles Sherman and Walter Schaffer, deserve special thanks for their willingness to answer questions, identify and locate materials, and discuss issues. The opinions expressed in this paper are the author's and do not reflect those of either the author's affiliation--Vanderbilt (University) Institute for Public Policy Studies, the Committee on Biomedical and Behavioral Research Personnel, or the National Research Council.

support activities, incorporating different selection procedures and educational strategies (e.g., "one-to-one" individually negotiated, student/mentor apprenticeships versus more formally structured degree programs within an institution). Consequently, previous evaluations in the NRSA program differ in terms of the specific program of interest, the target populations examined, the training activities involved, and the outcomes studied.

It also must be remembered that the NRSA program is but one benefactor of research training. Both the National Institutes of Health (NIH) and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) sponsor other programs with quite similar aims (i.e., increasing the supply of productive researchers in health-related areas). These programs either directly or indirectly sponsor research training and/or research career development (e.g., Research Career Development Awards, Clinical Investigator Awards, the Minority Biomedical Research Support program, and individual investigator RO-1 grants). In addition, other federal agencies and nonfederal organizations support research training in biomedical and behavioral research at some level (e.g., the National Science Foundation's Graduate Fellowship Program and the Robert Wood Johnson Clinical Scholars Program). Thus, the goals, functioning, outcomes and effects of NRSA programs must be viewed within the larger context of research training occurring in university departments, medical schools, faculty laboratories, and independent research centers.

## WHAT TYPES OF EVALUATION QUESTIONS ARE OF CURRENT INTEREST?

In reviewing the quality of NRSA evaluation efforts, a major issue concerns the extent to which previous evaluations have addressed questions posed by major constituencies. Given that evaluations are intended to provide useful results, studies should speak to the key concerns expressed by the various stakeholders involved with the program.

Four major constituencies for the NRSA program can be identified. They include (1) Congress, which is responsible for authorizing the program and appropriating funds; (2) NIH and ADAMHA, along with their individual institutes, which are in charge of administering the programs; (3) the individual fellows, trainees, and faculty involved in NRSA-supported training activities, and (4) other audiences with vested interests in training researchers (e.g., professional associations, scientific societies, and national "blue ribbon" committees concerned with research and science policy). In order to identify the issues of primary interest to these constituencies, relevant legislation and evaluation reports were reviewed. Individual interviews with congressional staff, federal agency personnel responsible for NRSA policies and evaluation activities, and individuals in charge of specific NRSA programs ($\underline{N}$ = 16) also yielded insight into the questions for which evaluative data are sought.[2]

It should be noted that neither is interest in these questions always generated independently by each constituency nor is each constituency equal in terms of the urgency with which its demands are accommodated. For example, questions dictated by reauthorizing legislation and formal requests from congressional oversight committees to federal agencies mandate a response; Congress indeed is the holder of the purse strings, and given limited time and resources, its requests often rank higher on the list of agency priorities for evaluation. In addition, the evaluation questions of most interest to a particular group often depend on the extent of its nature with the NRSA program. For example, agency staff whose major responsibility lies in administering institutional training grants may be most enthusiastic about collecting data that could improve their ability to monitor and guide programs; in contrast, scientific societies' demands may stem primarily from their desire to develop stronger arguments for increased NRSA funding in their respective discipline(s).

The major evaluation questions that have been and/or currently are of interest to key NRSA constituencies can be categorized into seven generic types. These include questions about[3]

- o the demand for the NRSA program (e.g., the adequacy of the current supply of for biomedical researchers);
- o levels of program participation, including numbers and characteristics of awardees;
- o characteristics of program operation and functioning, such as whether payback requirements affect the attractiveness of the NRSA program to qualified applicants;
- o program <u>outcomes</u> (e.g., the research career accomplishments of awardees) and/or program <u>effectiveness</u> (e.g., whether the subsequent success of awardees in obtaining federal grants is directly attributable to the program);

---

[2]A list of individuals interviewed is available from the author upon request.

[3]Because of time constraints, attention was focused on those constituencies most involved in setting priorities and administering policies for NRSA programs (i.e., Congress and federal agencies).

- outcomes and/or effectiveness of individual NRSA components (e.g., whether the Medical Scientist Training Program is more successful in training physician/investigators than extramural postdoctoral traineeships, intramural fellowships, physician/scientist awards, and/or a combination of training support mechanisms);

- cost-effectiveness of the NRSA program; and

- the development and maturation of scientific careers in general and the role of research training in this process (e.g., the components and determinants of scientific productivity).

Appendix A provides detailed examples of the questions that emerged during interviews with congressional staff, federal agency personnel, and others involved in research training activities and policy.

## DEMAND FOR THE PROGRAM AND LEVELS OF PARTICIPATION

Having definitive data on the need for research training support and levels of participation in NRSA programs are "bottom line" demands of all major stakeholders. For example, both the authorizing and reauthorizing legislation for the NRSA (e.g., P.L. 93-348 and P.L. 100-607) specify that awards are to be made only in areas/fields that have demonstrated a need for researchers. As such, Congressional appropriation committees traditionally have sought to base their fiscal decisions on this information, and actual "numerical recommendations" that indicate the number of training slots necessary to address shortages of researchers in specific areas have been frequently requested.[4]

Agency staff share this predilection for reasonably precise estimates of researchers needed in specific fields, disease categories, or problem areas. Other groups, including both those who lobby Congress regularly for NRSA funding in individual disciplines and those concerned with the overall health of the scientific enterprise, also clamor for better projections of supply and demand. Occasionally, these stakeholders have even launched their own data collection efforts in an attempt to obtain this information (e.g., Barries, 1986; Porter, 1979).

[4]In addition to specifying the number of training slots needed in a field or research area, there was a consensus among congressional staff that better explanations about the ways in which recommendations were derived (e.g., the assumptions underlying supply and demand models) were needed.

Related to supply and demand issues are questions about the "niche" occupied by NRSA programs in the overall landscape of research training support. All constituencies interviewed want to know the types of sponsors, the levels of their investment, and major priority areas for funding. Congress, in particular, wants such descriptive information so as to ascertain what the appropriate role of the federal government should be in the research training enterprise.

A third question of perennial interest, frequently arising at congressional hearings, centers around the distribution of NRSA programs and funds. All constituencies want an accounting of awarded fellowships and traineeships, the research fields supported (e.g., nursing or primary care research), and changes over time. Such data are perceived as crucial to determining whether NRSA expenditures are targeted at "shortage" areas, to ascertaining whether agencies have responded to specific congressional directives, and to identifying where changes in NRSA program priorities or policies may be warranted. Also viewed as important is information on the characteristics of awardees, typically in terms of their sex, race/ethnicity, and institutional affiliation. Many of these questions have been spurred by disappointment in the low rates of participation by women and ethnic minorities in science, coupled with a concern that the nation's pool of scientists and engineers may prove inadequate to meet future challenges (e.g. Office of Technology Assessment, 1985; Vetter, 1989).

## PROGRAM CHARACTERISTICS AND OPERATION

Of primary interest to federal agency staff who administer NRSA programs and policies are questions related to program functioning. These questions are quite diverse in their scope and content. They include requests for information on how institutional review groups (IRGs) make decisions about training grant awards, the amount and types of research training received by predoctoral and postdoctoral trainees, and whether faculty mentors indeed have active research programs in areas most relevant to an institute's goals and objectives.

Program officers, however, are not the only source of these questions. The legislation for the NRSA program itself (e.g., P.L. 100-607) speaks to the general need for program monitoring so as to "determine what modifications in the [NRSA] programs are required to meet the needs [for research personnel]." More explicitly framed "operational" questions also have been posed by Congress, including how the payback requirement and current stipends for NRSA awards affect participation in the program.

**PROGRAM OUTCOMES AND PROGRAM EFFECTIVENESS**

Questions related to both absolute outcomes (the accomplishments of NRSA trainees and fellows) and comparative outcomes (e.g., the performance of NRSA-funded predoctoral students in the life sciences as contrasted with those supported by the National Science Foundation) are specified clearly in the legislation. All authorizing and reauthorizing language states that National Academy of Sciences shall "identify the kinds of research positions available to and held by individuals completing [NRSA and other current training programs]" (e.g., P.L. 100-607, Part F, Section 489). Another example appears in the Health Research Extension Act of 1985 (P.L. 99-1258), which requested data on the "number of persons who receive NRSA awards and who engage in health research or training as a career."

On the other hand, questions related to program effectiveness (i.e., whether outcomes are directly attributable to NRSA-supported training) are much less frequent and clearly enunciated. A broad and relatively vague mandate for effectiveness data appears in the law (P.L. 100-607); NAS is directed to "assess [current NRSA programs] and other current training programs available for . . . such personnel." Aside from this fairly global injunction, however, being able to confidently link NRSA training with specific achievements ranks lower on Congress's list of evaluation priorities.

Agency staff members also express less interest in effectiveness questions, particularly if the attention paid to them is at the expense of other data collection efforts. What does generate enthusiasm among this group, is obtaining better information on program outcomes--both in absolute and comparative terms. Training officers at the various institutes want to know what happens to their awardees--for example, whether they remain active in research, whether their research is in the area of their NRSA training, and whether they have been instrumental in training other researchers. If these outcomes ultimately can be contrasted with the performance of individuals who received other types of research training that is supported by either their own agencies or by other sponsors, this would be an additional asset.

It is likely that the lower priority assigned to addressing effectiveness issues stems from an array of factors. For example, there is an awareness of the enormous difficulty and cost involved in obtaining unequivocal data on the effects of research training, given the current structure of these programs (e.g., the heterogeneity of training experiences and the lack of uniformly applied selection criteria). Related to this is the strong sentiment, based on the substantial erosion in NRSA training monies over the last decade, that the first priority for spending any additional funds appropriated by Congress must be to increase the number of training slots rather than to initiate rigorous impact evaluations. Further, in many programs staff members maintain that the necessary data for answering more basic questions about program demand and operation are not available and that this situation must be corrected before such "second-order" questions as program effectiveness are considered.

When questions as to the effectiveness of NRSA programs do surface, they typically center around issues of relative effectiveness. For example, data that can "tease out" the effects of NRSA programs in producing biomedical researchers relative to the performance of other research training programs with similar goals are deemed more salient than evaluations aimed at understanding whether NRSA training is more effective than no research training at all or research training that is entirely financed by the individual through loans or other personal sources.

## OUTCOMES AND/OR EFFECTIVENESS OF INDIVIDUAL PROGRAM COMPONENTS

As previously mentioned, the NRSA comprises a heterogeneous group of programs, many of which also have distinct program components. These include different funding mechanisms, different target populations, and different training philosophies and strategies. Outcomes associated with these individual components and their relative effectiveness have comprised the focus of congressional and administrative inquiries. For example, the Health Research Extension Act of 1985 requested a study on "the effectiveness of [the training grant] mechanism in encouraging individuals to engage in health research and training as a career." Of constant concern to agency staff is "what works best" among or within NRSA components. Illustrative of this interest are such questions as "Are M.D./Ph.D. programs or postdoctoral traineeships more efficacious in producing physician/investigators?" and "Is predoctoral training that is grounded in a particular discipline more successful for increasing the number of researchers attacking alcohol-related health problems than predoctoral training that incorporates several disciplinary perspectives and methodologies?"

### Cost-Effectiveness

For the most part, cost-effectiveness questions do not constitute a high priority among major constituencies. The few questions that emerged in the interviews pertained to identifying ways to "best use the training buck," particularly if research training funds continue to erode. Somewhat related to this concern are more global questions associated with the personal, disciplinary, and social costs incurred from having an insufficient amount of research monies available to support the number of high-quality applications for individual investigator awards from researchers who have been trained in NRSA programs.

### Development and Maturation of Scientific Careers

More frequently, the questions of interest to key NRSA constituencies are those that address research training, scientific productivity, and scientific career development in general rather than with regard to NRSA programs in particular. These questions span a variety of topics, including the relationship between research training and the quality of research, the factors governing an individual's choice to pursue a scientific career, and the resources required to most successfully maintain a productive research career. Also included in this category are those questions posed by agency staff about how to best measure relevant outcomes of research training (e.g., "active involvement in research" and "quality of research"). Although such questions are important for guiding and improving future evaluations of the NRSA program and can indeed be addressed by well-designed studies, it must be kept in mind that providing answers is neither the sole responsibility of NRSA nor the evaluation efforts connected with this program.

### Where Improvements in Evaluation Activities Are Needed?

In the previous section the major evaluation questions of current interest to key constituencies were identified. Although these questions covered all aspects of the program, the priority areas centered around those issues associated with demand for the program, levels of participation, characteristics of training and recipients, and program outcomes.

To date, past evaluation efforts have provided a wealth of data about National Research Service Awards, but many questions basic to understanding how these programs operate and what happens to awardees remain. Most individuals in charge of NRSA programs often continue to find themselves operating in almost a vacuum with regard to having sound, empirical data about how awardees are selected, the characteristics of participants, and the training environments and activities supported. Further, knowledge about the subsequent performance of awardees currently is confined to a limited set of indicators that vary considerably across individual NRSA programs and that incompletely characterize the intended outcomes. Improving this situation (i.e., "filling the gaps") is what must drive the individual items included in any portfolio of future evaluation activities.

Four major gaps exist in terms of having an adequate knowledge base about NRSA programs:

1. Basic questions about program participation and functioning, although of great interest to key constituencies, have remained inadequately addressed.

2. Our understanding of program outcomes, let alone program effects, is still limited.

3. Insufficient attention has been given to determining what works best across and within program components.

4. Evaluation efforts have been sparse in many fields and research problem areas.

The first three gaps focus on "points of slippage" between the types of evaluation questions currently of interest to major constituencies and those that have comprised the thrust of prior evaluation efforts. An examination of the generic questions addressed by the 16 evaluation studies/reports reviewed indicated that 56 percent (N = 9) addressed program participation issues, and 50 percent (N = 8) collected data on program characteristics and operation. The overwhelming majority (94 percent, N = 15) presented information on one or more outcomes for programs or program elements, and 38 percent (N = 6) attempted to address in some way the effectiveness of NRSA programs or distinct components. None of the studies reviewed dealt with issues of cost-effectiveness.[5]

At first glance these percentages suggest that many of the questions of interest to NRSA stakeholders (e.g., program outcomes) indeed have been addressed. However, it must be remembered that within each of these generic evaluation issues lie a variety of subquestions. For example, questions about program operation encompass the nature of the trainee selection process, characteristics of training, and the relationship of the payback requirement to participation levels. As shall be seen, the match between constituency priorities and evaluation efforts

---

[5]Given that studies could focus on more than one type of evaluation question, these percentages do not sum to 100 percent. Information on how the evaluations were chosen for review and on the broad categories of questions addressed by each is presented in Appendixes B and C, respectively.

In classifying these evaluations, distinguishing between "outcome" studies and "effectiveness" studies has been in many cases a matter of judgment. For the purposes of this paper, an effectiveness study is one that incorporated either methodological (e.g., matching) and/or statistical procedures so as to control for selectivity bias. Those studies that attempted to compare outcomes but did not include any real consideration of selectivity bias were designated, rightly or wrongly, outcome studies.

on these more specific questions is where the discrepancies surface. (See Appendix A for a detailed enumeration of the questions posed by constituencies and Appendix D for a listing of those addressed by evaluation activities.)

## GAPS IN UNDERSTANDING PROGRAM PARTICIPATION AND OPERATION

Although past evaluation efforts have addressed aspects of program participation and operation, several issues have escaped careful examination, including some that are basic to understanding any discrete program or intervention (e.g., program implementation). This situation is partly an outgrowth of the limited amount of resources that have been allotted for evaluation activities. Consequently, some programs (e.g., those sponsored by ADAMHA) have received little scrutiny. Another problem has concerned the fact that when evaluations were initiated, the short timelines imposed often dictated that the focus be on collecting outcome data (no mean feat by itself), with only secondary attention given to examining participation levels or program characteristics. As a result, summary profiles describing NRSA applicants, awardees, and program activities are either nonexistent, sketchy, or idiosyncratic in terms of the populations covered, the variables of interest, and the time periods examined.

### Needed Information on the Demand for the Program

Although development of better supply and demand indicators for biomedical and behavioral science research personnel is covered more thoroughly in the full committee report, one related component deserves special mention in this paper. This concerns the extent of our knowledge about the research training enterprise as a whole (e.g., the total amount of funds, training opportunities, and types of training provided by all sponsors). Congress' motivation for having such information stems from its desire to ascertain what its role should be in financing research training and then to apply this understanding when making decisions about the NRSA program. Similarly, the interest of NIH and ADAMHA staff arises from their wish to better understand their own agency's total involvement in research training, particularly by mechanisms other than those covered under the NRSA umbrella (e.g., research assistantships paid by grants to individual investigators).

Answers to these questions also are requested in the charge for evaluation specified in the authorizing and reauthorizing legislation for the NRSA program: to "assess current [NRSA programs] and other current training programs available for the training of such personnel" (P.L. 100-607, Part F, Section 489). To carry out this charge, a map of the geography and topography of non-NRSA funding sources and mechanisms for research training must be constructed.

Previous committee reports (e.g., National Research Council, 1977, 1981) have attempted to survey this terrain, but this is no easy task. Currently, the best sources of data are the annual Survey of Earned Doctorates (SED) conducted by the National Research Council and the National Science Foundation's (NSF) Survey of Graduate Science Students and Postdoctorals. However, each has certain limitations. For example, the NSF survey requires institutions to indicate only one source of support for a graduate student. Although respondents to the SED are instructed to identify all sources of support and estimate the percentage of support received from each source, their ability to reconstruct these data accurately is unclear.

Understanding the variety of research training activities sponsored by NIH and ADAMHA via non-NRSA mechanisms represents one step toward mapping the terrain, however. Of particular interest here is predoctoral and postdoctoral research training paid by research grants to individual investigators. Available data suggest that the use of this mechanism in supporting research training is not infrequent; research assistantships paid by federal and other grants were a source of predoctoral support for 16 percent of the 1987 Ph.D. recipients in the life sciences--an almost equal percentage to that reported for NIH traineeships (Coyle and Thurgood, 1989).

Developing this capacity to obtain detailed training support data for all Public Health Service agencies may be more feasible than one might expect, given earlier and more recent efforts by NIH and ADAMHA. Information on all paid and unpaid personnel working on research grants was collected on a sample basis for all PHS grants in 1963 and NIH grants in 1969; beginning in 1973, these data were again requested of all NIH grantees in the NIH Research Grants Manpower Survey. Unfortunately, this effort was abolished in 1980 despite a reasonably favorable evaluation (Williams, 1979).

ADAMHA staff members currently are investigating the feasibility of implementing a similar system for their own research grants and have already developed the system specifications, along with conducting some preliminary pilot tests (see Tjioe, 1989, for a description of this system). Variables in this system include name, social security number, role on the grant, type of position, sex, highest degree(s), year of degree, birth date, field (e.g., surgery), and research discipline (e.g., brain damage) for all personnel connected with awards made by ADAMHA institutes. If resources were available to establish and maintain this data base for all PHS awards, questions relating to the various ways in which research training is supported by these agencies, along with differences across fields and levels of training, could be addressed.

Moreover, the ability to address additional questions concerning research training and personnel, including detecting shifts over time, would be enhanced. For example, postdoctoral research associates supported on faculty research grants (R0-1s) could be identified by the data base at the start of their tenure; these individuals (or a sample) then could be surveyed about their training (e.g., the extent of their involvement in relevant activities and the nature of their relationships with advisors and other faculty) so as to determine whether and how these experiences may differ from those received by NRSA-funded individuals. If it is true that students are supported by various types of mechanisms, a more complete picture of each trainee's or fellow's total PHS-supported research training would be obtained. This would work towards acquiring a better sense of an individual's training support history and help improve efforts to elucidate the relationship of outcomes to types and length of research training.

**Needed Information on Participation Levels and Participant Characteristics**

Data relevant to questions asking for the numbers of awardees or positions are readily available. They are published annually in data books or other reports issued by NIH (1987, 1988) and ADAMHA (1989a), along with their individual institutes. They also appear in the majority of evaluation studies reviewed, along with each of the committee's previous reports (e.g., National Research Council, 1983, 1985). Depending on the specific report, information on the number of awardees may be disaggregated by type of training mechanism (individual fellowship versus traineeship), level of training (e.g., predoctoral versus postdoctoral), type of training (e.g., M.D. or Ph.D.), major field grouping, sponsoring institute (e.g., National Cancer Institute), or some combination of these variables.

Detailed profiles depicting even basic demographic characteristics of participants (e.g., sex, race/ethnicity, and educational background), however, are less frequently found in formal evaluation reports, although there are some exceptions. A few evaluation studies (i.e., Coggeshall and Brown, 1984; Garrison and Brown, 1986; National Institute of Dental Research Ad Hoc Consultant Panel, 1988; Velletri et al., 1985) did report descriptive data on awardees' educational backgrounds. Information on sex of awardees was included in Garrison and Brown's (1986) study of NIH postdoctoral appointments. Data on race/ethnicity typically have appeared only in internal program reports or evaluations of the Minority Access to Research Careers

(MARC) programs Garrison and Brown 1985; Primos, 1989a, b; Sherman, 1983b; Task Force on Minority Research Training, 1986).

Developing detailed profiles or time series of participation levels by one or more of these major indices from these documents also is not easy. In part, this is because there exists no agreed-upon format for how to best report these data. It is not uncommon for individual evaluation studies and agency publications to use slightly different lexicons for classifying research fields and specialties, to apply different schemes for aggregating individual disciplines into major field clusters, or to employ different counting strategies (e.g., number of individuals versus full-time equivalents). For example, in some instances MARC awards have been included under predoctoral awards, and in other documents they have been demarcated; sometimes awards made by the Fogarty Center have been included in the total counts (National Research Council, 1979), and sometimes they have not (National Institutes of Health, 1987). Similarly, in 1981 ADAMHA began using a different system for classifying the major fields of their trainees and fellows. Such discrepancies often may be difficult to detect, even for the most savvy user, because of the variability across studies in the use of detailed footnotes.

A more central problem lies in the fact that the accuracy of the information for some demographic variables in the major research training data bases is questionable. Look (1989) identified problems in the IMPAC data file maintained by NIH (the master data base from which the Trainee-Fellow File [TFF] is constructed); these included nonreporting and incorrect reporting of gender and race/ethnicity, along with inconsistent reporting of data on discipline, field, and specialty. It is the case that demographic data are available from other sources, which might then be used to augment gaps in existing agency files. For example, for some programs, particularly small programs such as those administered by the Minority Resources Branch at the National Institute of Mental Health (NIMH), demographic data on trainees and fellows funded by the MARC program are collected and maintained by program personnel (e.g., Primos, 1989a). Information on all Ph.D.s awarded by U.S. universities, which includes those who received NRSA support, also is collected by the SED. However, gaining access to and merging these files with PHS data require resources, and monies typically become available only when large-scale evaluation studies, which focus on outcomes, are commissioned. In addition, information on non-Ph.D. populations (e.g., M.D.s) has proved more difficult to obtain (see Carter, Robyn, and Singer, 1983, and Martin, 1986, for examples of problems with specific data bases).

In general, there remains a paucity of complete and accurate information available on the characteristics of NRSA participants. Although special analyses can be performed and some descriptive data can be "patched together" from previous evaluation studies, many audiences do not know answers to such simple questions as "What have been the trends in NRSA awards in the clinical sciences?" and "What percentage of women and ethnic minorities have received NRSA support?"

This same situation is even more characteristic of applicants to NRSA programs, although the issue is somewhat more complicated. "Applicants" include individuals who apply for NRSA fellowships and institutions that apply for NRSA training grants. Some information is available on the numbers and characteristics of individual applicants for fellowships, and although fellowships do not constitute the bulk of training slots funded, their numbers are not insubstantial (147 fellowships for NIMH in 1987 and 1,664 for NIH in 1986). However, individual applicants who might apply for training grant slots cannot be identified. These persons are selected by training grant directors/committees at individual institutions, and information is seldom reported on unsuccessful candidates, assuming that at least in some cases individuals do apply and some type of selection process occurs (most likely for postdoctoral traineeships). The selection processes inherent in the training grant mechanism thus make a comprehensive profile of all applicants unworkable without primary data collection efforts such as those initiated by Bickel et al. (1981) for the MST program. At the same time, determining whether the award selection process is working as intended and ultimately the effects of the training grant mechanism require that attention be paid to gathering applicant data. In addition, the characteristics of the institutions that apply for training grants are worthy of examination in terms of understanding the geographical distribution of applicants and awards, the factors correlated with awarding funds to a grant application, and so forth.

Increased attention to collecting data on participation issues would contribute to laying a firmer groundwork for understanding not only the demand for the program but also issues associated with how individual or program characteristics may be related to certain successful outcomes. For example, relationships have been found between sex and grant activity for postdoctoral fellows (e.g., Garrison and Brown, 1986) and NSF predoctoral fellows (Snyder, 1988) and between certain characteristics of training programs and subsequent grant application efforts of trainees (C. Roth, personal communication, June 1989).

**Needed Information on Program Characteristics and Activities**

There exists a plethora of questions about the characteristics of NRSA programs and their functioning (see Appendix A). These questions not only include ones raised by Congress, which typically focus on program procedures and regulations (e.g., the payback requirement), but also those generated by agency staff (e.g., the nature of the trainee selection process and the types of training activities carried out in funded programs).

About half (56 percent) of the 16 evaluation studies reviewed devoted some attention to program "operational" issues (see Appendix C). For example, descriptive statistics on the duration of NRSA support were contained in evaluations of NIH predoctoral appointments (Coggeshall and Brown, 1984; National Research Council, 1976; Velletri et al., 1985), NIH postdoctoral appointments (Garrison and Brown, 1986), and the National Institute of Drug Abuse and NIDR trainees and fellows (Clouet, 1986; NIDR Ad Hoc Consultant Panel, 1988). Related to this is the need for data describing the extent to which individual awardees receive multiple NRSA awards (e.g., predoctoral and postdoctoral). This aspect has been addressed in a few studies (e.g., ADAMHA, 1986, unpublished report; Coggeshall and Brown, 1984; Garrison and Brown, 1986). Such data are important in order to obtain a good sense of the "dosage" of training and how it might subsequently relate to measures of outcomes and effectiveness.

Other aspects of program operation also have received some scrutiny. In response to a specific congressional request, the influence of the payback requirement on the number and quality of applicants and the number of awardees who pursued health research or training careers was examined by NIH (1986). An in-depth exploration of NIH program "processes" was undertaken by the National Research Council (1978), with site visits to institutions with training grants conducted to obtain a better sense of how training monies were used and to suggest ways in which training policies might be improved. Garrison and Brown (1985), in their evaluation of MARC undergraduate training grants, also obtained qualitative information gleaned from site visits on such operational issues as the activities on which MARC funds were spent, departmental composition, and recruitment practices.

At the same time, a profile of how programs function in terms of recruitment, selection, and actual training activities--issues currently of interest to major constituencies--is not available for the majority of NRSA programs. The lack of these data is disturbing, not only because program modifications may then fall prey to being guided more by personal judgment and experience than by empirical data but also because the success of other evaluation efforts (i.e., outcome and effectiveness studies) hinges on understanding how participants are selected and the distinct types of training (if any) they receive.

Take, for example, the question about how trainees are assigned to training grants, particularly in terms of Ph.D.

predoctoral training and M.D. postdoctoral training. It has been speculated that NRSA support is simply viewed by departments as another "pot of money" that can be channeled to students who currently are not receiving other types of financial aid. If this practice is common, it may mean that traineeships, rather than being highly competitive, are reserved for those individuals judged as less qualified to compete for other sources of support (e.g., prestigious university fellowships or postdoctoral appointments). The situation is further compounded if NRSA stipends are lower than those offered by most other sponsors or if payback obligations are viewed as burdensome by the individuals who would be most qualified to receive NRSA support.

## GAPS IN UNDERSTANDING PROGRAM OUTCOMES AND EFFECTS

Questions concerned with program outcomes clearly are specified in the legislative authority for the NRSA program, along with being of significant interest to agency staff in charge of these training programs (see Appendix A). Here the focus is on knowing what happens to awardees (e.g., "Are they engaged in health research careers?"). As previously discussed, questions concerning program effectiveness also are implied in the legislative authority for the NRSA program and, although not the highest priority, do generate some enthusiasm among some agency personnel.

The questions on effectiveness generating the most interest, however, are not those of the breed "Are NRSA-supported predoctoral fellows more likely to be successful health researchers than those who receive no predoctoral training and/or support?" Rather, they address the issue of relative effectiveness. For example, congressional policymakers want to know if NRSA programs are more effective than research training programs administered by NSF. Agency staff want answers to such questions as "Is predoctoral or postdoctoral training more effective in producing researchers in the clinical sciences?"

Given the strong interest in outcomes, examining program achievements comprised a major emphasis in the overwhelming majority of evaluation studies reviewed (see Appendix C). For the most part, the unit of analysis for these studies was the individual awardee. The role of NRSA training support and the consequences of losing this support on departments were, however, explored in two of these efforts (National Research Council, 1978, 1981).

The bulk of studies focused on those outcomes that to varying degrees reflect involvement in research, given the legislative authority for the NRSA. There was some heterogeneity in terms of the number of outcomes examined, with a few studies

(ADAMHA, 1986, unpublished report; Schneider, 1980) providing data on essentially only one outcome (success in obtaining NIH/ADAMHA funding or type of employer). Across the various reports, there also was considerable variability in which outcomes were examined, depending on the training sponsor (e.g., NIH or ADAMHA), degree and field of training (e.g., Ph.D. or M.D. and biomedical or behavioral sciences), levels of training (e.g., predoctoral versus postdoctoral), and time periods examined.

For any given outcome, however, there was considerably less variation in terms of how it was measured. Awardee outcomes typically were operationalized as attainment of doctorate and pursuit of postdoctoral research training (for predoctoral award recipients); type of employment, usually academic employment; time spent in research; pursuit of and success in obtaining external research grants, particularly grants awarded by NIH and ADAMHA; and publication performance (numbers of publications and citations).

One primary reason for this is that only 4 of the 12 studies involved any primary data collection on awardees; Clouet (1986), the National Research Council (1977), and Sherman et al. (1981) all collected at least some data directly from awardees, and Bickel et al. (1981) surveyed medical school deans about students in their programs. Instead, the typical study has relied on archival data: data on demographic characteristics, educational history, and employment plans of individuals who have just earned their Ph.D. (the Doctorate Records File compiled from the SED); data containing information on all individuals who have applied and/or been awarded grants from PHS agencies (the Consolidated Grants Application File, along with similar data, where available, from such other funding sponsors as NSF); data from a biennial sample survey conducted by NSF on the employment activities of Ph.D.-holders in science and engineering fields (the Survey of Doctorate Recipients [SDR]); and employment data from reports submitted by awardees after completion of their NRSA appointments in order to fulfill payback requirements. This reliance on archival data has at least partly resulted from the constraints imposed by limited funding for evaluation, short timelines for reporting, and OMB regulations for data collection efforts contracted by federal agencies.

In addition to gathering information on outcomes for NRSA recipients, four studies (Coggeshall and Brown, 1984; Garrison and Brown, 1986; NIH, 1986; National Research Council, 1976) did address program effectiveness at some level. These studies were all efforts sponsored by the committee, and for the most part, the major focus was on evaluating NIH programs. [6]

Given that NRSA supports both predoctoral and postdoctoral training for M.D.s and Ph.D.s and that the training strategies for each of these groups are reasonably distinctive, evaluation activities will be discussed separately.

**Predoctoral Training for Ph.D.s.**

Three major studies examined outcomes associated with NRSA-sponsored predoctoral training (Coggeshall and Brown, 1984; National Research Council, 1976, 1977). In general, the results indicated that NIH awardees distinctly outperformed their comparisons in terms of greater involvement in research (e.g., receipt of additional postdoctoral research training, time spent in research, and grant application/award activity). These individuals also had somewhat better track records in carrying out high quality research (as measured by citations). Similar to the results of previous studies on the determinants of academic careers (Long, et al., 1979; McGinnis and Long, 1988), awardees did not experience any greater success in locating academic employment, once prestige of doctoral institution had been controlled. For each of these findings, however, the causal linkages between NRSA-funded training and these outcomes remain unclear.

For example, in its survey of 1971-1975 Ph.D. recipients in the biomedical and behavioral sciences, along with nurses who had earned their doctorates during the same time period, the National Research Council (1977) found distinct differences between those individuals who had received NRSA predoctoral support from ADAMHA/NIH/HRA (Group 1) and those who did not receive this support (Group 2). The size of these discrepancies, however, varied across the three broad fields. Looking at individuals'

---

[6]None of these studies had designs that could confidently support causal attributions, however. The majority did use multiple comparisons that embodied differing levels of selectivity, but in some cases, comparisons (e.g., predoctoral awardees versus those with no Ph.D. training) were so hopelessly confounded as to be meaningless. Those that paid some attention to issues of selectivity are the ones that are considered in this paper.

Even with this more narrow focus, summarizing the results of this smaller set of studies is difficult, given the diversity of programs examined, the variability in career patterns and research activities in different fields and specialties, and the influence that differing time periods of training may have on outcomes (e.g., the effect of labor market expansion and contraction on the availability of academic positions).

reported success in obtaining PHS support for their current research, the proportions of NRSA awardees versus those who received no NRSA support were 58.5 percent versus 43.4 percent in the biomedical sciences, 29.5 percent versus 16.0 percent in the behavioral sciences, and 38.1 percent versus 18.4 percent for nurses. NRSA awardees in all three fields also were slightly more likely to report greater time spent in research, although the differences were quite small; the average time reported was 59.8 percent for Group 1 members versus 52.0 percent for Group 2 members in the biomedical sciences, with corresponding percentages of 28.4 percent versus 22.6 percent for behavioral scientists and 15.5 percent versus 11.6 percent for nurse-researchers. Awardees in the biomedical sciences also were much more likely to have spent the first year after their doctorate in postdoctoral study (65.2 percent versus 47.9 percent) in contrast to behavioral science Ph.D.s (15.6 percent versus 10.4 percent). Comparisons of the percentages employed in academic environments for each group yielded little or no differences.

Better performance in research as a function of NRSA predoctoral support may, however, be simply a product of preexisting differences between the groups and differential training experiences. The groups examined in this study were those with NRSA support and those without NRSA support. This latter group is quite heterogeneous, comprised of individuals who received other types of federal support, university support, or no financial support for their graduate training. Further, the fact that a variety of predoctoral training experiences characterized this group, some that may or may not be similar to NRSA-supported training activities, makes interpretation of both the differences and the lack of differences between these two groups impossible.

Significantly better information is available from two studies that used comparison groups designed to help control for differences associated with heterogeneity of training experiences and, to a lesser degree, selectivity. The National Research Council (1976) compared the types of employment (e.g., academic and business) and time spent in various activities from 1968 to 1970 for three groups of Ph.D.s: (1) awardees of NIH predoctoral support; (2) those who received other non-NIH predoctoral support that could be identified (e.g., awards from NSF); and (3) those who received neither predoctoral support from NIH nor the other agencies covered in the study. In addition, the attempt was made to include only those individuals who had <u>not</u> engaged in postdoctoral study so as to reduce the possible influence of additional formal research training. Similar results were found. Looking at Groups (1) and (2), although the percentage employed in academic environments was almost identical (71.4 percent versus 71.6 percent, respectively), NIH awardees spent more time in research than those Ph.D.s supported by other sources (an average of 53 percent versus 41 percent).

A more recent study (Coggeshall and Brown, 1984) of NIH predoctoral awards also attempted to at least partially control for the heterogeneity of training experiences and selectivity. Looking at those individuals who received their Ph.D.s in the biomedical sciences between 1967 and 1981, three study groups were compared: (1) those who received at least 9 months of NIH predoctoral support; (2) those who earned their degree from the same departments as the first group but who received 0-8 months of NIH support; and (3) those who graduated from departments that did not have NIH training funds. This strategy permitted two important considerations: (1) those departments receiving NIH funds, often the top-ranked departments in the biomedical sciences, apply the same criteria to accept students, and thus those students enrolled in the doctoral program, regardless of their source of predoctoral support, may be more similar in terms of individual differences (e.g., abilities); and (2) that students who are in departments with NIH funded programs but who are not supported by these funds for an extended length of time may benefit from certain resources accruing to NIH-supported departments.

The most instructive comparison in this study for questions pertaining to specific outcomes associated with NRSA support is between those groups from the same set of departments. The findings of previous studies were confirmed. Although there emerged no differences in terms of subsequent academic employment between those with or without NIH support, distinctions did appear in variables related to research performance. For example, the percentage of NIH awardees who subsequently received NIH postdoctoral support was 34.4 percent as compared to only 20.7 percent of their departmental counterparts. Recipients of NIH training funds also were 32 percent more likely to apply for NIH research grants (30.5 percent versus 23.1 percent) and, if they applied, 13 percent as likely to be awarded them (62.3 percent versus 55.0 percent). In general, there was a stronger tendency for NIH awardees to have published at least one article (e.g., 86.3 percent of NIH-supported individuals versus 63.2 percent of those not receiving support for FY 1977 Ph.D.s) and to have more citations per article (e.g., 8.2 versus 6.0 for this same cohort). By examining data on length of NIH predoctoral support, Coggeshall and Brown provided additional insight into NRSA awards, indicating a clear relationship between length of NIH support and performance on several major outcomes.

While such results do suggest that NIH predoctoral support, at least in the biomedical sciences, increases the probability that an individual will have a research career in health-related areas, it nonetheless remains a small role. Analyses regressing years since Ph.D., the quality of the predoctoral institution, and total months of NIH predoctoral support on number of NIH research grant applications, average priority score awarded to

NIH grant applications, total number of articles published, and average number of citations per article published yielded $R^2$s of .08, .06, .03, and .07, respectively (Coggeshall and Brown, 1984). Thus, it is apparent that other factors (e.g., individual abilities, the nature of research in certain fields, subsequent postdoctoral research training, and available resources for research at the awardee's employment setting) remain plausible and key contributors to producing active researchers in these areas.

**Postdoctoral Research Training for Ph.D.s.**

Three major studies have focused on identifying the outcomes of NRSA-supported postdoctoral training, primarily those of biomedical scientists. In general, those with postdoctoral training, regardless of the sponsor, outperformed on all measures as compared to those who were supported for their predoctoral education but who did not choose to pursue additional postdoctoral study.

More recent examinations of NIH postdoctoral training in the biomedical sciences have been carried out for 1964-1977 Ph.D. recipients (National Institutes of Health, 1986) and for 1961, 1966, 1971, and 1976 Ph.D. recipients in the biomedical sciences (Garrison and Brown, 1986). Here the major comparison groups were (1) NIH postdoctoral trainees and fellows, (2) Ph.D.s who had likely received postdoctoral training from other sponsors, and (3) those who reported no plans for postdoctoral study at the time they received their degree. Substantial differences emerged between NIH postdoctoral awardees and those who indicated no plans for postdoctoral study; for example, Garrison and Brown (1986) found that NIH awardees were three times as likely as the "no plans" group to have applied for NIH/ADAMHA research grants (56.9 percent versus 19.6 percent) and four times as likely to have been awarded a grant (40.0 percent versus 9.2 percent). This latter difference was reduced somewhat when only those who applied for grants were considered (70.3 percent of NIH awardees versus 47.1 percent of "no plans" group). They also were more likely than those with no postdoctoral training to have obtained faculty appointments 8-9 years after their Ph.D. (66.7 percent versus 52.7 percent) and, depending on the specific cohort examined, to have published more articles and received more citations per article. A study by NIH (1986) revealed similar findings in terms of academic employment and research funding activity.

Given that individuals who choose to undergo the additional years of training involved in postdoctoral appointments may share certain characteristics that are distinct from those of individuals who do not engage in postdoctoral study, comparing outcomes of NIH postdoctoral awardees with those who had their postdoctoral training sponsored by other agencies is more appropriate for addressing questions related to the specific value of NRSA support. In the studies previously described, the advantages of NIH support remained, although the differences between the two groups were smaller. Comparing Ph.D.s with NIH postdoctoral appointments to those who had their postdoctoral study supported by another sponsor, the National Research Council (1976) found that the NIH study group spent substantially more time in research (an average of 61.2 percent) as compared to those with non-NIH appointments (40.0 percent) and published articles that were more frequently cited by their colleagues (e.g., 73.1 citations per person versus 62.0 for those investigators aged 31-40). In contrast, although both groups had high rates of academic employment, the non-NIH supported postdoctorates were somewhat more likely to be in universities and medical schools (90.5 percent versus 82.2 percent).

As Garrison and Brown (1986) found, NIH awardees continued to outperform in terms of grant application activity those individuals whose postdoctoral training was supported via another source (56.9 percent versus 34.5 percent); Also, they were more likely to have been awarded a grant (40.0 percent versus 22.3 percent); this disparity decreased substantially, however, when considering only those applying for such grants (70.3 percent versus 64.8 percent). There did appear to be some advantage in terms of academic employment; the percentage obtaining a faculty position was 66.7 percent for NIH awardees as compared to 56.7 percent for those with other types of postdoctoral training, but consonant with previous research (McGinnis, Allison, & Long, 1982), this relationship could be primarily accounted for by other factors (e.g., prestige of doctoral institution). Similar results were reported by NIH (1986).

Similar to the situation that exists for predoctoral training, however, multiple regressions on such outcome measures as grantsmanship that controlled for other factors (e.g., selectivity of baccalaureate institution and reputation of doctoral institution) yielded small multiple $R^2$s, ranging for the most part from .06 to .14 (Garrison and Brown, 1986). This reinforces the conclusion that several other factors contribute to fostering successful research career paths and achievements, although little is known about the exact nature and strength of the relationships.

**Postdoctoral Training for M.D.s.**

The role of postdoctoral training for M.D.s was examined by the three studies discussed in the preceding paragraphs. However, the difficulty in interpreting the results--resulting from problems encountered in drawing comparison groups resembling in both orientations and experiences M.D.s with NRSA-supported, postdoctoral research training--is exacerbated, given that the vast majority of physicians do not follow research careers. In addition, identifying reasonable comparison groups in these retrospective studies is further complicated by the fact that existing data bases for physicians typically are less complete that those for Ph.D. recipients.

Differences between M.D.s with postdoctoral appointments and those without postdoctoral training, some of which appear to be substantial, were found by the National Research Council (1976) for certain outcomes: employment in medical schools and universities (40.9 percent versus 7.4 percent, respectively); the average amount of time reported in conducting research (10.6 percent versus 2.6 percent); and numbers of publications and citations (e.g., 58.6 citations versus 10.3 citations per person for M.D.s aged 41-50). By the use of additional comparison groups, a strong relationship between the existence and length of formal research training and outcomes also appeared--a relationship that has been supported by analyses of more recent trainees (Levey et al., 1988; Sherman, 1983a, 1983b, 1989). In addition to the M.D. groups specified above, two other groups were identified: individuals who had earned both an M.D. and a Ph.D. and who had or had not received postdoctoral training. With the exception of average time spent in research, the results showed a ranking among these groups in line with the amount of research training received. For example, the proportions employed in academic settings were 67.5 percent for M.D./Ph.D.s with postdoctoral appointments, 60.4 percent for M.D./Ph.D.s who did not pursue postdoctoral study, 40.9 percent for M.D.s who had NIH-supported postdoctoral appointments, and 7.4 percent for M.D.s with neither a Ph.D. nor postdoctoral training. On each of the four measures used in the study, the performance of M.D./Ph.D.s, regardless of whether they had been engaged in postdoctoral study, was higher than for those M.D.s who did not possess a Ph.D.

The two remaining studies tried to draw comparison groups that addressed in some way selectivity issues. Rather than looking only at <u>all</u> M.D.s without postdoctoral training, Garrison and Brown (1986) also identified another group of M.D.s who received their degree in 1965 or 1974, who reported their primary activities to be "research" or "teaching," but who had not received postdoctoral research training. Looking at 1974 M.D.s only, there were differences between this group and NIH postdoctoral trainees and fellows. For example, those M.D.s with NIH-supported postdoctoral training also were slightly more apt to have applied for NIH/ADAMHA research grants (18.6 percent versus 12.0 percent) and subsequently been awarded funding (8.7 percent versus 5.5 percent).

A comparison of these outcomes between M.D.s who had NIH postdoctoral fellowships and those who had unsuccessfully <u>applied</u> for these fellowship was performed by the NIH (1986). Although both this study and the Garrison and Brown (1986) study demonstrated that NIH fellows comprise a small and select group of M.D.s with NIH postdoctoral awards, this comparison is instructive, although still equivocal, in that it attempts to address some issues of selectivity. Looking at 1968 and 1971 M.D. recipients, the National Institutes of Health found the NIH fellows consistently outperformed their unsuccessful applicant counterparts in terms of medical school faculty appointments (65.1 percent versus 43.5 percent) and NIH/ADAMHA application activity (27.4 percent versus 19.4 percent). Of those who applied for grants, 59.1 percent of the fellows versus 33.3 percent of the unsuccessful fellow applicants received an award.

In general, all of the previously described studies on predoctoral and postdoctoral training have contributed to our knowledge about certain accomplishments of NRSA awardees. Because of unresolved problems with selectivity and heterogeneity of training experiences, however, they have yet to yield strong evidence concerning the effects of NRSA training.

## GAPS IN UNDERSTANDING OUTCOMES AND EFFECTS OF NRSA PROGRAM COMPONENTS

The NRSA program is comprised of several different programs and/or components. For example, there are two basic award mechanisms--individual fellowships and departmental training grants. These mechanisms can be regarded as distinct components, given that they involve different selection procedures and possibly different training experiences. Evaluation studies have paid attention to identifying outcomes and, occasionally, effects associated with these two funding mechanisms, along with examining similar questions for other types of programs/components.

Differences in the outcomes for trainees versus fellows have been investigated by several efforts (ADAMHA, 1986, unpublished report; Clouet, 1986; Coggeshall and Brown, 1984; Garrison and Brown, 1986; NIDR Ad Hoc Consultant Panel, 1988; NIH, 1986; Velletri et al., 1985). For the most part, each has shown that fellows consistently outperformed their trainee counterparts on all measures of interest.

Garrison and Brown (1986) looked at several distinct programmatic strategies aimed at providing M.D.s with postdoctoral research training. These included training appointments for study at NIH and extramural awards (individual fellowships and training grants). Five groups were examined: (1) M.D.s who had extramural fellowships and NIH intramural appointments; (2) M.D.s who had extramural traineeships and NIH intramural appointments; (3) those who only had received NIH intramural appointments; (4) M.D.s who only received extramural fellowships; and (5) M.D.s who only received extramural traineeships. In general, the results indicated that individuals with appointments stemming from very competitive selection procedures (extramural fellowships and intramural appointees) were more likely to receive research support from NIH or other major biomedical research sponsors, to have academic appointments, and to exhibit better publication records. Those with <u>both</u> forms of training had the highest performance on every outcome measure.

Other distinct NRSA program components also have been evaluated. One program involves the institutional training grants supported by MARC (Honors Undergraduate Research Training Grants). Garrison and Brown (1985) conducted a study of those programs that were sponsored by NIH, gathering data on how the program was functioning, the outcomes of program graduates, and the impact on the institution. For example, they examined the types of activities supported by these programs, student satisfaction with the program, educational and occupational status of former program alumni, and institutional enrollments of ethnic minorities in the biomedical sciences. From these data, they found general satisfaction with the program, a substantial proportion of program alumni currently enrolled in another graduate or professional program, and a higher percentage of biology baccalaureates awarded by institutions that had MARC programs as compared to those with no MARC program. Limited outcome data on MARC awardees supported by ADAMHA also have been collected, with the most recent results indicating that 75 percent of students supported during the 1980-1986 period were enrolled in graduate school, along with another 4 percent in medical school (Primos, 1989a).

Another component--short-term research training in health profession schools that received support from NIH--was examined by Sherman (1984). This study involved two components: (1) an examination of whether short-term trainees who had been enrolled in the program during the 1961-1970 period had applied for and/or received NIH/ADAMHA research grants and (2) a comparison of career and research plans and preferences of 1982 and 1983 medical school graduates who had received short-term training and who had not. Although the data suggested a stronger research commitment among program participants as compared to those graduates who were not enrolled in this program, selectivity problems make interpretation of these results impossible.

One distinct and highly visible component is the Medical Scientist Training Program (MSTP). Only two small-scale studies of this program have been conducted aimed at training physician/scientists. Bickel et al. (1981) surveyed medical school deans to examine the adequacy of the number of MSTP training slots, given the pool of qualified applicants, and to compare attrition rates for students in MSTP with those for other M.D./Ph.D. programs. A second study by Sherman et al. (1981)

attempted to assess the effectiveness of MSTP in producing physician/scientists, along with other M.D. research training programs (NIH-supported postdoctoral fellowships and traineeships, NIH research associateships, and NIH clinical associateships). Individuals graduating from MSTP between 1968 and 1973 were matched with students who subsequently trained in one of the other three programs and who were similar in certain characteristics (e.g., type of medical school, age, and MCAT scores). The results suggested that MSTP graduates outperformed the other groups in terms of achieving faculty status and advancement, obtaining NIH research grants, and publication performance.

Since that time, evaluation activities of the MSTP have been minimal (i.e., a brief telephone survey of graduates by program staff and procuring additional analyses of the data collected in Coggeshall and Brown's 1984 study of NIH predoctoral trainees and fellows). To date, there are no firm plans for a more comprehensive evaluation, although the committee in 1983 recommended that a better "picture of costs, training completion rates, post-training employment histories, scientific accomplishments, etc." be developed (National Research Council, 1983).

## GAPS IN BASIC EVALUATIVE DATA FOR PROGRAMS IN SPECIFIC FIELDS/RESEARCH PROBLEMS

The large majority of evaluation studies have focused on NRSA programs administered by NIH. In contrast, evaluation of NRSA-supported training at ADAMHA are both few in number and restricted in scope and coverage. Of the three efforts reviewed, only one included all ADAMHA awardees, another focused on individuals supported by a single institute (NIDA), and the third examined one specific program within an institute (psychology research training sponsored by NIMH). To date, the primary source of evaluative information about programs in health services research (now residing in NCHSR) and those in nursing, administered by the Center for Nursing Research resides in the Committee's report issued in the late 1970s (National Research Council, 1977).

Even when evaluations of ADAMHA programs were carried out, these efforts were circumscribed in nature. Activities typically focused on only a small set of outcomes (e.g., success at obtaining PHS research grants or initial employment of predoctoral trainees after receipt of the doctorate). No attempts were made at examining program effectiveness.

This situation is in some ways not surprising. Since 1975, the bulk of training has been supported by NIH; about 9 of every 10 fellowships and training grants have been awarded by NIH programs (Biomedical and Behavioral Research Personnel Committee, 1989). At the same time, however, this skewed distribution of evaluation efforts has resulted in a lack of information about research training in many fields/research areas, given that ADAMHA, NCHSR, and the Center for Nursing Research provide much of the federal research training support in the behavioral sciences, health services, and nursing.

## RECOMMENDATIONS FOR FUTURE EVALUATION ACTIVITIES

The major "gaps" described above imply that any portfolio of future evaluation efforts for the NRSA program should consider the following issues:

- The quality of the major data bases on NRSA appointments should be assessed so as to ensure that information on program recipients covers the key characteristics of most interest, is accurate, and is collected uniformly on all NRSA components.
- A core set of evaluative data to be collected for all research training programs funded by NRSA should be identified.
- Future efforts should be targeted at gathering information on program characteristics and operation, including data on selection of fellows and trainees, the types of training activities received by individuals in the program, and how these may differ across various program components, fields, and so on.
- Increased attention should be paid to measuring the full range of program outcomes both in absolute and comparative terms.
- In order to facilitate the development of outcome assessments, basic research on scientific career development and maturation, scientific productivity, and the dynamics of training should be supported.
- Although interest in determining the effectiveness of NRSA programs is relatively circumscribed and many aspects of the program currently are not amenable to rigorous impact assessments, the feasibility of implementing effectiveness studies, at least for distinct components of the program, should be explored.

**ENSURING THE QUALITY OF DATA BASES ON NRSA APPOINTMENTS**

The Trainee-Fellow File (TFF), derived from the IMPAC file on all extramural awards made by PHS for research, training, and other activities, is the primary source of information on all persons applying for and receiving fellowships and/or traineeships from NIH, ADAMHA, and other PHS agencies. Thus, having accurate and up-to-date information is extremely important for identifying the populations and subgroups eligible for inclusion in evaluations of the NRSA program. In addition, the ability to use archival data successfully from other major data bases (e.g., the SED and the Consolidated Grants Application File) to augment the TFF would be enhanced.

Clouet's (1986) study of NRSA trainees and fellows supported by the National Institute of Drug Abuse identified misclassification problems in the TFF, at least for NIDA awardees. In the course of identifying NIDA grantees listed in this data base, she found that 67 percent of the student awards identified from agency files were not in the TFF and that over 1,000 people in NIDA nonacademic training programs were included. The extent to which these problems apply to data for other institutes administering NRSA programs is unclear, although interviews with NIH officials did not suggest that it was symptomatic of NIH programs. At the same time, however, Look's 1989 examination of the information available on trainees and fellows suggests that other problems (e.g., inconsistent and even inaccurate reporting on basic demographic characteristics) may exist. Thus, attention should be paid to assessing the quality of data in the TFF and resolving any problems, particularly in light of the following recommendations aimed at increasing evaluation efforts of NRSA programs.

**ENSURING A CORE SET OF EVALUATION ACTIVITIES FOR ALL NRSA PROGRAMS**

To date, evaluation activities for programs in certain fields or areas (e.g., the behavioral sciences and health services research) have been minimal and haphazard. For this reason there are no bodies of information for these fields comparable to that existing for the biomedical and clinical sciences. For example, the primary source of outcome data on health services research training support is the National Research Council's 1977 survey. In the case of the behavioral sciences, similar data were collected in the National Research Council's 1977 study, and although other evaluations did include the behavioral sciences in their populations, they were less than illuminating--either because the studies did not analyze the results for the behavioral sciences separately (e.g., Coggeshall and Brown, 1984) or because the measurement of outcomes was restricted to a single variable, with no incorporation of reasonable comparison groups (ADAMHA, 1986, unpublished report; Schneider, 1980).

This distinct gap in our understanding of the role played by NRSA-supported training for major broad fields of biomedical and behavioral research should be remedied. As found in previous studies of research training, distinct differences in patterns of study and accomplishments appear with respect to individual fields (National Research Council, 1981b; Snyder, 1988). Further, those fields that are most lacking in evaluative data are those that also make substantial contributions to addressing key health and mental health problems (e.g., National Research Council, 1985), along with other social concerns (e.g., Gerstein et al., 1988).

Moreover, there is some suggestion that there may be future shortages of qualified researchers in these areas. A recent analysis by ADAMHA (1989b) indicated that the pool of researchers working in alcohol abuse, drug abuse, and mental health areas is aging rapidly; in 1979, 26 percent of ADAMHA-funded investigators were 35 or younger, and this percentage has declined to 13 percent in 1987. The proportion of young applicants to ADAMHA has followed a parallel trend. In addition, the average age of Ph.D. principal investigators has increased 1.6 times faster than the average age of NIH-funded researchers.

The key point is that there should be a core set of evaluative data collected on all NRSA programs and research areas. Only in this manner can the scope and breadth of NRSA activities be examined and strategies developed to improve programs aimed at all fields, levels, and types of research training supported by NRSA funds.

**INITIATING EVALUATION ACTIVITIES THAT PROVIDE BETTER INFORMATION ON PROGRAM PARTICIPATION AND OPERATION**

At present, it is difficult to characterize what is happening in NRSA programs, regardless of the type, level, or field of training. For example, we cannot readily provide profiles on the basic demographic characteristics of predoctoral and postdoctoral fellows and trainees, the characteristics of institutions receiving NRSA awards, the types of training models and activities being supported, the components of NRSA training that distinguish it from other training programs aimed at producing researchers in the same fields, and so forth. This dearth of knowledge not only handicaps our ability to understand the basic components of the NRSA program, but it also hampers designing studies that can better assess outcomes (and, ultimately, the effects) of NRSA training and identify with greater confidence which types of training strategies may work better than others (e.g., multidisciplinary programs versus those that focus on one discipline or specialty).

Efforts need to be initiated to rectify this situation. Although retrospective studies of former trainees and fellows have provided useful insight into training gaps and deficiencies (e.g., Gentile et al., 1987), they are limited. Surveys of recent Ph.D.s and M.D.s/Ph.D.s upon receipt of their doctorate could provide some preliminary insight into the types of variables that should be examined and how they can be measured best (e.g., the phrasing of survey questions, if appropriate). The same can be said for surveys of NRSA-supported postdoctoral trainees and fellows who have just completed their postdoctoral appointment.

The most preferred strategy is to collect data from individuals at key intervals <u>during the training process</u> so as to avoid the problems of selective memory and to obtain a more complete picture of how and when training actually occurs. Such efforts also could provide data on many other issues. Program implementation could be examined (i.e., did individuals actually receive the training set forth in the funded application?). If information on trainee satisfaction were collected, the need for modifications in training policies or regulations also might be identified. Finally, if such efforts were extended to a sample of predoctoral students or postdoctoral associates without traineeships in the same academic program and possibly to students in similar programs in departments without training grants, we would begin to develop a better sense of the strength and integrity of various NRSA "treatments" (e.g., whether NRSA programs provide training experiences sufficiently different from those received by students in the same program or in other programs).

The manner in which trainees are selected is another important area worthy of close scrutiny. We know, at least for the biomedical sciences, that predoctoral trainees tend to receive their training in top-ranked institutions/departments; presumably, the training grant application review process is selecting the best programs, and the departments in which these programs are based are highly selective in terms of graduate student admitting policies.[7] However, we do not know whether

[7] These programs have been described as "contests" (Rossi, personal communication, May, 1989) in that the most promising applicants are chosen to receive awards. It appears that the main contestants are institutions; the top-ranked research universities clearly are given preference. However, within the departments that house the NRSA programs, it is not clear whether the contest rules are the same, given the problems described in

those who actually receive traineeships are awarded them because they are the best "match" in terms of the requirements of the NRSA-supported program, because it is a convenient source of money at a particular period of time, or for other reasons. The same problem plagues applicants for postdoctoral appointments, particularly in medical schools. Needless to say, the ability to ultimately provide answers to questions seeking evidence on effectiveness is hampered by our lack of knowledge about how individuals are selected to receive NRSA support.

## INCREASED ATTENTION TO ASSESSING PROGRAM OUTCOMES

Those associated with NRSA programs, particularly at the program administrative level, want to know what works best in research training, either in terms of specific training models (e.g., broad versus specialized training) or in terms of specific populations (e.g., predoctoral support versus postdoctoral support versus both types of support in producing physician/scientists). Previous large-scale evaluation efforts that have had to rely on retrospective assessment strategies have been unable to confidently provide unequivocal answers to these questions because of the problems associated with drawing solid comparison groups. Given this substantial interest in program outcomes, particularly comparative outcome data, concerted attention should be devoted to examining the full range of outcomes implied by the goals of the NRSA program (i.e., both research and teaching activities as specified in program announcements and payback requirements for NRSA programs) and to pilot testing new data collection strategies. The issues and variables warranting consideration are discussed by Fox (1983) and Gee (1989) and should be guided by previous research on scientific careers (e.g., Long et al., 1979; McGinnis et al., 1982; McGinnis and Long, 1988; Stephan, 1986). Not only would efforts to accurately measure outcomes and test various measurement strategies help to better examine scientific productivity and career development, but they also would allow an understanding of the marginal gain associated with using these measures as contrasted to those previous approaches that have had to rely on existing archival data bases.

Outcome assessments also could benefit from efforts geared at improving the use of existing archival data bases. For example, rather than using archival data bases to obtain a "snapshot" at one point in time of research grant success, the feasibility of using such data bases to track individuals over time, similar to that performed by Biddle et al. (1989) in

---

this paper. This area requires further investigation, modeled on the work by Carter et al. (1987) for Research Career Development Awards.

examining the career accomplishments of Research Scientist Development awardees, could be explored. Another possibility would be to assess the feasibility of using relevant data from other archival sources that could augment the information on traditionally used measures (e.g., research grant success) or provide data on outcomes that have received only minimal attention (e.g., teaching others to be researchers). For example, Yasumura (1986) found that recipients of the K0-6 awards were more likely to receive PHS training grants--a relevant outcome, based on the request made by Congress for obtaining information on the number of awardees who engage in health research or teaching as a career (P.L. 99-1258).

In addition, exploring relationships between program characteristics and outcomes also is needed. For example, recent data on NHLBI awardees suggest that those M.D.s who were postdoctoral trainees in programs designed to train both M.D.s and Ph.D.s were more likely to be receive a NIH grant than those who were postdoctoral trainees in programs only aimed at M.D.s (C. Roth, personal communication, June, 1989). More efforts of this kind can guide subsequent outcome studies in terms of identifying key programmatic variables that should be considered.

Another focus should be on designing outcome studies where the entire "dosage" of training is measured. At the very least, for outcome studies of ADAMHA, NIH, or other PHS support, this information should be available from the TFF currently in existence. However, given that NRSA research training is just one part of funding for scientific personnel development, measuring the length of training must take into account all types of research training and development (e.g., short-term research training, predoctoral and postdoctoral traineeships and fellowships, and career development awards). If the previously described efforts for developing and/or reinstituting data bases on personnel working on PHS grants also are initiated, this would provide additional information on dosage by getting data on predoctoral or postdoctoral support financed by faculty research grants from these agencies. The completeness and accuracy of this information then could be checked by actually contacting the individuals chosen for examining outcomes to determine what types of training support they actually received--both from PHS and other sponsors.

The final point is a call for evaluation efforts that incorporate a more longitudinal perspective. To date, the majority of the studies have been of the snapshot variety--career accomplishments at one point in time. Future evaluation activities ideally must track individuals from the time they begin training (or apply for training opportunities) through the completion of their training and during their scientific careers. Even more limited efforts would benefit from following the Carter et al. (1987) and Biddle et al. (1988) approach to measuring outcomes (e.g., PHS research grant activity) at several points after the individual has completed training. Only within a longitudinal perspective can a specific achievement or pattern of achievements be interpreted and yield information that can illuminate our understanding of these programs and research training in general.

## INCREASING THE AVAILABILITY OF RESOURCES FOR RESEARCH ON RESEARCH TRAINING

Developing better outcome strategies is at least partially dependent on better understanding the nature of scientific careers, what influences productivity, and other similar relationships. Currently, funding for research on scientific careers, resources, and so forth is not abundant. The NSF's Division of Science Resources devotes some funds to supporting studies in these areas (e.g., National Science Foundation, 1986). However, monies for this program are limited, and proposals often must include all scientific and engineering fields rather than those most relevant to NRSA programs or must focus on using preexisting survey data collected by NSF (data that typically includes, for example, only small numbers of NRSA-supported individuals in various categories).

Research on scientific career development and enhancement can go a long way toward addressing basic questions about productivity, motivations for choosing a scientific career, the factors that facilitate or inhibit research productivity, and so on. In conjunction with these research programs, efforts need to be targeted at developing new measures of scientific productivity and quality of research and testing their feasibility. Work on assessing the quality of training programs, apart from the accomplishments of their graduates, is another component that requires attention, particularly if future evaluation efforts aimed at judging the quality of training grant recipients (i.e., programs/departments within institutions) are considered worthy of exploration.

The list of research questions is endless, and reasonably strong arguments could be made for choosing any particular question or methodological strategy as the initial starting point. The major point is that resources for these efforts should be increased so that our ability to carry out evaluation activities that can both meet the needs of major NRSA constituencies and assist in the continual improvement of training activities in all fields of health-related research can be enhanced.

**EXPLORING THE FEASIBILITY OF EVALUATION DESIGNS FOR ASSESSING EFFECTIVENESS**

Given our lack of knowledge about the nature of the selection and training processes, improving the quality of future designs certainly represents a technical challenge. Perhaps the first step toward obtaining better estimates of the relative effectiveness of different types of research training programs and determining the training strategies or mechanisms that work best is to identify where the use of high quality designs can be implemented.

One way to begin this process is to focus initially on exploring opportunities that may exist for assessing specific components. For example, the MSTP appears to be a promising candidate for consideration, given that training program directors often vigorously recruit students. Thus, a pool of applicants may be available to use as a comparison group. The feasibility of "beating the bushes" to augment the size of the applicant pool, of convincing directors to collect necessary data on applicant characteristics, and of persuading them to adhere to a standard selection process and criteria should be explored. Even if random assignment is not possible (a likely event), the use of such designs as regression-discontinuity is worth examining. This latter type of design was used successfully in Carter et al.'s (1987) evaluation of the Research Career Development Award.

In addition, exploring the feasibility of implementing high quality evaluation designs for the MST program is attractive, given the level of interest in this component. For example, some concern has been expressed over the available supply of physician-researchers, and MD-Ph.D. training programs are being established by other agencies to resolve this problem. At the same time, however, the best way to produce physician-scientists is a point of controversy. Further, questions have been raised about the fact that the total expenditures per graduate for MST programs have been increasing over the last decade and are now significantly higher than for graduates supported by other predoctoral training grants (National Research Council, 1983). Well-designed evaluations of the MST program, focusing on the outcomes of its graduates and its effectiveness on certain outcomes and as compared to other research training alternatives for physician-scientists, would do much towards addressing these issues.

The lower level of interest in effectiveness questions expressed by constituencies, coupled with the difficulty in carrying out high quality studies to address these questions, does not, however, imply that exploring the feasibility of implementing rigorous designs for estimating program effects is not important. The long-term objective for NRSA evaluation efforts should be to gain an understanding of what works in training, which programs and program elements work better, and how training should be assessed. Small-scale, pilot tests of more rigorous approaches can be quite instructive in terms of identifying where more better designs can be implemented and ultimately yield an understanding of research training itself.

Many of the concerns being expressed today (e.g., future shortages of trained scientists, issues of scientific misconduct, and the lack of interdisciplinary research efforts in many major problem areas) have their core issues surrounding human resources and training. If we are successful in providing a better understanding of NRSA research training and how it operates and contributes to the development of outstanding researchers and research mentors, one step toward addressing these issues will have been taken.

## REFERENCES


Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA). 1989a. ADAMHA NRSA Research Training Tables FY 1988. Rockville, MD.

Alcohol, Drug Abuse, and Mental Health Administration. 1989b. Addressing the Shortage of Young ADAMHA Principal Investigators. (Available from Walter Schaffer, ADAMHA, Parklawn Building, 5600 Fishers Lane, Rockville, MD 20857.)

Barries, J. 1986. Projected supply and demand for Ph.D. Biologists. Paper presented at the meeting of the American Zoological Society, December 1986.

Bickel, J. W., C. R. Sherman, J. Ferguson, L. Baker, and T. E. Morgan. 1981. The Role of M.D.-Ph.D. Training in Increasing the Supply of Physician-Scientists. New England Journal of Medicine 304:1265-1268.

Biddle, A. K., G. M. Carter, J. S. Uebersax, and J. D. Winkler. 1988. Research Careers of Recipients of the Research Scientist and Research Scientist Development Award. Santa Monica, CA: The RAND Corporation.

Carter, G. M., A. Robyn, and A. M. Singer. (1983). The Supply of Physician Researchers and Support for Research Training: Part I of an Evaluation of the Hartford Fellowship Program (N-2003-HF). Santa Monica, CA: The RAND Corporation.

Carter, G. M., J. D. Winkler, and A. K. Biddle. 1987. An Evaluation of the NIH Research Career Development Award. Santa Monica, CA: The RAND Corporation.

Clouet, [illegible] H. 1986 The Career Achievements of Trainees and Fellows Supported by the National Institute of Drug Abuse. Rockville, MD: NIDA.

Coggeshall, P. E., and P. W. Brown. 1984. The Career Achievements of NIH Predoctoral Trainees and Fellows. Washington, D.C.: National Academy Press.

Coyle, S. L., and Thurgood, D. H. 1989. Summary Report 1987: Doctorate Recipients from United States Universities. Washington, D.C.: National Academy Press.

Fox, M. F. 1983. Publication Productivity Among Scientists: A Critical Review. Social Studies of Science 13:285-305.

Garrison, H. H., and P. W. Brown, 1986. Minority Access to Research Careers: An Evaluation of the Honors Undergraduate Research Training Program. Washington, D.C.: National Academy Press.

Gee, H. 1989. Productivity. Paper commissioned by the Biomedical and Behavioral Research Personnel Committee, National Academy of Sciences, Washington, D.C.

Gentile, N. O., G. S. Levey, P. Jolly, and T. H. Dial. 1987. Post-doctoral Research Training of Full-Time Faculty in Departments of Medicine. Washington, D.C.: Association of Professors of Medicine and Association of American Medical Colleges.

Gerstein, D. R., Luce, R. D., N. J. Smelser, and S. Sperlich (eds.). 1988. The Behavioral and Social Sciences: Achievements and Opportunities. Washington, D.C.: National Academy Press.

Levey, G. S., C. R. Sherman, N. O. Gentile, L. J. Hough, T. H. Dial, and P. Jolly. 1988. Postdoctoral research training of full-time faculty in academic departments of medicine. Annals of Internal Medicine 109:414-418.

Long, J. S., P. D. Allison, and R. McGinnis. 1979. Entrance into the academic career. American Sociological Review, 44:816-830.

Look, M. 1989. Final Report on the Proposed Data Matrix for the Biomedical and Behavioral Research Personnel Committee. Bethesda, MD: Quantum Research Corporation.

Martin, M. R. 1936. The Results of a Manual Look-up of 300 References from Non-foreign Medical School Faculty Clinicians from the B/I/D Journal Set and an Analysis of Report by the AAMC on the Number and Character of Physical Researchers. (Available from Division of Program Analysis, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20092.)

McGinnis, R., and J. S. Long. 1988. Entry into academia: Effects of stratification, geography, and ecology. In D. W. Brenneman and T. I. K. Youn (eds.) Academic Labor Markets and Careers, New York: Falmer Press.

McGinnis, R., P. D. Allison, and J. S. Long. 1982. Postdoctoral training in bioscience: Allocation and outcomes. Social Forces 60:701-723.

National Institute of Dental Research (NIDR) Ad Hoc Consultant Panel. 1988. The National Research Service Award Program: Report of the NIDR Ad Hoc Consultant Panel. Bethesda, MD: NIDR.

National Institutes of Health (NIH). 1986. Effects of the National Research Service Award Program on Biomedical Research and Teaching Careers. Washington, D.C.: NIH.

National Institutes of Health. 1987. NIH Data Book 1987. Washington, D.C.

National Institutes of Health. 1988. NHLBI Fact Book Fiscal Year 1988. Washington, D.C.: NIH.

National Research Council (NRC). 1976. Research Training and Career Patterns of Bioscientists: The Training Programs of the National Institutes of Health. Washington, D.C.: National Academy of Sciences.

National Research Council. 1977. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy of Sciences.

National Research Council. 1978. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy of Sciences.

National Research Council. 1979. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy of Sciences.

National Research Council. 1981a. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy of Sciences.

National Research Council. 1981b. Postdoctoral Appointments and Disappointments. Washington, D.C.: NRC.

National Research Council. 1983. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy Press.

National Research Council. 1985. Personnel Needs and Training for Biomedical and Behavioral Research. Washington, D.C.: National Academy Press.

National Research Council. 1989. Biomedical and Behavioral Research Scientists: Their Training and Supply. Washington, D.C.: NRC.

National Science Foundation. 1988. Project Summaries: FY 1988 (NSF 88-336). Washington, D.C.: U.S. Government Printing Office.

Office of Technology Assessment. 1985. Demographic Trends and the Scientific and Engineering Work Force (OTA-TM-SET-35). Washington, D.C.: U.S. Government Printing Office.

Porter, B. F. 1979. Transition: A follow-up study of 1973 postdoctorals. In the American Physical Society, The Transition in Physics Doctoral Employment 1960-1990. New York, NY: APS.

Primos, M. E. 1989a. Minority Access to Research Careers Program: Briefing on Program Activities, FY 1980 to 1986. (Available from M. Primos, Minority Research Resources Branch, NIMH, Parklawn Building, 5600 Fishers Lane, Rockville, MD 20857.)

Primos, M. E. 1989b. Minority Fellowship Program: Summary of Program Activities, FY 1973 to FY 1987. (Available from M. Primos, Minority Research Resources Branch, NIMH, Parklawn Building, 5600 Fishers Lane, Rockville, MD 20857.)

Schneider, S. 1980. Positions of Psychologists Trained for Research. American Psychologist 35:861-866.

Sherman, C. R. 1983a. Notes on the NIH Role in Support of Postdoctoral Research Training of Two Groups of Physicians. (Available from Charles Sherman, NIH, 9000 Rockville Pike, Bethesda, MD 20092.)

Sherman, C. R. 1983b. Training and Manpower Development. Presentation at the Meeting of the Advisory Committee to the Director, NIH, Bethesda, MD.

Sherman, C. R. 1984. Perspectives on Short-Term Training in Schools of the Health Professions. (Available from Charles Sherman, NIH, 9000 Rockville Pike, Bethesda, MD 20092.)

Sherman, C. R. 1989. The NIH Role in the Training of Individual Physician Faculty: A Supplementary Analysis. (Available from Charles Sherman, Office of Science Policy and Legislation, NIH, 9000 Rockville Pike, Bethesda, MD 20092.)

Sherman, C. R., H. P. Jolly, T. E. Morgan, E. J. Higgins, D. Hollander, T. Bryll, and E. R. Sevilla III. 1981. On the Status of Medical School Faculty and Clinical Research Manpower 1968-1990. Washington, D.C.: National Institutes of Health.

Snyder, J. 1988. Early Career Achievements of National Science Foundation Graduate Fellows, 1967-1976. Washington, D.C.: National Research Council.

Stephan, P. 1986. Age publishing patterns in science: Estimation and measurement issues. In Anthony F. J. van Raan (ed.), A Handbook of Quantitative Studies of Science and Technology. Amsterdam, Holland: Elsevier Science Publishing Division of 'th Holland.

Task Force on Minority Research Training. 1986. Minority Research Training Programs at the National Institute of Mental Health: Descriptions and recommendations. Rockville, MD: National Institute of Mental Health.

Tjioe, P. 1989. Researchers' Tracking System (RTS) User's Guide. Laurel, MD: General Sciences Corporation.

Velletri, P. A., C. R. Sherman, and G. Bowden. 1985. A Comparison of the Career Achievements of NIH Predoctoral Trainees and Fellows. Washington, D.C.: National Institutes of Health.

Vetter, B. M. 1989. Recruiting Doctoral Scientists and Engineers Today and Tomorrow (Occasional Paper 89-2). Washington, D.C.: Commission on Professionals in Science and Technology.

Williams, P. C. 1979. Evaluation of the Annual Manpower Survey of NIH Research Grants. Sumner, MD: Cooper-Williams Associates.

Yasumura, S. 1984. The Research Career Award (K06): A 20-Year Perspective and an Analysis of Research Productivity. Washington, D.C.: National Institutes of Health.

Zumeta, W. 1985. Extending the Educational Ladder: The Changing Quality and Value of Postdoctoral Study. Lexington, Mass.: D.C. Heath.

# APPENDIX A
## EVALUATION QUESTIONS ASKED BY MAJOR NRSA CONSTITUENCIES

| Type of Question | Congress | NIH/ADAMHA Staff |
|---|---|---|
| **Demand for the Program** | What is the Nation's overall need for biomedical and behavioral science personnel?*<br><br>In which fields and research areas are investigators most needed? How many training slots are needed to address these shortages?<br><br>Who are the major sponsors of research training?<br><br>What role does industry play in supporting research training? What should its role be?<br><br>If the NRSA program was terminated, would the remaining research training programs adequately meet the need for biomedical and behavioral scientists?*<br><br>Why should NRSA funding levels be increased when there are complaints as to the insufficient amount of research monies available to fund all high quality research grant applications? | How many clinical researchers are needed to address major health problems over the the next decade?<br><br>What is the need for scientists who work problems related to drug abuse? How many are needed to address particular issues (e.g., neuromechanisms of drug action?<br><br>How much postdoctoral research training is supported by mechanisms other than NRSA?<br><br>Are there more high quality NRSA applications that can be supported with existing funds? |

## APPENDIX A (Continued)
## EVALUATION QUESTIONS ASKED BY MAJOR NRSA CONSTITUENCIES

| Type of Question | Congress | NIH/ADAMHA Staff |
|---|---|---|
| **Program Participation** | How many individuals receive NRSA awards each year?**<br><br>What does the pool of applicants look like, and has it changed over time in terms of demographics and quality?<br><br>How many individuals are supported in each of the research fields targeted as shortage areas? | How many individuals receive NRSA fellowships vs. traineeships?<br><br>How many women and minorities are being supported by NRSA?<br><br>In what types of institutions are individuals being trained? |
| **Program Characteristics and Operation** | Does the payback requirement function as a disincentive to individuals who would be qualified for NRSA support? Has this requirement affected the number and quality of applicants?**<br><br>Does the payback requirement have any effect on keeping NRSA recipients active in research for longer periods of time?**<br><br>Do current stipend levels affect the attractiveness of NRSA awards to potential recipients? | Is the payback requirement structured appropriately? In particular, should not MDs who are supported for shorter periods of time (e.g., less than 12 months) be required to pay back NRSA funds, given that they may be less likely to be active researchers?<br><br>What criteria are actually applied by Institutional Review Groups in recommending training grant applications for funding? Are these criteria consistent with what we know about the characteristics of successful training programs? |

APPENDIX A (Continued)
EVALUATION QUESTIONS ASKED BY MAJOR NRSA CONSTITUENCIES

| Type of Question | Congress | NIH/ADAMHA Staff |
|---|---|---|
| **Program Characteristics and Operation (continued)** | What modifications are required in NRSA policies to meet the need for research personnel? | How are individual trainees actually selected by training grant directors? Is this prcess uniform across programs in the same field or research area?<br><br>What types of training are awardees actually receiving (e.g., formal courses and "hands-on" research experience?<br><br>Are the awardees actually being trained to be independent researchers or are they simply being utilized as "hired hands" in faculty laboratories? |
| **Program Outcomes and Effectiveness** | What are the kinds of research positions available to and held by NRSA awardees?*<br><br>How many NRSA awardees have careers in health research or teaching?**<br><br>How effective is the NRSA program compared to other research training programse, e.g., those operated by NSF? | What happens to NRSA awardees?<br><br>Do awardees go on to have active research careers in areas relevant to the mission of the sponsoring agency (e.g., drug abuse research)?<br><br>What works best in training physician-researchers (e.g., MD/PhD programs or extramural postdoctoral traineeships)? |

APPENDIX A (Continued)
EVALUATION QUESTIONS ASKED BY MAJOR NRSA CONSTITUENCIES

| Type of Question | Congress | NIH/ADAMHA Staff |
|---|---|---|
| Effectiveness of Program Components | How effective is the training grant mechanism in encouraging individuals to engage in health research and training careers?** | Is broad-based or specialty training the best way to produce researchers in a specific disease area?<br><br>Which types of programmatic strategies work best in training grants? |
| Cost-effectiveness of NRSA Programs | | How can an agency best spend its "training buck?"<br><br>What is the cost of training researchers and then not utilizing them due to problems with insufficient R01 grant funds? |
| "Research" on Research Training | What is the best way to produce clinical researchers?<br><br>What is the best way to attract undergraduates to pursue PhDs in a biomedical science? | How long can we expect a scientist to be actively involved in research?<br><br>What attracts a student to pursue a degree in a biomedical science?<br><br>What are the key "choice points" in deciding to pursue a career in science vs. in other areas? |

## APPENDIX A (Continued)
## EVALUATION QUESTIONS ASKED BY MAJOR NRSA CONSTITUENCIES

| Type of Question | Congress | NIH/ADAMHA Staff |
|---|---|---|
| "Research" on Research Training | What is the relationship between research training activities and producing quality investigators? | How can the "best and the brightest" into scientific careers? * |
| | | What are the components of scientific productivity and how do these change over time? What does this mean for how the outcomes of NRSA programs should be measured? |
| | | In evaluating research training programs, what are the most appropriate outcome measures to use? |

<u>Note</u>. Unless otherwise indicated, the source of these questions were interviews with Congressional staff and individuals in charge of training programs at NIH and ADAMHA (N=16 interviews).

*Question is specified in the language of P.L. 100-607.

**Question is specified in the language of P.L. 99-1258.

## APPENDIX B

For the purposes of this paper, certain basic criteria were used to identify the evaluation studies chosen for review:

- The focus of the study was on NRSA programs. Studies of research training in general (e.g., Zumeta, 1983), although their populations may have included NRSA awardees, were not included as formal evaluations of these programs.

- "Data books" such as those produced by NIH (1988) and ADAMHA (1989a) were not classified as formal evaluations depite the fact that they provide data on number of awards and fundings levels. This type of document, however, is considered when discussing the availability of information on major evaluation questions (e.g., levels of program participation).

- The results had to exist in a form that was available for at least limited dissemination. All studies presented in Appendixes C and D were either formal published reports or reports prepared for internal and/or limited distribution (upon requested to interested parties). Other evaluation-related activities for which no "write-up" was available (e.g., computer printouts for which results were not summarized or informal staff attempts to gather preliminary data on outcomes of awardees) were not considered. They are, however, discussed in the paper, where appropriate.

Sixteen individual evaluation efforts met these criteria, and the major characteristics of each are summarized in Appendixes C and D. Appendix C presents information related to the coverage of each study in terms of: (1) the agency administering the programs (i.e., NIH or ADAMHA); (2) whether the study focused on a particular program or specific institute (e.g., NIMH); (3) the major category of training evaluated (e.g., predoctoral vs. postdoctoral); (4) whether the focus of the program was on training MDs, PhDs, or undergraduates; (5) the major fields of scientific training incorporated (e.g., biomedical vs. clinical sciences); (6) the time period examined (e.g., 1978-81 awardees); and (7) the major questions addressed by the evaluation.

In Appendix D are described selected methodological characteristics so as to better depict how evaluations were conducted and identify how program outcomes and/or effects were examined, given that the majority of studies reviewed attempted to address these types of questions.

APPENDIX C

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM

Major Characteristic.

| Study and Year | Major NRSA | | | Focus on Specific Inst/Program | | Level of Training | | | Types of Degree Programs | | | Major Fields Included[a] | | | | Years Included (19__) | Major Questions[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NIH | ADAMHA | | Yes | No | Und | Pre | Post | PhD | MD | Other | Bio | Clin | Beh | Oth | | Part Chars | Prog Oper | Outc Eff | Comp Eff | Cost Eff |
| **EVALUATION ACTIVITIES BY NAS** | | | | | | | | | | | | | | | | | | | | | |
| Coggeshall & Brown | X | | | | X | | X | | X | | | X | | | | 67-81 PhD recipients | X | X | X* | | |
| Garrison & Brown, 1986 | X | | | | X | | | X | X | X | | X | X | | | 61,66,71, 76,81 PhD recipients 65,68,71,74, 77 MD recipients | X | X | X* | X | |
| National Research Council, 1976 | X | | | | X | | X | X | X | X | | X | X | X | X | 38-72 NIH awardees Scientists and MDs as of 1976 | X | X | X* | X | |
| National Research Council, 1977 | X | X | X | | X | | X | | X | | | X | | X | X | 71-75 PhD recipients NRSA awardees in health services research | | | X | | |
| National Research Council, 1978 | X | X | | | X | | X | | X | | | X | | X | | PhD granting programs as of 1975 | | X | X | | |

## APPENDIX C

### EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM (Continued)

| Study and Year | Major NRSA Sponsor: NIH | Major NRSA Sponsor: ADAMHA | Focus on Specific Inst/Program: Yes | Focus on Specific Inst/Program: No | Level of Training: Und | Level of Training: Pre | Level of Training: Post | Types of Degree Programs: PhD | Types of Degree Programs: MD | Types of Degree Programs: Other | Major Fields Included[a]: Bio | Major Fields Included[a]: Clin | Major Fields Included[a]: Bel | Major Fields Included[a]: Oth | Years Included (19__) | Major Questions[b]: Part Chars | Major Questions[b]: Prog Oper | Major Questions[b]: Outc Efft | Major Questions[b]: Comp Efft | Major Questions[b]: Cost Efft |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| National Research Council, 1981 | X | | | X | | X | | X | | | X | | | | PhD granting programs that lost NIH support in 79-80 | | | | X | |
| Garrison & Brown, 1985 | X | | X URHP | | X | | | | | X | X | | X | | 77-84 program participants and institutions | X | X | | X* | |
| **EVALUATION ACTIVITIES BY NIH** | | | | | | | | | | | | | | | | | | | | |
| NIH, Division of Program Analysis, 1986 | X | | | X | | | X | X | X | | X | X | | | 64-77 PhD recipients 68,71,74 MD recipients | X | X | X* | X | |
| Velletri et al., 1985 | X | | | X | | X | | X | | | X | | | | 67-74 PhD recipients | X | X | | X | |
| Sherman, 1984 | X | | X Short-Term | | | X | | | X | X | | X | | | 58-80 NIH awardees 82-83 senior year medical students | | | | X | |
| National Institute of Dental Research, 1987 | X | | X NIDR | | | | X | X | X | X DDS/DMD DDS/DMD-PhD | X | X | X | X | All NIDR awardees | X | | X | X | |

C-2

APPENDIX C

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM (Continued)

Major Characteristics

| Study and Year | Major NRSA | | Focus on Specific Inst/Program | | Level of Training | | | Types of Degree Programs | | | Major Fields Included[a] | | | | Years Included | Major Questions[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NIH | ADAMHA | Yes | No | Und | Pre | Post | PhD | MD | Other | Bio | Clin | Beh | Oth | (19__) | Part Chars | Prog Oper | Outc Efft | Comp Efft | Cost Efft |
| **EVALUATION ACTIVITIES BY ADAMHA** | | | | | | | | | | | | | | | | | | | | |
| Schneider, 1980 | | X | X NIMH | | X | X | X | | | | | | X Psych | | 68-78 PhDs from NRSA programs in psychology | | | X | | |
| Clouet, 1986 | | X | X NIDA | | X | X | X | | | | X | | X | X | 73-84 NRSA awardees | X | | X | X | |
| ADAMHA, 1986 | | X | | X | X | X | X | | | | X | X | X | X | 67-83 NRSA awardees | X | | X | | |
| **EVALUATION ACTIVITIES BY OTHER SPONSORS** | | | | | | | | | | | | | | | | | | | | |
| Sherman, et al., 1981 | X | | X MSTP | | X | X | | | | MD-PhD | | X | | | 68-73 MD recipients | | | | X* | |
| Bickel, et al., 1981 | X | | X MSTP | | X | | | | | X MD-PhD | | X | | | 79 MD-PhD recipients 79 MD-PhD students & "drop-outs". 76 MSTP applicants | | X | | X | |

[a] Although field definitions differ slightly across studies and NRSA sponsors, these field categories are generally described as follows: Biological sciences (includes fields in the basic biomedical sciences such as anatomy, genetics, and microbiology); clinical sciences (includes fields where the typical degree is a MD, DDS, and DMD such as internal medicine, psychiatry, or dentistry); behavioral sciences (e.g., psychology, sociology, or anthropology); and other (e.g., health services research and nursing research).

[b] Key to abbreviations:
- Part Chars - Questions concerning levels of participation in the program or characteristics of participants.
- Prog Oper - Questions about program operation (e.g., length of support provided, amount of support, and types of training received).
- Outc Efft - Questions concerning program outcomes/effectiveness (e.g., career outcomes of awardees and effectiveness of NRSA).
- Comp Efft - Questions about outcomes and/or effectiveness of specific components or programs within the overall NRSA program (e.g., programs administered by a certain NIH institute or distinct types of training programs such as MSTP).
- Cost Efft - Questions concerning cost effectiveness.

*Indicates an attempt to look at effectiveness through the use of either methodological or statistical procedures.

APPENDIX D

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM: SELECTED METHODOLOGICAL CHARACTERISTICS

| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
|---|---|---|---|---|---|
| | | | **EVALUATION ACTIVITIES FOR NAS** | | |
| Coggeshall & Brown, 1984 | o PhDs who had 9 or more months of NIH predoc support<br>o PhDs trained in departments with NIH predoc support but who received 0-8 months of NIH support<br>o PhDs trained in departments that had no NIH predoc funding | Multiple regressions for selected outcome variables, using:<br>o years since PhD<br>o prestige of PhD institution<br>o years of NIH predoc support | o Attainment of doctorate<br>o Postdoctoral training<br>o Early career employment (academic positions and involvement in research)<br>o Pursuit of NIH and NSF research grants<br>o Publications and citation | o Retrospective<br>o Based on existing data bases<br>o Population data for some outcomes | o How successful are NIH predocs in pursuing careers in biomedical research?<br>o What is the relative importance of NIH predoc support in explaining outcomes?<br>o Are there any "era" effects?<br>o Do outcomes vary among specific NIH institutions?<br>o Has the NRSA program changed over (e.g., numbers of awards made by institutions?) |
| Garrison & Brown, 1986 | For PhDs:<br>o NIH postdoc Fellows<br>o NIH postdoc Trainees<br>o PhD recipients for same time period who indicated plans for non-NRSA postdoc<br>o PhD recipients for same time period who indicated "no postdoc plans" | Multiple regressions for selected outcomes, using:<br>o Years to PhD<br>o Prestige of PhD granting department<br>o Selectivity of baclaureate institution<br>o Study group membership | o Employment (academic positions and involvement in research)<br>o Pursuit of research grants from NIH, NSF, VA, and selected private sponsors<br>o Publications and citations | o Retrospective<br>o Based on existing data bases<br>o Population data for some outcomes | o How successful are NIH postdocs with PhDs in pursuing biomedical research careers?<br>o Are there differences in outcomes for NIH trainees vs. NIH fellows?<br>o Are there "era" effects?<br>o Has the program changed over time?<br>o What are the characteristics of awardees? |
| | For MDs:<br>o NIH postdoc Fellows<br>o NIH postdoc Trainees<br>o Those with no NIH postdocs who reported involvement in teaching/research<br>(Continued on next page) | Multiple regressions for selected outcomes, using:<br>o Prestige of medical school granting MD<br>o Employment in medical school | o Medical school employment<br>o Pursuit of research grants from NIH, NSF, VA, and selected private sponsors)<br>o Publications and citations | o Retrospective<br>o Based on existing<br>o Population data for some outcomes | o How successful are NIH postdocs with MDs in pursuing biomedical research careers?<br>o Are there differences in in outcomes for NIH Trainees vs NIH Fellows?<br>(Continued on next page) |

APPENDIX D (Continued)

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM: SELECTED METHODOLOGICAL CHARACTERISTICS

| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
|---|---|---|---|---|---|
| Garrison & Brown, 1986 (Continued from previous page) | o Those with no NIH postdocs and who reported primary involvement in "other" activities<br><br>AND<br><br>o NIH Trainees who had intramural res. appt.<br>o NIH Fellows who had intramural res. appt.<br>o Intramural appointees only<br>o Extramural NIH Trainees only<br>o Extramural NIH Fellows only<br>o No NIH extramural or intramural apts. | o Study group membership | SAME AS ABOVE | | o Are there differences in outcomes for those with NIH intramural training?<br>o Are there any "era" effects?<br>o What are the characteristics of awardees? |
| National Research Council, 1986 | For PhDs:<br>o NIH postdocs<br>o NIH predocs who had no postdoc support<br>o PhDs with no pre- or postdoc support<br>o Those who did not receive the PhD<br>For MDs:<br>o MD-PhDs who also had NIH postdocs<br>o MD-PhDs who had no NIH postdocs<br>o MDs with NIH postdocs<br>o MDs with no PhD or postdoc training | None | o Attainment of doctorate<br>o Academic employment<br>o Primary work activity (e.g., research, training, or administration)<br>o Publications and citations | o Retrospective<br>o Use of existing data bases<br>o Site visits to departments | o How successful are NIH progress in encouraging and promoting biomedical research/training careers?<br>o Does the mechanism of support affect the outcomes?<br>o What are the characteristics of awardees?<br>o How have training monies been used by programs and what are need improvements?<br>o What are the characteristics of the programs (e.g., stipend levels)? |

APPENDIX D (Continued)

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM: SELECTED METHODOLOGICAL CHARACTERISTICS

| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
|---|---|---|---|---|---|
| National Research Council, 1977 | o PhDs who were NRSA awardees<br>o PhDs awarded during the same time period but who did not receive NRSA support<br>o Pre- and postdoc awardees of NRSA programs related to health services research | None | o Postdoctoral training<br>o Employment sector<br>o Research activity (time, spent in research, health-relatedness, and source of support)<br>o Time spent in other types of activities (e.g. teaching and service)<br>o Attitudes toward relevance of degree and research training | o Sample survey to PhD recipients | o What were the career outcomes of NRSA awardees in biomedical and behavioral sciences?<br>o What were the career outcomes for NRSA services research? |
| National Research Council, 1978 | o PhD granting departments | Responses were analyzed by institutional quality, control, type, department age, and existence of training grant. | o Changes in student population<br>o Changes in program activities<br>o Changes in quality of training activities<br>o Changes in departmental resources (e.g., student aid) | o Survey to PhD-granting departments<br>o Site visits | o What is the role of federal funds in graduate training?<br>o What is the impact of lost training support on departments?<br>o What is the state of the labor market? |
| National Research Council, 1981 | o Programs that had lost NIH support | Not applicable | o Changes in student population<br>o Changes in program activities<br>o Changes in quality of training | o Site visits | o What is the impact of lost training support on departments? |
| Garrison & Brown, 198[illegible] | o Institutions with MARC programs<br>o Those without MARC programs<br>o Career plans | | o Biological sciences degrees awarded by institutions<br>o Participant satisfaction<br>o MARC curriculum<br>o Career plans<br>o Career attainments | o Survey of former program participants<br>o Site visits to programs | o What are the characteristics of the Honors Undergraduate Research Training Program?<br>o What are the characteristics and outcomes of students?<br>o What is the program's impact on participating institutions, particularly in terms of improving science curricula? |

APPENDIX D (Continued)

EVALUATION STUDIES OF THE NATIONAL RESEARCH SERVICE AWARD PROGRAM: SELECTED METHODOLOGICAL CHARACTERISTICS

| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
|---|---|---|---|---|---|
| **EVALUATION ACTIVITIES FOR NIH** | | | | | |
| NIH, Office of Planning & Program Analysis, 1986 | For MDs:<br>o NIH postdoc Fellows<br>o NIH postdoc Trainees<br>o Unsuccessful applicants for NIH Fellowships<br>o Those who did not apply for NIH fellowships | None | o Medical school employment<br>o Pursuit of NIH/ADAMHA research grants | o Retrospective<br>o Use of existing data bases | o How did the number and quality of fellowship applicants change after instituting payback requirement?<br>o Did the NRSA program affect the number of persons in health research/teaching careers?<br>o Do NRSA recipients pursue health research/training careers? |
| | For PhDs:<br>o NIH postdoc Fellows<br>o NIH postdoc Trainees<br>o Those with "postdoc plans but no NIH postdoc appointments<br>o Those with "no plans" for postdocs | None | o Academic employment<br>o Time spent in research and teaching<br>o Pursuit of NIH/ADAMHA research grants | | o How effective are training grants in encouraging health research careers? |
| Velletri et al., 1985 | o NIH predoc Trainees<br>o NIH predoc Fellows | None | o Attainment of PhD<br>o Postdoctoral training<br>o Early career employment (academic positions and involvement in research)<br>o Pursuit of NIH and NSF research grants<br>o Publications and citations | o Retrospective<br>o Based on existing data bases<br>o Population data for some outcomes | o Are there differences in the outcomes of NIH trainees vs. NIH fellows?<br>o Do the results of these comparisons change over time? |

APPENDIX D

EVALUATION STUDIES OF NATIONAL RESEARCH SERVICE AWARD: METHODOLOGICAL CHARACTERISTICS (Continued)

| Study and Year | Major Characteristics: Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
|---|---|---|---|---|---|
| National Institute of Dental Research, 1937 | o NIDR postdocs Trainees<br>o NIDR postdoc Fellows<br><br>For NIDR Postdocs:<br>o DDS-PhD/DMD-PhD<br>o PhDs only<br>o DDS/DMD only | None | o Employment<br>o Pursuit of NIH and NIDR research grants<br><br>o Pursuit of NIDR research grants | o Retrospective<br>o Based on existing data bases<br>o Population data | o How successful are NIDR postdocs in pursuing research careers?<br>o Do the outcomes differ for Trainees vs. Fellows?<br>o What are the characteristics of awardees?<br>o What are the characteristics of the training sites for awardees?<br>o What is the relationship between type of degree and extent of postdoc training to obtaining subsequent NIDR research funding? |
| Sherman, 1984 | o Awardees of NIH short-term training funds<br>o Those who had not received NRSA short-term training | None | o Research experience in medical school<br>o Post-graduation employment and research training plans<br>o Career preferences and expectations<br>o Debt upon graduation<br>o Pursuit of NIH research grants | o Based on existing data bases | o How many short-term trainees have been supported?<br>o How many short-term trainees eventually applied for and/or received NIH grants?<br>o Do the research plans of former trainees differ from non-trainees? |
| **EVALUATION ACTIVITIES OF ADAMHA** | | | | | |
| Schneider, 1980 | o Predocs and postdocs who had received NRSA awards in psychology from NIMH | None | o Postdoctoral training<br>o Initial employment position | o Survey of training program directors | o What are the initial employment positions of former NRSA awardees from psychology programs? |

APPENDIX D

EVALUATION STUDIES OF NATIONAL RESEARCH SERVICE AWARD: METHODOLOGICAL CHARACTERISTICS (Continued)

| | Major Characteristics | | | | |
|---|---|---|---|---|---|
| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
| ADAMHA, 1986 | o ADAMHA predocs<br>o ADAMHA postdocs<br>o Those receiving both | None | o Receipt of ADAMHA/NIH research grants | o Retrospective<br>o Based on existing data bases | o How many individuals received training support?<br>o How many subsequently received ADAMHA/NIH grants?<br>o How many ADAMHA research grantees received training support? |
| Clouet, 1986 | o NIDA Trainees<br>o NIDA Fellows | None | o Career quality (composite) rating of promotion rate, employer prestige, and awards)<br>o Research quality (composite rating of papers, awards and presentation)<br>o Type of employer | o Survey of awardees<br>o Existing trainee files | o How successful are NIDA awardees in pursuing biomedical research careers? |
| **EVALUATION ACTIVITIES BY OTHER SPONSORS** | | | | | |
| Sherman et al., 1981 | o MSTP graduates<br>o NIH Intramural Clinical Associates<br>o NIH Intramural Research Associates<br>o NIH Extramural Postdocs | Comparison groups were matched matched on sex, age, type of medical school, period of training, and MCAT scores. | o Employment positions<br>o Pursuit of NIH research grants<br>o Publications | o Retrospective<br>o Based on existing data bases<br>o Survey to comparison group members | o How successful were graduates in pursuing medical research/teaching careers? |

APPENDIX D

EVALUATION STUDIES OF NATIONAL RESEARCH SERVICE AWARD: METHODOLOGICAL CHARACTERISTICS (Continued)

| | Major Characteristics | | | | |
|---|---|---|---|---|---|
| Study and Year | Study Groups | Consideration of Pre-Award Differences | Types of Outcomes Studies | Type of Study | Major Questions |
| Bickel, et al, 1981 | o Those with MST support<br>o Those in other MD-PhD programs | None | o Attrition from MD-PhD Programs<br>o Applications to MST program | o Survey of medical deans | o Were those in MST more likely than those in other MD-PhD programs to complete the program?<br>o What are the characteristics of applicants to MST programs?<br>o How does the number of applications MST programs comp with the number of available training slots? |

PRODUCTIVITY

*Helen Hofer Gee**

## INTRODUCTION

In 1986 NIH contracted with the Institute of Medicine to organize a conference on research training. A central, though not explicitly stated, purpose of the conference was to obtain guidance on how to continue to meet Congressionally mandated requirements for periodic reports on the role of and need for research training in the biomedical and behavioral sciences. Ostensibly, the conference was concerned with

> an examination of how successfully research training has been conducted, which program mechanisms produce the most suitable training, and what information is required to enable further assessments of national needs for researchers in the decade ahead. [Listed first among three categories of issues requiring attention was] . . . measuring the productivity of scientists in their research programs and as reflections of their training. The issue in productivity is how to improve the measurement of it; simply gauging productivity by the current popular methods is inadequate for the task at hand. (Institute of Medicine, 1986)

Anyone who has ever been faced with the task of having to select among individuals--for employment, advancement, funding, awards--has dealt with the issue of assessing productivity and has, implicitly or explicitly, weighed available evidence of previous performance. The difficulty and complexity of these decisions may well underlie the malaise that is apparent in the committee report. The committee's more explicitly reported concerns with such measures as success in obtaining research grants, citation counts that ignore differences among and possibly within disciplines, and studies that fail to consider work environments suggest that the real problem lies not in the measures of productivity per se that have been used, but in how the measures have been used--that is, in the designs of assessments of training support programs.

---

* The opinions expressed in this paper are the author's and do not necessarily reflect those of either the Committee on Biomedical and Behavioral Research Personnel or the National Research Council.

Unfortunately for those who seek quick solutions, concepts relevant to the measurement of productivity are inextricable from those concerning almost all other domains within the social study of science. Dealing directly with the problems of productivity measurement therefore requires cognizance of the state of the entire science. Any study, for example, that ignores differences among or within disciplines ignores more than three decades of intensive study of the entire social structure of science, not just the study of "productivity" per se.

> In the critical, scholarly essay, Gilbert (1978) noted that there is a reciprocal relationship between the theoretical framework which the social scientist brings to his work and the indicators which he will find most appropriate for his research . . . the adequacy of an indicator can only be assessed through a detailed study of the context in which the phenomena to be measured are embedded, and of the validity of the measurement theory on which it relies . . . this requirement is equivalent to the demand that we understand the functioning of the scientific community at a micro-level.

A small community of scientists (static in size in the United States since about 1980, but rapidly increasing in Western and Eastern Europe and Japan since the late 1970s) has been making significant progress in the direction Gilbert suggests (see Appendix and References). The most recent burst of research activity relevant to the assessment of productivity began when Martin and Irvine (1983) assessed basic research activity and programs (radio astronomy and physical sciences). Their papers specified "partial indicators" of scientific progress and investigated the extent to which these indicators "converge" to produce valid and reliable estimates of the productivity of designated groups of scientists. The work created a virtual storm of criticism, largely because it was so far-reaching (see Chubin, 1987). The continuing discussion has instilled new vigor into the development of the field.

The concept of multiple partial criteria was certainly not introduced by Irvine and Martin. Even Clark's study of the careers of psychologists (1957) incorporated the concept in a general sense. As noted by Jones et al. (1982), Weiss (1972) discussed them:

> At best each is a partial measure encompassing a fraction of the large concept . . . . Moreover, each measure [may contain] a load of irrelevant superfluities, "extra baggage" unrelated to the outcomes under study. By the use of a number of such measures, each contributing a different facet of information, we can limit the effect of irrelevancies

and develop a more rounded and truer picture of program outcomes.

However, as Chubin (1987) concludes in his discussion of Irvine and Martin's work, it is

> . . . also politically astute, serving scholarly and policy communities. It explicitly anticipates criticism and sources of error, disarms skeptics, and gets an analytical foot in the right doors--those shielding the offices of policymakers who have come to rely on participant scientists and their own imprecise and self-serving devices for making decisions about who gets and who doesn't.

Moravcsik (1986) hailed the extensive debate and critiques of the Martin and Irvine work as a welcome sign--

> since it shows that the field has reached the state of maturity when its applications concrete situations are sufficiently realistic to create a heated controversy, involving people from a variety of professional backgrounds.

He commented further that neither critics nor Irvine and Martin, in their response to critics, offered any specific suggestions for improvement. Moravcsik then proposed that some suggestions can be made and conclusions drawn concerning the need for future activities by relating the debate to another effort in science assessment--namely, a project organized by the United Nations Center for Science and Technology for Development (UNCSTD) and centered on a paper that Moravcsik was commissioned to write. Moravcsik reported further that at a meeting held in Graz, Austria, in May 1984 the paper was discussed:

> The UNCSTD project did not result in a set recipe for assessing science and technology. On the contrary, the project concluded that there is no such universal recipe, and hence that the aim should be to devise a process which, in any particular case, yields a methodology for an assessment.

## SUGGESTED OUTLINE FOR PLANNING STUDIES OF PRODUCTIVITY OR QUALITY

The proposed UNCSD process serves as a useful framework in which to present some thoughts about planning studies focusing on the assessment of productivity. The following list draws heavily and directly on Moravcsik's report:

1. <u>Identify the goals of science that are to be taken into account.</u> Moravcsik noted,

> . . . Science and technology have many different goals, aim and justifications, and in any particular case it must be specified which one (or which ones) of these are taken into account, and with what weight.

Studies of National Institutes of Health research training programs have ostensibly aimed at assessing a common goal of such programs--to wit, the production of trained scientists who will contribute to the advancement of the biomedical sciences. Prior to the mid 1970s, this was interpreted by some Institutes as including support for the clinical training of physicians in areas where the supply of expertise was felt to be inadequate. After it became clear that the majority of these individuals simply entered private practice, however, those programs were for the most part discontinued. Such discrepancies must, therefore, be given careful attention in planning studies of program outcomes. Since the mid '70s, NIH training programs in general have focused specifically and exclusively on research training. Assessment of the success of these research training programs have, however, interpreted the terms "contribute" and "advancement" quite narrowly. Teaching (either future researchers or practitioners), biomedical research administration, mentoring (i.e., guiding the graduate education of future researchers), and conducting research that does not seek external funding and research that cannot (because of the interests or concerns of the power structure within which it is conducted) be published have often been denied recognition as goal-relevant behavior. Consideration should be given to whether any of these professional activities should be explicitly recognized as contributing to the advancement of the biomedical sciences and, if so, studies should be designed to assess these kinds of productive endeavor.

2. <u>Recognize the multidimensionality of goals, of potential pathways to them, and of methods of measuring outcomes; specify which dimensions and connections of the system are to be taken into account.</u>

Once goals have been specified (and it is recognized that achieving those goals can and is likely to be expressed in different ways), study designs must allow for the varieties of pathways and outcomes that may occur. Cole and Cole (1973) set the stage for this type of inquiry in their cross-sectional analyses. The work of Long, McGinnis, and Allison (1979, 1981, 1982) examined many of the same "connections" as the Coles but, by following a cohort longitudinally, revealed a different sequence of career development. The Long and McGinnis work has been particularly notable in its pursuit of the significance of context, the multidimensionality of career pathways, and the changing significance of predictors in assessing productivity at stages in research career development. In another notable analysis of the NIH Research Career Development Program, Carter et al. (1987) examined both selection processes and outcomes using multivariate techniques to assess the significance of correlates and causal relations, as well as a sophisticated cohort selection procedure to control for disciplinary differences.

3. <u>If, as is usually the case, it is not feasible to study all aspects of a system, specify which aspects are to be included and which will be omitted and indicate clearly the implications of these decisions for the assessment process.</u>

Moravscsik provided an apposite illustration of one perspective:

> If, of two cars, one has a higher top speed, and the other a lower gasoline consumption per mile, it is not possible to say which is the 'better' car without ascribing some value judgment to high speeds versus economy in the use of fuel.

Two other examples come to mind: (1) if in planning a study of the effectiveness of a training program, it was decided that pursuit of a research career in the private sector was a favorable outcome but that assessing the performance of former trainees who followed that path was not feasible, they could be explicitly excluded from potential comparison cohorts; (2) if research administration is deemed a favorable outcome, those research administrators could be excluded from comparisons in which research publications were used as indicators and included where other measures of productivity, more suitable to their employment, were used. The guideline simply demands precise specification of the details of the design of an investigation.

4. <u>Specify how the results of the assessment are to be used.</u>

A study intended to assist program managers in their decisionmaking will seldom have the same design requirements as a study intended to inform policy decisions. If policy decisionmakers are to be informed, for example, the delineation of possible alternative indicators of productivity may be critical, whereas meeting program management needs may require more intensive analysis of only those that are the most direct manifestations of program goals. The key is to consider carefully the kinds of decisions that the study is intended to influence.

5. <u>Select a set of indicators that will satisfy the requirements of each of the study design considerations. Recognize and specify the limitations of each of the indicators.</u> To quote Moravscik,

> There are many types of indicators: input versus output; quantitative versus qualitative, indicators of activity, productivity, or progress; indicators of quality, importance or impact; there are functional and instrumental indicators; there are micro- and macro-indicators; there are "data-bases" and "perceptual" indicators; and so on. Some indicators are already 'on the shelf' and can be taken from it and used in new situations. More likely, however, the most appropriate indicators for a new situation need to be improvised for that particular situation. . . . Be reconciled to the fact that in any case, you will end up with a set of indicator measurements which, in general, cannot be reduced to a one-dimensional measure and hence to an unambiguous ranking.

It is apparent that the selection and/or development of indicators of productivity depend on the kinds of questions that are being asked and the perceived complexity of the system involved. An indicator that provides excellent explanatory data for one study may be useless in another context. Every measure, moreover, has limitations that may, under some conditions, obviate their utility and, in other circumstances, may be totally irrelevant. If a study plan is suitably mapped, it may not be feasible to use the same indicators of productivity for all individuals in a cohort. For example, if teaching undergraduate students is judged to be an acceptable outcome of research training, the productivity of an individual whose primary activity is teaching will not be appropriately assessed by counting that individual's production of research papers--but consideration might be given to using the production of review papers as one of several measures of performance in the educational domain. However, for some outcomes regarded as suitable expressions of the goals of an enterprise, no suitable approach to assessment "measurement" is available to evaluators. In such cases the individuals should be removed from comparison groups that are to be analyzed statistically rather than, as is often the case, counted a "failure" according to indicators that appropriately measure the productivity of other members of the group.

## MEASURES OF PRODUCTIVITY

The above overview should make it clear that any discussion of specific measures of productivity is necessarily superficial, simplistic, and incomplete because outside the context of the design of a specific study, there is not a great deal to be said about any particular measure. In addition, since productivity in one sense or another is the focus of most of the studies of the social science of science, a thorough literature review would require a fe years of effort. Nonetheless, various measures that might be used in studies of productivity are discussed below. The discussion is intended to draw attention to complexities, issues, and problems in the use of these measures, knowledge of which might aid in carrying out the kind of careful approach to study design outlined earlier.

**Publication Counts**

While it is generally agreed that the principal, or most prevalent, immediate outcome of the active research investigator's efforts is the preparation of papers published in professional journals and by 1982 a nearly 2600-item bibliography listing publications analysis items was available (Hjerppe, 1982), counts of publications continue to be derogated. Because the analysis of publications plays a dominant role in social studies of science, a complex, highly sophisticated methodology has been developed. The intellectual leader of modern-day social studies of science was Derek J. deSolla Price (1961, 1963). The early development of computer-based analytic methods, which have stimulated much of the sophisticated analysis characteristic of social analysis of science studies of the past two decades, resulted largely from the enterprise of two individuals: Eugene Garfield (1955) developed the *Science Citation Index* (SCI), on which most publications analysis work is dependent; Francis Narin and his associates at Computer Horizons took the lead in exploring and developing measures to maximize the utility of the wealth of information contained in the SCI. In 1969 (Narin, 1977) the area even acquired its own label--bibliometrics--to describe collectively quantitative, analytical studies of written communication.

In simplest terms, publication counts are no longer acceptable as a measure of productivity unless at least the following potential sources of error or misinterpretation are controlled or accounted for:

- differences among disciplines of cohort members,
- differences among journals in terms of measured influence (see section on journals page 131),
- differences in "quality" or "impact" as measured by citations or peer assessment (or journal influence),
- professional age of cohort members, and

- social context of cohort members.

Despite concerns about "loud noises from empty vessels," publication counts have been shown repeatedly to correlate positively with assessments of quality and to contribute useful independent variance to the assessment of productivity. Reported correlations between quantity and quality measures vary considerably among studies, between approximately r =.23 and r =.80; differences may relate to disciplines, characteristics of cohorts, or even to how quantity and quality of publications are measured.

In a series of studies conducted in the late 1970s (see Narin, 1983), numbers of publications by faculty and staff in universities and hospitals were shown to be extremely highly correlated with NIH funding (r = .90 to 95); and there were no economies or diseconomies of scale in the funding of research grants. Funding and publication relationships may appear to break down, however, when small aggregates of researchers or disciplines are assessed and especially when basic and clinical research publications are intermixed. Publication rates of basic scientists differ markedly from those of clinical scientists, who publish less frequently and whose research is usually very much more costly. When the funding and publication rates of small aggregates of subjects are investigated, the tendency is to ignore such disciplinary differences, thus ignoring an important moderating variable. With small aggregates other minor sources of error--such as idiosyncratic events that may affect the usual patterns of behavior of part of a group for a period of time--may also obscure an underlying relationship. When large aggregates and adequate time spans are employed, such obfuscating sources of error will usually cancel each other out, permitting stable, underlying relations to be revealed.

When a quick, inexpensive estimate of productivity is needed, large quantities of data are available, and the comparability of cohorts is established, a simple count of publications may well provide adequate information. Ordinarily, however, such a single measure is useful primarily as a means of setting the stage for a more comprehensive investigation of some aspect of science or scientific behavior.

**Weighted Counts:** The use of weighted counts of papers permits obtaining a preliminary estimate of quality without waiting for citations to become available; it is also an inexpensive means of obtaining an estimate of quality for large numbers of papers.

Each paper is weighted by an influence weight assigned to the journal in which the paper appears.[1]

**Paper Counts in the "Best" Journals:** Committees charged with evaluating group or individual scientific performance will sometimes request that publications be counted only in a selection of the "best" journals. Such a practice would be seriously inequitable, since scientists do not have equal access to journals. For example, those located in smaller institutions are more often published in less influential journals, as are younger, less well-established investigators; and regional differences abound in some disciplines. McAllister and Narin (1983) investigated these relationships in the publications of all U.S. medical schools, using average citation influence per paper measures: the average citation influence per paper increased with the total number of biomedical publications, even when institutional control (public and private), region, and areas of research emphasis were controlled. The positive relation between number of papers and citation influence was shown to hold within disciplines (biochemistry and internal medicine were analyzed in detail) and within research "level" (i.e., along basic and clinical research dimensions).

**Data Bases:** NIH-supported studies that have involved counts of published scientific papers have almost always depended on computerized data bases derived by CHI from Medline and the SCI. The source data bases require a great deal of preliminary massaging to consolidate information and correct inconsistencies; but once prepared, they make data available unobtrusively, make accessible several different quantitative measures of publication performance, avoid the increasingly restrictive problem of securing clearance from the Office of Management and Budget (involved, in studies of federal programs, in any attempt to go directly to the scientific community for information), and are more accurate than individual reports. An interesting departure from the use exclusively of the comprehensive data base was reported by V. L. Simeon et al. (1986), who had studied a large research institution in Yugoslavia. In their study several forms of publication and communication were employed in addition to SCI journals (e.g., papers in other scientific journals and congress proceedings, books and monographs, technical articles in

---

[1]The technique developed by Computer Horizons, Inc. (CHI), determines journal influence weights by the weighted number of citations each journal receives over a given period of time. See F. Narin, *Evaluative Bibliometrics: The se of Publication and Citation Analysis in the Evaluation of Scientific Activity* (Report to the National Science Foundation), 1976; and F. Narin, G. Pinski, and H. H. Gee, "Structure of the Biomedical Literature," *Journal of the American Society for Information Science* 27:25-45, 1976.

encyclopediae and popularizations, and presentations at scientific meetings). A multivariate analysis revealed interesting patterns of change among the several variables over time. This rather preliminary study, which was focused on change in publication behavior following the introduction of minimal criteria for promotion warranted no conclusions; but it suggested to this writer the possibility that some measures of these types might be useful in considering criteria suitable for assessing the productivity of individuals whose careers, though academic, are not directly focused on the production of original research.

**Activity Indexes:** In recent years the utility of a new approach to using publication counts, the "activity index," has been demonstrated, particularly in studies conducted by CHI for NIH. Activity indexes are ratios that make use of publication counts in a relational context, thus allowing comparisons to be made among groups while allowing each group to be described within its own context.[2] Describing NIH Institutes' relative investment in the support of research in different disciplines is a case in point (see Gee and Narin, 1986). Journal papers are more readily and accurately assigned to disciplines than are dollars, and a ratio that describes an Institute's investment in a discipline relative to both the Institute's investment in all disciplines and the "size" of the discipline among all others in a data set provides a great deal of information for comparison among [illegible] disciplines and Institutes. Schubert and Braun (1986) suggest several additional types of indexes that might be useful for different purposes.

**CITATIONS**

Ever since Clark's study of psychologists (1957), citation counts have been a favored measure for the assessment of productivity. In most cases, citations alone or in combination with publication counts are more closely correlated with subjective estimates of productivity than are any other measures. They are more universally applicable to the assessment of scientific research activity than are other measures because (1) publication is the most accessible means of expression available to all scientists, and (2) being published offers a broader audience to the scientist than any other medium.

---

[2]The percent of an organization's papers that are published in a given discipline is divided by the percent of all papers in the data set that represent that discipline. An index of "1.0" indicates that the level of publication activity of this group in this discipline is consonant with the level that discipline represents among all disciplines.

Rather than referring to citations as measures of "quality," as was common in the 1970s, the current practice is to refer to them as measures of "impact" or "utilization" or "influence." The implication is that before citations can be referred to as measures of the "quality" of research, the issue should be investigated in the given context of definition.

From an entirely different perspective, Moravcsik (1986), Chubin (1987), Cronin (1984), Vinkler (1987), and others have discussed and/or analyzed the functions and meaning of citations in terms of author motivation. Vinkler, whose contribution is most recent, has provided a concise review of the literature concerning definitions, classification, and roles that citations play in the scientific literature, concluding (in concert with Cronin) that the <u>information carrier</u> role is the most important. Vinkler distinguishes between "professional" (work is based on the cited work or uses part of it) and "connectional" (e.g., desire of an author to establish a connection with the cited author or work) reasons for citation. In Vinkler's study, a group of productive investigators rated each of the references they had listed in a selected recent paper, identifying which of eight professional and/or nine connectional reasons had motivated the decision to cite, and the strength of the motive. Most (81 percent) citations were made solely for professional reasons--that is, in a literature review for "completeness" or because the current work was based at least in part on the cited work, the cited work confirms or supports the work in the citing paper, or the cited work is criticized or refuted (at one of three levels). Citations made partially for professional nd partially for connectional reasons accounted for 17 percent; only two percent were made sol ly for connectional reasons. It was also found that two to three times as many papers are reviewed as are actually cited. Failure to cite was also investigated; the principal reason found was that a work was not considered important enough to the current effort to warrant citation. Second most important was the "obliteration" phenomenon--the origin so well known that citation was not needed. A citation threshold model has been developed, and data confirm that the threshold depends primarily on the professional relevance of the work potentially citable in a given paper.

Narin (1976), considered citations as an assessment mechanism:

> Citation counts may be used directly as a measure of the utilization or influence of a single publication or of all the publications of an individual, a grant, contract, department, university, funding agency or country. Citation counts may be used to link individuals, institutions, and programs, since they show how one publication relates to another. . . In addition to these evaluative uses, citations also have

important bibliometric uses, since the references from one paper to another define the structure of the scientific literature.

Narin has presented both literature reviews and extensive analytical evidence that substantiate the utility of citations in the assessment of productivity and provide a most valuable guide to technical aspects of their use (e.g., with respect to the consideration of such issues as time, differences among fields and disciplines, and the use of indexes to make possible cohort comparisons).

In 1982 Narin and McAllister prepared for NSF a complete set of counts of all U.S. papers listed in the SCI for the years 1973 through 1976 and distributions of the citation counts received in the first through the fourth years after publication (mean, median, and mode) by each of the 106 subfields listed in NSF's *Science Indicators*. The average (mean) number of citations received in the first four years by papers in biomedical subfields ranged from 1.8 for "Miscellaneous Clinical Medicine" papers to 15 for "Biochemistry and Molecular Biology" and "General Biomedical Research" (publications in journals such as *Nature, Science, Federation Proceedings)*. The data illustrate dramatically the importance of taking disciplinary differences into account.

As with publication counts, the day has long since passed that simple citation counts would be regarded as acceptable measures of performance. Even average numbers of citations per paper are useful measures only when all of the precautions cited for paper counts are observed--that is, controls are exercised for sources of difference, such as discipline and time (both publication date and citation count period), that are not part of the question addressed. A longitudinal study of NIH-supported research 1973-1980 (Gee and Narin, 1986) employed citation counts per paper that had been scaled in relation to time and transformed into a standardized score in relation to all papers in a given research level, field and subfield. In a more recent study, Gee (1988) employed counts of papers in which each paper was weighted by the "influence weight" of the journal in which it appeared. For early examples of carefully considered treatment of citation data, see Carter's 1974 and 1978 reports on the NIH Peer Review system and on comparison of program project with individual (RO-1) grants.

**Highly Cited Papers:** The percentage of papers produced by a cohort whose members' papers are among the most highly cited 1, 5, or 10 percent of all papers in a discipline or specialty has particular appeal as an assessment of quality. With a computer the data are easily obtained, and the technique admits the inclusion of an outstanding paper that has been published in a less touted journal, is not directly influenced by author institution, is free of the possible biases of peer ratings, and provides an appropriate means of comparing across many different group dimensions. It is disadvantageous only for the cohort so large that it virtually defines the distribution of citations received by a group of journals. (For example, papers supported by the National Cancer Institute and the citations that they receive virtually define the distribution of citations to "cancer" journals). Comparisons between citation averages and percent of papers among the most highly cited 10 percent offer added insight into the distribution of performance within individual cohorts.

Although the overwhelming majority of studies that have compared subjective ratings with citation counts have yielded strong positive correlations, not all have. Where the correspondence is weak, the data often serve to reveal characteristics of the peer judgments rather than indicating deficiency in the citation evidence (see, for example, Anderson et al., 1978).

**JOURNALS AND SUBFIELDS**

Some of the dimensions of difference among journals and subfields are described in Narin (1976): using the 1973 SCI, tables were developed showing the distributions of differences in types of publications (articles, notes, letters, etc.), numbers of journals, publications, references, citations, and ratios of these counts. Dimensions of differences drawn from these tables include the following:

- Discipline (or subfield) differences in numbers of references, citations and publications: In the 1973 SCI (articles, notes and reviews), citations per paper ranged from 2.4 for operations research and management science to 36.2 for physiology. In general, there are distinctive differences among fields and subfields: for example, mathematical scientists use few references, receive small numbers of citations, and produce relatively few papers while many basic biomedical scientists publish frequently and receive large numbers of citations.

- Variation in rates of growth and time distributions of citations: Rapidly growing subfields of science have higher fractions of references to recent papers than do slow-growing subfields. Rapidly growing subfields also tend to receive their modal numbers of citations earlier than do other subfields (e.g., in the second, instead of third, year after publication).

- Concentration and relative citedness: The concentration of publications and relative citedness varies both within and between fields of science. Some fields are characterized by relatively few, very large and influential journals while the literature of others is widely dispersed.

This kind of information was not widely discussed in the open literature until Moed et al. (1985) at the University of Leiden published their investigations that demonstrated the serious impact, especially on analyses or assessments within universities, of neglecting these characteristic differences among disciplines. The Moed study also revealed that operating on incomplete bibliometric data can have a serious impact on the outcome of an assessment. The characteristic differences among fields, subfields, and specialties of science have led to the development of a growing number of indexes that are aimed at making comparisons between cohorts justifiable. The most widely used are impact factors, influence measures, and relative citation rate and publication impact:

**Impact Factors:** The Institute for Scientific Information, publisher of the *Science Citation Index*, also publishes *Journal Impact Factors*, based on a 2-year accumulation of citations received divided by the number of papers published in the target year. These measures, while correcting for journal size, do not correct for characteristic differences in referencing and citation practice and, therefore, reflect different dimensions of citation behavior in different disciplines. Noma (1986) states:

> . . . there is no normalization for the different referencing characteristics of different segments of the literature: a citation received by a biochemistry journal, in a field noted for its large numbers of references and short citation times, may be quite different in value from a citation in astronomy, where the overall citation density is much lower and the citation time lag much longer. In addition, journals that publish longer papers, such as review journals, tend to have the highest impact factors.

Schubert and Glanzel (1983) have developed a method of estimating the reliability of mean citation rates per publication, computing a standard error based on the relative frequency of zero citations. They exemplify its use by computing a sample set of "corrected" ISI impact factors. They are, however, still left with the lack of comparability in other aspects of differences among disciplines.

**Influence Measures:** Pinski (1975), Anderson et al. (1978), and Noma (1986) developed three citation-based journal influence measures that offer greater breadth and precision of measurement than does the impact factor for capturing the information resident in the journals in which research is published. The three measures are:

- Influence weight, which is a size-independent measure of the weighted number of citations a journal received from other journals, normalized by the referencing practices of the field
- Influence per paper, which is essentially the weighted number of times an average paper in a journal is cited, where the weight is based on the influence of the citing journal; and
- Total influence (of a journal), which is the influence per publication times the total number of papers published over a given period of time.

Of these three measures, the influence per paper measure has proved most useful in studies of group publication performance. Total influence scores are, however, the scores most highly correlated with subjective judgmental ratings of university program quality (see Anderson et al., 1978). The influence measures offer a clear advantage over impact factors: the measures are determined within each of the fields of science, thus correcting for differences in citation practices and providing comparability across fields of science. In addition, citations from prestigious journals are weighted more than those from peripheral journals, thus introducing a quality concept, and the three different measures provide information from three different perspectives. Despite their superiority, influence measures have not been widely adopted. The iterative matrix manipulations involved are costly and cumbersome; as a result, it is not economical to revise them frequently, and only one revision--using 1982 publications--has been made of the original measures based on 1973 publications (see Noma, 1986). The number of journals in the 1982 set increased to over 3,000 from the 2,300 on which the 1973 measures were based. Changes in computational techniques were also made, so average measures are not directly comparable. Even so, Narin (1985) reports, that most correlations within fields ranged between r = .85 and r = .95; highly influential journals retained their high influence, and journals with relatively low influence ratings (e.g., r = .40 to .[illegible]0), tended to drift within that range. Depending, of course, on the aims of a study, it would appear that in many cases comparability of measures among disciplines should take precedence over recency of citation count.

**Relative Citation Rate and Publication Impact:** Schubert et al. (1983, 1986) and Vinkler (1986) have proposed two new "relative" citation indicators based on Garfield's SCI journal impact factors. Schubert's "relative citation rate" (RCR) compares actual citation counts with an expected citation rate based on the impact factors of the journals in which a set of papers has appeared, thus eliminatin characteristic differences in the publication and citation practices among disciplines. Schubert and Braun (1986) and Schubert, Glanzel, and Braun (1983) have used these indicators in developing relational charts on which the relative merit of nations' publications are compared.

Vinkler developed a "relative publication impact" ($P_I$) measure--which includes with a relative citedness indicator, numbers of publications and a "cooperation" measure based on coauthorships--to arrive at an index that includes both quantity and quality in its ratings (of departments, in this instance). Vinkler reports "good agreement" with subjective peer evaluations that are carried out regularly. None of these measures have yet been subject to critical review, nor have they been used outside the context in which they were developed. Further exploration does appear to be warranted.

**PEER ASSESSMENT**

Peer assessment, as the concept was developed by NIH to overcome the potential biases of individuals in decisionmaking situations, is now in the anomalous circumstance of being defended against intrusion by some biased individuals who oppose statistically based measurement techniques while at the same time being attacked by apparently politically motivated individuals who accuse peer assessment in the management of science of the very biases it was designed to overcome. Compounding the anomaly is the fact that the defenders fail or refuse to recognize that the quantitative analysis "intrusion" that they reject can and has provided, by far, the most convincing evidence available to prove the case for peer review. Longitudinal studies of NIH-supported publications (see Gee and Narin, 1986), for example, are unequivocal in their demonstration of the effectiveness of the NIH dual peer-review system of determining which grants shall be funded.

It is also something of an anomaly that we treat peer assessment as one among several different types of criteria that might be used to assess productivity, when in fact, almost all likely criteria are, at bottom, different representations of peer judgment. For the most part the different measures represent collections of judgments that are separated in time, in focus, and in method of combination. When a peer group is assembled to assess the productivity of an individual, a group, or a program, the outcome is based on the combined perceptions of any assortment of behaviors or perspectives its individual members may implicitly or explicitly agree upon at the time the judgment is delivered. Other measures, such as those derived from counts of citations to an accumulation of publications, represent a long series of separate peer and peer-group assessments ranging from acceptance to college and graduate schools, through awards of degrees, positions, funding, selection for publication and decisions to cite. The difference is that each of these assessments has been specifically focused on a related concept of merit, and the outcome statistics derive from the judgment of many more and more disparate peers. There is no question of which is the "better" type of measure; they represent different perspectives--sometimes only slightly and sometimes widely different. The quality and utility of either depends on the care taken to secure accuracy and to eliminate inappropriate or unfair considerations from the outcome and on the context in which the results are used. The significant point is not the superiority of one or the other measure but, rather, the extent to which either or both illuminate the questions at hand.

Those who are concerned with achieving a better understanding of how science functions seldom argue the point that peer review groups that can take into consideration not only an immediate product or situation, but also any extenuating circumstances that might alter the significance of any given piece of information are, as yet, better equipped to take into account all relevant factors in making a judgment about, for example, an individual grant application than any likely collection of statistics. (This is not to say that there do not exist experimental techniques that might well improve peer judgment procedures. It ignores also the issues of personal and group bias, which may seriously distort judgment). When very large groups of individuals or products--such as programs or large sets of publications or whole journals--are at issue, however, and combinations of statistical measures correspond imperfectly with peer judgments, it may now, after three decades of very active research, be prudent to conclude that here are two different kinds of evidence, each of which potentially offers useful and valuable information, and each of which should figure in the assessment process (assuming, of course, equally careful data gathering and handling).

**Citations and Peer Assessment:** Most studies that have involved both peer judgment and bibliometrics have been aimed at validating the utility of the bibliometric measures. The Wellspring Clark study (1957), however, which was conducted before the analysis of publications became an object of social scientific interest, simply noted that the multivariate combination of journal citations and offices held in a professional association together accounted for nearly 64 percent of the variance in numbers of votes received when active investigators were asked to identify significant contributors to the field. The single most important predictor was citations, which correlated r = .67 with the number of votes.

Carter's report (1974) was the first serious analytical investigation of the NIH research grant peer review process. Publication and citation data, dating from the late 1960s, were inadequate by present-day standards, but the kinds of considerations Carter brought to the problem still warrant the attention of anyone proposing to investigate these kinds of relations. Carter found a low correlation (r = .40) between initial and renewal priority scores, and that

> . . . at least for grant applications from most of the larger basic science and clinical departments of medical schools, the judgments of the peer review process are significantly related to an objective measure of research output derived from citations to articles describing the results of the grant.

Two reports issued by the National Academy of Sciences in 1982 represent nearly opposite perspectives on the use of peer judgment and objective measures in assessment. One, in which opinions but no data or objective evidence of any kind were presented, concluded, essentially, that peer review was the only mechanism needed to assess quality or productivity in scientific research (COSEPUP, 1982). The other, in sharply contrasting peer group performance, applied 16 "measures"--4 based on peer ratings and 12 on records of program composition, support, and faculty publication performance--applied to 32 disciplines in 200 doctoral degree-granting institutions.[3] In the four biological science areas of primary concern to biomedical research, total journal influence ratings of faculty publications accounted for 50-70 percent of the variation in subjective judgmental ratings of faculty scholarly quality and 40-60 percent of program educational effectiveness. Notably, no attempt was made to combine the different types of information; rather, each of the items was reported for each institution.

A recent analysis by Lawani and Bayer (1986) of relations between peer and bibliometric assessment of quality is of interest because it compared peer and bibliometric assessments of cancer research papers. Papers abstracted in the *Yearbook of Cancer*, a selection made by large numbers of "peers," were classified as of high quality and compared with (1) papers listed

---

[3]The Committee on an Assessment of Quality-Related Characteristics of Research-Doctorate Programs in the United States began its work in response to growing criticism within the academic and educational communities of existing subjective ratings of graduate programs in the United States. The reports of this study serve as models of planning and reporting in the development and application of program assessment methods. Evidence is presented at every stage of development of the study to support decisions and analytic methods.

but not abstracted and (2) a restricted random sample of papers listed in *Biological Abstracts*. While citation frequency increased significantly with peer rating, there were discrepancies in the distributions. Of the most highly cited 100 papers, 14 were from the random set; also, 16.8 percent of the highly rated papers received 4 or less citations in the five years following publication (2.3 percent received none). Whether some of these will turn out to be "late bloomers" or will represent poor or biased choices for inclusion in the yearbook abstracts is not known. Lawani also found that quantity and quality were highly correlated and that the larger the number of coauthors, the larger the proportion of papers included in the yearbook. Also, self-citations relative to total cites declined with increased quality but did not affect the level of agreement between peer assessment and citation count.

Porter, Chubin and Xiao-Yin-Jin (1986) recently compared Sloan Fellows' "most cited" papers with their own selection of which of their papers were their "best." For 1974 fellows, 35 percent of papers they perceived as "best" were also most highly cited; this percentage rose to 42 for 1984 fellows. Eighty-nine percent of papers were coauthored, but Fellows were much more likely than citations were to select as best those on which they were first author. The letters that had recommended the Fellows for appointment, and the citations to their publications tended to emphasize methodological contributions; fellows themselves tended to identify their theoretical and empirical papers as their best.

Research evidence suggests that, when possible, both peer and bibliometric data should be made available for consideration. If only one measure can be obtained, in any study involving large numbers of subjects and publications (e.g., >100 papers per subgroup), the investigator would be at least as well served with publication data as with peer judgments.

## GRANTS AND GRANT APPLICATIONS

NIH and the Veterans' Administration are probably the only two federally supported agencies that maintain data bases suitable for the analyses of scientist's grant application and award behavior. It has been possible in recent years to obtain some information about individuals to whom the National Science Foundation and several private foundations have awarded funds, but no information about applications has been available. The availability of accurate and complete information about grant applications and awards makes possible the investigation of many management and policy issues. Clearly also, the receipt of an award is itself an indicator of achievement, especially in recent years when only small percentages of grants are actually awarded.

The availability of longitudinal data on those who have become involved with NIH programs is unique.

As sources of information about NIH programs and policies, the grant information, in conjunction with the publication data available in NIH data bases, constitutes a treasury of resources for the social and cognitive study of science that is probably unmatched elsewhere in the world. It is therefore extremely unfortunate that little or no opportunity exists for exploiting these resources in the interest of further developing the theory and methodology needed to advance our understanding of how science functions.

One of the least valuable ends to which grant information can be put is as a single measure of the effectiveness of training support programs. The positive information that is yielded is, of course, directly informative, but when applications and approvals are compared with awards (with no control for disciplines or for type or location of appointment), the results can be misleading. Failure to apply for grants has been interpreted as a negative outcome, whereas non-application may be totally irrelevant to the productivity of both individuals who are pursuing research and those who are performing other services to science that can be regarded as successful. Some creative new attempts may be in order to design studies that will permit alternate patterns of successful outcome for such individuals as those engaged in research that is otherwise supported, as well as those whose administrative or teaching responsibilities preclude their applying for grants. This is not to say that grant applications and awards are not an important source of information about, for example, the success of training programs. It only cautions against its use exclusively, and without consideration of such limitations as disciplinary differences and the availability of funding.

## ACADEMIC RANK, RATE OF ADVANCEMENT, SALARY

These three measures are possible alternative measures of one aspect of advancement or productivity for individuals in different settings or with different types of appointments. Each must usually be qualified in terms of years since doctorate or since completion of training, and it should be possible to scale each of them so that comparability among individual members of a cohort would be achievable. Data about rank and rate of advancement are generally more readily available than salary information: private universities often refuse to release salary information, but will yield rank; industry, on the other hand, which often balks at releasing information about its investment in research and development, is usually less unwilling to report the salary of an individual employee. Problems with these measures are likely to be related to institutional size, policy, and prestige and caution is obviously needed to assure comparability. Even when all of these are taken into account, it is possible that constraints on salaries and promotions in different institutions may be such that the measures would be of marginal value.

**HONORS AND AWARDS**

It is intuitively desirable to be able to give "credit" for having won honors or awards. While information about them is generally available from the individuals who have received them, access to the individuals is rarely available in connection with federally supported studies because of the continuing effort of the Office of Management and Budget to restrict data gathering. As a result, obtaining the information is often a tedious exercise; and with the hundreds of different awards made among all of the sciences, the chances of missing some are not small. However, biographical sources carry this kind of information, and the awarding organizations are also receptive to inquiry.

The principal problems with awards and honors are the inequalities of opportunity and of significance to which they are prone. Again discipline is important, for access to awards varies widely among them. Location of employment may also militate against the opportunity to gain recognition, and the related gains of prestige and cumulative advantage. Some rather ingenious efforts have been made to overcome some of the disadvantages of these measures: The Coles (1973) had 300 physicists rate the visibility and prestige of 98 honorific awards and used their ratings to weight the available information. Others have set a mark--that is, some number of awards or honors may be set as a level of "success." As with grants, receiving awards may be a suitable indicator of achievement, but awards are very rare, and not receiving them is not an equitable indicator of their absence.

**MENTORING**

Mentoring has usually been studied as a predictor of the subsequent success of students (see, for example, Long et al., 1979), but as a measure of productivity it has generally been rejected because popularity as a mentor has been associated with the concept of operating "diploma mills." If suitable means of assessing the performance of individuals whose principal activities do not include active research participation or who devote only a small proportion of time to research are to be found, mentoring should be reconsidered. The mentor whose students become outstanding achievers may or may not deserve credit for having made an important contribution to science. Only studies that are able to obtain accurate assessments of ability, and to consider very long time spans are likely to be able to deal with such a measure (after completion of training and appointment to a university, it would probably require about 15 years' followup to obtain a minimum amount of useful information). An alternate measure to be considered could be assessment of a mentor's success in placing students on completion of the doctorate. If such a measure could be partialed out from its inevitable association with prestige of the mentor's institution, it should prove to be an appropriate measure. Placements could be "scored" by referring to the National Academy of Sciences *Assessment of Research Doctorate Programs* (Jones et al., 1982) for the appropriate disciplines.

**PATENTS**

While it can be assumed that any participant in the scientific community is willing, if not eager, to receive an award, the pursuit of patents is a specialized goal of a restricted subgroup. With the easing of federal restrictions on the ownership of patents, interest in obtaining them has undoubtedly increased in the population of scientists that was formerly restricted. To what extent the changing mores of society, which strongly encourage the deliberate pursuit of material reward, may also affect scientists' problem-selection behavior is not, to this writer's knowledge, known. But the current "state of the science" of many areas of research in the biomedical sciences, which has produced rapidly expanding opportunities for producing scientific advances that have significant commercial potential, is surely not without effect. Among scientists employed in the commercial sector, patents and salary are probably the two best potential measures of productivity available. But how patents should be assessed in the academic sector is not so clear, and whether academics will change their publication behavior when nearing a patentable advance, as many in the commercial sector have been forced to do, is also not clear.

Patents have become an international focus of attention in assessing the productivity of nations. Narin's report to NSF (1988) of comparisons between U.S. and Japanese patent activity has drawn widespread journalistic as well as scientific interest. Together with several present and former staff members, Narin has established a new dimension in bibliometrics through studies of relations between publications and patents. CHI, of which Narin is president, has developed a "Patent Citation Indicators Database" and "Full Text Patent Citation Data." The first makes available all information on the first pages of close to a million U.S. patents issued since 1971; and the second provides extracts from the full texts of patents issued between 1975 and 1983, including citations to U.S. and foreign patents. These resources are as yet untapped for studies of NIH research programs, although the potential for investigation of a wide variety of questions at the interface between science and technology is great (see, e.g., Noma (1986).

Keith Pavitt, a member of the prolific science studies group at the University of Sussex, has prepared an excellent review (1985). He notes early work in the 1960s, increased interest reflected in NSF's *Science Indicators*, and lists sources of data in Western Europe and the United States. The state of the field is discussed under the following headings:

- o analytical approaches (Narin, various economists, and NSF's *Science Indicators*);

- o types of activity measured (invention as distinct from innovation, relation to R&D expenditures, relative superior performance of small firms, protection against imitation, skewed distribution of monetary value);

- o international comparisons (summaries of several empirical studies in such areas as relations between per capita capitol expenditures on R&D and patent activity, problems of data-gathering in foreign countries, national differences in propensity to patent);

- o comparisons among industrial sectors (technology gap theory, difficulties of allocating patent classes uniquely to product based industry classes, relation of this problem to accuracy of estimates in different fields);

- o comparisons among technical fields (classification problems in attempting to relate patenting to rates of technical innovation, citation rates of "significant" patents, links between patents and scientific literature, technical profiles of industrial firms);

- o comparisons among industrial firms (relations between R&D and patenting, skewed distribution of value of patents and propensity to patent, inverse relation between propensity to patent and size of R&D programs); and

- o comparisons over time (increasing share of U.S. patents that are of foreign origin, possibility of increased concentration in relation to diminishing ratio of patents to R&D size).

Pavitt concludes with a list of areas in which systematic inquiry is needed. He contends that the elimination of sector and firm-specific biases will require more comprehensive and accurate information about the nature and determinants of patenting behavior within firms. Systematic sample survey data are required on the following subjects:

- the sources of the innovative activities that lead to patenting in particular, the intersectoral variance in the relative importance of R&D, production engineering small firms, and other sources;

- the time distribution of patenting activities over the life cycle of an innovation (in particular, does patenting typically reach a maximum at the time of commercial launch?);

- the propensity to patent the results of innovative activities: in particular, sector specific factors related to the effectiveness of patenting as a barrier to imitation, compared to alternatives; firm-specific factors related to perceptions of the costs and benefits of patenting; and country-specific factors relating to the costs and benefits of patenting; and

- the judgment of technological peers on the innovative performance of specific firms and countries, and on the relative rate of technological advance in specific fields: in particular, the degree to which these judgments are consistent with the patterns shown by patent statistics.

Finally, Pavitt calls for improved classification schemes, such that established patent classes can be matched more effectively, on the one hand to standard industrial and trade classifications and, on the other, to technically coherent fields of development.

## SUMMARY

There are, simply, no easy, ready-made solutions to the problems of identifying measures that will be useful in the assessment of productivity. There is need for the development and application of creative approaches to improving the utility of the kinds of information that can be obtained. The development, for example, of indexes that may increase the equitability of some measures. And there is need as well, in many cases, for increased attention to detail in designing studies and analyzing data.

The two sources of information that have the broadest potential value in the assessment of academic scientific performance are peer assessment and the analysis of publications, though there are circumstances in which neither may be appropriate. (For analyses involving the commercial sector, patent analysis--when used as an extension of publication analysis--should probably be added.) From the perspective that they tend to be fairly highly correlated, each contributes somewhat to confidence in the other, and to the extent that they are not correlated the need for both kinds of information is greater in the given measurement situation. Because peer assessment is so extremely costly, time consuming, and difficult to employ equitably, it may be necessary or worthwhile, especially in large-scale studies, to investigate whether there are records available about--for example, program operation, faculty activity, support, student outcomes, and resources (in addition to publication data)--that might be able to account for a large proportion of the variation in peer judgments of, program quality.

On the other hand, the use of publication and citation measures as the sole consideration in the assessment of the individual scientist's productivity can be rejected on a purely rational basis. As a means of confirming a positive subjective judgment of individual performance, there is no problem, but the opposite does not hold because there are myriad alternative explanations for low numbers of publications and for few or no citations. One of the more significant misjudgments that can result is the case in which few or no citations are received by highly significant papers that either are ahead of their time or are published in obscure journals. No imperfect tool that may be used to the disadvantage of the single individual (including peer judgment) can be justified. The caution warrants repeating (and appears fairly frequently in the bibliometric literature) that bibliometric measures are most appropriately employed in group comparisons in which aggregates of publications are large--just how large depends on how closely comparison groups can be matched. Correspondingly, peer assessments are most appropriately employed when peers are equally informed about all of the assessment targets and when self-serving competitive interests are absent. Perhaps the single most important factor in planning investigations of productivity is the need to employ multiple measures and to apply them selectively to the appropriate targets.

## REFERENCES


Anderson, Richard C., Francis Narin, and Paul McAllister. 1978. Publication ratings versus peer ratings of universities. *Journal of the American Society for Information Science* 29(2):91-103.

Carter, Grace M. 1974. Peer Review, Citations, and Biomedical Research Policy: NIH Grants to Medical School Faculty (Rand Report R-1583-HEW). Washington, D.C.: Rand Corporation.

_____, Clara S. Lai, Carolyn L. Lee. 1978. A Comparison of Large Grants and Research Project Grants Awarded by the National Institutes of Health (Rand Report R-2228-1-NIH). Washington, D.C.: Rand Corporation.

_____, John D. Winkler, and Andrea K. Biddle. 1987. An Evaluation of the NIH Research Career Development Award. Washington, D.C.: Rand Corporation.

Chubin, Daryl E. 1987. Research evaluation and the generation of big science policy. *Knowledge* 9(2):254-277.

_____, and Soumyo D. Moitra. 1975. Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science* 5:423-441.

Clark, Kenneth E. 1957. America's Psychologists: A Survey of a Growing Profession. Washington, D.C.: American Psychological Association, Washington, D.C.

Cole, Jonathan R., and Stephen Cole. 1973. Social Stratification in Science. Chicago: University of Chicago Press.

Committee on Science, Engineering, and Public Policy (COSEPUP). 1982. The Quality of Research in Science Methods for Postperformance Evaluation of Research in the National Science Foundation. Washington, D.C.: National Academy Press.

Cronin, B. 1984. The Citation Process. London: Taylor Graham.

Fox, Mary Frank. 1983. Scientists' publication productivity. *Social Studies of Science* 13(2):298-329.

Garfield, E. 1955. Citation indexes for science. *Science* 167.

_____. 1972. Citation analysis as a tool in journal evaluation. *Science* 178:471-479.

Gee, Helen Hofer. 1988. An Analysis of NIH Intramural Research Publications, 1973-1984 (Report to the Committee to Study Strategies to Strengthen the Scientific Excellence of the NIH Intramural Research Program). Washington, D.C.: National Academy Press.

_____, and Frances Narin. 1986. An Analysis of Research Publications Supported by NIH 1973-76 and 1977-80 (Publication No. 86-2777), Washington, D.C.: NIH.

Gilbert, G. Nigel. 1988. Measuring the growth of science. *Scientometrics* 1(1):9-34.

Hjerppe, R. 1982. Supplement to a "Bibliography of bibliometrics and citation indexing & analysis." *Scientometrics* 4(3):241-273.

Institute for Scientific Information. 1963. *Science Citation Index*. Philadelphia, PA: ISI.

Jones, Lyle V., Gardner Lindzey, and Porter Coggeshall (eds.). 1982. An Assessment of Research-Doctorate Programs in the United States. Washington, D.C.: National Academy Press.

Lawani, Stephen M., and Alan E. Bayer. 1983. Validity of citation criteria for assessing the influence of scientific publications: New evidence with peer assessment. *Journal of the American Society for Information Science* 34(1):59-66.

Leydesdorf, Loet. 1987. Various methods for the mapping of science. *Scientometrics* 11(5-6):295-324.

_____, and Peter van der Schaar. 1987. The use of scientometric methods for evaluating national research programs. *Science and Technology Studies* 5(1):22-31.

Long, J. Scott, Paul D. Allison, and Robert McGinnis. 1979. Entrance into the academic career. *American Sociological Review* 44(5):816-830.

_____, and Robert McGinnis. 1981. Organizational context and scientific productivity. *American Sociological Review* 46:422-442.

Martin, Ben R., and John Irvine. 1983. Assessing basic research. *Research Policy* 12:61-90.

McAllister, Paul R., and Francis Narin. 1983. Characterization of the research papers of U.S. medical schools. *Journal of the American Society for Information Science* 34(2):123-131.

McGinnis, Robert, and J. Scott Long. 1982. Postdoctoral training in bioscience: allocation and outcomes. *Social Forces* 60(3):701-722.

Moed, H. F., J. M. Burger, J. G. Frankfort, A. F. J. Van Raan. 1985. A comparative study of bibliometric past performance analysis and peer judgment. *Scientometrics* 8:3-4.

Moravcsik, Michael J. 1986. Assessing the methodology for finding a methodology for assessment. *Social Studies of Science* 16:534-39.

Narin, F. 1976. Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity (Report to the National Science Foundation). (Now available only through the National Technical Information Service (NTIS no. PB 252339/AS).

_____. 1983. Subjective vs. Bibliometric Assessment of Biomedical Research Publications (NIH Program Evaluation Report). (Unpublished report available from the NIH Office of Program Planning and Evaluation or from the author.)

_____. 1985. Measuring the Research Productivity of Higher Education Institutions using Bibliometric Techniques. Paper presented at a Workshop on Science and Technology Measures in the Higher Education Sector, OECD, Paris, France.

_____. 1988. Indicators of Strength: Excellence and Linkage in Japanese Technology and Science. Paper presented at the National Science Foundation, June 21, 1988 (See also F. Narin and E. Noma, Is technology becoming science?, *Scientometrics* 7(3):369-381, 1985.)

_____, and J. K. Moll. 1977. Bibliometrics: *Annual Review of Information Science and Technology* 12:32-58.

_____, G. Pinski, and H. H. Gee. 1976. Structure of the Biomedical Literature, *Journal of the American Society for Information Science* 27:25-45.

National Science Foundation. 1974. *Science Indicators*. Washington, D.C.: U.S. Government Printing Office (this report is published annually).

_____. 1982. *Studies of Scientific Disciplines: An Annotated Bibliography*. Washington, D.C.: U.S. Government Printing Office.

Noma, Elliot. 1986. *Subject Classification and Influence Weights for 3000 Journals*. Haddon Heights, NJ: Computer Horizons, Inc.

Pavitt, K., 1985. Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics* 7(1-2):77-99. (Pavitt cites B. Basberg, Technological change in the Norwegian whaling industry: A case study of the use of patent statistics as a technology indicator, *Research Policy* 11(3):163-171, 1982.)

Pinski, Gabriel. 1975. Subject Classification and Influence Weights for 2300 Journals (NSF Final Task Report). Haddon Heights, NJ: Computer Horizons, Inc.

Porter, A. L., D. E. Chubin, and Xiao-Yin-Jin. 1986. Citations and Scientific Progress: Comparing Bibliometric Measures with Scientist Judgments. *Scientometrics* 13(3-4):103-124.

Price, Derek de Solla. 1961. Science Since Babylon. New Haven, Conn.: Yale University Press.

_____. 1963. Little Science, Big Science. New York: Columbia University Press.

Reskin, Barbara. 1979. Review of the Literature on the Relationship Between Age and Scientific Productivity. In Committee on Continuity in Academic Research Performance, Research Excellence Through the Year 2000: The Importance of Maintaining a Flow of New Faculty into Academic Research (Appendix C: 189-207). Washington, D.C.: National Academy of Sciences.

Schubert, A. 1985. Quantitative studies of science: A current bibliography. *Scientometrics* 9(5-6)293-304.

_____. 1986. Quantitative studies of science: A current bibliography. *Scientometrics* 8(1-2):137-140.

_____, and T. Braun. 1986. Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5-6)281-291. (See also T. Braun, W. Glanzel, A. Schubert, One more version of the facts and figures on publication output and relative citation impact of 107 Countries, 1978-1980, *Scientometrics* 11(1-2):9-15, and (3-4):127-140.)

Schubert, A., and W. Glanzel. 1983. Statistical reliability of comparisons based on the citation impact of scientific publications. *Scientometrics* 5(1):59-74.

_____, W. Glanzel, and T. Braun. 1983. Relative citation rate: A new indicator for measuring the impact of publications. In D. Tomov and L. Dimitrova (eds.), Proceedings of the First National Conference with International Participation on

Scientometrics and Linguistics of Scientific Text, Varna, pp.80-81.

Simeon, V. L., et al. 1986. Analysis of the bibliographic output from a research institution in relation to the measures of scientific policy. *Scientometrics* 9(5-6):223-230.

Stowe, Robert C. 1986. Annotated Bibliography of Publications Dealing with Qualitative and Quantitative Indicators of the Quality of Science (A technical memorandum of the quality indicators project). Cambridge, MA: Harvard University.

Van Heeringen, A., and P. A. Dijkwel. 1987a. Age, mobility and productivity: I. *Scientometrics* 11:267-280.

_____. 1987b. Age, mobility and productivity: II. *Scientometrics* 11:281-293.

Vinkler, P. 1986. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* 10(3-4):157-177.

_____. 1987. A quasi-quantitative citation model. *Scientometrics* 12(1-2):47-72. (See also B. Cronin, *The Citation Process*, London: Taylor Graham, 1984; D. Chubin and S. D. Moitra, Content Analysis of References: Adjunct or alternative to citation counting? *Social Studies of Science* 5:423, 1975; and M. J. Moravscik and P. Murugesan, Some results on the function and quality of citation, *Social Studies of Science* 5:86-92, 1975.)

Weiss, C. H. 1972. Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, NJ: Prentice-Hall Inc.

## APPENDIX: SCIENCE STUDIES RESOURCES

Nearly three-quarters of a century has passed since Cole and Eales in 1917 reported their international comparison of counts of books and papers in comparative anatomy published between 1543 and 1880 (Narin, 1977). In 1926 Lotka demonstrated that the distributions of publications in a discipline (physics) is widely skewed and that most scientific papers are published by a small minority of scientists (Fox, 1983). So began inquiries into the use of publications measures in the assessment of productivity and the closely related concept of eminence. Rapid advancement, however, became feasible only when computers became readily accessible and inexpensive in the 1950s.

In a landmark empirical study conducted between 1954 and 1957, a committee of the American Psychological Association conducted an extensive inquiry into the correlates of productivity of all doctorates granted in the field of psychology between 1930 and 1944 (Clark, 1957). The study was significant in employing publication and citation measures as correlates of peer assessments of productivity and in recognizing the importance of investigating differences among subdisciplines and of taking into account variations in background, social, and psychological characteristics as correlates and potential predictors of eventual professional accomplishment and status. The study was also noteworthy in its use of computer-implemented quantitative methods to describe and compare the most productive with other members of the profession. In this sense it marks the empirical beginning of what has become a worldwide effort on the part of both theoretical and empirical investigators to achieve a better understanding of how science and scientists function and thrive in the society of our time.

Comprehensive theoretical and methodological as well as empirical studies of the sociology, psychology, and economics of science and scientists did not begin to appear in large numbers until the 1960s. Derek de Solla Price (1963) is appropriately credited with sparking the present-day intellectual development of inquiry into the assessment of research quality and eminence. Since then studies have proliferated rapidly in depth, breadth, and complexity as well as in number. Hjerppe (1982) added 518 items to an over 2,000-item "Bibliography of Bibliometrics and Citation Indexing & Analysis" published in Sweden in 1980. More directly relevant to the present inquiry are bibliographies that are being developed to assist groups of interested and involved scientists in their attempts to keep up with research aimed at achieving better understanding of how science and scientists function. Although it is not feasible to attempt a comprehensive review of all bibliographies that might be helpful to those concerned with the analysis of productivity and its essential correlates, a brief description of some publications that cover a great deal of the relevant research effort to about 1980 may be useful.

*** Jonathan and Stephen Cole, *Social Stratification in Science* (1973): The Coles conducted several different cross-sectional studies of academic physicists in their investigation of the social stratification system in science. The Coles staunchly defended the view that science functions as a meritocracy and concluded that physics is a universalistic and rational discipline in which quality of work (as measured by citations) is the chief determinant of ultimate status. (A recent personal communication indicates that J. Cole delivered a paper at American Sociological Association meetings that partially recants earlier views on universalism.) For more up-to-date, longitudinal analyses of scientists in biochemistry that result in a different conclusion, see Long et al. (1979), Long and McGinnis (1981), and McGinnis and Long (1982). The Coles examined multivariate interrelationships among departmental rank, number and assessed prestige of honorific awards, membership status in professional societies, geographical location, number and "quality" (citation counts) of publications in exploring the development of professional visibility, and eminence. The book also contains a brief historical account of the development of research in the social science of science.

*** Francis Narin, *Evaluative Bibliometrics* (1976): Narin cited 140 papers in providing a brief historical account of the development of techniques of measuring publications and citations, in reviewing a number of empirical investigations of the validity of bibliometric analyses, and in presenting details of the characteristics of and differences among scientific fields and subdisciplines. (The Annual Review of Information Science and Technology published a bibliography entitled "Bibliometrics" by Narin and Moll (1977), which contains many, but not all of the same references that are in *Evaluative Bibliometrics*.) The book, prepared for the National Science Foundation, contains explicit details of how several indices of journal influence are calculated and how variations within a field of science differ from variations within a subdiscipline. Three different influence measures are provided for each of the 2,250 journals in the 1973 *Science Citation Index*. [(New influence indices have since been calculated for some 3000 journals in the 1982 SCI (see Noma, 1986).] Some two dozen studies are cited that deal with the correspondence between literature-based and other methods of assessing the quality of scientific output.

*** NSF Division of Planning and Policy, *Social Studies of Scientific Disciplines*, (1982): This annotated bibliography "makes accessible to the managers and practitioners of science and engineering the findings from the social studies of science in a form that will be useful to them." The bibliography covers studies conducted up to the mid 1980s and reports on the work of nearly 300 authors, most with multiple entries. Although only one subsection is entitled "Productivity," it is not an exaggeration to estimate that at least 90 percent of entries in the work deal with material relevant to the measurement of this concept. An approximately similar percentage describe investigations that employ publications measures in their investigations of 23 identifiable but related subjects as dealt with in studies of 13 disciplines. A total of 285 studies yield nearly 500 entries in the bibliography, many studies having dealt with multiple disciplines. Subject categories in the bibliography include:

- Attitudes and Values
- Career Patterns
- Competition
- Development of Disciplines
- Discipline Comparisons
- Discipline Organization
- Discovery Process
- Education, Grad. Educ.
- Funding of Research
- Information Exchange
- National Comparisons
- Paradigm Characteristics
- Performance of research
- Productivity
    - Productivity - age
    - Professional Associations
- Publication practices
- Recognition and reward
- Social stratification
- Structure of the literature
- Structure of literature--
    - Specialty groups
    - Citation rates
    - Journal influence
- University Ratings

*** Mary Frank Fox, "Scientists' Publication Productivity," *Social Studies of Science* (1983): In this critical review, Fox discusses publication productivity in relation to psychological characteristics of individuals such as motivation, ego strength, cognitive style, personality and interests, and IQ, noting the restricted range of ability among scientists and the corresponding low correlation with measures of productivity as well as the fact that creativity does not exist in a vacuum. Citing Pelz and Andrews, she states, "Rather, social factors so affect the translation of creative ability into innovative performance that measured creativity is virtually unrelated to either the innovativeness or the productiveness of scientists' output." The importance of environmental characteristics such as institutional prestige and organizational freedom are summarized, including the important findings of Long and McGinnis, whose longitudinal studies point to the stronger effect of location on productivity than of productivity on subsequent location as had been previously reported in studies using cross-sectional designs. An interesting discussion of the closely entwined concepts of cumulative advantage and reinforcement is also included in this review of approximately a hundred different studies.

*** A. Schubert, "Quantitative Studies of Science: A Current Bibliography," *Scientometrics* (1985 and 1986). Close to 100 papers are listed in each year, and the list does not include those published in *Scientometrics* itself. The vast majority deal with empirical and methodological papers on bibliometric topics. While no country exceeds the United States in number of papers listed, the total number of foreign papers, not including Canada and the United Kingdom, was nearly twice the number of United States publications.

*** Robert C. Stowe, *An Annotated Bibliography of Publications Dealing with Qualitative and Quantitative Indicators of the Quality of Science (Including a bibliography on the access of women to participation in scientific research)* (1986). In addition to a list of core books, annotated entries are made under the following headings:

I. Bibliometric indicators of the quality of scientific research
   - Citations and publications as indicators of quality
   - Critiques of citation analysis
   - Citation Context Analysis

II. Qualitative approaches to and more general works on research evaluation

III. Works dealing specifically with "science indicators"

IV. Forecasting and research priorities

V. Peer review

VI. Quality and quantity in the history of science and philosophy

VII. Education

VIII. Issues involving quantity and quality in particular disciplines, including papers on social indicators

IX. Sociology of science

X. Methodological papers and bibliographies

XI. Access of women to participation in scientific research

## HEALTH SERVICES RESEARCH PERSONNEL: DEMAND, SUPPLY, AND ADEQUACY OF TRAINING RESOURCES

*Elizabeth McGlynn**

### INTRODUCTION

Health services research is an applied field that has as its domain all aspects of the health care delivery system. The ultimate goal of health services research is to provide the information base necessary to design a health care delivery system capable of maximizing the health of the population within the resource constraints imposed by the public and private sectors. The subjects of this research field include the training of health professionals; the process by which services are delivered; outcomes of care; methods by which patients pay and health professionals are reimbursed for care; the quality, efficacy, effectiveness, and appropriateness of services; and the interrelationships of these components.

Health services studies may be methodologic, descriptive, analytic, or experimental, which implies a range of requirements for trained personnel. The demands for individuals trained in health services research come from academe, government, and the private sector. Given the complexities of the health care delivery system, the field relies upon researchers trained in numerous disciplines, as well as those capable of bringing disciplines together to work cooperatively. Figure 1 illustrates broadly the domains of health services research.

| U.S. Population Health Needs --------> | Service Delivery System --------> | Outputs |
|---|---|---|
| --Demographics | --Efficacy | --Expenditures |
| --Culture | --Effectiveness | --Health transitions |
| --Morbidity | --Efficiency | --Satisfaction |
| --Mortality | --Equity | |
| | --Quality | |

FIGURE 1. A framework for health services research.

* The opinions expressed in this paper are the author's and do not necessarily reflect those of either the Committee on Biomedical and Behavioral Research Personnel or the National Research Council.

Much of health services research looks at the relationship between the needs of the population and the delivery of services. The gaps between the demands for and supply of health technology often are studied in order to formulate approaches to closing the gaps.

Finally, health services research examines the output of the service delivery system as an indicator of system performance. The two outputs that have been the principal focus of research are expenditures on care and the health transitions (increments and decrements to health) that result from the application of service delivery system technologies. Others have argued that patient satisfaction with care also should be assessed. Economists, statisticians, and econometricians have developed tools for examining expenditures, whereas clinicians and psychologists have studied the health transitions and satisfaction.

The diversity of questions formulated and analyzed by health services researchers makes it difficult to develop a neat taxonomy of the field. In an effort to explain through example, the remainder of this paper describes the training of individuals engaged in health services research, the various pathways used to produce these researchers, the current levels of support for training, and future events that might affect the demand for health services research personnel. At the conclusion of the paper, the question of the adequacy of these mechanisms for producing enough well-trained health services researchers for the near term is discussed. Throughout, many questions are raised for which there are not adequate data to arrive at an answer. These questions are raised not to frustrate the reader but to suggest avenues for a more systematic inquiry into the demand for and supply of health services research personnel.

## WHO CONDUCTS HEALTH SERVICES RESEARCH?

Defining the requirements for health services research personnel is a complex task, for by its nature health services research is an interdisciplinary field requiring individuals trained in a variety of disciplines, as well as individuals who are capable of bringing disparate disciplines together to examine questions about the delivery of health services. Because there are no certification or licensing requirements, it is difficult to estimate the number and types of individuals who currently are conducting health services research studies. In this section four indicators of the types of individuals who constitute the health services research community are provided.

One indicator comes from the membership of the Association for Health Services Research (AHSR), the professional society for such personnel. Table 1 gives the distribution of AHSR's membership by discipline. As the table demonstrates, the membership includes a wide variety of disciplines; these have been categorized for presentation purposes into more aggregate groupings and originally included more than 40 named disciplines. The top three disciplines are medicine (18.1 percent), economics (13.2 percent), and public health (12.7 percent). No single discipline is dominant.

The educational background of the membership provides another illustration of the training received by those currently practicing in the field and is displayed in Table 2. About half of the members hold a research doctorate degree; about 21 percent are physicians or other health professionals; almost 20 percent have masters degrees; and the remainder have bachelors degrees, are in training, or did not specify their training in the membership survey.

The institutional affiliations of members illustrate the variety of professional settings in which health services research personnel operate. Table 3 provides a distribution of the membership by the institutional affiliation of the individual. The greatest proportion of members are located in university settings (46.5 percent), followed by health care delivery or private health industry organizations (almost 27 percent). Other private organizations (e.g., consulting firms) and government employees make up 10.8 and 5.2 percent, respectively, of the membership.

Finally, the publications that comprise health services research demonstrate the variety of topics considered and disciplines required to conduct such studies. The AHSR has an award each year for the best article in the field. Table 4 shows a listing of the titles, journals,(first author, and discipline of the articles thit were nominated for 1988 Article of the Year. Health services research is conducted by persons who have been trained in a variety of disciplines. Most of the studies are interdisciplinary in nature and require a team of individuals who can work together to solve the complex methodologic and conceptual issues posed by the subject matter. The articles shown in Table 4 provide one indicator of the variety and type of topics addressed by leading health services researchers and some sense of the disciplines involved. The extent to which the field is interdisciplinary is understated because only the first author is shown. Many of these projects included physicians, social scientists, and methodologists (e.g., statisticians).

## TRAINING FOR HEALTH SERVICES RESEARCH

As suggested from the above discussion, the training currently received by those who are conducting health services research covers a variety of disciplines and degrees; therefore, there are a variety of formal and informal ways in which individuals become trained. Thus, the discussion of training considers each of these approaches to training separately:

- disciplinary degree programs (e.g., programs that grant a professional or research degree in a traditional discipline such as economics);
- interdisciplinary training programs (e.g., public policy, health services research);
- postdoctoral training programs; and
- on-the-job training (e.g., socialization of research personnel into the field of health services research).

Because the scope of health services research is broad, most individuals probably are trained in a narrow aspect of the field and then receive socialization into broader issues through some combination of additional work (e.g., postdoctoral studies) or by learning on the job. Whether this is the most efficient means of producing such personnel remains an open question.

## DISCIPLINARY DEGREE PROGRAMS

Currently, most individuals engaged in health services research have received their training primarily from disciplinary degree programs, including economics, quantitative methods (e.g., statistics, operations research, epidemiology), sociology, psychology, management (e.g., business, accounting, planning), and so on. It seems likely that for the foreseeable future, the field of health services research will continue to draw upon individuals with disciplinary degrees for a majority of its personnel. Clearly, the degree requirements for these programs already are well established, and the principal issues for this group of potential researchers are as follows:

- the type and amount of training necessary to acquire health services research skills and
- the extent of migration between disciplinary and interdisciplinary research and methods for influencing migration.

Although health services research is interdisciplinary and depends upon the skills and knowledge of investigators who have received different training, as the field has become more mature, it has developed its own language and information base. For health services research to be conducted efficiently and effectively (i.e., to avoid reinventing the wheel), individuals who wish to engage in this research must be socialized into the field in some manner. The two most common mechanisms--

postdoctoral training and on-the-job training--are discussed below. Most formal training programs take about 2 years, whereas on-the-job training has a less specific time frame. The amount of time required probably depends upon an individual's prior training, experience, and aptitude for learning new applications.

One of the difficulties in estimating the number of persons currently engaged in or capable of performing health services research is that a majority of individuals have the potential of migrating in and out of the field. Further, some may spend only a portion of their professional time in health services research, while spending the remaining time in some other activity (e.g., medical practice or other disciplinary research). Migration may have the advantage of providing a nlexible work force capable of expanding and contracting to meet the demands of the marketplace. The impact of migration on the supply of personnel is an issue that is particularly relevant to interdisciplinary research. Because little is known about the effects of such labor-market patterns, an investigation of the role of migration on the demand and supply of health services researchers should include attention to the following questions:

- What is the minimum amount of time that must be spent in health services research in order to produce high-quality research?

- Is migration between health services and other research fields desirable?

- What are the implications of migration for the quality of research produced?

Having this potentially large source of human energy on which to draw for new health services research talent raises the question of the mechanisms by which such individuals might be attracted into the field. The attractiveness of any career path depends upon at least two factors: the financial viability of that avenue as compared to competing alternatives and the relative importance of the problems under study.

The funding for health services research is small relative to biomedical research; thus, among those who have biomedical research as an alternative, health services research probably is not as attractive. However, those who are competing for research funding in the behavioral sciences may find health services research relatively attractive. If research funding in this area expands, one might expect greater numbers of persons to be drawn into the field. Other mechanisms, such as Career Development Awards, that provide a stable source of funding for new entrants to the field while they develop a research track record provide an appropriate means of attracting new talent into this field. Such awards are not widely available for health services research in general; most fall within more specialized areas of inquiry and tend to be aimed at physician researchers.

Health services research also may be attractive to some because of the policy significance of the issues addressed. There is considerable interest in government and the private sector regarding issues of patient health outcomes, the appropriateness and quality of medical care, and the cost-containment potential of alternative delivery systems. Because both the public and private sectors are seeking answers to complex questions, the opportunities for working in an area of high visibility have expanded. This will appeal to those who are interested in applied rather than more theoretical fields of inquiry.

**INTERDISCIPLINARY TRAINING PROGRAMS**

Interdisciplinary training is a relatively new--and increasingly important--avenue for producing health services researchers. Perhaps the most familiar programs are in public policy, public health, and health services or health policy research. In these programs students are introduced to a variety of methodologic approaches and taught the strengths and weaknesses of different methods for answering particular research or policy questions. This type of training introduces the individual to the contributions that can be made by different disciplines, which may be necessary in order to manage a team of researchers with different backgrounds.

The interdisciplinary training may be general (e.g., public policy) or specialized (e.g., health services research). Because there are few such programs, each one is almost a unique case, and it is difficult to generalize about the advantages and disadvantages of each approach. Perhaps it is most important to know that both paths exist and that both are likely to produce some high-quality individuals capable of conducting health services research projects. As these fields develop, however, this may become a more efficient and effective means of producing health services research personnel. For those in interdisciplinary programs, the major issues are the following:

- the availability of such training programs relative to the number of applicants;
- the ability of such programs to attract high quality students; and
- career opportunities that are open to graduates of these programs.

Currently, there is no information on any of these three issues. What would be required in order to answer these questions is a comprehensive listing of interdisciplinary training programs and a survey of them. Admissions data combined with interviews of program directors could be used to answer the first two questions. In particular, it would be helpful to know what types of individuals are likely to pursue these programs (e.g., recent college graduates, individuals with related job experience, individuals making career changes) and how they compare with typical applicants to graduate programs in the disciplines from which health services research draws.

The third question might be best addressed through a survey of alumni of these interdisciplinary training programs. Because these programs are relatively new, it is unlikely that one could piece together a career history (e.g., one that follows individuals from graduation through retirement), but early career paths could be documented. It is important to note that a variety of career opportunities are available to individuals trained in interdisciplinary research. Because many research results may suggest needed changes in policy, there are benefits to having trained individuals who are responsible for policy formulation and implementation.

**POSTDOCTORAL TRAINING PROGRAMS**

Individuals trained in either disciplinary or interdisciplinary programs may require additional career development such as that offered in postdoctoral training programs. These programs may serve different purposes, depending upon the background of the individuals entering the program, but all should enhance the quality of research produced and provide research experience that will advance the careers of these persons. For those with professional degrees (e.g., medicine, law), the postdoctoral program may serve to introduce the individual to research methods. For persons with doctoral training in a specific discipline, the postdoctoral training program may provide an opportunity to specialize in health services research. For those with interdisciplinary degrees, the postdoctoral program may offer additional experience in designing and managing research projects. Each of these functions is important, and because of the multiple pathways into health services research, each will continue to serve a purpose in the future. The issues in this area are:

- What are the key components of successful postdoctoral training in health services research?
- What is the demand for positions in these programs relative to the availability?
- Do such programs influence the career paths of their graduates?

Despite the multiple pathways into health services research and the variety of disciplines that participate, there may be a basic set of skills necessary in order to have a successful career in this field. Two important components in many health services research projects are a clinical perspective and methodologic expertise. These reflect the two levels at which analysis generally is conducted in such projects: at the doctor-patient level (clinical) and at the level of a particular group (policy).

Physicians, by virtue of their medical training and experiences, can bring a clinical perspective to health services research, but, in order to contribute to the formulation of research questions and study design, they also must be trained in research methods. Physicians are trained to approach each patient as a unique case, rather than to take the aggregate approach to formulating questions and collecting data that are required for conducting research. Thus, most physicians require training in research methods to acculturate them to the language and framework for problem solving.

Methodologic experts, on the other hand, provide a perspective on the design and conduct of research projects, but they may require socialization or training about the important clinical issues. For example, much interest has been expressed by Congress and others in devoting additional resources to patient-outcome research. A methodologic expert might be able to frame the questions but would require assistance in designing a medical records abstraction form, including making decisions about the critical values (e.g., what diagnostic test result justifies performing coronary artery bypass surgery). Alternatively, a cardiologist could suggest the key outcomes of bypass surgery that might indicate whether high-quality care was being delivered, but he or she might require assistance from an epidemiologist or statistician to design a study to estimate the prevalence of quality problems.

Estimating the demand for entry to postdoctoral programs should be approached in a manner similar to that suggested for interdisciplinary training programs. A comprehensive list of the postdoctoral programs and a survey of program directors that included information from admissions records would provide information on, for example, the number of qualified applicants turned down each year. It would also be useful to know what alternatives applicants to these programs are considering.

Finally, although there seems to be logic in developing postdoctoral training programs, it would be useful to know the outcomes of these programs and whether they have any substantial influence on the quality and success of graduates:

- Does postdoctoral training improve the ability of individuals to obtain research funding?

- Are researchers who have been through such programs more likely to have articles accepted for publication?
- Are they able to publish in higher-quality journals?
- Are they more likely to stay in research careers?

Answering these questions would require a survey of alumni and an examination of the backgrounds of successful and unsuccessful applicants for health services research funds.

**ON-THE-JOB TRAINING**

The final pathway into health services research is on-the-job training. Probably anyone who is conducting health services research has had some on-the-job training because it is a common means of career development. In the university setting such opportunities are provided through research assistantships or internships. Students may form mentor relationships with research supervisors, who in turn take responsibility for schooling the student in the methods and knowledge central to the chosen field of study. In all research settings this may be the method by which disciplinary-trained individuals are socialized into health services research (e.g., by serving as the economist on a health services research project). Reliance on this method of producing health services researchers is risky at best because the success of the approach depends upon the quality and commitment of those serving in a supervisory capacity. If the supervisor perceives his or her role as a teacher or mentor and is capable in either of those capacities, then the individual in training is more likely to have a successful experience. Many research projects, however, are constrained by tight budgets that may not allow for adequate learning opportunities. Motivated individuals may be able to learn even under such circumstances, but we would hardly want to depend upon this approach as the sole means by which such researchers are produced.

**HEALTH SERVICES RESEARCH REQUIREMENTS**

The field of health services research would be advanced by establishing some guiding principles about the requirements to become a health services researcher. From the taxonomy proposed in Figure 1, a first step could be taken to recommend areas of exposure. While there is no one pathway that should be required, it is reasonable to consider how we would want to train the next generation of researchers in this area. A proposed curriculum is outlined in Table 5 as an example of the areas of study recommended for those in health services research. Presumably, some researchers would be experts in one of the fields listed in the table, but they would also have been exposed through a combination of formal course work and research experience in some proportion (say two-thirds) of the other disciplines that contribute to leading health services research projects. The knowledge of these other fields is important both in formulating research strategies and managing projects that utilize multiple disciplines.

## SUPPORT FOR HEALTH SERVICES RESEARCH TRAINING

From the above discussion, it should come as no surprise that the field of health services research draws upon a variety of sources of funding for training. These sources include both public and private support. Those more easily identified are the sources dedicated to health services research training fellowships; clearly, this does not capture all of the funding sources for training those who become health services researchers. The AHSR maintains a directory of Health Services Research Organizations. The 1988 directory provides profiles on more than 80 health services research centers, including information on the number and funding for training fellowships. The appendix to this paper contains a listing of the predoctoral, postdoctoral, and other training support based on information provided in the directory.

As can be seen from Table 6, the number of fellowships available for training in health services research is quite small. Although this number most likely underestimates the true number, it probably captures the majority of available fellowships in major institutions. It also should be emphasized that these numbers represent fellowships specific to health services research and do not include individuals receiving disciplinary training who may eventually enter health services research through other channels. It is also worth noting that these programs exist in only a few states (15 or 16), which suggests that access to these programs is limited geographically. California and Massachusetts have the largest number of institutions offering predoctoral and postdoctoral training programs. Financial support for these training programs comes from both government and private sources.

### FEDERAL SUPPORT

A number of agencies within the federal government provide funding for both research and training activities. The most prominent agencies are within the Department of Health and Human Services (DHHS) and include the National Center for Health Services Research and Health Care Technology Assessment, the Health Care Financing Administration, the National Institute of Mental Health, and the National Institutes of Health. Outside DHHS, health-related research activities (although not necessarily health services research) are funded by the

Departments of Agriculture, Commerce, Defense, Education, Energy, Interior, Labor, and Transportation, as well as the Consumer Product Safety Commission, Environmental Protection Agency, Agency for International Development, National Aeronautics and Space Administration, the National Science Foundation, and the Veterans Administration. It is difficult to estimate what proportion of the health-related research conducted by these other agencies falls into the category of health services research. Such an estimate would require a separate study that has not yet been done, although it was recommended in the last report to Congress.

The National Center for Health Services Research and Health Care Technology Assessment (NCHSR-HCTA) is one of the two agencies primarily responsible for funding health services research. Table 7 shows the allocation of funding for NCHSR-HCTA since 1986. Although current dollars remain stable, the real dollars available for such research have been declining since the inception of the center. Further, new money is being added in specific research areas (e.g., patient outcomes and AIDS research) rather than being made available for general research funding. In fiscal year 1988 NCHSR-HCTA funded 40 new grants in the following areas:

- patient outcomes research;
- AIDS-related projects;
- rural hospitals;
- emergency medical and intensive care;
- prenatal care;
- technology assessment;
- role of market forces;
- home health care; and
- other policy concerns.

Grants were awarded for periods of 1 to 5 years, although most grants were for 2 years. The total amount obligated was $8.6 million.

NCHSR-HCTA also provides support for predoctoral and postdoctoral training. The agency has authority to award 10 to 20 dissertation grants annually to support individuals pursuing degrees in fields related to health services research while they write dissertations. The grants may be up to $20,000 annually and must be related to the current funding priorities of the agency. NCHSR-HCTA also has authority to award individual

National Research Service Awards for postdoctoral training related to preparation for an academic career as well as for experience in applying research methods to the study of the organization, financing, and delivery of health services. Awards may be for periods ranging from 1 to 3 years. In fiscal year 1988, about $1.3 million was available for such grants; approximately 25 percent is devoted to individual awards, and the remainder supports institutional awards. The institutional awards are designed to help institutions provide the training support that is required for both predoctoral and postdoctoral programs. Thus, some of the funds may be used for faculty development activities or program support costs, but the major focus is on providing direct training support for postdoctoral candidates. Institutional awards may be made for periods of up to 5 years and may be renewed, but individuals receiving support through institutional programs are subject to the 5-year (predoctoral) and 3-year (postdoctoral) limits imposed on individual awardees.

The other major source of federal funding for health services research is the Health Care Financing Administration's Office of Research and Demonstrations (HCFA-ORD). Table 8 shows funding levels for 1986 to 1990. HCFA-ORD has indicated that it currently supports the following seven areas of primary interest for funding research activities:

- access to quality care under Medicare and Medicaid and improved methods for measuring quality and effectiveness of care;
- refinement of the current Medicare physician payment methodology through the study of the causes for the growth in Medicare outlays for physician services and the development of cost-effective approaches to controlling such growth;
- increased competition and consumer choice and continued growth of Medicare capitated systems;
- continued improvement in the current Medicare hospital prospective payment system and the study of the outpatient delivery system;
- analysis of other Medicare and Medicaid program services and issues leading to increased efficiency in health care delivery and financing, particularly in program areas that have a significant impact on program and beneficiary expenditures (high-cost, high-volume services), including such areas as clinical laboratories, home health and long-term care, and treatment of acquired immunodeficiency syndrome;

- refinement of the current long-term care delivery and payment systems under Medicare and Medicaid; and

- studies related to the Medicare Catastrophic Coverage Act of 1988, particularly issues related to the impact of this legislation on program implementation and beneficiary behavior.

HCFA-ORD funds research through both contract and grant/cooperative agreement mechanisms. The above priorities reflect those for the grant/cooperative agreement category. Both contract and grant award studies are conducted through a competitive bidding process. HCFA also maintains four policy research centers that are designed to assist HCFA in conducting short-term policy analyses and other analyses that support HCFA's mission. Although HCFA-ORD does not support the training of health services researchers directly, the projects funded through the agency provide research opportunities for individuals pursuing both predoctoral and postdoctoral training. A study of the adequacy of training support should include the use of such research funding to provide assistantship positions for persons in training.

The Veterans Administration (VA) has two programs that provide training support. The VA Health Services Research and Development Field Program is a predoctoral training program run in nine centers nationally. Funding levels for 1987 are shown in Table 9. The program provides support for individuals pursuing a Ph.D. in numerous disciplines, including epidemiology, nursing, social work, and psychology. Candidates are nominated by the local field program, and final selection is made by the central office. Most awards are for 1 year, although an additional year of funding is possible through approval of a renewal application. A small training stipend (approximately $9,300) is provided to offset costs of research and living expenses. The VA also runs a postdoctoral training program for physicians who are working on a masters in public health focused on health services research training. There are about 12 such fellows participating in the program each year. The postdoctoral program is designed as a 2-year program and provides some support for research projects ($1,000 to $4,000 annually). For both the predoctoral and postdoctoral programs, candidates must demonstrate that their research is relevant to issues faced by the VA in the delivery of services.

## PRIVATE FUNDING

The private sector, particularly philanthropic foundations, also plays an important role in funding health services research and training. The private sources of funding listed in the AHSR 1988 directory include the following: The Bush Foundation, The John A. Hartford Foundation, The William and Flora Hewlett

Foundation, The Robert Wood Johnson Foundation, The Henry J. Kaiser Family Foundation, The W. K. Kellogg Foundation, The Andrew W. Mellon Foundation, The Pew Charitable Trusts, and the Retirement Research Foundation. These foundations contribute to training both directly, through the support of predoctoral and postdoctoral programs, and indirectly, through funding research.

The two most established direct training programs are The Robert Wood Johnson Clinical Scholars program, which provides postdoctoral training for physicians, and The Pew Charitable Trusts Health Policy Program, which funds predoctoral and postdoctoral training in health policy. In 1987 The Robert Wood Johnson Foundation provided $1.89 million to support 53 physicians training in 6 programs around the country. The Clinical Scholars program has operated since 1969 and has produced 471 alumni from 12 programs (including the 6 that are currently funded). Table 10 shows the funding levels for 1987.

Since 1982 The Pew Charitable Trusts has funded an interdisciplinary health policy fellowship program. Table 11 shows the programs along with current funding levels, the number of fellows, alumni, and type of program. Over the 4-year period of the current awards, Pew is devoting $7.4 million to this program. Each of the programs provides a slightly different focus. Boston University and Brandeis University jointly offer three programs in health policy studies. The doctoral program is a 2-year multidisciplinary program leading to a Ph.D. in health or social policy at the participating institutions. The Fellows Program is designed to develop new leadership in national health policy from the corporate sector. The Associates Program is designed to assist corporate and local government leaders from communities nationwide to develop and evaluate alternative policies for health care cost containment. The University of Michigan program is a nonresidential doctoral program in public health directed at individuals who have experience in health policy and will continue working full time while pursuing their degree. Candidates attend classes for the first 2 years during one weekend each month (Thursday through Sunday). During the third year, fellows spend three such weekends on campus. The RAND/UCLA program provides predoctoral support for individuals enrolled at either The RAND Graduate School of Public Policy Studies or the UCLA School of Public Health and runs a 1-year midcareer program. The University of California at San Francisco offers 2-year postdoctoral fellowships to social scientists, physicians, and other health professionals. The program is designed to prepare such individuals for leadership roles in health policy and health policy research.

## FUTURE EVENTS

In the absence of a crystal ball, it is difficult to predict what future events might affect the demand for and supply of health services researchers. What seems clear today, however, is that we are spending an enormous amount of money on a variety of services about which we know surprisingly little. Understanding which health care services are likely to benefit whom will require a considerable commitment of research resources and a group of researchers with specialized skills. Only a few of those currently conducting health services research have those skills. To bound the problem, this section summarizes some of the information about current rates of growth in research and training funds and provides broad estimates of what would be required for a large-scale commitment of resources.

Both the demand for and supply of health services research personnel have grown in the past decade. Although there is no systematic information giving a baseline figure and demonstrating the rates of growth over time, there is evidence of growth from information collected by AHSR. Membership in the professional organization has increased from 60 members in 1982 (its first year in operation) to more than 1,000 individual and 70 institutional members in mid-1988. In 1983 the directory listed 37 university-based health services and policy research organizations and 8 VA Health Services Research and Development Field Programs. By 1988 the number of university-based centers had grown to 54; there were 22 public and private policy analysis and research organizations and 9 VA programs.

After a substantial decline through the late 1970s and early 1980s, research funding through NCHSR-HCTA has increased 258 percent between 1986 and the President's 1990 budget. The 1-year increase between 1989 and 1990 is estimated to be 46 percent. Two policy areas are driving these increases: patient outcomes and AIDS-related research. The funding increases signal congressional interest in devoting resources to studying some of the complex problems facing health services delivery. As the field matures and becomes more sophisticated, demands for more highly trained individuals are likely to increase. Thus, it seems likely that demands for more and better trained researchers will continue.

Whether demand will outstrip supply can only be speculative at this point. The real question is whether the public and private sectors will make a commitment to understanding how best to spend scarce health resources. Work force estimates for this field would be useful but require some systematic effort to collect the necessary data. One way of looking at the demand is to estimate the number of full-time equivalent (FTE) researchers required to conduct research under the 1990 HCFA and NCHSR budget estimates. If an FTE is valued at $100,000 annually (salary, fringe, and overhead), then the 1990 budget (about $98 million) would require about 988 FTEs to conduct research. This includes individuals within the government administering the research as well as those actually conducting research. Current AHSR membership is about 7 percent higher than that FTE estimate.

Although such membership does not represent the universe of those participating in such research, it also includes individuals who are not engaged directly in research. It is safe to guess that demand and supply are about in equilibrium. However, there is evidence that the political will exists to expand such research, which suggests that demand may be starting to exceed supply. For example, Congress has allocated funds to study the effectiveness of the most commonly used procedures. If $1 million each was allocated to examining the 50 most common procedures, 500 researchers would be required for this effort. The majority of these individuals would need to be specially trained and thus new entrants.

Another way of estimating future demand is to envision expenditure about 1 percent of total health care expenditure on health services research. In 1987 health care expenditures were $500.3 billion, of which $443 billion was spent on personal health services. Spending 1 percent would suggest a $4.4 billion health services research budget, which implies about 44,000 FTEs. An increase of this magnitude seems unlikely, but it is equally clear that the demand for trained individuals is increasing and will continue to grow. Some quick response will be possible through migration of researchers out of other areas of inquiry and into health services, but too great an influx of researchers untrained in the field may cause an inefficient use of research funds.

Finally, it is fair to speculate that the problems facing the health care delivery system are not likely to be solved soon or easily. The population will continue to age and suffer from chronic diseases; AIDS will affect greater numbers of individuals; substance-abuse treatment costs will continue to climb; and the gaps between the rich and poor in terms of access to care will continue to grow. These are but a few examples of the types of issues health services researchers will be asked to study in the coming decade. While there does not seem to be a crisis in the availability of research personnel currently, the pipeline required to produce such individuals is sufficiently long to warrant conducting more sophisticated work force estimates now rather than when a problem is upon us.

## CONCLUSION

Any serious effort to address the goal of using scarce health care resources effectively will require a significant increase in the number of trained research personnel. Little systematic information is available on the adequacy of training for health services researchers. The field is both new and developing and would benefit from a serious examination of many of the questions that have been raised in this paper. Given the interest expressed by Congress and the private sector in obtaining answers to complex questions about the quality and availability of health services, the demand for individuals with such training is likely to continue to grow. Any examination of the need for individuals who are trained in health services research should include both the public and private sectors.

The field of health services research requires the participation of multiple disciplines, thus contributing to the difficulty in estimating the adequacy of supply either currently or in the future. Although the field is likely to continue to draw on multiple sources, centers that provide the necessary interdisciplinary training are becoming increasingly important and are essential for any large-scale effort to define practice guidelines and evaluate the effectiveness of medical care delivery. In the future, preference should be given to supporting these interdisciplinary or multidisciplinary training programs.

The number of fellowships available for both predoctoral and postdoctoral training is relatively small: 216 predoctoral and 133 postdoctoral fellowships (including the Robert Wood Johnson clinical scholars program) were identified. Assuming 2-year postdoctoral and 5-year predoctoral programs, this suggests that an additional 110 researchers per year are being produced. This number of fellowships is too small to ensure that an adequate supply of researchers is available in the future. Assuming that there are currently about 1,000 researchers and that fellowships ensure 100 net additions annually (and also assuming that about 10 per year retire from the field), it would take about 10 years to double the supply. A systematic study is necessary, however, in order to estimate the support that would be required to meet future demands.

The delivery of health services represents a major expenditure for both the private and public sectors. Training individuals who are capable of analyzing the most effective and efficient means of delivering such services should be a high priority. The problems addressed by health services research are complex and not likely to be resolved in the foreseeable future. The benefits of training researchers and policy makers capable of studying these problems are likely to exceed the costs. The private sector has demonstrated its willingness to participate in this effort by developing predoctoral and postdoctoral training programs. The private sector also is an important player in funding health services research. Without a substantial investment in future research and training, we have little hope of understanding, in a scientifically valid manner, how to target health care spending effectively. Many of the necessary analytic tools are in place; what we require is the political will to tackle the problem in the form of increased funding for research and training.

TABLES

126

TABLE 1: Membership in the AHSR, by Discipline, as of May 1988

| Primary Discipline | Percent of Members (n = 1042) |
|---|---|
| Medicine | 18.1% |
| Economics | 13.2 |
| Public health | 12.7 |
| Sociology | 7.4 |
| Business/management | 6.8 |
| Health services research | 5.3 |
| Quantitative methods | 4.8 |
| Other health professions | 4.8 |
| Public administration/ policy | 4.7 |
| Psychology | 3.6 |
| Other professional | 1.7 |
| Education | 1.5 |
| Other | 15.4 |

SOURCE: AHSR, personal communication.

TABLE 2: Educational Background of AHSR Members, May 1988

| Type of Degree | Proportion of Members (n = 1042) |
|---|---|
| Ph.D. | 50.4% |
| M.D./other health professional | 21.1 |
| Masters | 19.5 |
| Bachelors | 2.4 |
| Students in training | 1.3 |
| Other/not specified | 5.3 |

SOURCE: AHSR Reports, Spring 1988.

TABLE 3: Institutional Affiliation of AHSR Members, May 1988

| Institutional Affiliation | Proportion of Members (n = 1042) |
| --- | --- |
| University | 46.5% |
| Health care delivery | 15.6 |
| Health industry | 11.3 |
| Private organization | 10.8 |
| Government | 8.2 |
| Private foundation | 1.2 |
| Other | 6.4 |

SOURCE: AHSR, personal communication.

129

TABLE 4: 1988 Nominations for AHSR Article of the Year

| First Author | Title | Degree and Discipline | Journal |
|---|---|---|---|
| Luft, H. S.* | The Volume-Outcome Relationship: Does Practice Make Perfect? | Ph.D. Economics | Health Serv. Res. |
| Shortell, S. M. | The Effect of Regulation, Competition, and Ownership on Hospital Mortality | Ph.D. Management | N. Engl. J. Med. |
| Dubois, R. W. | Hospital Inpatient Mortality: Is It a Predictor of Quality? | M.D./Ph.D. Public Policy | N. Engl. J. Med. |
| Chassin, M. R. | Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? | M.D. | JAMA |
| Wennberg, J. E. | Use of Claims Data Systems to Evaluate Health Care Outcomes: Mortality and Reoperation Following Prostatectomy | M.D. | JAMA |
| Goldberg, H. I. | A Randomized Controlled Trial of Academic Group Practice: Improving the Operation of the Medicine Clinic | M.D. | JAMA |
| Jencks, S. F. | Bringing Excluded Psychiatric Facilities Under the Medicare Prospective Payment System | M.D. | Med. Care |

*Winner of 1988 Article of the Year.

| First Author | Title | Degree and Discipline | Journal |
|---|---|---|---|
| Soumerai, S. B. | Payment Restrictions for Prescription Drugs Under Medicaid: Effects on Therapy, Cost, and Equity | Sc.D. | N. Engl. J. Med. |
| Feder, J. F. | How Did Medicare's Prospective Payment System Affect Hospitals? | Economics | N. Engl. J. Med. |
| Shy, K. K. | Evaluating a New Technology: The Effectiveness Electronic Fetal Heart Rate Monitoring | Ph.D. | Ann. Rev. of Pub. Health |
| Sacks, H. S. | Meta-Analyses of Randomized Controlled Trials | Ph.D. | N. Engl. J. Med. |
| Verbrugge, L. M. | Exploring the Iceberg: Common Symptoms and How People Care for Them | Ph.D. | Med. Care |
| Young, M. J. | Do Cardiologists Have Higher Thresholds for Recommending Coronary Arteriography than Family Physicians? | M.D. | Health Serv. Res. |
| Rossi, P. H. | The Urban Homeless: Estimating Composition and Size | Ph.D., Sociology | Science |
| Warner, K. E. | Health and Economic Implications of a Tobacco-Free Society | Ph.D., Economics | JAMA |

SOURCE: Brook, R. H., 1989. "Health Services Research: Is It Good for You and Me?", *Academic Medicine*.

TABLE 5: Proposed Curriculum for Health Services Research Training

| Methods | Substantive Knowledge |
| --- | --- |
| Economics | Technology assessment |
| Quantitative methods (statistics, econometrics, operations research) | Health professions training and supply |
| Social science research methods (survey research, psychometrics) | Quality of care |
| Epidemiology | Health care financing |
| Demography | Organization and service delivery |
| | Outcomes assessment |

132

TABLE 6: Institutions Participating in Health Services Research Training

| Category of Training | Number of Fellowships | Number of Institutions | Number of States |
|---|---|---|---|
| Predoctoral | 216[1] | 26 | 15 |
| Postdoctoral | 80[1] | 23 | 16 |
| Other[2] | -- | 24 | 15 |

[1] The number shown is a minimum; some programs indicate that the actual number varies from year to year.

[2] Includes internships, midcareer training, master's level programs, ad hoc fellowships, research assistantships, independent study, and visiting scholars.

TABLE 7: Funding for NCHSR-HCTA (in thousands of dollars)

| Category | 1986 | 1987 | 1988 | 1989 | 1990* |
|---|---|---|---|---|---|
| Research | | | | | |
| Federal funds | $15,740 | $17,277 | $16,624 | $17,112 | $12,153 |
| Trust funds | 1,050 | 1,070 | 1,005 | 1,037 | 1,037 |
| Program support | 1,050 | 1,568 | 1,632 | 1,600 | 1,693 |
| Outcomes research | -- | -- | 1,945 | 5,879 | 28,000 |
| AIDS-related research | -- | -- | -- | 6,859 | 8,576 |
| One-percent funds (NMES) | | 16,000 | 15,318 | 10,033 | 11,154 |
| NRSA | | | 1,296 | 1,323 | 1,300 |
| Total | $17,840 | $35,915 | $37,820 | $43,843 | $63,913 |

* President's budget.

SOURCE: AHSR.

TABLE 8: Funding for the HCFA-ORD (in thousands of dollars)

| Category | 1986 | 1987 | 1988 | 1989 | 1990[1] |
|---|---|---|---|---|---|
| Federal funds[2] | $15,310 | $10,000 | $9,574 | $9,880 | $14,000 |
| Trust funds[3] | 14,370 | 18,000 | 18,000 | 17,233 | 19,760 |
| Total | $29,680 | $28,000 | $27,574 | $27,113 | $33,760 |

[1] President's budget.
[2] Appropriated by Congress.
[3] Set aside from trust fund collections by congressional action.

SOURCE: AHSR.

TABLE 9: VA Health Services Research and Development Field Programs, 1987

| Region | Projected Spending, 1987 | Research Staff (FTEs) | Affiliated Universities (no.) | VA Medical Centers (no.) |
|---|---|---|---|---|
| Northeast | $1,500,000 | 22.4 | 3 | 23 |
| Mid-Atlantic | 1,078,603 | 8.0 | 2 | 1 |
| Southeast and Southwest (3, 7) | 846,794 | 18.6 | 5 | 4 |
| Great Lakes (4) | 1,116,500 | 15.5 | 4 | 5 |
| Medical District 17 | 1,263,893 | 16.2 | 5 | 3 |
| Medical District 22 | 690,027 | 10.2 | 2 | 4 |
| Medical District 23 | 220,776 | 2.0 | 3 | 1 |
| Far West (6) | 1,600,000 | 9.5 | 2 | 3 |
| Northwest (6) | 544,414 | 5.1 | 5 | 6 |
| Total | $8,861,007 | 123.0 | 31 | 50 |

SOURCE: AHSR, 1988 Directory.

TABLE 10: Robert Wood Johnson Foundation Clinical Scholars Program, 1987

| Program | Funding | Number of Scholars | Number of Alumni | Year Began |
|---|---|---|---|---|
| UCLA | $272,299 | 8 | 54 | 1975 |
| UCSF/Stanford | 274,223 | 11[1] | 72 | 1969 |
| UNC-Chapel Hill | 256,370 | 7 | 53 | 1974 |
| University of Pennsylvania | 575,818 | 11[2] | 54 | 1974 |
| University of Washington | 374,187 | 9[3] | 61 | 1975 |
| Yale University | 137,123 | 7[3] | 47 | 1974 |
| Total | $1,890,020 | 53 | 341 | |

[1] Includes four VA fellows.

[2] Includes one VA fellow.

[3] Includes two VA fellows.

SOURCE: Robert Wood Johnson Foundation, personal communication.

TABLE 11: The Pew Charitable Trusts Health Policy Program, 1987-1991

| Program | Current Grant (millions of $) | Current Fellows | Number of Alumni | Program Type* |
|---|---|---|---|---|
| Boston University/ Brandeis | $2.3 | 13 | 25 | 1,3,4 |
| Michigan | 1.9 | 23 | 9 | 1 |
| RAND/UCLA | 1.6 | 19 | 28 | 1,5 |
| UC-San Francisco | 1.6 | 9 | 39 | 2 |
| Total | 7.4 | 64 | 101 | |

* 1 = predoctoral; 2 = postdoctoral; 3 = corporate fellows; 4 = associate; and 5 = midcareer.

TRAINING OF PHYSICIAN/SCIENTISTS

*Lloyd H. Smith, Jr.**

This memorandum is directed to certain topics relating to the role of physician/scientists in the conduct of the nation's biomedical research program and how national policy should be shaped to ensure that a sufficient number of physician/scientists are available to play that role in the future. The memorandum does *not* concern itself with the training and supply of Ph.D. scientists, although this is clearly an increasingly convergent problem. Nevertheless, many of the points raised here may have more universal applicability to all scientists who direct their attention to biomedical investigation. The following questions are addressed briefly:

1. Is there a need for physician/scientists as opposed to or in addition to basic scientists in the conduct of biomedical science, both in our academic health science centers and wherever else that form of science is conducted?

2. Are there deficiencies in number or quality in the current and projected supply of physician/scientists?

3. How many physician/scientists should be trained?

4. What are the best methods for recruitment and training of physician/scientists?

5. How can the effectiveness of training programs be evaluated?

Other questions could be formulated. If answers to the above five questions could be approximated with a reasonable degree of accuracy, however, they might serve as a basis for a rational national policy.

**SPECIAL ROLE OF THE PHYSICIAN/SCIENTIST**

A generation ago "medical research" was carried out largely by physicians, most of whom had relatively little formal training

* The opinions expressed in this paper are the author's and do not necessarily reflect those of either the Committee on Biomedical and Behavioral Research Personnel or the National Research Council.

in science beyond that offered in the medical school curriculum. Problems were identified in clinical medicine and then pursued to whatever depth the investigator was able to penetrate from previous experience or from techniques acquired for that specific purpose. Few physicians sought rigorous training in biological science *per se* paralleling the demanding schedules required for obtaining and maintaining clinical competence. The science taught in specialty divisions of departments of medicine, pediatrics, neurology, and so on tended to be both goal-oriented and superficial. For a few clinical investigators of genius, this approach to medical research could be highly productive, but most medical investigators were poorly prepared for sustained scholarship. For some, NIH offered both a sanctuary and a postgraduate school for learning science in depth, and the success of that experience is evident everywhere today in American academic medicine. The "doctor's draft" of the past brought with it, almost as a gratuitous byproduct, a great improvement in the quality of medical research in the United States.

During the last 30 years, the scope, scale, and sophistication of research into human (and eukaryotic) biology have been enhanced enormously. As a result, it has become the domain of the professional scientist rather than of the inspired amateur. The study of human biology in the broadest sense has spread far beyond hospitals and even medical schools and is being pursued intensively and successfully in universities and institutes with no direct commitment to medical care (e.g., MIT, Cal Tech, Harvard College, and University of California-Berkeley) and by individuals who have not been trained in medicine. These trends are clearly evident and can be documented easily by analyses of the flow of NIH and NSF funds.

What do these trends mean for the application of current revolutionary advances in biology to the prevention and treatment of human diseases? This, after all, is the basic concern of the citizens who support this science. In partial answer, several points are worthy of emphasis:

1. The remarkable progress in basic biology brings with it parallel new opportunities for human application in "clinical investigation." The seminal retrospective studies of Cumroe and Dripps (1976), for example, showed quite clearly that technical advances directly applicable to the care of patients with cardiovascular and respiratory diseases depended heavily on basic research (about 40 percent in their arbitrary definition). In effect, there is no real discontinuity between the most fundamental basic biological science and the technology of modern medicine.

2. Clinical research is not simply that which is performed by a physician. Many basic scientists carry out excellent clinical research, and some of the research done by physicians is clearly basic science. It is difficult to supply an unambiguous definition of clinical research. As a beginning, one can say that it is research directed toward the elucidation and, therefore, the control of a human disease either directly or with a reasonable degree of intellectual continuity.

3. The physician/scientist is more likely to tilt in the direction of clinical research, as defined above, than is the basic scientist. This is stated as an article of faith, based on the following driving forces:

   a. He or she is more likely to be aware of the existing problems in human health and disease and also what is important to do and what is feasible to do. This awareness is not only the product of medical education but also of the imperatives of daily experience.

   b. He or she is more likely to have a primary commitment to this type of applied research because of intrinsic interest and also because of the reward systems in the environment in which he or she pursues a profession (department, school of medicine, national peer network, and societies).

4. The physician/scientist is of particular importance in medical schools in order to translate the current role and future potential of basic science during the education of future practitioners. Those currently in medical school and in postdoctoral training will have their main professional experiences in the twenty-first century. That fact implies enormous changes in the technological basis of medical practice in ways that cannot now be foreseen. Only those educated in the spirit of scientific inquiry can hope to remain abreast of this continuously changing frontier.

5. Uniquely in medicine, the university is involved directly in operating a major industry--that of health care. In all other disciplines the university is divorced from the direct practical application of what it learns and what it teaches. In this type of "industrial setting" that exists in the clinical departments of all medical schools, it is of particular importance to have those who can serve on the interface between what is best in science (the university role) and what is best in medical practice (the industrial

role). This can best be done by those who have authentic credentials in both domains.

For all of these reasons, and perhaps others as well, it is very important to maintain a strong contingent of physician/scientists in American medicine, recognizing that they may have to compromise somewhat in both areas of activity. They are unlikely to be the best clinicians in their respective hospitals because clinical skills are not honed in the laboratory, and they may not be the most productive scientists because they will need to commit extra time for maintaining basic clinical competency. Unless some are successful in such an amphoteric existence, however, medical schools will fragment into "two societies," as C. P. Snow described in another context some years ago.

**SUPPLY OF PHYSICIAN/SCIENTISTS**

Do we have a sufficient number of physician/scientists at the present time? The general consensus is that we do not, although this is an area in which it is difficult to obtain data. First of all, the physician/scientist is not a standard unit. Perhaps we should invent a new scale of "competency units" in which to attempt such measurements. The diminishing effectiveness of the physician (in comparison to the Ph.D.) in competition for NIH support has been well-publicized. Perhaps of greater concern is the difficulty encountered by every search committee that seriously attempts to identify physician/scientists for academic appointments within even our most prestigious, research-intensive medical schools. Within given disciplines (e.g., cardiology, gastroenterology, etc.), the number of physician/scientists who have attained some clinical competence and who could also meet the stringent criteria for scholarship for appointment in a basic science department is incredibly small. It is widely perceived that the paucity of such individuals who are competent in modern science and some phase of clinical medicine is the main limitation in our ability to move ahead quickly in modern clinical research. This is the most valid reason for the current move to create institutes or departments of "molecular medicine," or some variation on that theme, in order to obtain the climate in which such work can prosper. But the number of currently available physician/scientists who can do distinguished work in these areas is limiting. In summary, we may have a reasonable number of physicians who have participated in some form of research, but we do not have enough who have undergone rigorous preparation for scientific careers.

**HOW MANY PHYSICIAN/SCIENTISTS SHOULD BE TRAINED?**

An excess. This is not a facetious answer. Scientific ability does not parallel intelligence and hard work. Some very bright people are relatively ineffective as working scientists;

others less intellectually gifted are highly competent as scientists. Creativity in science cannot be predicted with any confidence; it can only be demonstrated over time. The cost of scientific training is comparatively modest when balanced against projected expenditures during a career in science. For all of these reasons, it is a better strategy to train an "excess" of scientists and then to continue to support only the best for subsequent research. The basic purpose of science policy, after all, is not to employ scientists but to ensure that the nation's goals in research and education be met. Therefore, the focus should be on the results obtained rather than on the number of scientists employed. There are many secondary gains for expanded training of physician/scientists, even for those who do not remain in this arena of activity. For science itself they furnish useful workers who contribute to the completion of the investigative projects in which they are being trained. Of equal importance, the individual benefits from learning further about scientific method and thereby is better prepared to evaluate medical progress and to learn to adapt to a career in the practice of medicine for the twenty-first century. Extra scientific training for the physician should not, therefore, be considered as a waste of resources.

The ability of our nation to predict future work force needs has not been impressive in science or in medicine. It is patently impossible to predict with any degree of precision the needs for a system that is undergoing vast fluctuations. Because of this, it is generally more prudent to overshoot rather than undershoot, particularly in view of the fact that the process of training a physician/scientist is a very long one, perhaps eight to ten years. This process is, therefore, inherently sluggish in its response to market forces or attempted social engineering.

Even if it were possible to predict with some precision the number of future available positions for physician/scientists in academia, industry, public service, and other venues, this figure would not furnish a rational basis for the number to be trained. As noted previously, investigators vary widely in productivity and creativity and therefore do not represent a standard product. There is probably a finite pool of candidates who are capable of the highest level of sustained creativity. The real challenge in scientific work force policy is to identify the candidates and to bring them to fruition as scientists, realizing that this can be done only by training many and choosing the best from among them.

**Methods for Recruiting and Training Physician/Scientists**
**Current Barriers**

It might be useful first to consider briefly the current barriers to attracting the best young physicians into scientific careers. Since there are few useful data concerning this theories abound. Some are as follows:

1. The deterrent effects of personal indebtedness. Here data are available concerning the increasing indebtedness of medical school graduates (about $30,000-$40,000 for public schools and about $40,000-$60,000 for private schools, with wide ranges beyond these averages). Such debts clearly reduce the ability of the student to electively undertake yet further training at low compensation directed toward a career with modest long-term fiscal rewards.

2. The increasing stretch in trying to maintain competency in both science and medicine as both become more demanding. This creates continuing tensions for the individual who attempts a career that bridges both domains.

3. The perceived insecurities of academic life based on the vagaries of short-term extramural funding for research. Current investigators transmit these insecurities to those who are considering similar careers.

4. The current "industrial turmoil" in clinical departments as they contend with the demands for health care delivery and financial viability in a time of change. The resulting harassment is not lost upon the young.

5. Possibly changing levels of expectation. More students are married or have other elective obligations than in the past and may be less willing to undergo the longer periods of relative privation to prepare for a research career.

6. The attractiveness of medical practice. Some of this attractiveness relates to the immediacy of personal gratification in helping other human beings as opposed to the qualitatively different and often deferred rewards of medical research.

There are many other factors as well, including the absence or paucity of suitable role models, jam-packed medical school curricula based on fact engorgement rather than on problem solving, and the end of the doctor's draft. Obviously, all of these deterrents will occur in a different silhouette for each student.

**General Principles of Training**

There is no single best method of training the physician/scientist (i.e., best for each individual and for every setting).

It is noted that most scientists tend to recommend the type of pathway that guided their own careers. This topic can be considered usefully under four headings:

1. **Entrapment of the young:** Ideally, there should be a period of research prior to or during medical school in order for the student to decide whether he or she enjoys this activity and is good at it (the two factors generally coexist). Time should be made available for this in the medical school curriculum or in arrangements for summer laboratory experiences. It is important for every student to learn something in depth--to penetrate to the frontier of knowledge, however narrow the subject may be. For some the intense trial period for research may be extended usefully for up to a year. Such a period is necessary in order for both the student and the sponsoring agency or department to decide whether a more extensive investment of time and money is warranted in his or her training.

2. **Training in depth:** It is imperative that the serious physician/scientist receive training in depth in a scientific discipline relevant to medicine. It is both inaccurate and arrogant to assume that the intensive professional training of a physician prepares him or her to compete in modern biological science with a scientist who has undertaken the rigorous discipline of a Ph.D. degree. Rarely can this type of training in science be achieved in a specialty division of a clinical department (although there are some exceptions here). Whether this training fulfills the formal requirements for a Ph.D. degree probably is immaterial, but it should be comparable to such a program in rigor and scope. During this time, the physician should not attempt to carry out parallel clinical duties, which will only serve as a diversion from his or her critical opportunity to become a serious scientist (or to demonstrate to himself or herself and to others that such a career is not his metier).

3. **Coordination of scientific and clinical training:** The coordination of clinical and scientific training is a vexing problem that defies an easy solution. This simply presages, however, the future problems that the physician/scientist will have throughout his or her subsequent career in coordinating a role as both a physician and a scientist. In general, there are two approaches, each with some advantages and disadvantages:

a. *Pre-M.D. scientific training*: The prototype here is the M.D./Ph.D. program. The advantages are an earlier commitment to science in depth and continuity with the standard basic science of the medical school curriculum. Another advantage is that the individual has this extra scientific competency at his or her command during the intensive clinical experiences of the senior year of medical school and the years of residency training. The obvious disadvantage is the enforced hiatus (usually at least three years) between the completion of thesis work and the chance to return to full-time science. Many of these same considerations would hold for more extensive non-Ph.D.-directed scientific experience during medical school or for the individual who has obtained a Ph.D. prior to entering medical school.

b. *Post-M.D. scientific training*: The prototype here is NIH or NIH-equivalent experience that many current investigators obtained in the past. The advantages are that at this stage of maturity the individual is perhaps more likely to know his or her future plans and also that the work can continue without serious interruption following the training period. A theoretical disadvantage is that the individual may by now have firmly decided upon a clinical career without having had a serious look at the alternative of research. In addition, accumulated debts may by this time have diminished the feasibility of investing the necessary additional years for research training.

   Either model can be effective as demonstrated by those in academic medicine today. In this brief outline above, little consideration was given to the financial implications of these alternatives, since these would vary depending on the existing means of support--for example, the full support provisions of the federally sponsored MD/PhD program.

4. **Launching a career:** The beginning years of a career as a physician/scientist are of vital importance as the individual attempts to establish an independent program as a scholar. During this transitional period, such individuals require protected time and should not be required to participate as heavily in departmental activities (patient care, teaching, and university service) as do those who are not making a similar effort to bridge basic science and medical research.

This topic is directly pertinent to the future supply and viability of physician/scientists in clinical departments, but it is not discussed here further.

## MODELS FOR TRAINING IN DEPTH

As noted previously, there are many possible models for training for research for the physician/scientist. Individuals with exceptional motivation and talent can develop successful scientific careers through any one of these or other pathways.

1. **M.D./Ph.D. programs:** In the M.D./Ph.D. programs the student pursues a regular Ph.D. degree, usually after the second year of the standard medical school curriculum, in one of the basic science departments of his or her institution. The Ph.D. training often can be somewhat truncated (three years) because of the student's previous course in science, but it is usually not shaped specifically for the physician-to-be. These programs have been judged to be highly successful as judged by the number of participants who have remained active in science and who have achieved independent research funding through peer review mechanisms. The pros and cons of this mode of training for research have been listed previously. Data concerning the outcome of NIH programs are readily available. It is the general opinion that this pathway should be retained and expanded. Not least of the positive features is that the trainee usually is able to avoid or attenuate the burden of personal debt. Since these programs are well-defined and have been analyzed thoroughly, the M.D./Ph.D. pathway is not considered further here. It seems clear that such programs should continue to receive high national priority.

2. **Postdoctoral fellowships for the physician/scientist:** At least three years of rigorous training in modern biological science usually is necessary for most individuals, however gifted, to arrive at a stage of independence as an investigator (Goldstein, 1986). In fact, the period of time may be longer and require the equivalent of the "post-doc" experience that is now *de rigueur* for Ph.D. recipients. For those who are training for a career in research, this time should not be diluted with simultaneous clinical responsibilities, which inevitably serve to divert attention and energies elsewhere. Preferably, this experience should be in an active basic science laboratory that is on the cutting edge of some discipline that is ultimately applicable to medical research and that is in the usual predoctoral, postdoctoral climate of competitive ideas and productivity. Since the boundaries of modern

biological science are pervious, it is not particularly important whether the nomenclature of the department is that of cell biology, molecular biology, biochemistry, or microbiology. The same powerful methodologies of working with peptides, proteins, receptors, DNA, RNA, monoclonal antibodies, and so on can be acquired and will have ultimate applicability to virtually any problem relating to the pathogenesis and treatment of human disease. Beyond the methodologies--and even more vital--the nascent physician/scientist will be honed in the rigors of defining problems and learning how they can be approached rationally. The physician in training to become a physician/scientist should not be concerned that the problems being pursued seem excessively fundamental and far from human biology. It is always easier to move from that which is very basic to that which is applicable than to try to learn selective elements of experimental biology as the need arises in clinical investigation.

One approach has been to establish a program of national fellowships for the postdoctoral training of the physician/scientist. It is useful to analyze the current program of the Markey Foundation (which is scheduled to be phased out over the next few years). In this program provision is also made for some assistance for the physician/scientist at the entry level of beginning faculty membership. Similarly, the Physician/Scientist Award program of NIH has endorsed this approach over the past five years, so that experience is beginning to accumulate about its effectiveness.

3. **Postdoctoral training programs for the physician/ scientist:** The M.D./Ph.D. program and the individual fellowship program for training in science were described briefly above. In general, there has been considerable experience with both modes of training. Formalized *training programs* for physicians usually have been established as appendages to specialty divisions of clinical departments (e.g., in gastroenterology, cardiology, endocrinology, etc.). The scientific training that such divisions can supply rarely approaches that available in basic science departments. Also, it is diluted frequently by simultaneous training in the parent medical discipline. It is not surprising, therefore, that many of the physicians so trained quickly leave investigative careers to enter the practice of their respective specialties. Although this process has continued to ensure a supply of scientifically trained clinicians, teachers, and clinical investigators, it has not

sufficed to develop physicians who are sufficiently well-grounded in the fundamental sciences for sustained scholarship.

These considerations have led to proposals for new forms of institutional postdoctoral training programs that would be shaped more specifically for the production of physician/scientists. Such a successful program should contain the following elements:

a. It should represent a true consortium between the clinical and preclinical departments of the institution with equal responsibility for the design and administration of the program.

b. Selection of the trainees should be made during the senior year of medical school, based on evidence of some previous experience in research and overall promise. Selection and planning then can be coordinated for both basic clinical training and subsequent scientific training. It would be advantageous to have the basic science departments participate in the selection process in order to ensure their commitment to those who enter the coordinated program.

c. Formal course work in the physical and biochemical sciences should be an integral part of any such program so that its graduates command a theoretical background comparable to that obtained by those with graduate degrees in the biological sciences. The extent of the required course work can be individualized based upon the level of prior training, but it should be rigorous and at the graduate level in the relevant disciplines (biochemistry, genetics, molecular biology, cell biology, etc.).

d. The training program should be for not less than three years, of which most will be invested in direct research experience under the supervision of a mentor.

   At the completion of this training period, which often may be extended beyond the formal three-year program, most of the physician/scientists will rejoin their respective clinical departments for subspecialty clinical training in their chosen disciplines. Some will elect to remain in basic science and will enrich those disciplines with their breadth of training and interest in human biology.

## EVALUATION OF PROGRAMS

Science is what scientists do. The logical approach would be to evaluate the scientific achievements of those who have completed a given type of training. Obviously, this is much more difficult to do than it seems at first. It is usual, therefore, to use as a surrogate a number of indirect indices, such as the number who remain in an academic environment, those who have successfully attained an RO-1 grant, or those who have remained active in research (as judged usually by grant activity) after some arbitrary period of time. None of these measures is very satisfactory since the quality of science is not appraised, other than its success at securing competitive funding. In a sense, the appraisal has been transferred to the respective study sections.

One can use other indices of peer recognition, such as election to elitist scientific societies, or one can use citation criteria or arbitrary assignments of value to publication in certain of the more stringently reviewed journals. Ultimately, most evaluations are highly subjective. Furthermore, they cannot be readily made except over extended periods of time in view of the natural history of a scientific career.

There must, therefore, be an element of faith. A program that enhances the exposure of students to science as a creative process (rather than a repository of inert facts) and that encourages and supports those who wish to pursue rigorous scientific training in depth must be assumed to increase the chance that those who are most qualified will more likely pursue careers as physician/scientists. Intuitively it seems so. NIH experience of the years of the doctors draft seems to support this supposition. It is not a rigorous method of assessment, but it may have to suffice. Furthermore, as noted previously, those who fall out of the system, that is, do not have sustained careers as productive physician/scientists--cannot be judged flatly as failures since they may well be more capable of adapting to the complexities of the practice of medicine in the twenty-first century.

## SUMMARY

In a sense, this brief position paper is in itself a summary of an opinion without formal documentation. Several points seem worthy of emphasis in future policy concerning the training of physician/scientists:

1. There is a special need for physician/scientists in the conduct of the nation's biomedical research. The increasing interest and involvement of Ph.D. scientists in medical investigation are of great importance and

benefit, but this does not replace the need for physician/scientists.

2. The number of well-qualified physician/scientists is in short supply now and will continue to be limiting in current projections. As a result, there are missed opportunities for the rapid application of modern biology to medical problems.

3. Our ability to predict work force needs is inexact, based on past record, and the responsiveness of the production system is sluggish inherently because of the long pipeline and other variables. Therefore, it is wiser to overshoot rather than undershoot in projections of work force needs.

4. The physician/scientist is not a unit of scientific competency and productivity. The nation should be less interested in the exact number of FTEs employed than in the pace of scientific progress. A wise policy therefore would be as follows:

    a. to train larger numbers of physician/scientists but retain only the best in scientific careers; and

    b. to accept the fact that this investment in training gives value received even for those who do not remain in science:

        o The cost of training is low in comparison with the ultimate investment in the scientific work of those supported; hence, it pays to allow for the choice of the best, based on performance in research.

        o The "trainees" are, in fact, modestly reimbursed laboratory workers who contribute great value to the direct conduct of research during their so-called training periods.

        o Those who leave science return to medicine better prepared to apply more critically the science of the future in their respective professional careers.

5. Four phases of training that can be defined arbitrarily:

    a. *Entrapment*: Early research experience for the medical student will help to define those who have interest and competency in research.

b. *Training in depth*: This requirement is noted below.

c. *Coordination with clinical training*: The physician/scientist must balance proficiency in both domains.

d. *Launching of an independent scientific career*: The early years of a scientific career require special consideration and protection.

6. Training in depth usually requires at least three years of experience in some phase of basic biological science. This is absolutely necessary if the physician is going to be able to serve effectively on the frontier between the best in science and the realm of medical research. Rarely can this be achieved in a clinical department as they are now constituted.

7. A number of pathways have been successful in offering to physicians this kind of training for proficiency in science. All of these models, which were described briefly here, can offer training in depth and should be continued in a balanced national program for training physician/scientists:

   a. the M.D./Ph.D. programs;

   b. national fellowship programs; and

   c. postdoctoral training programs for physician/scientists.

   The last category of specific training programs probably warrants the most attention, since such programs currently are less well-defined.

**REFERENCES**

Cumroe Jr., J. H., and R. D. Dripps. 1976. Scientific basis for the support of biomedical science. *Science* 192:105-111.

Goldstein, J. L. 1986. On the origin and prevention of PAIDS (Paralyzed Academic Investigator's Disease Syndrome): Presidential address delivered before the 78th annual meeting of the American Society for Clinical Investigation, Washington, D.C., May 3, 1986. *Journal of Clinical Investigation* 78:848-854.

# BIOMEDICAL/BEHAVIORAL COHORT MODEL: A TECHNICAL PAPER

*Joe G. Baker**

## INTRODUCTION

The legislation prompting this study requires the assessment of ". . . the nation's overall need for biomedical and behavioral research personnel. . . ."[1] Past committees have defined this "need" in labor market terms--that is, how many biomedical and behavioral researchers will be "needed" in the future? The purpose of this paper is to document the model used to project the future labor market for biomedical and behavioral scientists. Future labor market conditions are defined in terms of improvements or deteriorations from historical market conditions, and future demand conditions are examined in the context of the appropriate National Research Service Awards Acts variable--that is, Ph.D. production and postdoctoral study. Following a descriptive overview of the model used are detailed discussions of the various model components, data, and coefficient estimates.

## THE MODEL: AN OVERVIEW

Past committee projections of the future need for biomedical and behavioral scientists have focused on academic demand. Job openings were developed based upon growth in academic positions and openings resulting from faculty death, retirement, and field switching. These projections were developed for the near term, the 1985 committee report included projections to 1990.

This model expands the earlier analysis in several ways:

1. In almost every biomedical and behavioral field, the major source of historical and projected employment growth is in nonacademic sectors, primarily private industry. Thus, this model includes industry, government, hospital, and other nonacademic sources of demand for biomedical and behavioral scientists.

---

* The opinions expressed in this paper are the author's and do not necessarily reflect those of either the Committee on Biomedical Research Personnel or the National Research Council.

---

[1] Section 489 of P.L. 99-158.

2. The focus of the NRSA program is research personnel. This analysis develops separate projections for the labor market in general and that subsection of the labor market associated with scientists whose primary work activity is research and development (R&D) or the management of R&D.

3. Given concern over "graying" of the work force, this analysis includes a demographic/economic model for estimating scientists' attrition due to death, retirement, and net occupational movement based upon the age and experience structure of the scientific work force.

4. This analysis includes a model of labor supply.

5. Given that the median time to complete the biomedical sciences Ph.D. has grown from seven years in the late 1970s to eight years in 1987, the 1997 biomedical scientist labor market will be influenced by student decisions and NRSA policy in 1989. In the behavioral sciences, median time to Ph.D. has increased from approximately 8.5 years to 10.5 years during the same period. These types of lags argue for a longer time horizon of analysis; the current study projects labor market variables to the year 2000.[2]

Figure 1 is a schematic of the labor market assessment model. The stock of scientists in time period *t* is characterized by biological age (years since birth) and career age (years since degree). Historical data provide estimates of the number of deaths and retirements by biological age; these scientists are removed from the scientist stock. Those who do not retire or die can leave the field for other employment; this is assumed to be a function of career age. These estimates of outmigration are net of inmigration from other fields and are estimated from historical data. The surviving scientist stock is available for employment in period *t*+1. The required scientist stock in period *t*+1 is estimated from submodels that link demand for scientists to the demand for the good or service that scientists produce--for example, R&D or graduate students. These demand submodels vary by discipline (biomedical, behavioral, clinical), work duties (R&D and non-R&D), and sector (academic, industrial, governmental). The difference between the surviving scientist stock and the required scientist stock in period *t*+1 is the job openings (vacancies) that must be filled by new entrants. These

[2]See National Research Council, *Summary Report 1987 Doctorate Recipients from United States Universities* (Washington, D.C.: National Academy Press, 1989), Table P.

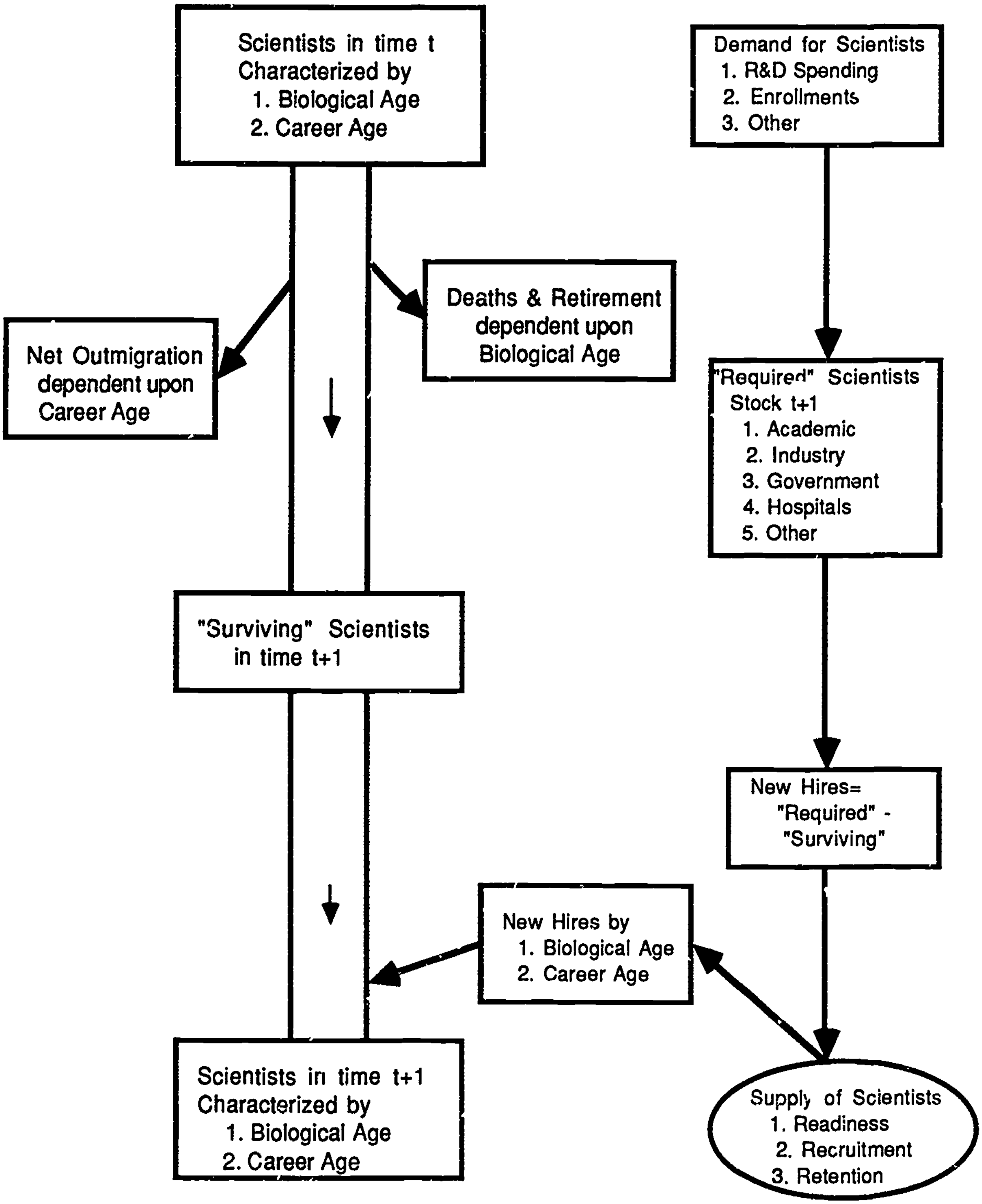



**Figure 1.** Schematic of the Labor Market Assessment Model

job vacancies are compared to supply to compute "vacancy ratios" --that is, the number of vacancies per new Ph.D. or postdoctorate. Increases or decreases in future vacancy ratios from historical ratios give one a sense of tightening or loosening of the projected demand/supply balance in the scientist labor market. This same basic analysis is replicated for the R&D subsector.

## SCIENTIST ATTRITION: DEATH, RETIREMENT, AND NET MOBILITY

The model used to estimate attrition is a demographic/ economic cohort survival model that follows closely the work of Kuh, Radner, and Fernandez.[3] The nature of a cohort survival model is to essentially "march" each cohort of scientists through time and apply age-specific rates of death, retirement, and net outmobility as the cohort ages. The population described here is that of basic biomedical scientists. Separate coefficients were estimated for behavioral scientists and are included in the Appendix tables.

Three types of annual scientist attrition are addressed by the model: death, retirement, and net outmobility. The annual number of deaths and retirements are assumed to be a function of an individual's "biological" age (B). Annual net outmobility is defined as the number of scientists who leave biomedical science for other occupations minus the number of workers who enter biomedical science from other occupations. Net outmobility is assumed to be a function of a scientist's "career" age (C). It is unlikely, for example, to see a large net outmobility rate in the younger career ages because scientists are unlikely to move out of an occupation that has consumed an average of 8 years of training. Later in their careers, a segment of the scientists will more likely move out of science and into administration or management.

In addition to career patterns, scientist mobility could be influenced by labor market characteristics. For example, if the labor market deteriorated and job opportunities and wage growth were depressed, scientist outmobility would probably increase. However, historical data indicate very little variation in scientist unemployment and underemployment rates through time. Also, it could be argued that scientists would be unlikely to

---

[3]Charlotte Kuh and Roy Radner, *Mathematicians in Academia: 1975-2000* (Washington, D.C.: Conference Board on the Mathematical Sciences, 1980); and Luis Fernandez, *Project on Quantitative Policy Analysis Models of Demand and Supply in Higher Education* (Washington, D.C.: Carnegie Council on Policy Studies in Higher Education, 1978).

leave a career in reaction to short term cyclical economic conditions. In the model described here, these labor market influences are assumed to be small compared to career patterns as determinants of net mobility, i.e., they are ignored.

Each scientist in the model is defined by biological age and career age. Although scientists can enter the system at any career age, those who enter at career ages other than one are incorporated into the net outmobility rates. For practical purposes, all new entrants enter the system at career age one. The parameters of the model include:

- Initial biological and career age distribution,
- Biological age-specific death and retirement rates, and
- Career age-specific net outmobility rates.

Obviously, there is a high degree of correlation between biological age and career age. Because of this, one can estimate fairly accurately the biological age distribution of a group of scientists given their career ages. If one assumes that the biological age distribution of newly minted Ph.D.s is stable, one can construct a model whose states (biological age and career age) depend only upon the career age distribution. At any time, the biological age distribution can be estimated from the career age distribution and the model can be simplified.[4]

**Data**

Except where noted, all data are from the Survey of Doctorate Recipients (SDR) sponsored by the National Science Foundation and the National Institutes of Health. These data are collected by the National Research Council biannually from a cross-section sample of approximately 10 percent of the U.S. Ph.D. scientist population.

**Initial Career and Biological Distribution**

Table 1 is a cross-tabulation of the distribution of biomedical scientists in 1987 by career age and biological age. Given that a biomedical scientist is $C_{<=5}$ ; then $P(B_{<30}) = .115$. These probabilities allow the estimation of biological age from a given career age.[5]

---

[4]See Fernandez, *op. cit.*, pp.132-33.

[5]Career and biological age cross-tabulations for behavioral scientists are contained in the appendix.

Table 1. Career Age and Biological Age Distribution of Biomedical Scientists, 1983-1987

| | Career Age | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Bio Age | <=5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | >=41 | Total |
| <30 | 1918 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1939 |
| 30-34 | 9463 | 4650 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | 14156 |
| 35-39 | 3748 | 8547 | 4908 | 8 | 0 | 0 | 0 | 0 | 0 | 17211 |
| 40-44 | 977 | 3298 | 8618 | 5161 | 0 | 0 | 0 | 0 | 0 | 18054 |
| 45-49 | 280 | 678 | 2535 | 6340 | 2357 | 57 | 0 | 0 | 0 | 12247 |
| 50-54 | 124 | 204 | 907 | 2084 | 2943 | 1151 | 0 | 0 | 0 | 7413 |
| 55-59 | 81 | 101 | 274 | 744 | 1330 | 2037 | 1228 | 77 | 0 | 5872 |
| 60-64 | 3 | 61 | 77 | 181 | 511 | 1048 | 1702 | 474 | 6 | 4063 |
| 65-69 | 10 | 0 | 16 | 84 | 91 | 95 | 376 | 566 | 219 | 1457 |
| >=70 | 0 | 0 | 8 | 0 | 13 | 22 | 57 | 174 | 88 | 362 |
| Total | 16604 | 17560 | 17386 | 14602 | 7245 | 4410 | 3363 | 1291 | 313 | 82774 |

Relative Distribution

| | Career Age | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Bio Age | <=5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | >=41 |
| <30 | 0.1155 | 0.0011 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30-34 | 0.5699 | 0.2648 | 0.0024 | 0 | 0 | 0 | 0 | 0 | 0 |
| 35-39 | 0.2257 | 0.4867 | 0.2822 | 0.0005 | 0 | 0 | 0 | 0 | 0 |
| 40-44 | 0.0588 | 0.1878 | 0.4956 | 0.3534 | 0 | 0 | 0 | 0 | 0 |
| 45-49 | 0.0168 | 0.0386 | 0.1458 | 0.4341 | 0.3253 | 0.0129 | 0 | 0 | 0 |
| 50-54 | 0.0074 | 0.0116 | 0.0521 | 0.1427 | 0.4062 | 0.2609 | 0 | 0 | 0 |
| 55-59 | 0.0048 | 0.0057 | 0.0157 | 0.0509 | 0.1835 | 0.4619 | 0.3651 | 0.0596 | 0 |
| 60-64 | 0.0001 | 0.0034 | 0.0044 | 0.0123 | 0.0705 | 0.2376 | 0.5060 | 0.3671 | 0.0191 |
| 65-69 | 0.0006 | 0 | 0.0009 | 0.0057 | 0.0125 | 0.0215 | 0.1118 | 0.4384 | 0.6996 |
| >=70 | 0 | 0 | 0.0004 | 0 | 0.0017 | 0.0049 | 0.0169 | 0.1347 | 0.2811 |
| Total | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

SOURCE: 1983, 1985 and 1987, Survey of Doctorate Recipients, NRC.

Table 2. Death and Retirement Rates

| Bio Age | Retire Rate | Death Rate |
|---|---|---|
| <30 | 0 | 0.0011 |
| 30-34 | 0 | 0.0017 |
| 35-39 | 0 | 0.0027 |
| 40-44 | 0 | 0.0038 |
| 45-49 | 0 | 0.0063 |
| 50-54 | 0 | 0.0114 |
| 55-59 | 0.0026 | 0.0179 |
| 60-64 | 0.0753 | 0.0271 |
| 65-69 | 0.1714 | 0.0378 |
| >=70 | 1 | 0.05 |

SOURCE: Charlotte V. Kuh and Roy Radner, Mathematicians in Academia: 1975-2000, A Report to the National Science Foundation, Washington, D.C.: Conference Board of the Mathematical Sciences, 1980, pp. 84-86.

## Death and Retirement Rates

Table 2 contains estimates of death and retirement rates. These rates are for academicians *in toto*.

## Net Mobility Rates

Conceptually, the net outmobility rates of scientists are simply the gross occupational exits minus the gross occupational entrants by career age. Practically, these longitudinal movements are extremely difficult to measure given existing data sets. Therefore, a shortcut method based upon a cross-section of scientists by career age was used[6] to estimate these movements. As shown in Table 3 the values of each cell are the sum of SDR survey years 1983, 1985, and 1987. For example, the number of biomedical scientists of $C_{<=5} = 31,538$: $C_{<=5}$ for 1987 (10,823) + $C_{<=5}$ for 1985 (10,720) + $C_{<=5}$ for 1983 (9,995). Putting aside the debate of whether postdoctorates are students or employed and counting them as employed, the total pool of biomedical scientists $C_{<=5}$ is 48,847 (31,538 + 17,309). The total number of Ph.D.s in all fields except humanities and engineering with $C_{<=5}$ is 202,517; this is assumed to be the primary supply source for biomedical scientists. Because these estimates are for attrition other than death and retirement, all scientists who will retire must be removed from the labor supply pool (column 6, Table 3), giving a labor supply pool of 202,378 for $C_{<=5}$. Thus, for every 1,000 scientists of $C_{<=5}$, 241 are biomedical scientists [(31538 + 17309)/(202378)]. For every 1,000 scientists of $C_{6-10}$, 231 are biomedical scientists. Assuming that this cross-section reflects the longitudinal movement of scientists, then the net outmobility rate of $C_{<=6}$ is 4.1 percent [(241-231)/241). This rate is for a five-year period and must be annualized: assuming that the average time of transition for a given scientist $C_{<=6}$ to $C_{6-10}$ is 2.5 years, then the annualized net outmobility rate is 1.79 percent (i.e., 98.21 percent of biomedical scientists of $C_{<=6}$ "survive" into the next year, while 1.79 percent move into nonbiomedical fields).[7]

## Model Structure

Given the above parameters, the operation of the model is fairly straightforward. At a point in time *t*, a group of scientists S are identified by their career age C. Death and retirement are estimated from this group by the following

---

[6]This shortcut method was suggested by Robert Dauffenbach, a member of the Committee on Biomedical and Behavioral Research Personnel.

[7]Of course, this survival rate is the difference between larger gross outmobility and gross immobility.

equation:

(1) $P(DR_r) = \Sigma(n) \; P(B_n|C_r)*P(DR_n|B_n)$

Where

$P(DR_r)$ = the probability of death or retirement given career age r;
$P(B_n|C_r)$ = the probability of biological age = n given career age = r;
$P(DR_n|B_n)$ = the probability of death or retirement given biological age = n;

and $\Sigma(n)$ sums these probabilities across all biological ages n.

Given that an individual does not retire or die, he or she can now leave the biomedical work force for other occupations. The total attrition rate for career age = r is thus:

(2) $P(A_r) = P(OM_r)*[1-P(DR_r)]$

where

$P(A_r)$ = conditional probability of total attrition given career age r;
$P(OM_r)$ = simple probability of net outmobility for career age = r
$[1-P(DR_r)]$ = probability of *not* retiring or dying given career age r;

. . . the total attrition estimate for all employed biomedical scientists is:

(3) $TA = \Sigma(r) \; S_r*P(A_r)$

where

TA = total attrition from death, retirement, and net outmobility and
$S_r$ = the number of scientists career age r.

Thus, for each group of scientists career age r, the estimated attrition leaves a surviving group of scientists in period t+1 that moves to the next career age r+1. Summing across all career ages gives the total attrition for the stock of scientists S and allows for the computation of the total surviving scientists in period *t+1* (Figure 1).

## DEMAND FOR SCIENTISTS

As shown in Figure 1, the demand for scientists is derived from the services that they produce--for example, R&D and teaching graduate students. Other factors, in particular wage rates, can also affect the level of scientist demand. The demand models used in this analysis--simple forms of this conceptual model[8] estimated for each field (biomedical and behavioral), sector (academic, industrial, hospital,) and work activity (R&D and non-R&D)--are included in the Appendix. Table 4 contains the detailed projections for biomedical scientists; the Appendix contains the detailed tables for nonclinical psychologists and other behavioral scientists. The projection model assumptions were developed by the Committee on Biomedical and Behavioral Research Personnel.

Table 3. Estimated Quit Rates for Biomedical Scientists, 1983-1987

| Career Age | (1) Biomed | (2) Biomed R&D & Mgt R&D | (3) Postdoc Biomed | (4) Retired | (5) Total Ph.D.s | (6) Total Ph.D.s -retire | (7) Bio Sci per 1000 Sci | (8) Annual Quit Rate | (9) Annual R&D Quit Rate |
|---|---|---|---|---|---|---|---|---|---|
| <=5 | 31538 | 19695 | 17309 | 139 | 202517 | 202378 | 241 | | |
| 6-10 | 48171 | 29975 | 2419 | 373 | 219649 | 219276 | 231 | -1.79% | -8.18% |
| 11-15 | 49162 | 26230 | 531 | 990 | 216355 | 215365 | 231 | 0.00% | -6.69% |
| 16-20 | 34190 | 16674 | 135 | 2279 | 156309 | 154030 | [illegible] | -1.38% | -5.06% |
| 21-25 | 18257 | 8406 | 24 | 4041 | 90762 | 86721 | 211 | -2.20% | -4.52% |
| 26-30 | 12726 | 5383 | 36 | 6428 | 66917 | 60489 | 211 | 0.03% | -3.21% |
| 31-35 | 8567 | 3790 | 24 | 13162 | 55630 | 42468 | 202 | -1.67% | 0.10% |
| 36-40 | 2743 | 1095 | 7 | 11772 | 26291 | 14519 | 189 | -2.60% | -6.51% |
| 41+ | 1295 | 495 | 0 | 6507 | 11286 | 4779 | 271 | 15.40% | 13.24% |
| Total | 206649 | 111743 | 20485 | 45691 | 1045716 | 1000025 | | | |
| | | | | | Average annual quit rate | | | -0.8% | |

SOURCE: 1983, 1985, and 1987 Survey of Doctorate Recipients, National Research Council.

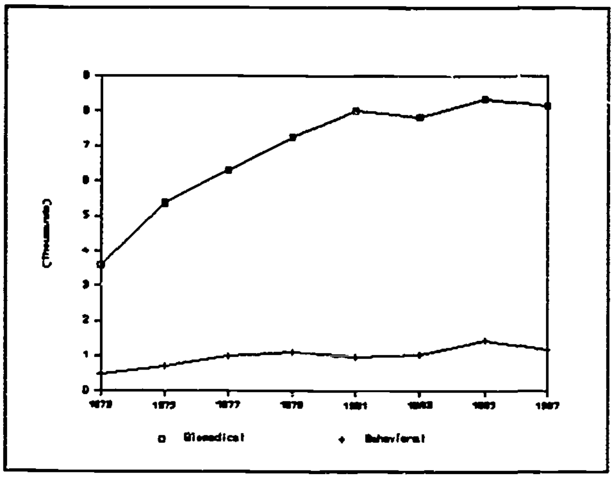
### Postdoctorates

There are no explicit projections for postdoctorates. The level of postdoctorate employment is assumed to be an institutional variable determined by public policy and, as a simplifying assumption, the level of postdoctorate employment was

[8]For a model of biomedical scientist demand that includes factor prices in both the demand and supply equations, see Joe G. Baker, "The Ph.D. Supply Crisis: A Look at the Biomedical Sciences," paper given at the June 21, 1989, Western Economics Association Meeting, Lake Tahoe, Nevada. It should be noted that the inclusion of factor prices did not change the salient results of the analysis.

held constant during the time frame of analysis. Thus, the level of postdoctoral employment does not affect the level of total employment; and given that most postdoctorates are young, very little labor supply is assumed to be lost from death, retirement, or outmobility.

If the postdoctorate "pool" is assumed constant with little or no attrition, then one can further assume that the annual exits from this pool equal entrants, and the net effect on labor supply to positions outside the pool is small. If the pool is changing in size, then entrants and exits will not be equal. As can be seen in Figure 2, the postdoctoral pool has been flat since 1981 at approximately 8,100 biomedical scientists. In the behavioral sciences, postdoctoral employment has been flat since 1979 at approximately 1,100 scientists.




SOURCE: Survey of Doctorate Recipients.

**Figure 2.** Postdoctoral Employment in Biomedical and Behavioral Sciences, 1973-1987

**Analysis: Vacancy Ratics**

The model discussed above estimates annual job openings from death, retirement, net outmobility, and growth. As a means to summarize the relationship between these job openings and labor supply and to provide information about the future labor market, the concept of "vacancy ratios" was developed. Simply put, the vacancy ratio is the average number of job openings per new Ph.D. For the period 1983-1987, the number of job openings from death, retirement and growth averaged 4846 annually (Table 5). The average number of new Ph.D.s in the biomedical sciences produced annually for this same period was 3862. Thus, the "vacancy ratio" was 1.25 openings per each new Ph.D. An "R&D vacancy ratio" is calculated by comparing R&D job openings to postdoctorate production.

Obviously, not all new Ph.D.s in biomedical science go into the biomedical field; also the field draws Ph.D.s from other areas (e.g., physical sciences and other life sciences) including foreign scientists. However, the primary source of new Ph.D.s is U.S. graduates in the field and, thus, the vacancy ratio gives one a sense of the historical relationship between this supply source and demand. No value judgments are made in terms of what the "correct" vacancy ratio is; the projected vacancy ratios simply provide information about the relative state of the labor market under differing assumptions.

Table 4. Summary Projections in Biomedical Sciences

Biomedical Sensitivity Model

| Model Assumptions | High | Mid | Low |
|---|---|---|---|
| 1. Federal Health R&D Funding Growth | 4.0% | 2.7% | 0.2% |
| 2. Private Health R&D Funding Growth | 13.0% | 9.0% | 5.0% |
| 3. Other Health R&D Funding Growth | 4.0% | 3.0% | 2.0% |
| 4. Grad and Undergrad Biomed Enrollment | 1.0% | 0.0% | -1.0% |
| 5. "Other" biomed R&D Employment Growth | 3.5% | 2.5% | 1.5% |
| 6. "Other" biomed non-R&D Employment Growth | 10.0% | 8.0% | 5.0% |

Projected Employment of Biomedical Scientists, 1973-2000
(in 100s of workers)

| | Low Case | | | Mid Case | | | High Case | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | % R&D |
| 1973 | 396.4 | 197.8 | 49.9% | 396.4 | 197.8 | 49.9% | 396.4 | 197.8 | 49.9% |
| 1975 | 447.2 | 211.8 | 47.4% | 447.2 | 211.8 | 47.4% | 447.2 | 211.8 | 47.4% |
| 1977 | 479.2 | 237.9 | 49.6% | 479.2 | 237.9 | 49.6% | 479.2 | 237.9 | 49.6% |
| 1979 | 543.8 | 284.5 | 52.3% | 543.8 | 284.5 | 52.3% | 543.8 | 284.5 | 52.3% |
| 1981 | 600.1 | 316.2 | 52.7% | 600.1 | 316.2 | 52.7% | 600.1 | 316.2 | 52.7% |
| 1983 | 626.4 | 322.1 | 51.4% | 626.4 | 322.1 | 51.4% | 626.4 | 322.1 | 51.4% |
| 1985 | 703.2 | 366.7 | 52.1% | 703.2 | 366.7 | 52.1% | 703.2 | 366.7 | 52.1% |
| 1987 | 762.6 | 437.6 | 57.4% | 762.6 | 437.6 | 57.4% | 762.6 | 437.6 | 57.4% |
| 1988 | 786.4 | 451.2 | 57.4% | 802.4 | 464.0 | 57.8% | 816.6 | 475.3 | 58.2% |
| 1989 | 791.6 | 459.0 | 58.0% | 823.3 | 484.1 | 58.8% | 851.6 | 506.5 | 59.5% |
| 1990 | 811.1 | 466.8 | 57.6% | 858.7 | 504.2 | 58.7% | 901.7 | 538.1 | 59.7% |
| 1991 | 825.0 | 474.6 | 57.5% | 888.8 | 524.4 | 59.0% | 947.8 | 570.9 | 60.2% |
| 1992 | 839.3 | 482.4 | 57.5% | 919.9 | 545.2 | 59.3% | 996.8 | 605.9 | 60.8% |
| 1993 | 855.2 | 490.4 | 57.3% | 953.6 | 566.9 | 59.4% | 1050.7 | 643.9 | 61.3% |
| 1994 | 869.6 | 498.5 | 57.3% | 987.2 | 589.8 | 59.7% | 1107.5 | 685.7 | 61.9% |
| 1995 | 886.5 | 506.8 | 57.2% | 1024.6 | 614.1 | 59.9% | 1171.5 | 732.1 | 62.5% |
| 1996 | 894.9 | 515.4 | 57.6% | 1055.3 | 640.1 | 60.6% | 1232.7 | 783.7 | 63.6% |
| 1997 | 913.2 | 524.3 | 57.4% | 1097.9 | 668.1 | 60.8% | 1310.3 | 841.4 | 64.2% |
| 1998 | 928.8 | 533.5 | 57.4% | 1140.0 | 698.3 | 61.3% | 1392.5 | 906.0 | 65.1% |
| 1999 | 947.4 | 543.1 | 57.3% | 1187.7 | 731.0 | 61.6% | 1486.0 | 978.4 | 65.8% |
| 2000 | 973.6 | 553.1 | 56.8% | 1245.6 | 766.5 | 61.5% | 1596.4 | 1059.6 | 66.4% |
| Growth Rates: | | | | | | | | | |
| 73-87 | 4.8% | 5.8% | | 4.8% | 5.8% | | 4.8% | 5.8% | |
| 87-91 | 2.0% | 2.0% | | 3.9% | 4.6% | | 5.6% | 6.9% | |
| 87-20 | 1.9% | 1.8% | | 3.8% | 4.4% | | 5.8% | 7.0% | |

NOTE: This table does not include postdoctoral employment or unemployment.
SOURCE: Estimated by National Research Council.

Table 4. (continued) Projections of Biomedical Employment by Sector, 1987-2000
(Workers in 100s)

| Sector | 1987 Total | 1987 R&D | 1987 % R&D | 1991 Total | 1991 R&D | 1991 % R&D | 2000 Total | 2000 R&D | 2000 % R&D | 87-91 Growth Total | 87-91 Growth R&D | 87-20 Growth Total | 87-20 Growth R&D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Academic** | | | | | | | | | | | | | |
| Low Case | 430.3 | 227.5 | 52.9% | 435.3 | 235.6 | 54.1% | [illegible] | 226.1 | 55.4% | 0.3% | 0.9% | -0.4% | 0.0% |
| Medium Case | 430.3 | 227.5 | 52.9% | 461.5 | 253.6 | 55.0% | [illegible] | 260.8 | 55.7% | 1.8% | 2.8% | 0.7% | 1.1% |
| High Case | 430.3 | 227.5 | 52.9% | 484.3 | 268.0 | 55.3% | [illegible] | 297.0 | 55.7% | 3.0% | 4.2% | 1.7% | 2.1% |
| **Industrial** | | | | | | | | | | | | | |
| Low Case | 185.6 | 126.9 | 68.3% | 229.8 | 153.1 | 66.6% | 366.0 | 234.6 | [illegible] | 5.5% | 4.8% | 5.4% | 4.8% |
| Medium Case | 185.6 | 126.9 | 68.3% | 253.7 | 177.0 | 69.8% | 509.8 | 378.6 | 74.3% | 8.1% | 8.7% | 8.1% | 8.8% |
| High Case | 185.6 | 126.9 | 68.3% | 280.4 | 203.7 | 72.6% | 733.1 | 601.8 | 82.1% | 10.9% | 12.6% | 11.1% | 12.7% |
| **Government** | | | | | | | | | | | | | |
| Low Case | 70.5 | 46.7 | 66.2% | 72.9 | 47.0 | 64.5% | 79.9 | 47.8 | 59.9% | 0.8% | 0.2% | 1.0% | [illegible] |
| Medium Case | 70.5 | 46.7 | 66.2% | 79.3 | 53.4 | 67.4% | 108.7 | 76.7 | 70.5% | 3.0% | 3.4% | 3.4% | 3.9% |
| High Case | 70.5 | 46.7 | 66.2% | 83.1 | 57.2 | 68.9% | 135.7 | 103.6 | 76.4% | 4.2% | 5.2% | 5.2% | 5.3% |
| **All Other Biomedical** | | | | | | | | | | | | | |
| Low Case | 76.3 | 36.6 | 47.9% | 87.1 | 38.8 | 44.6% | 119.2 | 44.4 | 37.2% | 3.4% | 1.5% | 3.5% | 1.5% |
| Medium Case | 76.3 | 36.6 | 47.9% | 94.4 | 40.4 | 42.8% | 158.4 | 50.4 | 31.8% | 5.5% | 2.5% | 5.8% | 2.5% |
| High Case | 76.3 | 36.6 | 47.9% | 100.1 | 42.0 | 41.9% | 194.2 | 57.2 | 29.4% | 7.0% | 3.5% | 7.5% | 3.5% |
| **Total** | | | | | | | | | | | | | |
| Low Case | 762.6 | 437.6 | 57.4% | 825.0 | 474.6 | 57.5% | 973.6 | 553.1 | 56.8% | 2.0% | 2.0% | 1.9% | 1.8% |
| Medium Case | 762.6 | 437.6 | 57.4% | 888.8 | 524.4 | 59.0% | 1245.6 | 766.5 | 61.5% | 3.9% | 4.6% | 3.8% | 4.4% |
| High Case | 762.6 | 437.6 | 57.4% | 947.8 | 570.9 | 60.2% | 1596.4 | 1059.6 | 66.4% | 5.6% | 6.9% | 5.8% | 7.0% |

NOTE: Excludes postdocs and unemployed.

SOURCE: Projections by NRC.

Table 5. Historical and Projected Vacancy Ratios, Biomedical Sciences, 1973-2000

| | Annual Average | | | | | |
|---|---|---|---|---|---|---|
| | Total | | | R&D | | |
| Year | Vacancies | Ph.D.s | Ratio | Vacancies | Post-docs | Ratio |
| 1973-78 | 3455 | 3499 | 0.99 | 2344 | 2868 | 0.82 |
| 1978-83 | 3957 | 3763 | 1.05 | 2863 | 3788 | 0.76 |
| 1983-87 | 4846 | 3862 | 1.25 | 4157 | 4072 | 1.02 |
| 1987-95 | | | | | | |
| Low | 4047 | 3662 | 1.11 | 3086 | 3900 | 0.79 |
| Mid | 5955 | 3969 | 1.50 | 4626 | 3900 | 1.19 |
| High | 8063 | 4298 | 1.88 | 6386 | 3900 | 1.64 |

NOTE: Assumes that Ph.D. production changes in proportion to enrollment assumptions and postdoctoral production remains constant.

SOURCE: National Research Council.

**************************************************************************************

BIOMEDICAL PH.D. R&D PROJECTION MODEL

Mid Case R&D

**************************************************************************************

| Bio Age | Projected Years 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 | Growth Rates (base = 1987) 1991 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <30 | 1025 | 1903 | 2173 | 2453 | 2426 | 2489 | 2572 | 3282 | 20.7% | 12.2% | 9.4% |
| 30-34 | 7484 | 11520 | 12769 | 14065 | 14597 | 15419 | 16344 | 21127 | 14.3% | 10.3% | 8.3% |
| 35-39 | 9099 | 10019 | 10312 | 10635 | 11730 | 12725 | 13766 | 18926 | 3.2% | 5.3% | 5.8% |
| 40-44 | 9545 | 9297 | 9227 | 9173 | 9383 | 9546 | 9733 | 13053 | -0.8% | 0.2% | 2.4% |
| 45-49 | 6475 | 6846 | 6941 | 7026 | 6958 | 6888 | 6828 | 7126 | 1.8% | 0.7% | 0.7% |
| 50-54 | 3919 | 4545 | 4708 | 4852 | 4921 | 4975 | 5020 | 5062 | 4.7% | 3.1% | 2.0% |
| 55-59 | 3104 | 3368 | 3437 | 3497 | 3698 | 3868 | 4009 | 4407 | 2.6% | 3.2% | 2.7% |
| 60-64 | 2148 | 2116 | 2102 | 2085 | 2212 | 2313 | 2392 | 2821 | -0.5% | 1.4% | 2.1% |
| 65-69 | 770 | 657 | 630 | 603 | 626 | 621 | 614 | 697 | -4.9% | -2.8% | -0.8% |
| >=70 | 19[illegible] | 150 | 140 | 131 | 139 | 136 | 132 | 148 | -7.5% | -4.5% | -2.0% |
| Total | 43760 | 50420 | 52440 | 54520 | 56690 | 58980 | 61410 | 76650 | 4.6% | 4.3% | 4.4% |

Percent Distribution

| | 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|
| <30 | 2.3% | 3.8% | 4.1% | 4.5% | 4.3% | 4.2% | 4.2% | 4.3% |
| 30-34 | 17.1% | 22.8% | 24.3% | 25.8% | 25.7% | 26.1% | 26.6% | 27.6% |
| 35-39 | 20.8% | 19.9% | 19.7% | 19.5% | 20.7% | 21.6% | 22.4% | 24.7% |
| 40-44 | 21.8% | 18.4% | 17.6% | 16.8% | 16.6% | 16.2% | 15.8% | 17.0% |
| 45-49 | 14.8% | 13.6% | 13.2% | 12.9% | 12.3% | 11.7% | 11.1% | 9.3% |
| 50-54 | 9.0% | 9.0% | 9.0% | 8.9% | 8.7% | 8.4% | 8.2% | 6.6% |
| 55-59 | 7.1% | 6.7% | 6.6% | 6.4% | 6.5% | 6.6% | 6.5% | 5.7% |
| 60-64 | 4.9% | 4.2% | 4.0% | 3.8% | 3.9% | 3.9% | 3.9% | 3.7% |
| 65-69 | 1.8% | 1.3% | 1.2% | 1.1% | 1.1% | 1.1% | 1.0% | 0.9% |
| >=70 | 0.4% | 0.3% | 0.3% | 0.2% | 0.2% | 0.2% | 0.2% | 0.2% |
| =>55 | 14.2% | 12.5% | 12.0% | 11.6% | 11.8% | 11.8% | 11.6% | 10.5% |

Mean Biological Age

| | 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|
| | 43.7 | 42.5 | 42.2 | 41.9 | 41.8 | 41.7 | 41.5 | 40.9 |

| | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 | Average Annual Vacancies (base = 1990) 1991 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Total Vacancies** | | | | | | | | | | |
| D&R | 804 | 782 | 764 | 748 | 798 | 803 | 910 | 793 | 784 | 857 |
| Quits | 1529 | 1586 | 1642 | 1699 | 1765 | 1833 | 2247 | 1557 | 1675 | 1888 |
| Growth | 2010 | 2020 | 2080 | 2170 | 2290 | 2430 | 3550 | 2015 | 2167 | 2780 |
| Total | 4343 | 4388 | 4486 | 4617 | 4852 | 5071 | 6707 | 4365 | 4526 | 5525 |
| **Attrition Rates** | | | | | | | | | | |
| D&R | 1.7% | 1.6% | 1.5% | 1.4% | 1.4% | 1.4% | 1.2% | | | |
| Quits | 3.2% | 3.1% | 3.1% | 3.1% | 3.1% | 3.1% | 3.1% | | | |
| Total | 4.8% | 4.7% | 4.[illegible]% | 4.5% | 4.5% | 4.5% | 4.3% | | | |

**********************************************************************************************

BIOMEDICAL PHD PROJECTION MODEL

Mid Case

**********************************************************************************************

| Bio Age | Projected Years 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 | Growth Rates (base = 1987) 1991 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <30 | 1786 | 2577 | 2831 | 3105 | 3073 | 3257 | 3337 | 4223 | 12.2% | 8.1% | 6.8% |
| 30-34 | 13042 | 16684 | 17830 | 19061 | 19605 | 20793 | 21827 | 27810 | 8.1% | 6.6% | 6.0% |
| 35-39 | 15857 | 16777 | 17050 | 17353 | 18531 | 19326 | 20579 | 26777 | 1.8% | 3.3% | 4.1% |
| 40-44 | 16633 | 16735 | 16808 | 16910 | 17197 | 17283 | 17571 | 21448 | 0.3% | 0.7% | 2.0% |
| 45-49 | 11283 | 12522 | 12910 | 13290 | 13307 | 13312 | 13373 | 14089 | 3.4% | 2.1% | 1.7% |
| 50-54 | 6830 | 8441 | 8926 | 9386 | 9683 | 9957 | 10223 | 10890 | 6.9% | 5.2% | 3.7% |
| 55-59 | 5410 | 6323 | 6615 | 6898 | 7478 | 8001 | 847[illegible] | 10162 | 5.2% | 5.8% | 5.0% |
| 60-64 | 3743 | 4088 | 4188 | 4280 | 4635 | 4949 | 5235 | 6873 | 2.8% | 4.3% | 4.8% |
| 65-69 | 1342 | 1389 | 1390 | 1381 | 1488 | 1489 | 1491 | 1867 | 0.9% | 1.3% | 2.6% |
| >=70 | 334 | 335 | 332 | 326 | 361 | 354 | 349 | 422 | -0.1% | 0.6% | 1.8% |
| Total | 76260 | 85870 | 88880 | 91990 | 95360 | 98720 | 102460 | 124560 | 3.9% | 3.8% | 3.8% |

Percent Distribution

| | 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|
| <30 | 2.3% | 3.0% | 3.2% | 3.4% | 3.2% | 3.3% | 3.3% | 3.4% |
| 30-34 | 17.1% | 19.4% | 20.1% | 20.7% | 20.6% | 21.1% | 21.3% | 22.3% |
| 35-39 | 20.8% | 19.5% | 19.2% | 18.9% | 19.4% | 19.6% | 20.1% | 21.5% |
| 40-44 | 21.8% | 19.5% | 18.9% | 18.4% | 18.0% | 17.5% | 17.1% | 17.2% |
| 45-49 | 14.8% | 14.6% | 14.5% | 14.4% | 14.0% | 13.5% | 13.1% | 11.3% |
| 50-54 | 9.0% | 9.8% | 10.0% | 10.2% | 10.2% | 10.1% | 10.0% | 8.7% |
| 55-59 | 7.1% | 7.4% | 7.4% | 7.5% | 7.8% | 8.1% | 8.3% | 8.2% |
| 60-64 | 4.9% | 4.8% | 4.7% | 4.7% | 4.9% | 5.0% | 5.1% | 5.5% |
| 65-69 | 1.8% | 1.6% | 1.6% | 1.5% | 1.6% | 1.5% | 1.5% | 1.5% |
| >=70 | 0.4% | 0.4% | 0.4% | 0.4% | 0.4% | 0.4% | 0.3% | 0.3% |
| =>55 | 14.2% | 14.1% | 14.1% | 14.0% | 14.6% | 15.0% | 15.2% | 15.5% |

Mean Biological Age

| | 1987 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|
| | 43.7 | 43.4 | 43.3 | 43.3 | 43.4 | 43.3 | 43.3 | 42.9 |

| | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 2000 | Average Annual Vacancies (base = 1990) 1991 | 1995 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Total Vacancies** | | | | | | | | | | |
| D&R | 1564 | 1582 | 1593 | 1601 | 1788 | 1816 | 2154 | 1573 | 1657 | 1859 |
| Quits | 837 | 883 | 927 | 977 | 994 | 1040 | 1230 | 860 | 943 | 1034 |
| Growth | 3540 | 3010 | 3110 | 3370 | 3360 | 3740 | 5790 | 3275 | 3355 | 4665 |
| Total | 5941 | 5475 | 5630 | 5948 | 6142 | 6596 | 9174 | 5708 | 5955 | 7558 |
| **Attrition Rates** | | | | | | | | | | |
| D&R | 1.9% | 1.8% | 1.8% | 1.7% | 1.9% | 1.8% | 1.8% | | | |
| Quits | 1.0% | 1.0% | 1.0% | 1.1% | 1.0% | 1.1% | 1.0% | | | |
| Total | 2.9% | 2.9% | 2.8% | 2.8% | 2.9% | 2.9% | 2.8% | | | |

APPENDIX

Table A-1. Academic Projection Model's Biomedical Sciences

| | Employment | | | | Projected | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | Total | Non-R&D | R&D | R&D$ | Wgt Enrll | Total Fac. | R&D | Non-R&D | % R&D |
| Assumption | | | 3.0% | 0.0% | | | | | |
| 73 | 25471 | 15556 | 9915 | 1184.3 | 158698 | 23609 | 9190 | 14419 | 38.9 |
| 75 | 28332 | 18040 | 10292 | 1192.6 | 180485 | 27048 | 9825 | 17223 | 36.3 |
| 77 | 30384 | 17861 | 12523 | 1259 | 195353 | 31259 | 12884 | 18375 | 41.2 |
| 79 | 33566 | 18551 | 15015 | 1346.9 | 197091 | 33545 | 15006 | 18539 | 44.7 |
| 81 | 36482 | 19854 | 16628 | 1488.3 | 191996 | 35974 | 16397 | 19577 | 45.6 |
| 83 | 36963 | 20053 | 16910 | 1571.3 | 190546 | 37403 | 17111 | 20291 | 45.7 |
| 85 | 41032 | 21425 | 19607 | 1828.6 | 190291 | 41439 | 19802 | 21637 | 47.8 |
| 87 | 43025 | 20274 | 22751 | 2142.3 | 189218 | 44106 | 23323 | 20783 | 52.9 |
| 88 | | | | 2206.5 | 189218 | 44495 | 23712 | 20783 | 53.3 |
| 89 | | | | 2272.7 | 189218 | 44841 | 24058 | 20783 | 53.7 |
| 90 | | | | 2340.9 | 189218 | 45147 | 24364 | 20783 | 54.0 |
| 91 | | | | 2411.1 | 189218 | 45416 | 2463[illegible] | 20783 | 54.2 |
| 92 | | | | 2483.5 | 189218 | 45650 | 24867 | 20783 | 54.5 |
| 93 | | | | 2558.0 | 189218 | 45852 | 25069 | 20783 | 54.7 |
| 94 | | | | 2634.7 | 189218 | 46027 | 25244 | 20783 | 54.8 |
| 95 | | | | 2713.8 | 189218 | 46176 | 25393 | 20783 | 55.0 |
| 96 | | | | 2795.2 | 189218 | 46302 | 25519 | 20783 | 55.1 |
| 97 | | | | 2879.0 | 189218 | 46409 | 25626 | 20783 | 55.2 |
| 98 | | | | 2965.4 | 189218 | 46498 | 25715 | 20783 | 55.3 |
| 99 | | | | 3054.4 | 189218 | 46572 | 25789 | 20783 | 55.4 |
| 100 | | | | 3146.0 | 189218 | 46634 | 25851 | 20783 | 55.4 |

NOTE: The Acedemic Projection Model uses the Singer model and assumes the following:

1. The non-R&D faculty to student ratio (.1098 in 1987) will remain constant through projection period; therefore taking .1098 of wgt. enrollment gives non-R&D employment.
2. R&D faculty employment is simply total-non-R&D.
3. The Singer model is used to project total faculty employment:

$$F/WS = (K-C)\exp(-e^{(a-bM)})+C$$

N= 10
Std Er= .00580
$R^2$= .965

Where K and C are scaling constants
a,b are model parameters
M= biomedical R&D expenditures
F= biomedical faculty
WS= .25*(undergrads)+.75*(grad students)

| Coef. | Est. | Std.Er. | t |
|---|---|---|---|
| a | 2.391 | 2.6316 | 0.9085 |
| b | 0.0019 | 0.0024 | 0.8182 |
| C | 0.1005 | 0.0454 | 2.2121 |
| K | 0.2519 | 0.1607 | 1.5674 |

Table A-2. Industrial Projection Models, Biomedical Sciences

| Year | Employment Total | Non-R&D | R&D | Priv R&D$ | BLS DRG Outp | Projected Total Ind | R&D | Non-R&D | % R&D |
|---|---|---|---|---|---|---|---|---|---|
| Assumption | | | | 4.0% | | | | | |
| 73 | 5887 | 1757 | 4130 | 991 | 165.3 | 6340 | 4657 | 1682 | 73.5 |
| 75 | 7502 | 2501 | 5001 | 1037 | 178.4 | 7081 | 4860 | 2221 | 68.6 |
| 77 | 7755 | 2626 | 5129 | 1139 | 203.8 | 8575 | 5309 | 3266 | 61.9 |
| 79 | 9633 | 3276 | 6357 | 1265 | 218.1 | 9718 | 5863 | 3854 | 60.3 |
| 81 | 11785 | 4463 | 7322 | 1497 | 230.0 | 11228 | 6885 | 4344 | 61.3 |
| 83 | 13729 | 5239 | 8490 | 1871 | 245.8 | 13525 | 8531 | 4994 | 63.1 |
| 85 | 15960 | 6296 | 9664 | 2249 | 250.9 | 15398 | 10195 | 5203 | 66.2 |
| 87 | 18562 | 5875 | 12687 | 2768 | 281.8 | 18954 | 12480 | 6474 | 65.8 |
| 88 | | | | 2879 | 293.1 | 19514 | 13174 | 6340 | 67.5 |
| 89 | | | | 2994 | 287.4 | 19787 | 13681 | 6105 | 69.1 |
| 90 | | | | 3114 | 312.7 | 21354 | 14208 | 7146 | 66.5 |
| 91 | | | | 3238 | 325.4 | 22425 | 14757 | 7668 | 65.8 |
| 92 | | | | 3368 | 338.5 | 23534 | 15327 | 8207 | 65.1 |
| 93 | | | | 3502 | 354.7 | 24794 | 15920 | 8874 | 64.2 |
| 94 | | | | 3642 | 367.0 | 25916 | 16537 | 9380 | 63.8 |
| 95 | | | | 3788 | 383.9 | 27253 | 17178 | 10075 | 63.0 |
| 96 | | | | 3940 | 381.2 | 27809 | 17845 | 9964 | 64.2 |
| 97 | | | | 4097 | 399.2 | 29243 | 18539 | 10704 | 63.4 |
| 98 | | | | 4261 | 410.2 | 30417 | 19260 | 11157 | 63.3 |
| 99 | | | | 4432 | 426.6 | 31842 | 20011 | 11831 | 62.8 |
| 100 | | | | 4609 | 458.0 | 33914 | 20791 | 13123 | 61.3 |

NOTE: The Industrial Projection Model uses an R&D and output model and assumes the following:

1. Industrial biomedical R&D employment is a function of private R&D expenditures as per regression Equation 1. This equation coefficient was applied to change in Priv R&D$ to determine projected employment.

2. Industrial biomedical non-R&D employment is a function of drug industry ~nstant dollar output as per Equation 2. This equation coefficient was applied to change in BLS drug output to determine projected employment.

| EQUATION 1 | |
|---|---|
| (R&Demp)=f(R&D$) | |
| Regression Output: | |
| Constant | 294.76 |
| Std Err of Y Est | 426.51 |
| R Squared | 0.9808 |
| No. of Observations | 8 |
| Degrees of Freedom | 6 |
| X Coefficient | 4.4021 |
| Standard Error of Coefficient | 0.2512 |

| EQUATION 2 | |
|---|---|
| (non-R&Demp)=f(drgoutp) | |
| Regression Output: | |
| Constant | -5117. |
| Std Err of Y Est | 639.62 |
| R Squared | 0.8786 |
| No. of Observations | 8 |
| Degrees of Freedom | 6 |
| X Coefficient | 41.133 |
| Standard Error of Coefficient | 6.2410 |

Table A-3. Government Projection Models, Biomedical Sciences

| Year | Total | Employment Non-R&D | R&D | Pub R&D$ | BLS Gov Output | Total Gov | Projected R&D | Non-R&D | % R&D |
|---|---|---|---|---|---|---|---|---|---|
| Assumption | | | | 2.7% | | | | | |
| 73 | 4338 | 1244 | 3094 | 2103 | 110.4 | 4479 | 2929 | 1550 | 65.4 |
| 75 | 4517 | 1410 | 3107 | 2226 | 109.6 | 4559 | 3071 | 1489 | 67.4 |
| 77 | 4568 | 1632 | 2936 | 2397 | 109.5 | 4760 | 3279 | 1481 | 68.9 |
| 79 | 5080 | 1759 | 3321 | 2612 | 111.0 | 5157 | 3561 | 1596 | 69.1 |
| 81 | 5398 | 1869 | 3529 | 2484 | 115.0 | 5292 | 3390 | 1902 | 64.1 |
| 83 | 5988 | 2380 | 3608 | 2504 | 119.0 | 5625 | 3416 | 2208 | 60.7 |
| 85 | 6479 | 2661 | 3818 | 2942 | 122.6 | 6526 | 4042 | 2484 | 61.9 |
| 87 | 7049 | 2379 | 4670 | 3120 | 124.5 | 6957 | 4328 | 2629 | 62.2 |
| 88 | | | | 3204 | 125.2 | 7256 | 4823 | 2433 | 66.5 |
| 89 | | | | 3291 | 124.8 | 7388 | 4986 | 2402 | 67.5 |
| 90 | | | | 3380 | 126.4 | 7684 | 5159 | 2524 | 67.1 |
| 91 | | | | 3471 | 127.2 | 7929 | 5343 | 2586 | 67.4 |
| 92 | | | | 3565 | 128.0 | 8186 | 5539 | 2647 | 67.7 |
| 93 | | | | 3661 | 128.9 | 8464 | 5748 | 2716 | 67.9 |
| 94 | | | | 3760 | 129.7 | 8747 | [illegible] | 2777 | 68.3 |
| 95 | | | | 3861 | 130.7 | 9061 | 6207 | 2854 | 68.5 |
| 96 | | | | 3965 | 130.6 | 9307 | 6461 | 2846 | 69.4 |
| 97 | | | | 4072 | 131.7 | 9662 | 6732 | 2930 | 69.7 |
| 98 | | | | 4182 | 132.3 | 9998 | 7022 | 2976 | 70.2 |
| 99 | | | | 4295 | 133.3 | 10386 | 7333 | 3053 | 70.6 |
| 100 | | | | 4411 | 135.3 | 10873 | 7667 | 3206 | 70.5 |

NOTE: The Government Projection Model uses an R&D and output model and assumes the following:

1. Government biomed R&D Employment is a function of Federal R&D expenditures as per regression Equation 1. This equation coefficient was applied to change in Fed R&D$ to determine projected Employment.

2. Government biomed non-R&D Employment is a function of federal government constant dollar output as per Equation 2. This equation coefficient was applied to change in BLS government output to determine projected Employment.

| EQUATION 1 | | EQUATION 2 | |
|---|---|---|---|
| Ln(R&Demp)=F(fed R&D$) | | (Non-R&Demp)=F(fed output) | |
| Regression Output: | | Regression Output: | |
| Constant | 7.1749 | Constant | -6903. |
| Std Err of Y Est | 0.0751 | Std Err of Y Est | 213.73 |
| R Squared | 0.7803 | R Squared | 0.8477 |
| No. of Observations | 8 | No. of Observations | 8 |
| Degrees of Freedom | 6 | Degrees of Freedom | 6 |
| X Coefficient | 0.0003 | X Coefficient | 76.565 |
| Standard Error of Coefficient | 0.0000 | Standard Error of Coefficient | 13.248 |

Table A-4. All Other Employment, Biomedical Sciences

| Year | Total (Employment) | Non-R&D (Employment) | R&D (Employment) | Total Other (Projected) | R&D (Projected) | Non-R&D (Projected) | % R&D (Projected) |
|---|---|---|---|---|---|---|---|
| Assumption | | | | | 2.5% | 8.0% | |
| 73 | 3941 | 1296 | 2645 | 3941 | 2645 | 1296 | 67.1 |
| 75 | 4368 | 1589 | 2779 | 4368 | 2779 | 1589 | 63.6 |
| 77 | 5215 | 2015 | 3200 | 5215 | 3200 | 2015 | 61.4 |
| 79 | 6102 | 2344 | 3758 | 6102 | 3758 | 2344 | 61.6 |
| 81 | 6348 | 2204 | 4144 | 6348 | 4144 | 2204 | 65.3 |
| 83 | 5964 | 2766 | 3198 | 5964 | 3198 | 2766 | 53.6 |
| 85 | 6852 | 3271 | 3581 | 6852 | 3581 | 3271 | 52.3 |
| 87 | 7626 | 3970 | 3656 | 7626 | 3656 | 3970 | 47.9 |
| 88 | 0.0482 | 0.0832 | 0.0233 | 8035 | 3748 | 4287 | 46.6 |
| 89 | | | | 8472 | 3842 | 4630 | 45.3 |
| 90 | | | | 8938 | 3938 | 5000 | 44.1 |
| 91 | | | | 9437 | 4036 | 5400 | 42.8 |
| 92 | | | | 9969 | 4137 | 5833 | 41.5 |
| 93 | | | | 10540 | 4240 | 6299 | 40.2 |
| 94 | | | | 11149 | 4346 | 6803 | 39.0 |
| 95 | | | | 11802 | 4455 | 7347 | 37.7 |
| 96 | | | | 12502 | 4566 | 7935 | 36.5 |
| 97 | | | | 13250 | 4681 | 8570 | 35.3 |
| 98 | | | | 14053 | 4798 | 9255 | 34.1 |
| 99 | | | | 14913 | 4918 | 9996 | 33.0 |
| 100 | | | | 15836 | 5041 | 10796 | 31.8 |

NOTE: The "All Other" sector is a simple trend model that projects assumed rates of growth for the R&D and non-R&D sectors through the year 2000.

Table A-5. Academic Projection Models, Nonclinical Psychology

| Year | Total | Employment Non-R&D | R&D | Total Beh Sc Fac. | Nonclinical Psych Projected Total Fac. | R&D | Non-R&D | % R&D |
|---|---|---|---|---|---|---|---|---|
| Assumption | | | | -2.0% | | | | |
| 73 | 9452 | 7380 | 2072 | 19900 | 9452 | 7380 | 2072 | 21.9% |
| 75 | 10863 | 9016 | 1847 | 23600 | 10863 | 9016 | 1847 | 17.0% |
| 77 | 10905 | 8799 | 2106 | 25600 | 10905 | 8799 | 2106 | 19.3% |
| 79 | 11538 | 8729 | 2809 | 26900 | 11538 | 8729 | 2809 | 24.3% |
| 81 | 12586 | 9836 | 2750 | 28200 | 12586 | 9836 | 2750 | 21.9% |
| 83 | 12404 | 9785 | 2619 | 29800 | 12404 | 9785 | 2619 | 21.1% |
| 85 | 13221 | 10655 | 2566 | 31700 | 13221 | 10655 | 2566 | 19.4% |
| 87 | 13058 | 9972 | 3086 | 31800 | 13058 | 9972 | 3086 | 23.6% |
| 88 | | | | 31732 | 13026 | 2745 | 10281 | 21.1% |
| 89 | | | | 31635 | 12986 | 2740 | 10246 | 21.1% |
| 90 | | | | 31444 | 12908 | 2724 | 10184 | 21.1% |
| 91 | | | | 31166 | 12794 | 2699 | 10094 | 21.1% |
| 92 | | | | 30806 | 12646 | 2668 | 9978 | |
| 93 | | | | 30369 | 12466 | 2630 | 9836 | |
| 94 | | | | 29860 | 12257 | 2586 | 9671 | 21.1% |
| 95 | | | | 29283 | 12021 | 2536 | 9484 | 21.1% |
| 96 | | | | 28644 | 11758 | 2481 | 9277 | 21.1% |
| 97 | | | | 27945 | 11472 | 2420 | 9051 | 21.1% |
| 98 | | | | 27192 | 11162 | 2355 | 8807 | 21.1% |
| 99 | | | | 26388 | 10832 | 2286 | 8547 | 21.1% |
| 100 | | | | 25536 | 10483 | 2212 | 8271 | 21.1% |

NOTE: The Academic Projection Model uses a quadratic regression of total behavioral faculty = f (behavioral graduate students):

1. Growth assumptions for behavioral student enrollment provided by the committee.

2. The portion of total behavioral faculty that is clinical psychology, nonclinical psychology, and "Other Behavioral Sciences" remains fixed at 1987 levels.

3. The portion of nonclinical psychology in R&D, which has averaged 21.1 percent from 1973-1987, remains fixed.

Table A-6. Industrial Projection Models, Nonclinical Psychology

| Year | Total | Employment Non-R&D | R&D | Projected Total Ind | R&D | Non-R&D | % R&D |
|---|---|---|---|---|---|---|---|
| Assumption | | | | 3.0% | | | |
| 73 | 1322 | 811 | 511 | 1322 | 511 | 811 | 38.6 |
| 75 | 1619 | 984 | 635 | 1619 | 635 | 984 | 39.2 |
| 77 | 1787 | 1031 | 756 | 1787 | 756 | 1031 | 42.3 |
| 79 | 1676 | 950 | 726 | 1676 | 726 | 950 | 43.3 |
| 81 | 2731 | 1765 | 966 | 2731 | 966 | 1765 | 35.4 |
| 83 | 3586 | 2481 | 1105 | 3586 | 1105 | 2481 | 30.8 |
| 85 | 3194 | 2140 | 1054 | 3194 | 1054 | 2140 | 33.0 |
| 87 | 3400 | 2511 | 889 | 3400 | 889 | 2511 | 26.1 |
| 88 | | | | 3502 | 1264 | 2238 | 36.1 |
| 89 | | | | 3607 | 1302 | 2305 | 36.1 |
| 90 | | | | 3715 | 1341 | 2374 | 36.1 |
| 91 | | | | 3827 | 1381 | 2445 | 36.1 |
| 92 | | | | 3942 | 1423 | 2519 | 36.1 |
| 93 | | | | 4060 | 1466 | 2594 | 36.1 |
| 94 | | | | 4182 | 1510 | 2672 | 36.1 |
| 95 | | | | 4307 | 1555 | 2752 | 36.1 |
| 96 | | | | 4436 | 1601 | 2835 | 36.1 |
| 97 | | | | 4569 | 1650 | 2920 | 36.1 |
| 98 | | | | 4706 | 1699 | 3007 | 36.1 |
| 99 | | | | 4848 | 1750 | 3098 | 36.1 |
| 100 | | | | 4993 | 1802 | 3191 | 36.1 |

NOTE: The Industrial Projection Model is a simple trend model that grows 1987 Employment to the year 2000 using committee assumptions of:

1. Growth of 2 percent, 3 percent and 4 percent.

2. The average proportion of total industry nonclinical psychology employment that is engaged in R&D for the 1973-1987 period (.06%) will remain fixed throughout the projection period.

Table A-7. Government Projection Models, Nonclinical Psychology

| | Employment | | | Projected | | | |
|---|---|---|---|---|---|---|---|
| Year | Total | Non-R&D | R&D | Total Gov | R&D | Non-R&D | % R&D |
| Assumption | | | | | | | |
| 73 | 1083 | 426 | 657 | 1083 | 657 | 426 | 60.7 |
| 75 | 1170 | 345 | 825 | 1170 | 825 | 345 | 70.5 |
| 77 | 1404 | 495 | 909 | 1404 | 909 | 495 | 64.7 |
| 79 | 1164 | 334 | 830 | 1164 | 830 | 334 | 71.3 |
| 81 | 1235 | 487 | 748 | 1235 | 748 | 487 | 60.5 |
| 83 | 1320 | 670 | 650 | 1320 | 650 | 670 | 49.3 |
| 85 | 1189 | 511 | 678 | 1189 | 678 | 511 | 57.0 |
| 87 | 1724 | 711 | 1013 | 1724 | 1013 | 711 | 58.7 |
| 88 | | | | 1741 | 1072 | 669 | 61.6 |
| 89 | | | | 1759 | 1083 | 675 | 61.6 |
| 90 | | | | 1776 | 1094 | 682 | 61.6 |
| 91 | | | | 1794 | 1105 | 689 | 61.6 |
| 92 | | | | 1812 | 1116 | 696 | 61.6 |
| 93 | | | | 1830 | 1127 | 703 | 61.6 |
| 94 | | | | 1848 | 1139 | 710 | 61.6 |
| 95 | | | | 1867 | 1150 | 717 | 61.6 |
| 96 | | | | 1884 | 1160 | 723 | 61.6 |
| 97 | | | | 1901 | 1171 | 730 | 61.6 |
| 98 | | | | 1918 | 1181 | 736 | 61.6 |
| 99 | | | | 1935 | 1192 | 743 | 61.6 |
| 100 | | | | 1952 | 1203 | 750 | 61.6 |

NOTE: The Government Projection Model is a simple trend model that uses the BLS estimates of growth in total psychologists (1 percent from 1987-1995; .09 percent from 1995-2000). The average proportion of scientists involved in R&D for the 1973-87 period (61.6 percent) was assumed constant over the projection period.

Table A-8. All Other Employment, Nonclinical Psychology

| Year | Employment Total | Non-R&D | R&D | Projected Total Other | Projected R&D | Projected Non-R&D | % R&D |
|---|---|---|---|---|---|---|---|
| Assumption | | | | 3.4% | -0.5% | 7.2% | |
| 73 | 1043 | 407 | 636 | 1043 | 636 | 407 | 61.0 |
| 75 | 1159 | 520 | 639 | 1159 | 639 | 520 | 55.2 |
| 77 | 1221 | 684 | 537 | 1221 | 537 | 684 | 44.0 |
| 79 | 1409 | 833 | 576 | 1409 | 576 | 833 | 40.9 |
| 81 | 1433 | 875 | 558 | 1433 | 558 | 875 | 38.9 |
| 83 | 1413 | 1022 | 391 | 1413 | 391 | 1022 | 27.7 |
| 85 | 1497 | 1111 | 386 | 1497 | 386 | 1111 | 25.8 |
| 87 | 1662 | 1075 | 587 | 1662 | 587 | 1075 | 35.3 |
| 88 | 0.0338 | 0.0718 | -0.005 | 1736 | 584 | 1152 | 33.6 |
| 89 | | | | 1795 | 581 | 1235 | 32.4 |
| 90 | | | | 1856 | 578 | 1324 | 31.2 |
| 91 | | | | 1918 | 575 | 1419 | 30.0 |
| 92 | | | | 1983 | 572 | 1521 | 28.9 |
| 93 | | | | 2050 | 569 | 1630 | 27.8 |
| 94 | | | | 2120 | 567 | 1747 | 26.7 |
| 95 | | | | 2191 | 564 | 1872 | 25.7 |
| 96 | | | | 2265 | 561 | 2007 | 24.8 |
| 97 | | | | 2342 | 558 | 2151 | 23.8 |
| 98 | | | | 2421 | 555 | 2305 | 22.9 |
| 99 | | | | 2503 | 553 | 2471 | 22.1 |
| 100 | | | | 2587 | 550 | 2648 | 21.2 |

NOTE: The "All Other" sector is a simple trend model that projects assumed rates of growth for the R&D and non-R&D sectors through the year 2000. The rates of growth are the 1973-1987 rates for non-R&D (7.2 percent per annum) and R&D (-0.5 percent per annum).

Table A-9. Academic Projection Models, Other Behavioral Sciences

| | Employment | | | | Other Behavioral Sciences Projected | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Total | Non-R&D | R&D | Total Beh Sc Fac. | Total Fac | R&D | Non-R&D | % R&D |
| Assumption | | | | -2.0% | | | | |
| 73 | 6135 | 5201 | 934 | 19900 | 6135 | 5201 | 934 | 15.2 |
| 75 | [illegible] | 6559 | 1062 | 23400 | 7621 | 6559 | 1062 | 13.9 |
| 77 | [illegible] | 7784 | 1455 | 25600 | 9239 | 7784 | 1455 | 15.8 |
| 79 | 9568 | 7915 | 1653 | 26900 | 9568 | 7915 | 1653 | 17.3 |
| 81 | 9477 | 8074 | 1403 | 28200 | 9477 | 8074 | 1403 | 14.8 |
| 83 | 11002 | 9546 | 1456 | 29800 | 11002 | 9546 | 1456 | 13.2 |
| 85 | 11009 | 9530 | 1479 | 31700 | 11009 | 9530 | 1479 | 13.4 |
| 87 | 10767 | 9079 | 1688 | [illegible] | 10767 | 9079 | 1688 | 15.7 |
| 88 | | | | [illegible] | 10741 | 1602 | 9139 | 14.9 |
| 89 | | | | 31635 | 10708 | 1596 | 9113 | 14.9 |
| 90 | | | | 31444 | 10644 | 1586 | 9058 | 14.9 |
| 91 | | | | 31166 | 10550 | 1572 | 8978 | 14.9 |
| 92 | | | | 30806 | 10428 | 1554 | 8874 | 14.9 |
| 93 | | | | 30369 | 10280 | 1532 | 8748 | 14.9 |
| 94 | | | | 29860 | 10108 | 1506 | 8602 | 14.9 |
| 95 | | | | 29283 | 9912 | 1477 | 8435 | 14.9 |
| 96 | | | | 28644 | 9596 | 1445 | 8251 | 14.9 |
| 97 | | | | 27945 | 9459 | 1409 | [illegible] | 14.9 |
| 98 | | | | 27192 | 9204 | 1371 | [illegible] | 14.9 |
| 99 | | | | 26388 | [illegible] | 1331 | 7601 | 14.9 |
| 100 | | | | 25536 | 8644 | 1288 | 7356 | 14.9 |

NOTE: The Academic Projection Model uses a quadratic regression of total behavioral faculty = f (behavioral graduate students).

1. Growth assumptions in behavioral student enrollment provided by the committee.

2. The portion of total behavioral faculty that is clinical psychology, nonclinical psychology, and "Other Behavioral Sciences" remains fixed at 1987 levels.

3. The portion of other behavioral sciences in R&D, which has average 14.9 percent from 1973-1987, remains fixed.

Table A-10. All Other Employment, Other Behavioral Sciences

| | Employment | | | Projected | | | |
|---|---|---|---|---|---|---|---|
| Year | Total | Non-R&D | R&D | Total Ind | R&D | Non-R&D | % R&D |
| Assumption | | | | 3.0% | 0.07 | 0.1491 | |
| 73 | 456 | 179 | 277 | 456 | 277 | 179 | 60.8 |
| 75 | 579 | 201 | 378 | 579 | 378 | 201 | 65.4 |
| 77 | 837 | 328 | 509 | 837 | 509 | 328 | 60.9 |
| 79 | 1277 | 450 | 827 | 1277 | 827 | 450 | 64.8 |
| 81 | 1343 | 524 | 819 | 1343 | 819 | 524 | 61.0 |
| 83 | 1476 | 1028 | 448 | 1476 | 448 | 1028 | 30.3 |
| 85 | 1680 | 1308 | 372 | 1680 | 372 | 1308 | 22.2 |
| 87 | 1969 | 1254 | 715 | 1969 | 715 | 1254 | 36.3 |
| 88 | 0.1101 | 0.1491 | 0.0700 | 2206 | 765 | 1441 | 34.7 |
| 89 | | | | 2474 | 819 | 1656 | 33.1 |
| 90 | | | | 2779 | 876 | 1903 | 31.5 |
| 91 | | | | 3124 | 937 | 2186 | 30.0 |
| 92 | | | | 3515 | 1003 | 2512 | 28.5 |
| 93 | | | | 3960 | 1073 | 2887 | 27.1 |
| 94 | | | | 4465 | 1148 | 3317 | 25.7 |
| 95 | | | | 5040 | 1229 | 3812 | 24.4 |
| 96 | | | | 5695 | 1315 | 4380 | 23.1 |
| 97 | | | | 6440 | 1407 | 5033 | 21.8 |
| 98 | | | | 7289 | 1505 | 5783 | 20.7 |
| 99 | | | | 8256 | 1611 | 6646 | 19.5 |
| 100 | | | | 9360 | 1723 | 7637 | 18.4 |

NOTE: The "All Other" projection model is a simple trend model that grows 1987 employment to the year 2000 assuming that 1973-1987 historical growth rates in the R&D (7 percent) and non-R&D (14.9 percent) would continue through 2000.

Table A-11. Behavioral Enrollment Faculty, All Behavioral Sciences

| Assumption | -2.0% | -2.0% | | |
|---|---|---|---|---|
| **Year** | **Est Behav. Undergrad Enroll** | **Est Behav. Grad Enroll** | **Estimated Total Fac** | **Total Student** |
| 1972 | 705.8 | 44.7 | | 750.5 |
| 1973 | 695.6 | 48.2 | 19.9 | 743.8 |
| 1974 | 672.6 | 51.7 | 21.7 | 724.3 |
| 1975 | 668.1 | 55.4 | 23.6 | 723.5 |
| 1976 | 671.8 | 59.1 | 24.6 | 730.9 |
| 1977 | 663.5 | 58.3 | 25.6 | 721.8 |
| 1978 | 648.9 | 63.8 | 26.2 | 712.7 |
| 1979 | 612.8 | 63.8 | 26.9 | 676.6 |
| 1980 | 604.8 | 63.8 | 27.5 | 668.6 |
| 1981 | 593.6 | 64.8 | 28.2 | 658.4 |
| 1982 | 589.8 | 64.3 | 29.0 | 654.1 |
| 1983 | 574.5 | 63.1 | 29.8 | 637.6 |
| 1984 | 563.0 | 63.7 | 30.7 | 626.7 |
| 1985 | 551.7 | 65.6 | 31.7 | 617.3 |
| 1986 | 540.7 | 65.3 | 31.7 | 606.0 |
| 1987 | 529.9 | 65.1 | 31.8 | 595.0 |
| 1988 | 519.3 | 63.8 | 31.7 | 583.1 |
| 1989 | 508.9 | 62.5 | 31.6 | 571.4 |
| 1990 | 498.7 | 61.3 | 31.4 | 560.0 |
| 1991 | 488.8 | 60.0 | 31.2 | 548.8 |
| 1992 | 479.0 | 58.8 | 30.8 | 537.8 |
| 1993 | 469.4 | 57.7 | 30.4 | 527.1 |
| 1994 | 460.0 | 56.5 | 29.9 | 516.5 |
| 1995 | 450.8 | 55.4 | 29.3 | 506.2 |
| 1996 | 441.8 | 54.3 | 28.6 | 496.1 |
| 1997 | 433.0 | 53.2 | 27.9 | 485.2 |
| 1998 | 424.3 | 52.1 | 27.2 | 476.4 |
| 1999 | 415.8 | 51.1 | 26.4 | 4[illegible].9 |
| 2000 | 407.5 | 50.1 | 25.5 | 457.6 |

NOTE: Growth assumptions regarding enrollments in behavioral sciences were provided by the committee.

Faculty = f (behavioral graduate students)

| Regression Output: | | |
|---|---|---|
| Constant | | 3.5676[illegible]22079 |
| Std Err of Y Est | | 0.0207556027 |
| R Squared | | 0.9032859059 |
| No. of Observations | | 15 |
| Degrees of Freedom | | 12 |
| | | |
| X Coefficient | 0.0001001 | -0.0000000014 |
| Standard Error of Coefficient | 0.0001713 | 0.0000000088 |

Table A-12. Summary Projections for Nonclinical Psychology

| Nonclinical Sensitivity Model | | | |
|---|---|---|---|
| Model Assumptions (in percent) | High | Mid | Low |
| 1. Graduate Student Enrollment | 1.0 | 0.0 | -1.0 |
| 2. Industrial Employment Growth | 4.0 | 3.0 | 2.0 |

Projected Employment of Nonclinical Psychologists, 1973-2000
(in 100s of workers)

| | Low Case | | | Mid Case | | | High Case | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | % R&D |
| 1973 | 129.0 | 38.8 | 30.0 | 129.0 | 38.8 | 30.0 | 129.0 | 38.8 | 30.0 |
| 1975 | 148.1 | 39.5 | 26.6 | 148.1 | 39.5 | 26.6 | 148.1 | 39.5 | 26.6 |
| 1977 | 153.2 | 43.1 | 28.1 | 153.2 | 43.1 | 28.1 | 153.2 | 43.1 | 28.1 |
| 1979 | 157.9 | 49.4 | 31.3 | 157.9 | 49.4 | 31.3 | 157.9 | 49.4 | 31.3 |
| 1981 | 179.9 | 50.2 | 27.9 | 179.9 | 50.2 | 27.9 | 179.9 | 50.2 | 27.9 |
| 1983 | 187.2 | 47.7 | 25.5 | 187.2 | 47.7 | 25.5 | 187.2 | 47.7 | 25.5 |
| 1985 | 191.0 | 46.8 | 24.5 | 191.0 | 46.8 | 24.5 | 191.0 | 46.8 | 24.5 |
| 1987 | 198.4 | 55.7 | 28.1 | 198.4 | 55.7 | 28.1 | 198.4 | 55.7 | 28.1 |
| 1988 | 200.0 | 54.9 | 27.5 | 200.3 | 55.0 | 27.5 | 200.7 | 55.2 | 27.5 |
| 1989 | 195.5 | 54.0 | 27.6 | 202.2 | 55.5 | 27.4 | 204.9 | 56.1 | 27.4 |
| 1990 | 195.2 | 53.9 | 27.6 | 204.0 | 55.9 | 27.4 | 209.2 | 57.1 | 27.3 |
| 1991 | 194.9 | 53.9 | 27.6 | 206.0 | 56.3 | 27.4 | 213.6 | 58.1 | 27.2 |
| 1992 | 194.8 | 53.8 | 27.6 | 207.9 | 56.8 | 27.3 | 218.2 | 59.2 | 27.1 |
| 1993 | 194.7 | 53.8 | 27.6 | 210.0 | 57.2 | 27.3 | 223.0 | 60.2 | 27.0 |
| 1994 | 194.7 | 53.8 | 27.6 | 212.1 | 57.7 | 27.2 | 227.9 | 61.3 | 26.9 |
| 1995 | 194.8 | 53.7 | 27.6 | 214.2 | 58.2 | 27.2 | 232.9 | 62.5 | 26.8 |
| 1996 | 194.9 | 53.7 | 27.6 | 216.4 | 58.7 | 27.1 | 238.1 | 63.6 | 26.7 |
| 1997 | 195.1 | 53.7 | 27.5 | 218.7 | 59.2 | 27.0 | 243.4 | 64.8 | 26.6 |
| 1998 | 195.4 | 53.8 | 27.5 | 221.0 | 59.7 | 27.0 | 249.0 | 66.1 | 26.5 |
| 1999 | 195.8 | 53.8 | 27.5 | 223.4 | 60.2 | 26.9 | 254.7 | 67.4 | 26.5 |
| 2000 | 199.2 | 54.5 | 27.3 | 225.9 | 60.7 | 26.9 | 260.6 | 68.7 | 26.4 |
| Growth Rates (in percent): | | | | | | | | | |
| 1973-1987 | 3.1 | 2.6 | | 3.1 | 2.6 | | 3.1 | 2.6 | |
| 1987-1991 | -0.4 | -0.9 | | 0.9 | 0.3 | | 1.9 | 1.0 | |
| 1937-2000 | 0.0 | -0.2 | | 1.0 | 0.7 | | 2.1 | 1.6 | |

NOTE: This table does not include postdoctoral employment or unemployment.

SOURCE: Estimated by National Research Council.

Table A-12. (continued) Projections of Nonclinical Psychologists by Sector, 1987-2000
(Workers in 100s)

| Sector | 1987 | | | 1991 | | | 2000 | | | 1987-1991 Growth | | 1987-2000 Growth | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | Total | R&D |
| Academic | | | | | | | | | | | | | |
| Low Case | 130.6 | 30.9 | 23.6 | 121.0 | 25.5 | 21.1 | 109.8 | 23.2 | 21.1 | -1.9% | -4.6% | -1.3% | -2.2% |
| Medium Case | 130.6 | 30.9 | 23.6 | 130.6 | 27.6 | 21.1 | 130.6 | 27.6 | 21.1 | 0.0% | -2.8% | 0.0% | -0.9% |
| High Case | 130.6 | 30.9 | 23.6 | 136.7 | 28.9 | 21.1 | 158.5 | 33.5 | 21.1 | 1.2% | -1.7% | 1.5% | 0.6% |
| Industrial | | | | | | | | | | | | | |
| Low Case | 34.0 | 8.9 | 26.1 | 36.8 | 11.5 | 31.3 | 44.0 | 13.8 | 31.3 | 2.0% | 6.7% | 2.0% | 3.4% |
| Medium Case | 34.0 | 8.9 | 26.1 | 38.3 | 12.0 | 31.3 | 49.9 | 15.6 | 31.3 | 3.0% | 7.8% | 3.0% | 4.4% |
| High Case | 34.0 | 8.9 | 26.1 | 39.8 | 12.5 | 31.3 | 56.6 | 17.7 | 31.3 | 4.0% | 8.8% | 4.0% | 5.5% |
| Government | 17.2 | 10.1 | 58.7 | 17.9 | 11.1 | 61.6 | 19.5 | 12.0 | 61.6 | 1.0% | 2.2% | 1.0% | 1.3% |
| All Other Clin. Psych. | 16.6 | 5.9 | 35.3 | 19.2 | 5.[illegible] | 30.0 | 25.9 | 5.5 | 21.2 | 3.7% | -0.5% | 3.5% | -0.5% |
| Total | | | | | | | | | | | | | |
| Low Case | 198.4 | 55.7 | 28.1 | 194.9 | 53.9 | 27.6 | 199.2 | 54.5 | 27.3 | -0.4% | -0.9% | 0.0% | -0.2% |
| Medium Case | 198.4 | 55.7 | 28.1 | 206.0 | 56.3 | 27.4 | 225.9 | 60.7 | 26.9 | 0.9% | 0.3% | 1.0% | 0.7% |
| High Case | 193.4 | 55.7 | 28.1 | 213.6 | 58.1 | 27.2 | 260.6 | 68.7 | 26.4 | 1.9% | 1.0% | 2.1% | 1.6% |

NOTE: Excludes postdoctorals and unemployed.

SOURCE: Projections by National Research Council.

Table A-13. Estimated Quit Rates for All Behavioral Scientists, 1983-1987

| Career Age | (1) Behav | (2) Behav R&D & Mgt R&D | (3) Postdoc Behav | (4) Retired | (5) Total Ph.D.s | (6) Total Ph.D.s -Retire | (7) Beh Sci per 1000 Sci | (8) Annual Quit Rate | (9) Annual R&D Quit Rate |
|---|---|---|---|---|---|---|---|---|---|
| <=5 | 41948 | 5607 | 1834 | 139 | 201127 | 200988 | 218 | | |
| 6-10 | 43488 | 6282 | 343 | 373 | 217900 | 217527 | 201 | -3.1% | -7.5% |
| 11-15 | 36561 | 4612 | 127 | 990 | 215219 | 2142[illegible]9 | 171 | -6.3% | -12.0% |
| 16-20 | 23923 | 2915 | 68 | 2279 | 156711 | 1544[illegible]2 | 155 | -3.8% | -5.3% |
| 21-25 | 13410 | 1463 | 51 | 4041 | 91497 | 87456 | 154 | -0.4% | -4.3% |
| 26-30 | 10323 | 1271 | 38 | 6428 | 67207 | 60779 | 170 | 4.2% | 9.1% |
| 31- | 6155 | 950 | 8 | 1[illegible]2 | 55931 | 42769 | 144 | -6.5% | 1.6% |
| 36-40 | 1456 | 261 | 0 | 117[illegible] | 26355 | 14583 | 100 | -13.7% | -8.6% |
| 41+ | 303 | 43 | 0 | 6507 | 11240 | 4733 | 64 | -16.3% | -23.8% |

SOURCE: 1983, 1985, and 1987 Survey of Doctorate Recipients, National Research Council.

Table A-14. Historical and Projected Vacancy Ratios, Nonclinical Psychology, 1973-2000

| | Annual Averages | | | | | |
|---|---|---|---|---|---|---|
| | Total | | | R&D | | |
| Year | Vacancies | Ph.D.s | Ratio | Vacancies | Post-docs | Ratio |
| 1973-78 | 1189 | 1592 | 0.75 | 385 | 196 | 1.96 |
| 1978-83 | 1386 | 1555 | 0.89 | 359 | 226 | 1.[illegible] |
| 1983-87 | 1291 | 1435 | 0.90 | 422 | 261 | 1.61 |
| 1990-95 | | | | | | |
| Low | 995 | 1260 | 0.79 | 320 | 300 | 1.07 |
| Mid | 1252 | 1366 | 0.92 | 380 | 300 | 1.27 |
| High | 1551 | 1479 | 1.05 | 449 | 300 | 1.50 |

NOTE: Assumes that Ph.D. production changes in proportion to enrollment assumptions and postdoctoral production remains constant.

SOURCE: National Research Council.

Table A-15. Summary Projections for Other Behavioral Scientists

| Model Assumptions | | High | Mid | Low |
|---|---|---|---|---|
| 1. Graduate Student Enrollment | | 1.0% | 0.0% | -1.0% |
| 2. "All other" Employment Growth | | | | |
| | R&D | 7.0% | 5.0% | 3.0% |
| | NonR&D | 0.15 | 0.1 | 0.07 |

Projected Employment of Other Behavioral Scientists, 1973-2000
(in 100s of workers) (Grad enrollment model)

| | Low Case | | | Mid Case | | | High Case | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | % R&D |
| 1973 | 65.9 | 12.1 | 18.4 | 65.9 | 12.1 | 18.4 | 65.9 | 12.1 | 18.4 |
| 1975 | 82.0 | 14.4 | 17.6 | 82.0 | 14.4 | 17.6 | 82.0 | 14.4 | 17.6 |
| 1977 | 100.8 | 19.6 | 19.5 | 100.8 | 19.6 | 19.5 | 100.8 | 19.6 | 19.5 |
| 1979 | 108.5 | 24.8 | 22.9 | 108.5 | 24.8 | 22.9 | 108.5 | 24.8 | 22.9 |
| 1981 | 108.2 | 22.2 | 20.5 | 108.2 | 22.2 | 20.5 | 108.2 | 22.2 | 20.5 |
| 1983 | 124.8 | 19.0 | 15.3 | 124.8 | 19.0 | 15.3 | 124.8 | 19.0 | 15.3 |
| 1985 | 126.9 | 18.5 | 14.6 | 126.9 | 18.5 | 14.6 | 126.9 | 18.5 | 14.6 |
| 1987 | 127.4 | 24.0 | 18.9 | 127.4 | 24.0 | 18.9 | 127.4 | 24.0 | 18.9 |
| 1988 | 126.8 | 23.2 | 18.3 | 129.0 | 23.6 | 18.3 | 131.4 | 24.0 | 18.2 |
| 1989 | 126.4 | 23.2 | 18.3 | 130.7 | 23.9 | 18.3 | 135.8 | 24.7 | 18.2 |
| 1990 | 126.1 | 23.1 | 18.4 | 132.6 | 24.3 | 18.3 | 140.6 | 25.6 | 18.2 |
| 1991 | 125.9 | 23.2 | 18.4 | 134.7 | 24.7 | 18.4 | 145.9 | 26.4 | 18.1 |
| 1992 | 125.8 | 23.2 | 18.4 | 137.0 | 25.2 | 18.4 | 151.6 | 27.4 | 18.1 |
| 1993 | 125.9 | 23.2 | 18.4 | 139.5 | 25.6 | 18.4 | 158.0 | 28.4 | 17.9 |
| 1994 | 126.0 | 23.3 | 18.5 | 142.2 | 26.1 | 18.4 | 165.0 | 29.4 | 17.8 |
| 1995 | 126.4 | 23.3 | 18.5 | 145.1 | 26.6 | 18.3 | 172.8 | 30.5 | 17.6 |
| 1996 | 126.8 | 23.4 | 18.5 | 148.3 | 27.1 | 18.3 | 181.5 | 31.7 | 17.4 |
| 1997 | 127.4 | 23.5 | 18.4 | 151.8 | 27.7 | 18.2 | 191.1 | 32.9 | 17.2 |
| 1998 | 128.2 | 23.6 | 18.4 | 155.7 | 28.3 | 18.2 | 201.9 | 34.2 | 16.9 |
| 1999 | 129.1 | 23.7 | 18.4 | 159.9 | 28.9 | 18.1 | 213.9 | 35.6 | 16.6 |
| 2000 | 130.2 | 23.8 | 18.3 | 164.4 | 29.5 | 18.0 | 227.4 | 37.1 | 16.3 |
| Growth Rates: | | | | | | | | | |
| 73-87 | 4.8% | 5.0% | | 4.8% | 5.0% | | 4.8% | 5.0% | |
| 87-91 | -0.3% | -0.9% | | 1.4% | 0.7% | | 3.4% | 2.4% | |
| 87-20 | 0.2% | -0.1% | | 2.0% | 1.6% | | 4.6% | 3.4% | |

NOTE: This table does not include postdoctoral employment or unemployment.

SOURCE: Estimated by National Research Council.

Table A-15. (continued) Projections of Other Behavioral Scientists by Sector, 1987-2000
(Workers in 100s)

| Sector | 1987 | | | 1991 | | | 2000 | | | 1987-1991 Growth | | 1987-2000 Growth | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | % R&D | Total | R&D | Total | R&D |
| **Academic** | | | | | | | | | | | | | |
| Low Case | 107.7 | 16.9 | 15.7 | 101.4 | 15.1 | 14.9 | 89.4 | 13.3 | 14.9 | -1.5% | -2.7% | -1.4% | -1.8% |
| Medium Case | 107.7 | 16.9 | 15.7 | 107.7 | 16.0 | 14.9 | 107.7 | 16.0 | 14.9 | 0.0% | -1.3% | 0.0% | -0.4% |
| High Case | 107.7 | 16.9 | 15.7 | 114.6 | 17.1 | 14.9 | 133.0 | 19.8 | 14.9 | 1.6% | 0.3% | 1.6% | 1.2% |
| **All Other** | | | | | | | | | | | | | |
| Low Case | 19.7 | 7.2 | 36.3 | 24.5 | 8.0 | 32.9 | 40.7 | 10.5 | 25.8 | 5.6% | 3.0% | 5.7% | 3.0% |
| Medium Case | 19.7 | 7.2 | 36.3 | 27.1 | 8.7 | 32.1 | 56.8 | 13.5 | 23.8 | 8.3% | 5.0% | 8.5% | 5.0% |
| High Case | 19.7 | 7.2 | 36.3 | 31.3 | 9.4 | 29.9 | 94.4 | 17.2 | 18.3 | 12.3% | 7.0% | 12.8% | 7.0% |
| **Total** | | | | | | | | | | | | | |
| Low Case | 127.4 | 24.0 | 18.9 | 125.9 | 23.2 | 18.4 | 130.2 | 23.8 | 18.3 | -0.3% | -0.9% | 0.2% | -0.1% |
| Medium Case | 127.4 | 24.0 | 18.9 | 134.7 | 24.7 | 18.4 | 164.4 | 29.5 | 18.0 | 1.4% | 0.7% | 2.0% | 1.6% |
| High Case | 127.4 | 24.0 | 18.9 | 145.9 | 26.4 | 18.1 | 227.4 | 37.1 | 16.3 | 3.4% | 2.4% | 4.6% | 3.4% |

NOTE: Excludes postdoctorals and unemployed.

SOURCE: Projections by National Research Council.

Table A-16. Historical and Projected Vacancy Ratios, Other Behavioral Sciences, 1973-2000

| | Annual Averages | | | | | |
|---|---|---|---|---|---|---|
| | Total | | | R&D | | |
| Year | Vacancies | PhDs | Ratio | Vacancies | Post-docs | Ratio |
| 1973-78 | 1152 | 1211 | 0.95 | 289 | 97 | 2.97 |
| 1978-83 | 930 | 1093 | 0.85 | 120 | 124 | 0.97 |
| 1983-87 | 875 | 943 | 0.93 | 185 | 120 | 1.54 |
| 1987-95 | | | | | | |
| Low | 649 | 814 | 0.80 | 142 | 120 | 1.18 |
| Mid | 927 | 882 | 1.05 | 191 | 120 | 1.59 |
| High | 1353 | 955 | 1.42 | 250 | 120 | 2.08 |

NOTE: Assumes that Ph.D. production changes in proportion to enrollment assumptions and postdoctoral production remains constant. Grad model.

SOURCE: National Research Council.

National Academy Press

The National Academy Press was created by the National Academy of Sciences to publish the reports issued by the Academy and by the National Academy of Engineering, the Institute of Medicine, and the National Research Council, all operating under the charter granted to the National Academy of Sciences by the Congress of the United States.