Feasibility, Design and Planning Study for Evaluating the NIH Career Development Awards

Final Report

January 2, 2007

Prepared for:

National Institutes of Health
Bethesda, Maryland

Prepared by:

WESTAT
Rockville, Maryland

Feasibility, Design and Planning Study for Evaluating the NIH Career Development Awards

Final Report

Authors:

Joan Michie
Xiaodong Zhang
John Wells
Liam Ristow
Georgine Pion
Atsushi Miyaoka
Joy Frechtling

January 2, 2007

Prepared for:

National Institutes of Health
Bethesda, Maryland

Prepared by:

WESTAT
Rockville, Maryland

1. INTRODUCTION AND BACKGROUND

Overview

The National Institutes of Health (NIH) asked Westat to conduct a feasibility study for evaluating the Career Development Awards, a series of grant mechanisms sometimes called the “K-series” because of their activity code. The feasibility of an evaluation is defined as the extent to which an evaluation is appropriate and practical for implementation (Joint Committee on Standards for Educational Evaluation, 2003). For the current feasibility study, this meant determining if:

- The desired results of the Career Development Awards, across the Institutes and Centers (ICs) and award mechanisms can be identified;
- The indicator variables and outcomes to measure these desired results can be identified;
- Appropriate data to measure some of these outcomes are already available and of sufficient quality;
- New data collection for the other outcomes can be obtained without undue cost or burden; and
- An appropriate comparison group can be identified.

Westat conducted this study over a 9-month period between March and December 2006.

In this chapter, we provide the background and motivation for the feasibility study. We then outline the feasibility study research questions and methodology employed in chapter 2. We present the documentation of the NIH Career Development Awards program in chapter 3 and summarize the additional feasibility findings in chapter 4. In chapter 5, we present the suggested evaluation design.

Need for an Evaluation and General Background

The NIH has a long history of efforts to develop the nation’s workforce to conduct research in the health sciences. Extramural support has been provided through two similar programs, Research Training and Career Development. Beginning in 1937, the Research Training programs have supported students seeking doctoral degrees and postdoctoral training experiences. They have been reviewed and evaluated on a fairly regular basis.

Career Development Awards have been used since 1957 to guide and sustain the careers of individuals who have already completed various kinds of professional and research training. Since the inception of the program, more than 16,000 awards have been made at a total cost of almost $5 billion. Yet, evaluations of the program have been sporadic and have generally concentrated on specific kinds of awards or awards made by specific NIH Institutes and Centers. The purpose of this feasibility, design, and planning study was to prepare for a comprehensive evaluation of the Career Development programs.

Altogether there have been 24 different types of K awards, each with its own requirements. Over the years, some types of awards have been modified or dropped, while others have been added to reflect the changing needs of the workforce and information about program features that were considered to be most effective. (The types of awards are numbered in sequence from K01 through K30, with a few gaps in the sequence.) At present, 13 different types of awards are being offered.

The awards fall into two major classes—those in which the candidate has a mentor, and those in which the candidate has reached independence as a researcher and no longer needs additional mentoring. In addition, while most awards are made to individuals, several awards are made to institutions, whose representatives decide which individuals will be supported.

The main feasibility study activities were the following:

- Document the Career Development Award programs going back to 1957;
- Identify indicator variables and appropriate outcomes;
- Identify possible comparison groups;
- Identify data sources; and
- Develop a design for the evaluation.

2. FEASIBILITY STUDY RESEARCH QUESTIONS AND METHODS

In this chapter, we present the key search questions addressed in the feasibility study and the specific data collection methods and analysis approaches we employed to answer them.

Key Research Questions

Following are three key research questions and their subparts:

1. What are the characteristics of the Career Development Awards program going back to 1957?

   - What are the various NIH Career Development activities?
   - What are the explicit purposes and goals of each activity?
   - Which ICs utilized which K-series activities and in which years?
   - What have been the costs of each activity in current and constant dollars?
   - How many applications were reviewed and how many awards were made for each K-series activity?
   - What gaps are there in the NIH records?
   - What percent of the NIH budget and of the individual ICs has been allocated to K-series awards in each year?
   - Do the K-series activities have clear, measurable goals?
   - Are there recognized standards of performance in program announcements and agreed upon by relevant stakeholders that can be used to assess success?

2. What are the appropriate indicator variables and outcomes of interest?

   - Of the various performance measures, which outcomes are feasible to measure and which will most effectively reveal whether program goals are being or have been achieved?
   - What are theoretical or practical comparison groups that can be used in the determination of program success?

3. What data sources should be used to evaluate the program?

   - To what extent can existing data sources be used?
   - If it is determined that there is a need for primary data collection, what is the best way to collect the data?

Methodology for Answering the Research Questions

Assistance in all phases of the feasibility study was provided by the K-Series Evaluation Oversight Committee, which included representation across many of the ICs. Two formal meetings of the Committee were held, on April 18 and December 4, 2006. In addition, telephone discussions and e-mail exchanges were held with individual committee members on various topics. The members of the committee are shown in Exhibit 2-1.

Exhibit 2-1.—K-Series Evaluation Oversight Committee

Committee member	Affiliation
Walter Schaffer	OD, Project Officer
Robin Barr	NIA
Juliana Blome	OD
Genevieve deAlmeida-Morris	NIDA
David Eckstein	NCI
Sarah Glavin	NIDCR
Milton Hernandez	NIAID
James Hyde	NIDDK
Paul Johnson	NICHD
Henry Khachaturian	OD
Steve Klein	NICHD
Linda Kupfer	FIC
Bill McGarvey	OD
Robert Moore	OD
Carl Oberholtzer	NCI
James Onken	NIGMS
Katrina Pearson	OD
Carl Roth	NHLBI
Daniel Sklare	NIDCD
Jennifer Sutton	OD
Madeline Turkeltaub	NIAMS
Marina Volkov	NIMH
David Wilde	NCRR

The approaches used to address the individual research questions are described below.

What are the characteristics of the Career Development Awards program going back to 1957?

To document the K-series activities, we relied primarily on two sources of data: the Consolidated Grant Application File (CGAF) and the NIH Office of Extramural Research (OER) website. The CGAF contains the records for all applications for NIH grants and contracts. After receiving the appropriate security clearances, Westat obtained the records for all K-series applications, those awarded and those that were unsuccessful. These records included information such as the name of the principal investigator, the type of application, status of application, period of support, type of K-series, and IC. We had understood that an individual ID number had only been established within the past 5 years and the data file that we received was presented on an event basis rather than by individual. Therefore, we created an individual-based file for awardees and nonawardees by combining several variables in the CGAF to create individual IDs. Particular attention was given to examining the records of people with similar names and those with many records. (At the end of the feasibility study, we learned that another variable in the CGAF can be used as a unique ID number. However, program data that we generated matched those provided to us by NIH.)

We used the OER website to search for program announcements (PAs) and related documents for each of the K-series activities across all years it existed. The OER website's section on Funding Opportunities and Notices includes a searchable database that retrieves PAs and related documents from 1970 to present (http://grants1.nih.gov/grants/guide/index.html). PAs dating back to 1993 are available on this site and can be retrieved by release date, announcement number, title, issuing organization, activity code, or title. The OER website also provides links to Historical NIH Guide files, which contain PAs and related documents from 1970 to 1992 (http://grants1.nih.gov/grants/guide/historical/index.html). The NIH Guide files were initially distributed in hard copy from April 30, 1970 to January 10, 1992. These documents, historically published on an as-needed basis, were scanned to PDF files for public use.

Also, we located a limited number of relevant historical documents from NIH program officers, institute directors and staff, and the NIH library staff. For example, much of the information we obtained concerning the K06 award was taken from a report to NIH that summarized its 20 years of history at the time of its publication (Yasumura, 1984).

Additional information was obtained about the five types of institutional awards that NIH has offered over the years. These awards were made to institutions, which then distributed funds to individual trainees. For institutional awards, only the principal investigator is included in the CGAF; the individuals supported by the grants who are the focus of the evaluation are not. Individual ICs maintain additional records regarding the institutional awards.

To determine what types of individual records are available about individuals supported under an institutional award, NIH sent a request to the ICs for the following information on any type of institutional career development award at any point in the past:

- Description of the database that containing the following information:
    - Can you identify the number of individuals who received support each year from the K12 or other institutional career development awards?
    - Can you determine and report the identity of individuals who received support from such awards?
    - Is there information on candidates considered by the institutions but not appointed?
    - Is there information on the duration of appointments?
    - Are there other relevant variables in the database? If so, can you please list them?
- How can a contractor access the database?
- Who is an appropriate contact?
- In the absence of a database, do you have hardcopy records of individuals supported by institutional career development awards?
- What types of information are included in the hardcopy records?

- Who is the contact for such records?

Those who had access to such information were asked to provide the following:

- The number of new candidates for each year of the program.
- The duration of a typical appointment.

What are the appropriate indicator variables and outcomes of interest?

A literature review and examination of reports of evaluations of similar programs were used for determining indicator variables, outcomes of interest, and comparison group options. Based on the literature review, a logic model of the Career Development Award program was developed, which guided the development of indicators and measures and showed the relationships among them.

The selection of comparison groups was guided by a review of design theory and evaluations of similar programs. The pros and cons of different options were considered in terms of comparability and feasibility in data collection. In light of the selected comparison approach, potential issues with regard to sample size and sampling strategy, as well as data analysis approach, were explored.

One other strategy used to address this question was to obtain information about the evaluation of the K22 program that is currently underway. In addition, a meeting was held with the Federation of American Societies for Experimental Biology (FASEB), a group that is interested in obtaining information on clinician scientists who had received particular K awards tailored to this group.

It should be noted that in documentation activity, all K awards were included, but several K awards were excluded in the other phases of the feasibility study. The Transition Career Development Award (K22) and the Clinical Research Curriculum Award (K30) were excluded because separate evaluations of them are being conducted. Similarly, the Mentored Clinical Scientist Development Awards (K12) being evaluated by the National Institute of Child Health and Human Development, which are the Building Interdisciplinary Research Careers in Women's Health (BIRWCH), Women's Reproductive Health Research (WRHR), and the Roadmap K12, will not be included in the overall K evaluation. The Academic Career Award (K07) was going to be excluded because it can be somewhat different from the other K awards, but the Oversight Committee decided at their final meeting that it should be included.

What data sources should be used to evaluate the program?

An expert consultant, Georgine Pion, who has conducted prior research similar to the career award evaluation, reviewed the extant data sources that would be available for conducting the evaluation. She examined sources of information on both M.D.s and Ph.D.s and assessed the pros and cons of each source.

For outcomes that are not addressed in extant data sources, we examined approaches for primary data collection, including the conduct of a survey and bibliometrics.

3. DOCUMENTATION OF THE NIH CAREER DEVELOPMENT AWARDS PROGRAM

Westat has collected and analyzed data from various sources and developed a summary of the Career Awards program since its inception nearly 50 years ago. Our analysis includes information on all K awards, both active and inactive.

First we describe the gaps that exist in NIH records. Then we provide detailed information at the program level (i.e., across all K-series activities) on a year-by-year basis. The program-level data include the following:

- Total number of applications and new awards across all K activities;
- Total number of applications and new awards by K category;
- Total number of applications and new awards by IC; and
- Total dollars (current and in adjusted 2004 dollars) spent on all K awards, including total dollars spent on all K awards as proportion of total NIH budget.

Next, we provide a set of information on each of the individual K-series activities. Information specific to each of the individual K-series activities is organized numerically and includes the following:

- In which years the specific K award was utilized;
- Which ICs have utilized the specific K award and in which years;
- Summary of the purpose and goals of the specific K award as indicated in program announcements (PAs) and any changes in purpose or goals over time;
- Summary of other information, including award amounts, duration of awards, and eligibility, as stated in PAs and any changes in those areas over time;
- Number of applications and new awards for the specific K award by year;
- Number of applications and new awards for the specific K award by IC and by year; and
- Total dollars spent on the specific K award by IC and by year.

It should be noted that the program-level tables show both new and competing continuation awards, including change of institution (competing continuation), because these are the categories that we are recommending for the evaluation sample. In the tables showing individual K awards, only new awards are shown because the number of competing continuation awards for any individual K award is minimal in most cases.

Finally, we provide information about the institutional awards.

Gaps in Existing Records

As stated above, one purpose of the feasibility study was to identify gaps in NIH records. Therefore, it should be noted that our search of the NIH Office of Extramural Research website's database resulted in limited information on many of the specific K awards with respect to PAs and other historical information. As Exhibit 3-1 shows, we located a complete record of program announcements for only 10 of the 24 K-series awards. Seven of these awards were relatively new awards, originating since 1999. The only three awards that originated in prior decades for which we were able to locate a complete record were the K06, which was initiated in 1957, and the K11 and K12, both of which were initiated in 1984. We located only a partial record for another 10 awards, and this was especially the case for K-series activities that were initiated during the 1960s and 1970s. In many of these instances, the earliest records we found were from the 1980s. Our search of the database and other available historical documents did not result in any program announcements for four of the K-series awards.[1]

Exhibit 3-1.—Summary of available information on K-series awards

K award	Complete record	Partial record	No record	Earliest record	Earliest award
K01		X		1985	1968
K02		X		1987	1968
K03			X	NA	1962
K04		X		1974	1968
K05		X		1985	1970
K06	X			NA	1957
K07		X		1977	1974
K08		X		1977	1974
K10			X	NA	1974
K11	X			1984	1984
K12	X			1984	1984
K14		X		1988	1985
K15		X		1989	1985
K16		X		1988	1985
K17			X	NA	1993
K18	X			2003	2003
K20		X		1993	1989
K21			X	NA	1989
K22	X			1999	1999
K23	X			1999	1999
K24	X			1999	1999
K25	X			2000	2000
K26	X			2000	2000
K30	X			1999	1999

NA = not available.

[1] The only information we were able to locate on the K03, K10, K17, and K21 consisted of brief descriptions of each of these awards in documents entitled, "Activity Codes, Organization Codes, and Definitions Used in Extramural Programs." These documents, which are published annually and are called "activity code books," contain code definitions used in the IMPAC (Information for Management, Planning, Analysis, and Coordination) and CRISP (Computer Retrieval of Information on Scientific Projects) systems to identify various aspects of extramural research activities of NIH.

Summary of Program Level Information

As Figure 3-1 demonstrates, the number of applications for new and competing awards and the number of new and competing awards has increased steadily over time. Whereas in the 1960s the number of applications rarely exceeded 500, that number has exceeded 1,000 during every year since 1993. The number of new and competing awards has followed a similar trend.

Figure 3-1.—Total number of applications for new and competing continuation Career awards, and number of awards, by year

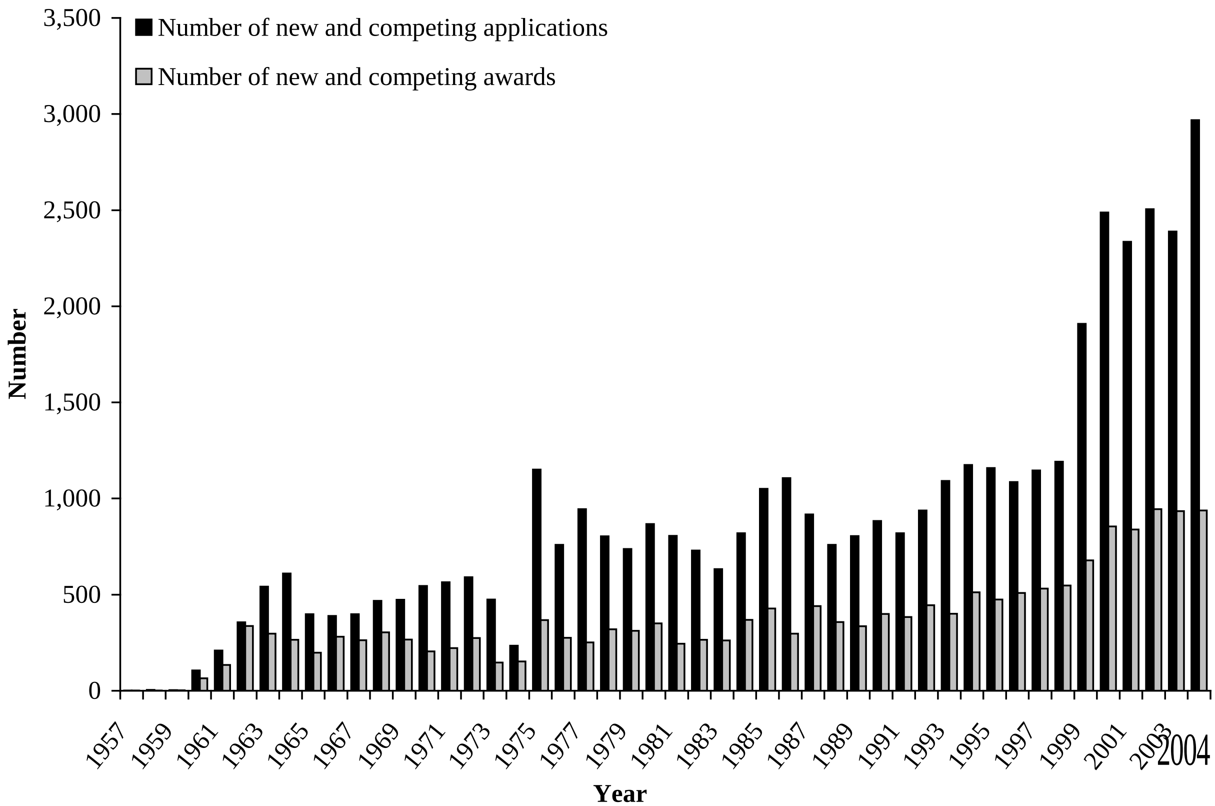


Figure 3-2 shows the success rate for applicants for new and competing awards for the entire history of the K-series activities. Over time, the level of competitiveness has steadily increased, with the exception of wide variations from year to year during the 1960s and 1970s. Not since 1975 has the success rate for applicants exceeded 50 percent, and the rate has been below 40 percent during every year since 2000.

Figure 3-2.—Success rate for applications for new and competing continuation Career awards, by year

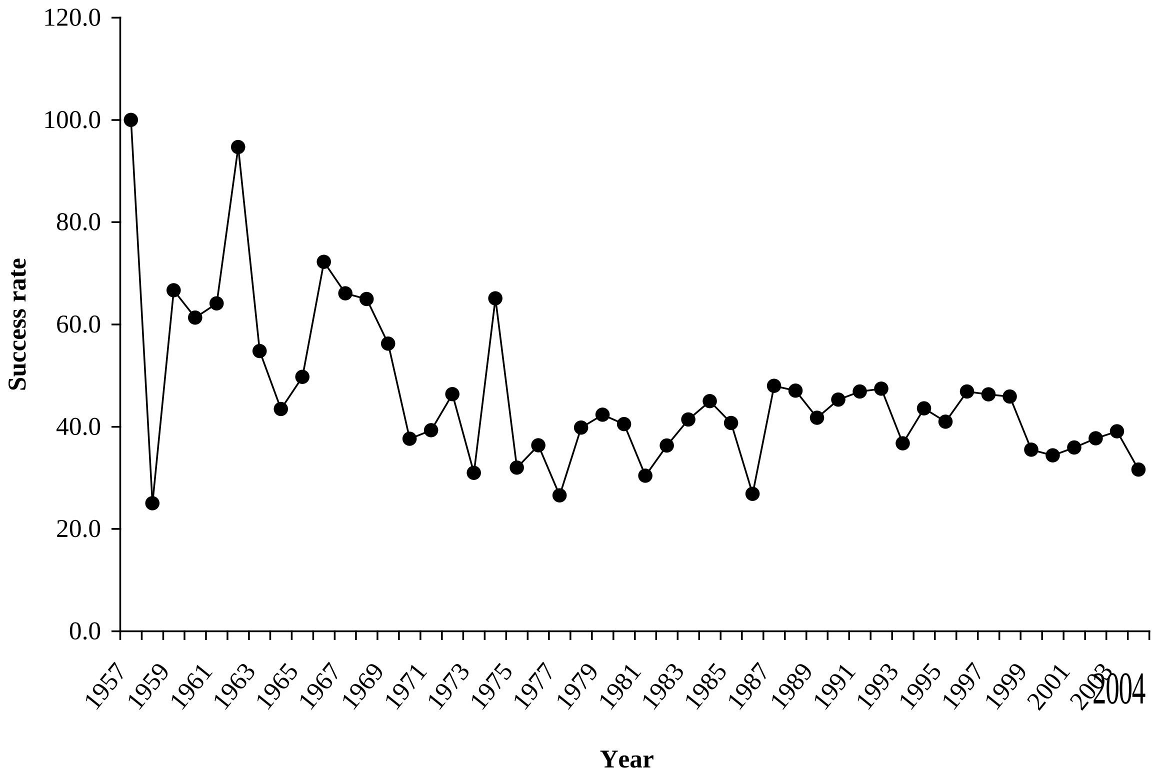


Figure 3-3 shows the numbers of new and competing awards provided by K activity across the entire history of the Career Awards program. The K04 and K08 activities were the leading categories with respect to the most individual awards. In addition, more than 12,000 of the total of more than 17,000 awards, or more than two-thirds, were provided by five K activities, the K01, K03, K04, K08, and K23.

Figure 3-3.—Total number of new and competing continuation awards, by K category

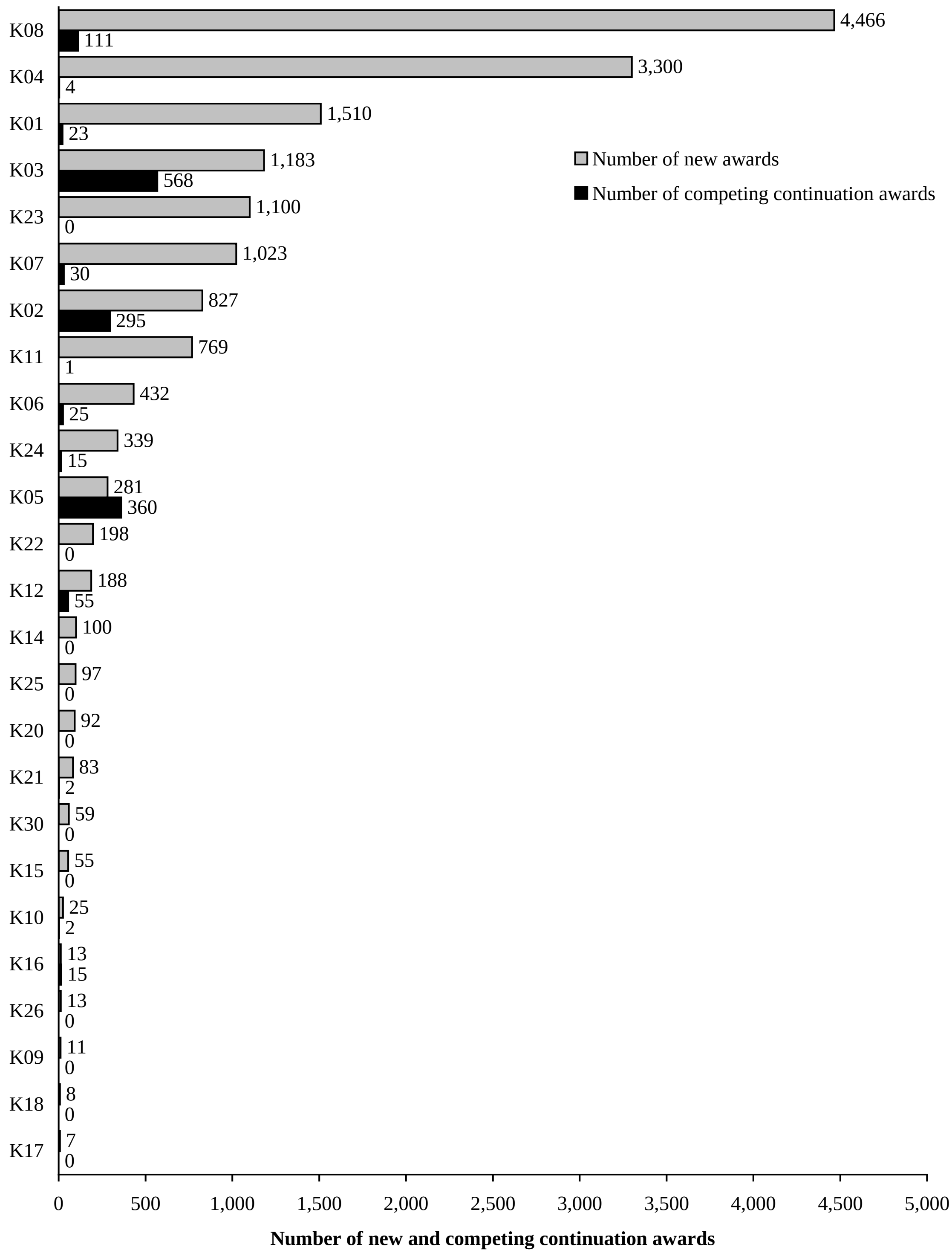
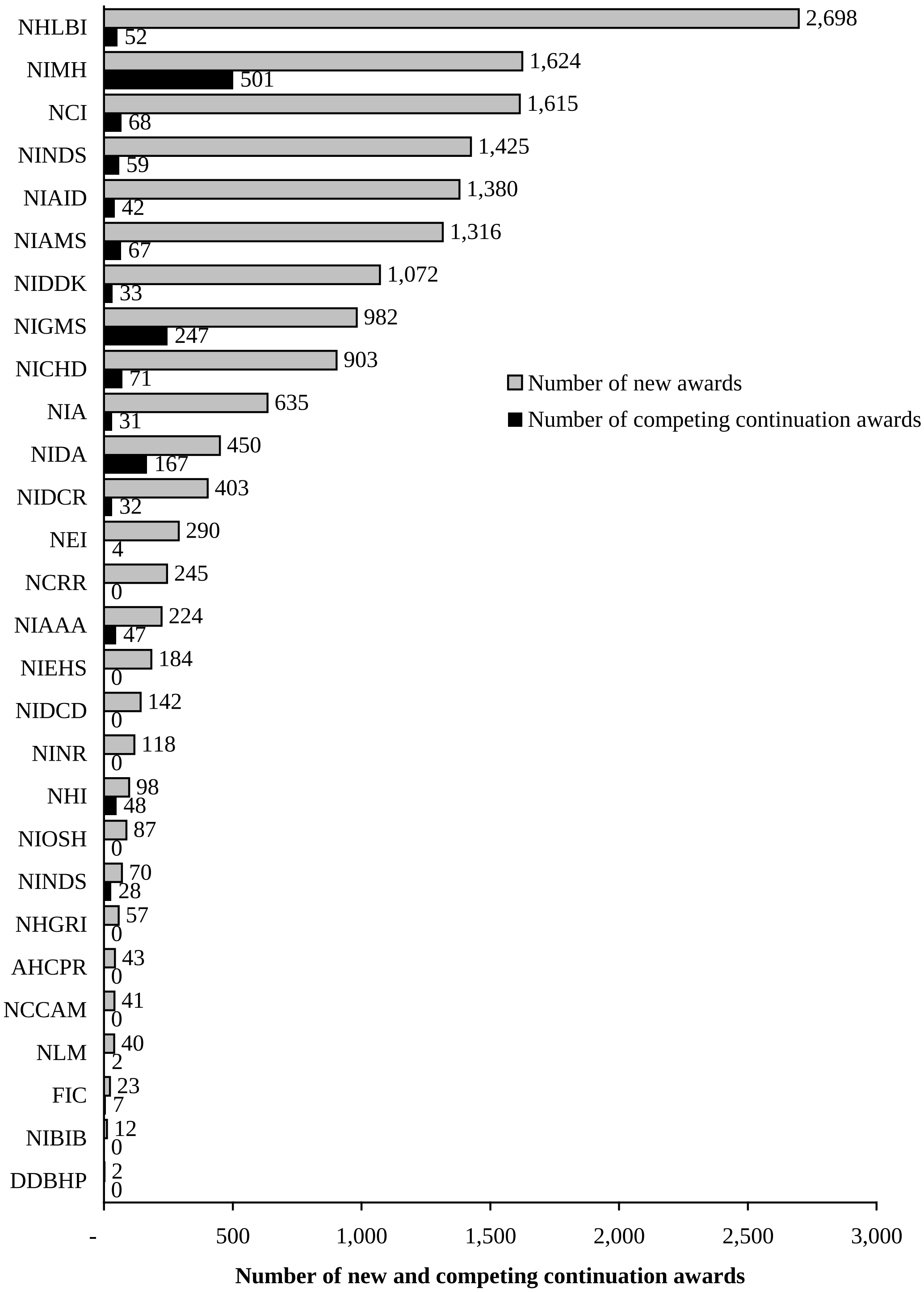


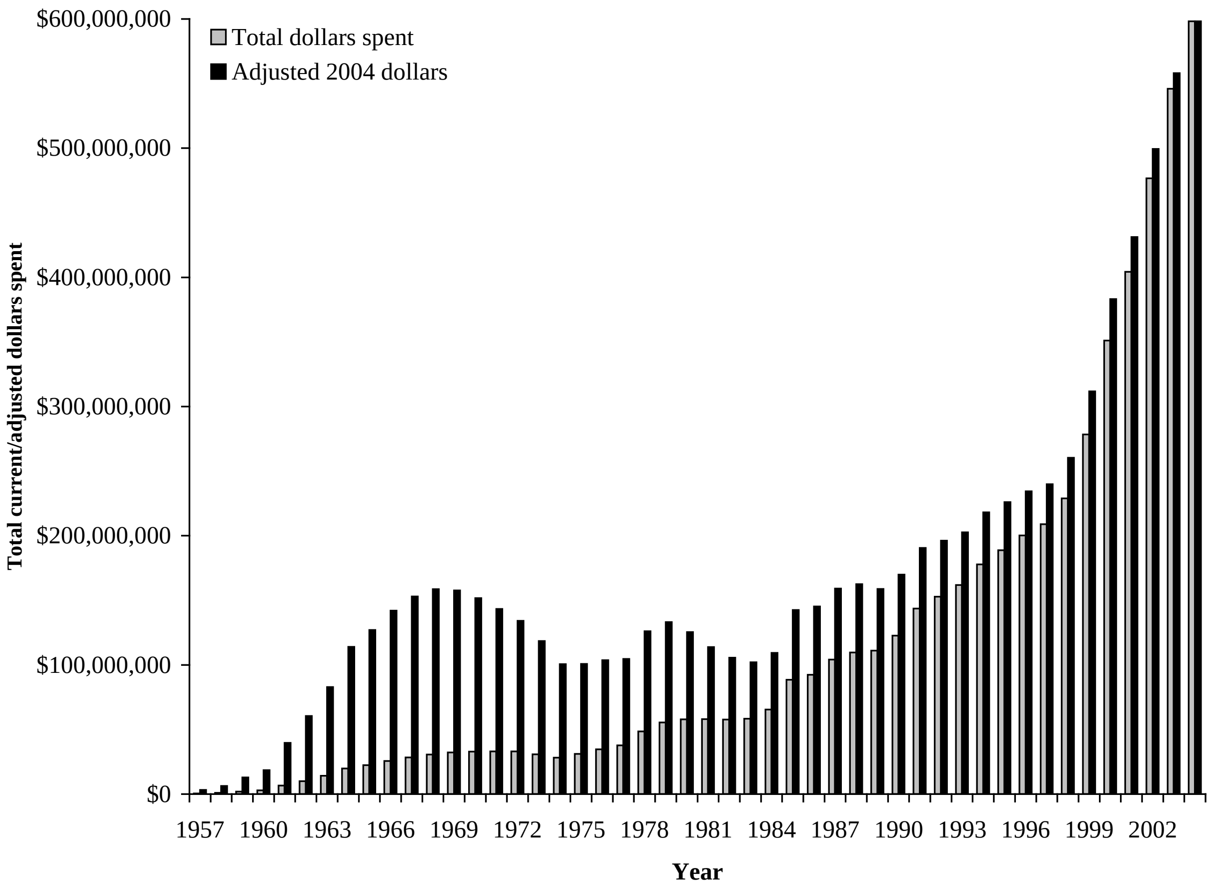
Figure 3-4 shows the number of new and competing awards provided by IC across the program's history. The awards are more evenly distributed across the various ICs than across K activities, as shown in Figure 3-3, with the leading IC (NHLBI) comprising approximately 15 percent of the awards. Only one other ICs account for more than 10 percent of the total awards—NIMH provided approximately 12 percent of the total awards.

Figure 3-4.—Total number of new and competing continuation Career awards, by IC



Finally, Figure 3-5 displays the annual expenditures on the Career Awards program, in current and adjusted dollars. As the data show, annual expenditures have increased steadily over time, with two exceptions during the periods of the late 1960s and late 1970s, in which relatively substantial increases occurred and were followed by modest declines in expenditures in subsequent years.[2] As a proportion of the total NIH budget, annual expenditures on the Career Awards program have ranged from less than 1 percent during the initial years of the program prior to 1960 to just over 3 percent in 1970. From the early 1970s to the late 1990s, expenditures as a proportion of the total NIH budget remained steady, comprising between 1.4 and 1.8 percent. In each year since 2000, the proportion has remained steady at approximately 2 percent.

Figure 3-5.—Total dollars spent on Career awards, current and adjusted 2004 dollars, by year



Summary of Information at the K-Award Level

Exhibit 3-2 summarizes available information for each of the individual K awards offered since 1957.[3] The exhibit presents information obtained from program announcements and other historical documents regarding the purpose of the award, award eligibility criteria, funding level, duration, and

---

[2] Program expenditures were adjusted to fiscal year (FY) 2004 dollars by using the federal budget composite deflator, which is based on the Consumer Price Index (CPI).

[3] The sole exception is the K09. According to CGAF data, there were 14 K09 awards. However, they were unlike all other K awards and were listed in activity code books as a support mechanism for the "chairman of the review committee" and "operation of the review group" for conducting internal evaluation activities. Based on this information, we have excluded the K09 from our historical review.

Exhibit 3-2.—Summary profile of all K awards

K-series	Award Name	Purpose	Years offered	Mentored award?	Eligibility criteria	Funding offered[1]	Award duration
K01	Mentored Research Scientist Development Award	Promotes research independence for new scientists or those entering new field.	1968-present	Yes	Research doctorate degree; level of experience dependant on IC and type of award.	Salary: $50,000-$180,000 per year; research support: up to $50,000 per year.	3-5 years
K02	Independent Scientist Award	Supports research development for newly independent scientists.	1968-present	No	Research doctorate degree; full-time academic employment; independent research support; within ~5 years of training.	Salary: up to $75,000 per year; limited research support.	3-5 years
K03	Research Career Development Award	Supported research development for newly independent scientist (predated K04)	1962-1968	?	?	?	?
K04	Research Career Development Award	Supported research development for newly independent scientists	1968-1996	No	Research doctorate degree; full-time academic employment; independent research support; within ~5 years of training.	Salary: up to $50,000 per year (as of 1991); limited research support.	Up to 5 years
K05	Senior Scientist Award	Supports research development for established scientists.	1968-present	No	Research doctorate degree; full-time academic employment, independent research support, distinguished research record.	Salary: up to federal salary limit; limited research support.	Up to 5 years
K06	Research Career Award	Supported established mid-career research scientists.	1957- 1965	No	Research doctorate degree; full-time academic employment independent research support; substantial number of working years ahead.	Salary: $25,000 per year for duration of recipient's career (as of 1964).	Duration of career
K07	Academic Career Award	Supports faculty members who wish to develop their research expertise and/or build or improve upon research curricula at their institution.	1971-present	Yes (for development awardees)	Research doctorate degree; full-time academic employment; other criteria varies by type of award.	Salary: up to $75,000 per year; research support: $50,000 per year.	Up to 5 years

Exhibit 3-2.—Summary profile of all K awards—continued

K-series	Award Name	Purpose	Years offered	Mentored award?	Eligibility criteria	Funding offered[1]	Award duration
K08	Mentored Clinical Scientist Development Award (individual)	Supports mentored research training and experience for clinicians.	1972-present	Yes	Clinical doctorate degree; full-time academic employment.	Salary: up to $85,000 per year; research support: $50,000 per year.	3-5 years
K10	Special Scientific Projects	Supported the utilization and to increase the understanding of record information in fields related to health.	1972-1992	?	?	?	?
K11	Physician Scientist Award (individual)	Supported mentored research training and experience for clinicians.	1984-1996	Yes	Clinical doctorate degree; full-time academic employment; some research experience required to forgo training component.	Salary: up to $40,000 per year; research support: $10,000-$20,000 per year (both as of 1990).	Up to 5 years
K12	Mentored Clinical Scientist Development Award (institution)	Supports institutional didactic research training programs for clinicians.	1984-present	Yes	Institutions must have adequate faculty/staff capacity to support program; participants must hold clinical doctorate degree.	Salary for participants: up to $75,000 per year; Research support for participants: $20,000-$30,000 per year.	Up to 5 years for training
K14	Minority School Faculty Development Award	Supported the development of faculty investigators at minority institutions.	1985-2003	Yes	Research doctorate degree; full-time employment at minority institution.	Salary: up to $50,000 per year; research support: $20,000 per year (both as of 1991).	Up to 5 years
K15	Dentist Scientist Award (individual)	Supported supervised clinical research training and experience for dentists.	1985- 1996	Yes	?	Salary: up to $50,000 per year (as of 1990).	?
K16	Dentist Scientist Award (institution)	Supported institutional clinical research training programs for dentists.	1985-1996	Yes	Institution required to have adequate faculty/staff capacity to support program; participant qualifications unknown.	Salary for participants: up to $50,000 (as of 1990).	Up to 5 years for training
K17	Research Career Re-entry Program	Supported basic or clinical scientists who planned to reenter their fields as active investigators after an absence.	1993-1996	?	?	?	?
K18	Career Enhancement Award	Supports short-term mentored research training for new and established investigators	2003-present	Yes	Research doctorate degree; level of experience required varies by award.	Salary: up to federal salary limit; research support: $50,000 per year.	6 months to 1 year
K20	Scientist Development Award for Clinicians	Supported research training and development experiences for clinicians.	1989-1996	Yes	Clinical doctorate degree; at least 2-3 years of postdoctoral research experience.	Salary: up to $75,000 per year; research support up to $50,000 per year.	Up to 5 years

Exhibit 3-2.—Summary profile of all K awards—continued

K-series	Award Name	Purpose	Years offered	Mentored award?	Eligibility criteria	Funding offered[1]	Award duration
K21	Scientist Development Award	Supported the development of outstanding biological or behavioral scientists pursuing alcoholism, drug abuse, or mental health research.	1989-1996	Yes	?	?	?
K22	Transition Career Development Award	Award is designed to ease the transition of new investigators from postdoctoral trainee to independent scientist.	1998-present	Yes (for two-stage awards)	Research or clinical doctorate degree; level of experience varies by type of award.	Salary: up to $140,000 per year; research support: up to $100,000 per year (both vary by type of award and institute).	Up to 5 years
K23	Mentored Patient-Oriented Research Career Development Award	Supports the career development of new clinical investigators pursuing patient-oriented research.	1998-present	Yes	Clinical doctorate degree with all specialty and subspecialty training completed; commitment to patient-oriented research.	Salary: $75,000-$180,000 per year; research support: up to $50,000 per year.	3 to 5 years
K24	Midcareer Investigator Award in Patient-Oriented Research	Supports career development for established clinician investigators involved in patient-oriented research.	1999-present	No	Clinical doctorate degree; full-time employment at associate professor level; commitment to patient-oriented research.	Salary: up to federal salary limit (for 25-50% effort); research support: up to $50,000 per year.	3 to 5 years
K25	Mentored Quantitative Research Development Award	Supports research skill development for scientists with strong quantitative backgrounds.	2000-present	Yes	Advanced degree in quantitative science or engineering; demonstrated research record.	Salary: $75,000-$180,000 per year; research support" up to $50,000 per year.	3 to 5 years
K26	Midcareer Investigator Award in Biomedical and Behavioral Research	Supports career development of established biomedical or behavioral scientists.	2000-present	No	Research doctorate degree; within 15 years of training.	Salary: up to $92,000 per year; research support: up to $25,000 per year.	3 to 5 years
K30	Clinical Research Curriculum Award	Supports institutional clinical research training programs.	1999-present	Yes	Institutions must have adequate faculty/staff to support program; participant eligibility unknown.	Total program costs: $300,000 per year.	2 years for training

[1] When available, approximate salary contribution limits set by NIH ICs are given. In other cases, salary is typically specified in program announcements as no higher than the federal limit, which is currently set at roughly $180,000 per year.

NOTE: The data shown in the exhibit represent the most current information available.

SOURCE: CGAF data file (for years offered), program announcements (PAs and PARs) and requests for applications (RFAs) issued by the NIH and its ICs. Some information also drawn from program descriptions appearing in NIH activity code booklets.

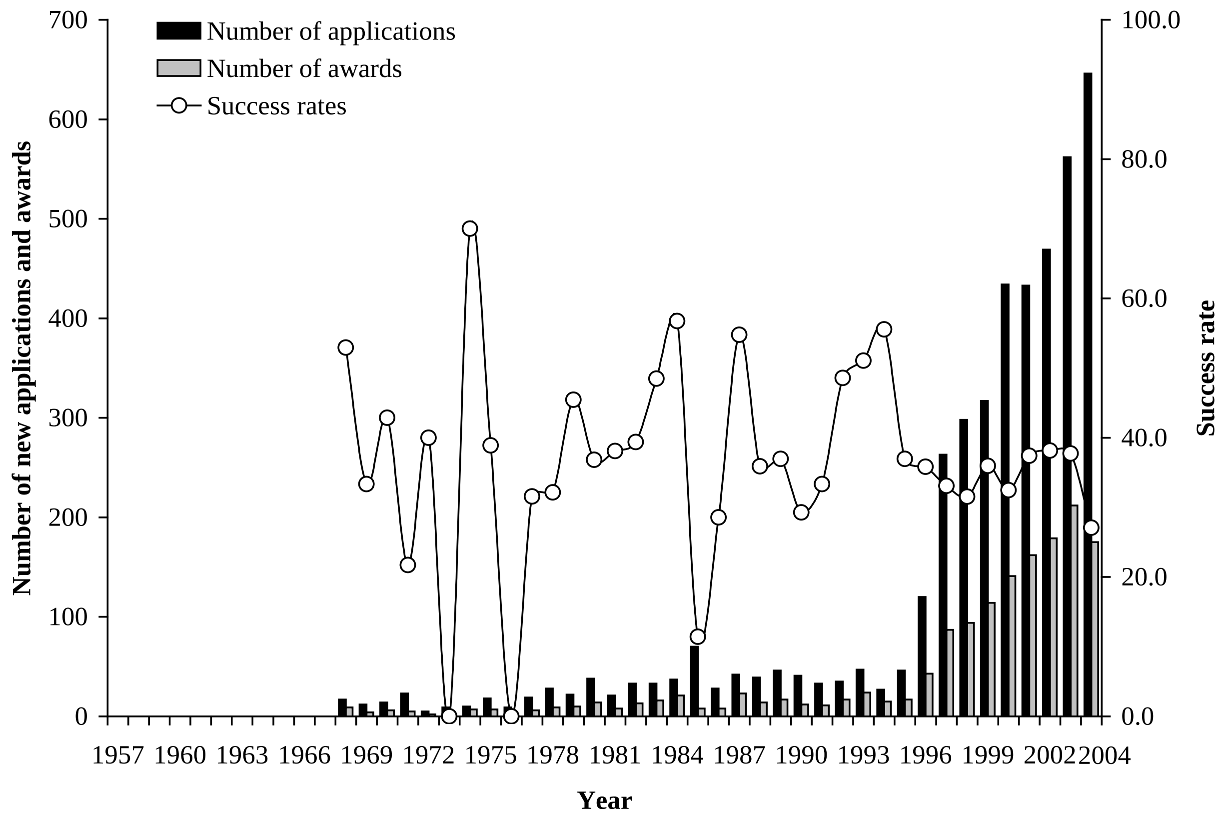
other characteristics. The exhibit also specifies the period of time in which the award was offered, based on CGAF data. For several awards, 1996 was the last year in which they were offered because they were consolidated with other awards. Overall, 14 awards were consolidated into 6 new career development awards during the mid-1990s at the time that the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) institutes were merged back into NIH. Specifics about which awards were consolidated are contained in the descriptions of the K awards.

Descriptions for each of the individual K awards, organized numerically, make up the remainder of this chapter. These detailed descriptions were developed through a review of program announcements and analysis of CGAF data. In addition to presenting detailed information on the aforementioned characteristics, each description addresses (where relevant) how the award has changed over time and identifies instances in which an award was merged with others. Many of the individual K award descriptions are accompanied by figures that display data on the number of new applications and awards, as well as total expenditures, on a year-by-year basis. A full set of K-level figures generated from the CGAF may be found in Appendix A, which is in a separate file.

K01: Mentored Research Scientist Development Award[4]

In existence since 1968, the Mentored Research Scientist Development Award (K01) supports the career development of scientists in the biomedical, behavioral, and clinical science fields. The program aims to help investigators achieve independent status by providing salary support for "protected" time free from teaching and administrative responsibilities so that awardees may focus on building their research programs. Ultimately, the program is intended to give researchers the tools they need to be competitive for prestigious research grants (such as the NIH R01 awards) and pursue crucial avenues of scientific inquiry. Figure 3-6 plots annual application and award data along with corresponding application success rates. As can be seen, K01 applications began to rise sharply in the 1990s, while awards increased at a more modest pace. As a result, application success rates have moderated to between 30 and 40 percent in recent years, with the notable exception of 2004, when award rates dropped to around 25 percent.

Figure 3-6.—Number of new K01 applications and awards and success rate, by year

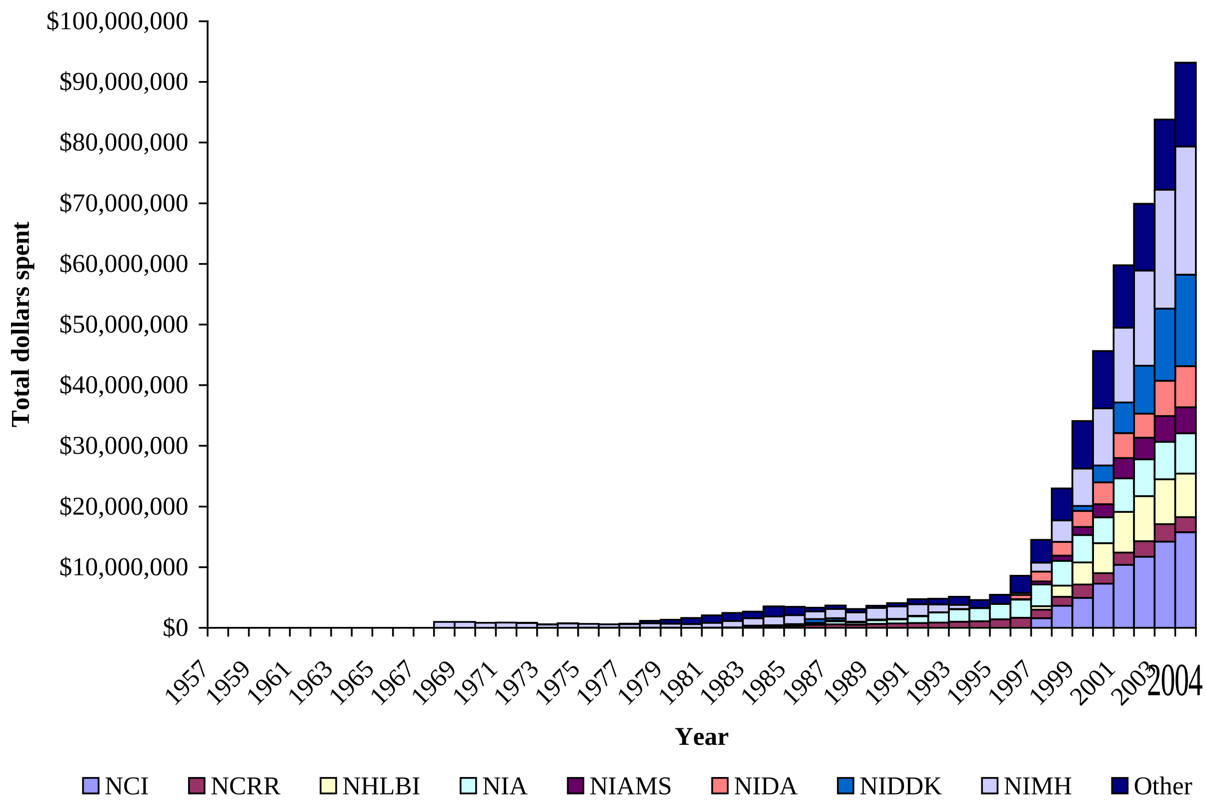


[4] In the mid-1990s, the K01 replaced the K01 (old version), K14 (Minority School Faculty Development Award), K17 (Research Career Re-entry Award), and K21 (Scientist Development Award).

The K01 is offered in various forms, although all share the same basic goals described above and use similar structures. As indicated in the title of the award, all versions support mentored research and training experiences for scientists—either to support new investigators or established researchers taking their career in new directions.[5] In the early and mid-2000s, the NIH offered variations on the K01 that supported scientists interested in pursuing research ethics and establishing research programs focused on the developing world. All K01 awards support investigators for a period ranging from 3 to 5 years and offer both salary support and funds to cover research and training expenses. Limits on salary support vary widely across institutes, currently ranging from $50,000 to $180,000 per year. K01 award announcements from the early 1980s suggest that salary support limits have roughly kept pace with or slightly exceeded inflation.[6] Research and training support funds have increased as well, now standing at $20,000 to $50,000 per year.

The K01 has been implemented widely by NIH ICs, especially in recent years. As shown in Figure 3-7, total spending on K01 awards has increased substantially in the past decade. In recent years, the National Institute of Mental Health has led K01 spending, followed by the National Cancer Institute and the National Institute of Diabetes and Digestive and Kidney Diseases.[7]

Figure 3-7.—Total dollars spent on K01 award by IC, by year



---

[5] The mentored feature of the award dates to at least the mid-1980s. See, for example, the National Institute on Aging Special Emphasis Research Career Award in Behavioral Geriatrics offered in 1983.

[6] A 1983 announcement for an NIA K01, for example, offered up to $30,000 in salary support, roughly $61,000 in current dollars.

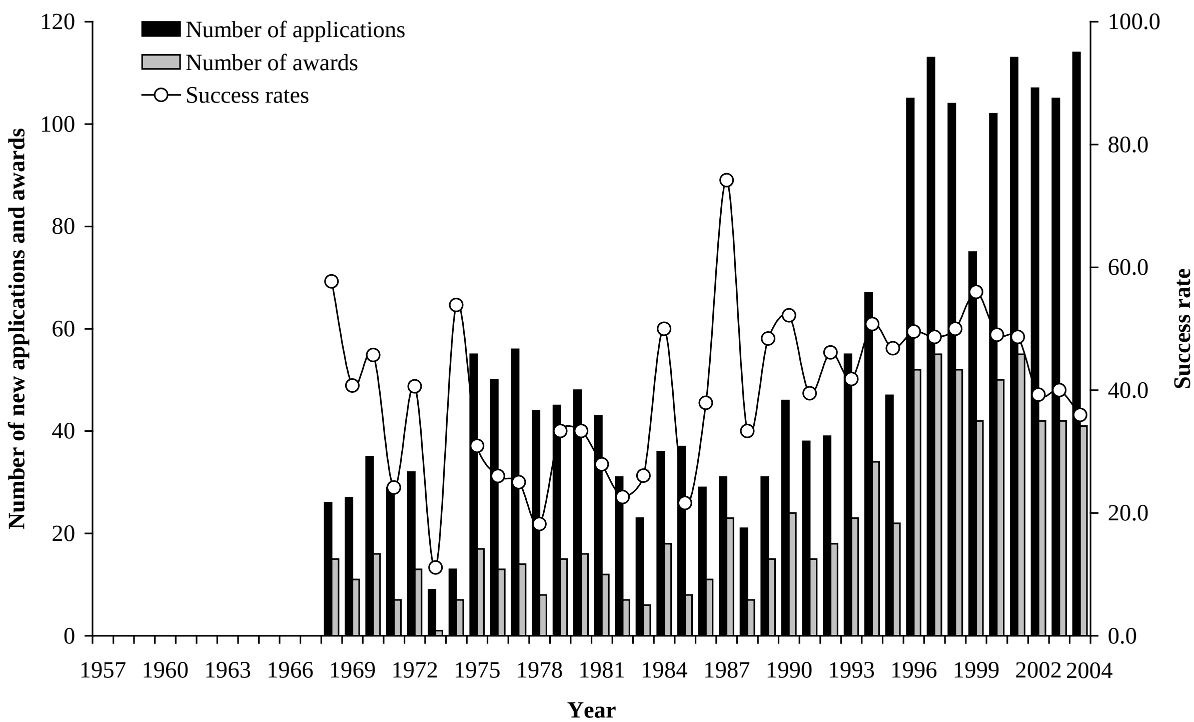
[7] For visual clarity, figures in this section showing total expenditures disaggregate only the eight top-spending ICs—in terms of total spending over the history of the award. The remaining ICs, if any, are grouped in the "other" category. Cases in which numerous ICs participate can therefore lead to relatively large combined expenditures in the "other" category.

Although institutes have generally adhered to the characteristics described above, they have used the award to support their specific scientific aims. A National Cancer Institute K01 known as the Howard Temin Award, for example, is intended to support the career development and transition to independent research of promising cancer researchers. Eligibility for the K01 is mostly determined based on an institute's target applicant pool. For example, institutes offering K01s to scientists who wish to make career changes typically require more experience than institutes using K01 funds to support new scientists (generally those with 2 or 3 years of postdoctoral training).

K02: Independent Scientist Award[8]

The K02 award supports protected time for independent scientists who require a period of intensive research focus as a means of enhancing their careers. The award is intended to foster the career development of promising scientists through 3 to 5 years of salary support, during which time recipients are expected to make a commitment of at least 75 percent effort to research activities. It is expected that the recipient will have full-time employment at an academic institution and independent research support at the time of the award (e.g., an R01 grant), so research funds are only provided on a limited, case-by-case basis. In existence since the late 1960s, the K02 goals and structure have remained fairly stable since 1985, the earliest year records are available.[9] As shown in Figure 3-8, K02 applications have risen substantially since the early 1990s, while new awards grew modestly in the 1990s before falling slightly in recent years.

Figure 3-8.—Number of new K02 applications and awards and success rate, by year

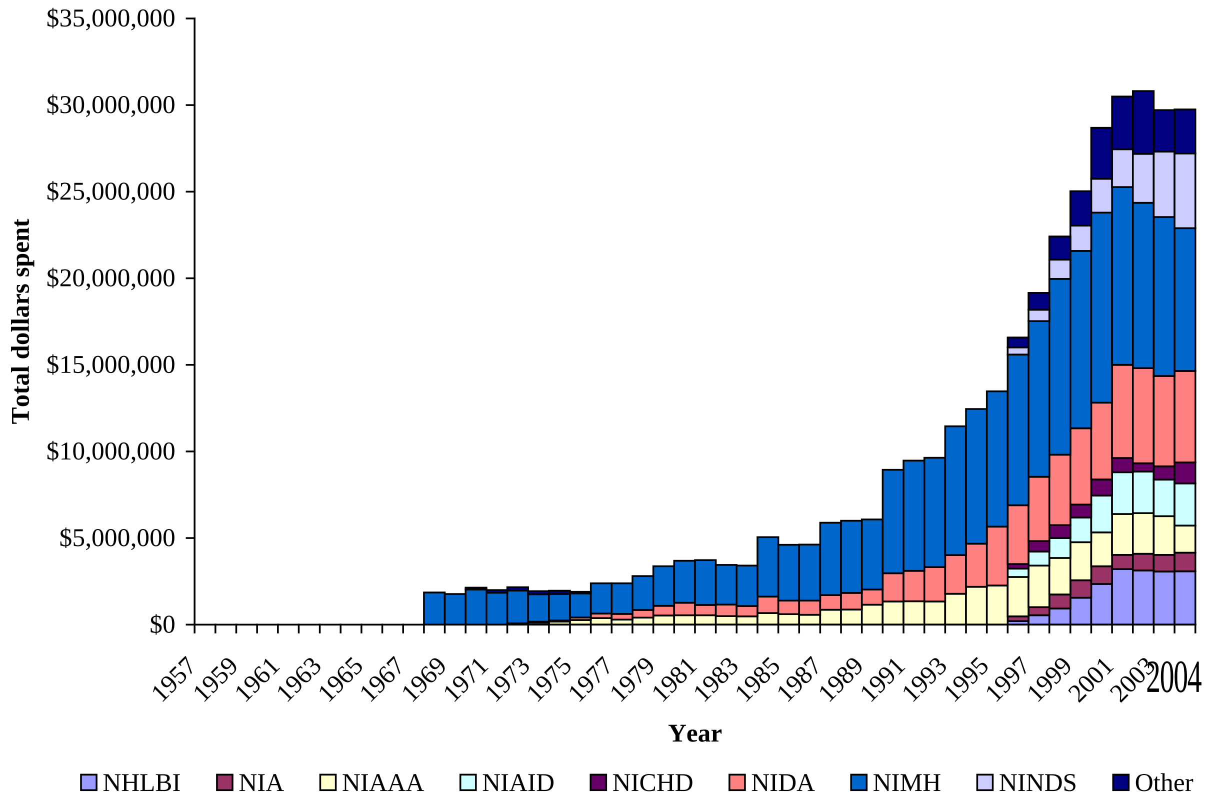


[8] In the mid-1990s, the K02 replaced the Research Scientist Development Award (old K02) and Research Career Development Award (K04).

[9] However, a 1985 announcement for a NIDA/NIMH K02 indicated that applicants would be eligible if they had completed a K01 grant. This criterion is not mentioned in current K02 announcements.

The K02 is offered widely across NIH ICs, although historical grant announcements at the IC level are somewhat limited. The Agency for Health Care Policy Research (AHCPR) has offered the K02 in support of health sciences researchers, and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) has offered the K02 in concert with other scientist development awards (e.g., the K01, K20, and K21). As of 1999, the AHCPR K02 offers new scientists (those within 5 years of completing their training) up to $75,000 in salary support. The same maximum salary support level was offered by the ADAMHA K02 in 1991. As shown in Figure 3-9, spending on the K02 rose steadily through the 1990s but has reached a plateau in recent years. A large share of K02 spending in recent years has been made by the National Institute of Mental Health, although this institute's K02 expenditures decreased slightly between 2001 and 2004.

Figure 3-9.—Total dollars spent on K02 award by IC, by year



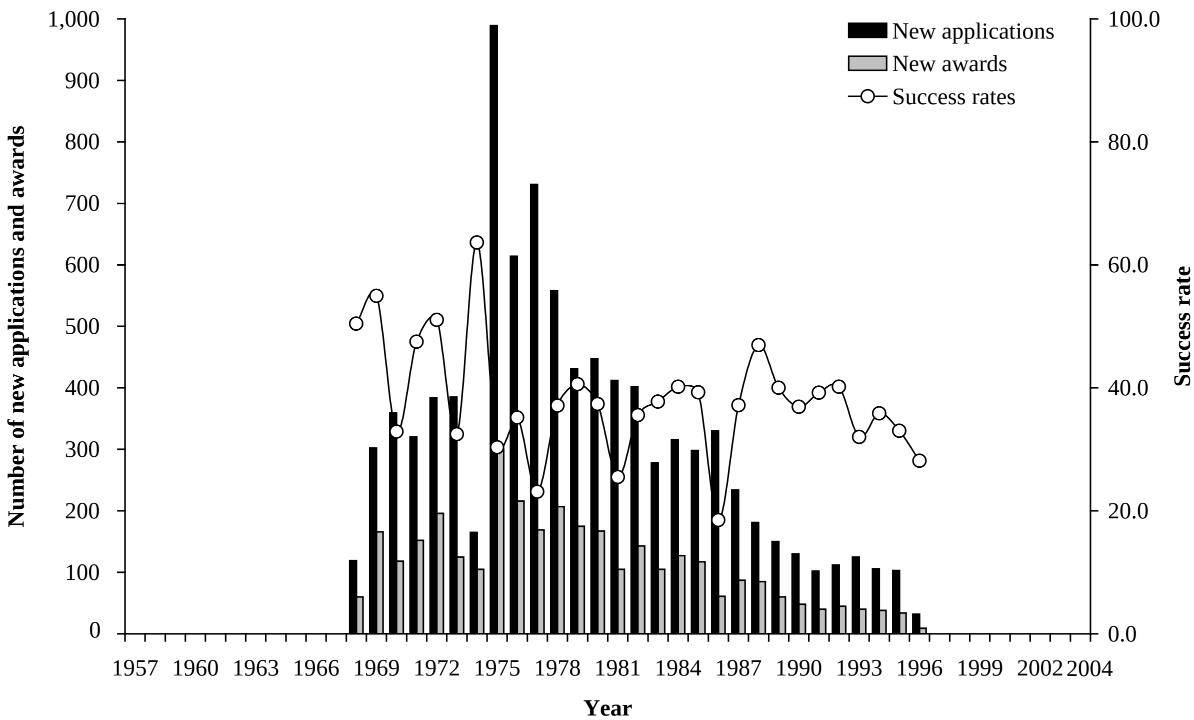
K03: Research Career Development Award

A predecessor of the K04 RCDA, the K03 award was intended to enhance the career development of outstanding scientists who required additional training and experience in a productive scientific environment to prepare them for independent research careers. NIH application data show that the number of new K03 applications and awards peaked in 1963—1 year after the program was introduced. Applications declined thereafter, and the program ended in 1968 (figures not shown). Additional historical information about the K03 is not available.

K04: Research Career Development Award

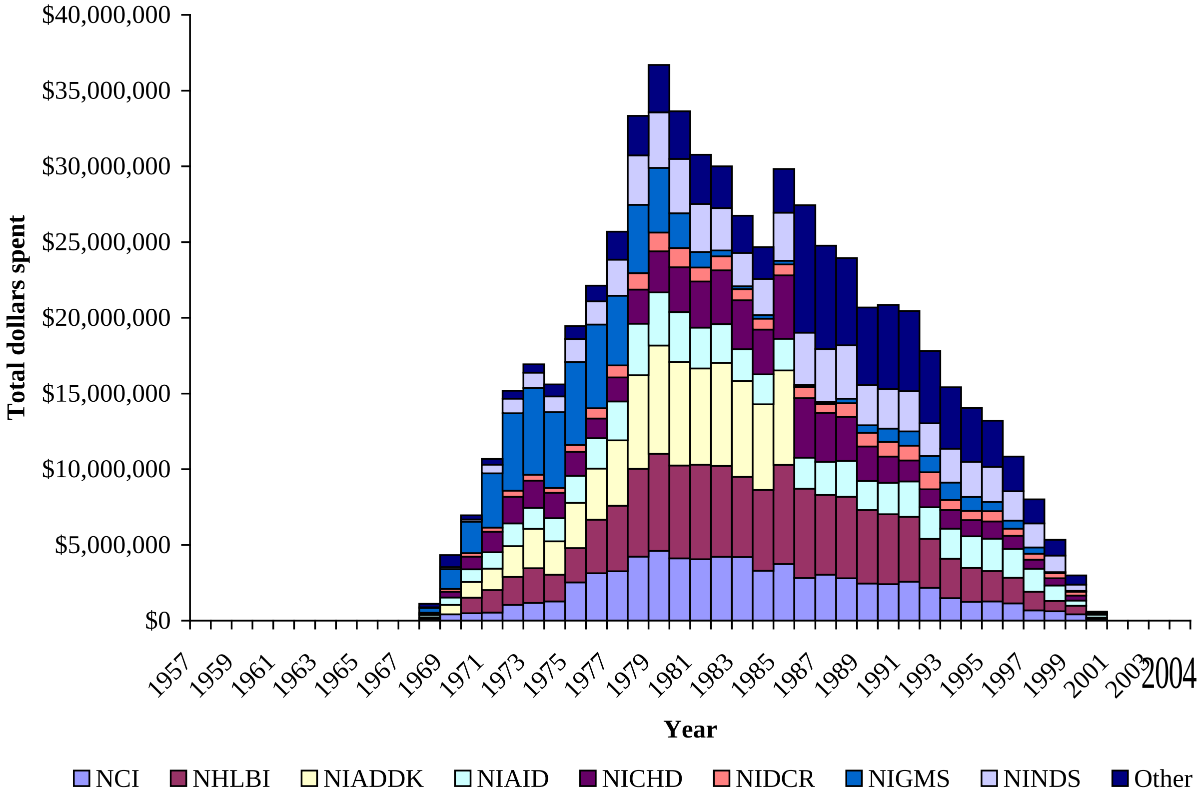
The K04 award, which was replaced in the mid-1990s by the Independent Scientist K02, provided salary support to outstanding scientists for the purpose of providing “protected” time for research activities. K04 recipients were to devote essentially full time to research and research-related (e.g., skill development) activities for the 5-year duration of the grant. The award was targeted at scientists at the junior faculty level with at least 5 years of postdoctoral research experience, including 2 years in which the recipient was supported by independent research grant support. More senior scientists with long research histories and/or tenure were not eligible for the K04 award. As shown in Figure 3-10, the K04 was offered from 1968 to 1996, with applications and awards peaking in 1975.

Figure 3-10.—Number of new K04 applications and awards and success rate, by year



Annual maximum salary support under the K04 ranged from $25,000 in the mid-1970s (shortly after the program began) to around $50,000 in the early 1990s.[10] Early versions of the award were renewable, a feature that was discontinued in 1971 when it was determined that 5 years was adequate for new faculty to solidify their research programs. K04 awards issued prior to the 1980s also allowed slightly less experienced investigators to apply, requiring at least 3 years of postdoctoral training instead of 5. However, early K04 records show that the program's basic aim of allowing newly independent scientists maximum time for research development to be fairly constant throughout the program's history. As shown in Figure 3-11, the K04 award was offered by a range of ICs, with large financial contributions offered in the mid-1980s by the National Heart, Lung, and Blood Institute and the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK, later renamed NIDDK).[11]

Figure 3-11.—Total dollars spent on K04 award by IC, by year



---

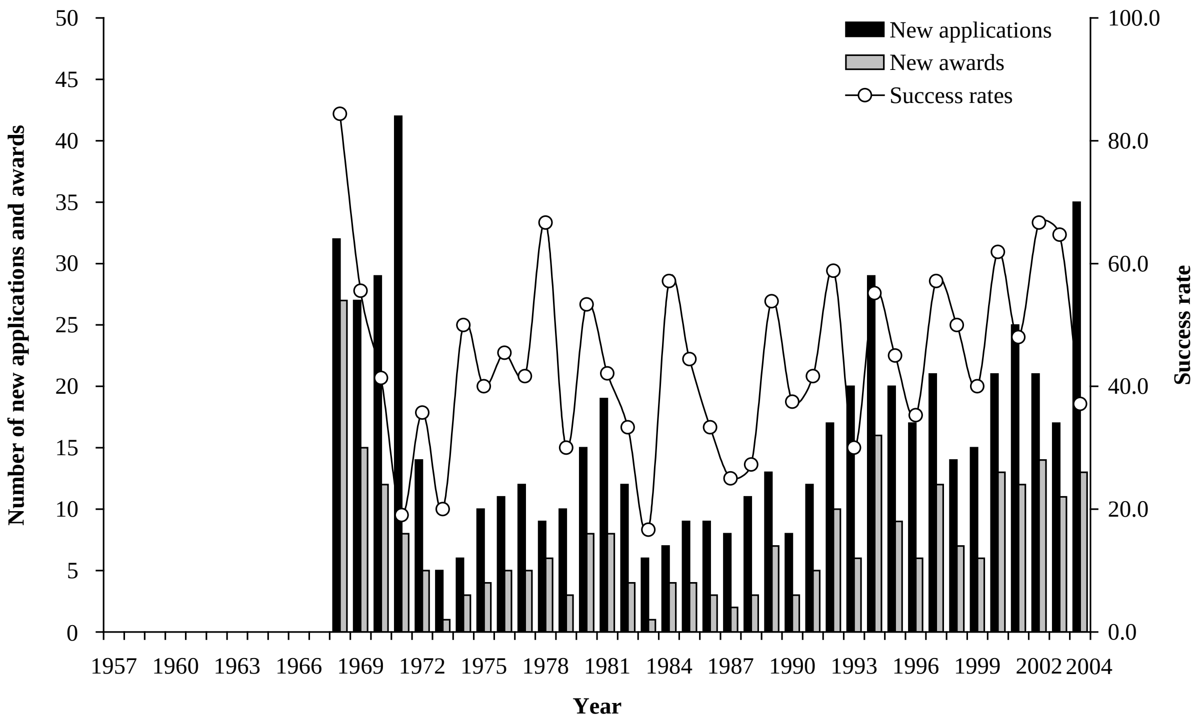
[10] While the K04 salary ceiling doubled in size between 1974 and 1991, the increase fell short of keeping pace with inflation.

[11] Due to the name change, NIDDK K04 expenditures appear in the "other" category after 1985 in Figure 3-11.

K05: Senior Scientist Award

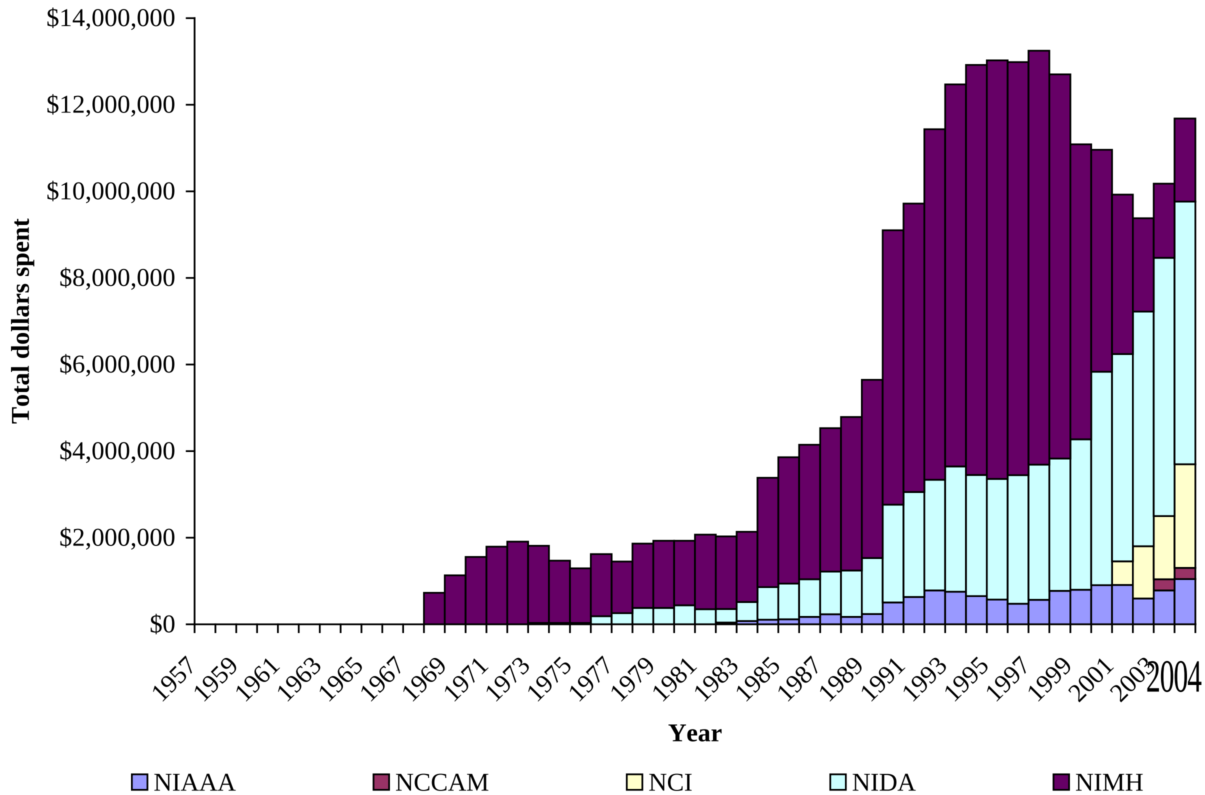
The K05 award supports established research scientists who are well recognized for making significant contributions to their field. The K05 supports senior scientists through salary contributions to the recipient's institution, allowing for protected time free of teaching or administrative responsibilities. Figure 3-12 shows that K05 applications and awards have varied significantly across years, as has the application success rate.

Figure 3-12.—Number of new K05 applications and awards and success rate, by year



Recent K05 awards provide up to $125,900 in salary support. The award also allows institutes to offer discretionary research support in some circumstances, although these funds are not adequate to support a large research program, requiring the recipient to also hold independent research support (such as an R01 grant). Recipients are expected to devote substantial time to research, training, and mentoring activities during the course of the award, which lasts up to 5 years. The basic aims of the award have remained stable since at least the mid-1980s, although limited program information prior to this time prevents a full historical account of the program, which began in 1968. Detailed information on historical salary limits is similarly not available. In 2004, the K05 was offered by five ICs, with the National Institute on Drug Abuse accounting for the majority of award spending (Figure 3-13).

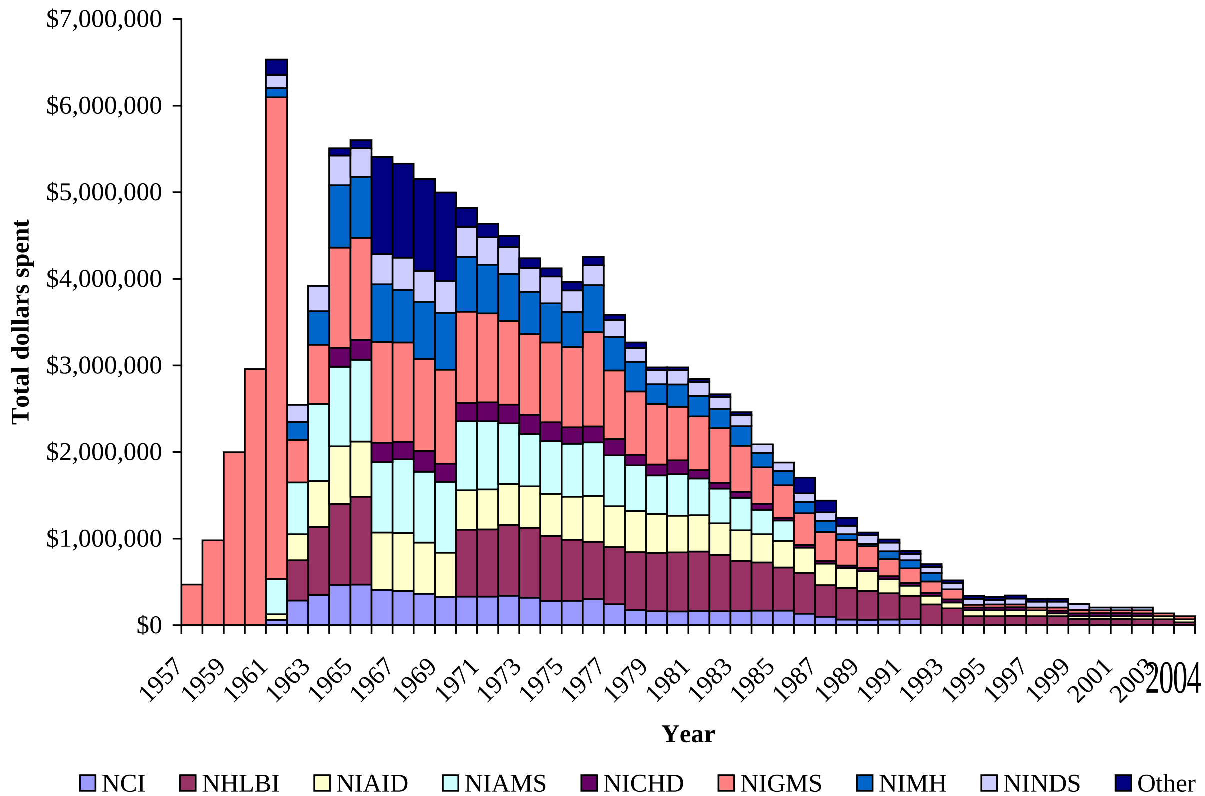
Figure 3-13.—Total dollars spent on K05 award by IC, by year



K06: Research Career Award

The Research Career Award (K06) was offered by the NIH from 1957 to 1965 in support of scientists with potential to make significant research contributions.[12] The award was intended for scientists with proven research records who still had a substantial number of working years ahead of them. As implied by the title, K06 awards offered support for the duration of a scientist's career in the form of salary contributions meant to offset teaching and administrative obligations. Although new awards were not made after 1965, some recipients continued to receive support through at least 2004. NIH data show that K06 awards peaked in 1962 at 118, while applications reached a high of 204 two years later (figures not shown). The award offered salary support of $25,000 per year made in renewable 5-year increments. Eligible applicants included scientists with a proven and productive research record and independent research funding who were approximately 45 years or younger. No significant changes were made to the program during the short period of time in which awards were given. As shown in Figure 3-14, spending on K06 awards continued until at least 2004.

Figure 3-14.—Total dollars spent on K06 award by IC, by year

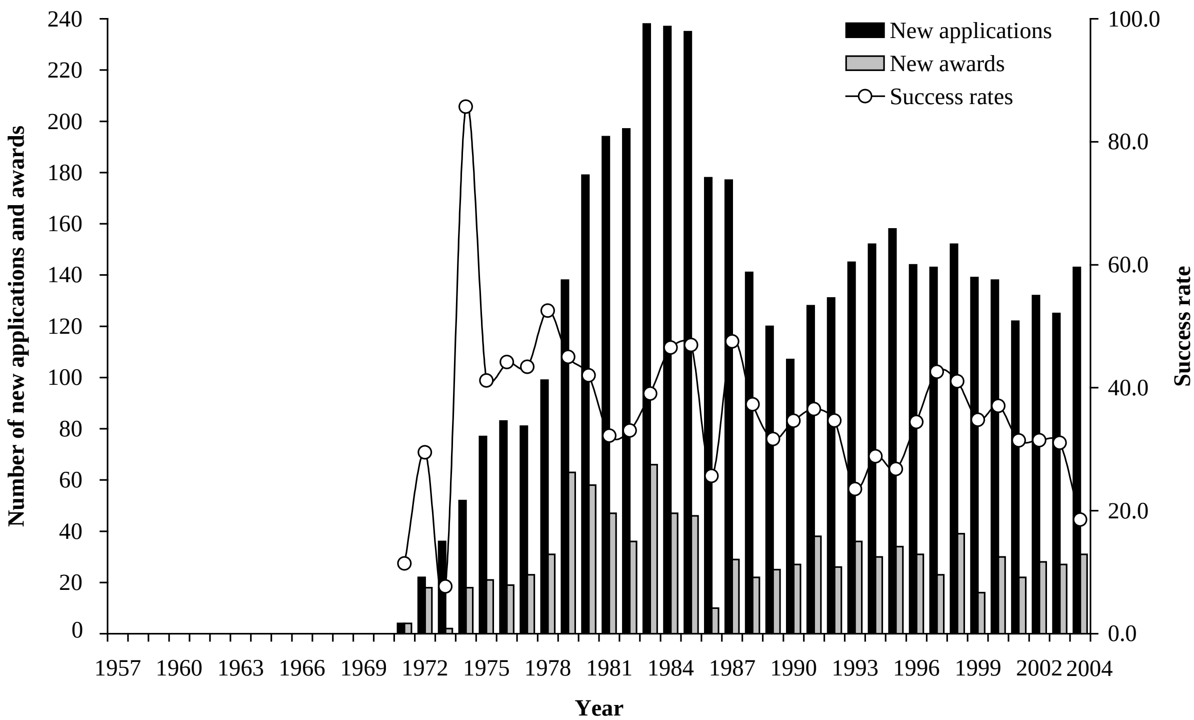


[12] In 1961, the K06 merged several earlier grant programs directed at established researchers, including the Career Research Professorship, Senior Fellowship Grant, and Special Fellowship Grant Program (see Yasumura, 1984).

K07: Academic Career Award

The Academic Career Award is intended to support scientific faculty members who wish to enhance their research expertise and/or build or improve upon science curricula at their home institution. Through two types of grant activities, the K07 award program aims to support newly independent investigators and enhance the educational or research capacity of grantee institutions. Both types offer salary support and contributions toward research, training, and program development costs. Figure 3-15 shows strong application rates for K07s in the early and mid-1980s, although the number of awards made has remained fairly stable since 1987.

Figure 3-15.—Number of new K07 applications and awards and success rate, by year

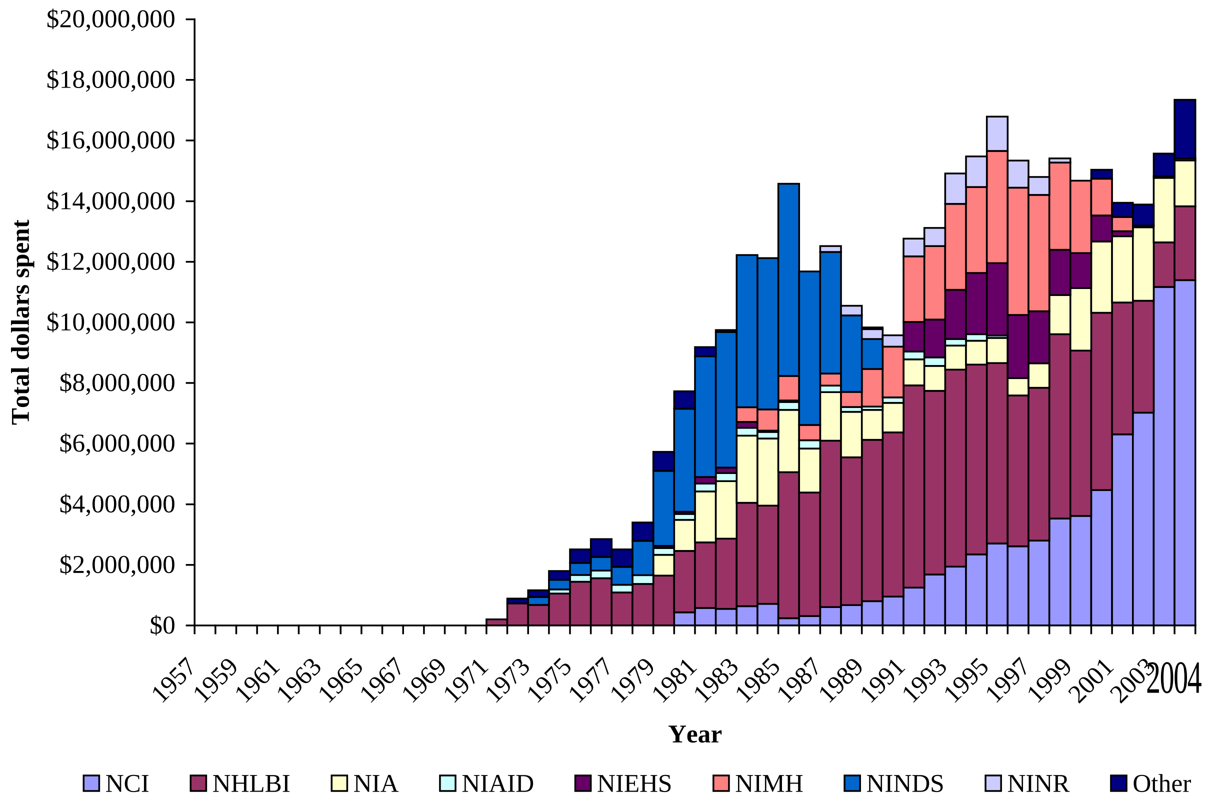


The first type of K07 activity—known as "development"—targets junior faculty who wish to build their own knowledge, skills, and leadership in a particular scientific area as a means of increasing the pool of qualified investigators in that area. Under the guidance of a mentor, recipients of development K07 awards conduct research, teach, and build curricula in the supported area. Development awardees must devote at least 75 percent of their time to grant-supported activities over the course of a 5-year award period.

The second type of K07 award—known as "leadership"—is aimed at more senior faculty who are in a position to have substantial influence on curricular improvement or expansion in a given research area at their institution. The goal of leadership K07 awards is to increase the visibility and overall research capacity for the supported research area within the medical and health research communities. Over a 5-year grant period, recipients of leadership K07 awards are expected to devote 25 percent to 50 percent of their time to grant-supported activities, which may include building new curricula, enhancing existing instructional programs, and conducting research in the supported area.

Several variants of the K07 have been offered by the NIH and its ICs. For example, the NIH has in recent years offered K07s in support of interdisciplinary research programs and behavioral and social science training programs in medical schools. Both of these awards, however, use only the leadership K07 activity. K07 awards offered by ICs have varied somewhat over time in the types of support offered and targeted recipients. While most of these awards share the similar goal of enhancing research capacity at an institution through support to an individual faculty member, some, such as the 1981 National Institute on Aging Academic Award, appear more directed at enhancing the recipient's own research capabilities.[13] Most K07 awards made by ICs provide the standard 5 years of support, sometimes with the option of competitive renewal. K07 salary contributions have risen steadily since the early 1980s, the first time period for which information on award amounts is available. Salary limits were set at around $30,000 per year in 1981, while some more recent K07 awards have capped salary contributions at $75,000 per year. Research support has also risen from around $5,000 per year in the early 1980s to $30,000 per year for recent K07s.[14] The K07 has been offered by a variety of ICs over time, although in recent years only a handful have made expenditures on the award (Figure 3-16).

Figure 3-16.—Total dollars spent on K07 award by IC, by year



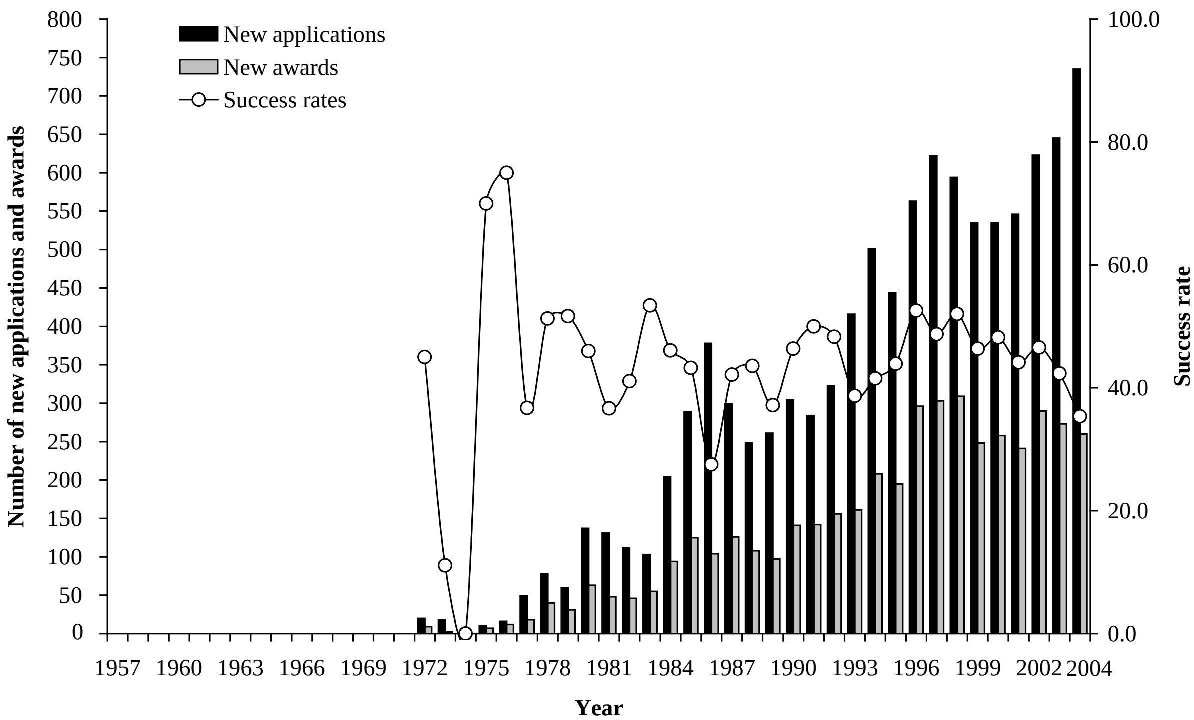
[13]Many earlier K07 grants use the title "Academic Award."

[14]See, for example, the 1981 NIA K07 and the 2006 NCI Cancer Prevention, Control, Behavioral and Population Sciences K07. These increases slightly outpaced the rate of inflation between 1981 and 2006.

K08: Mentored Clinical Scientist Development Award (individual)[15]

The Mentored Clinical Scientist Development Award represents the latest in a long series of NIH funding opportunities that support didactic study and mentored research experiences for individuals with clinical doctoral degrees. The award provides salary contributions to support protected time, allowing recipients to focus on a supervised research career development experience and reduce their teaching and administrative commitments. Although their primary aim is to support mentored career development, K08 awards also provide some support for research and training expenses. Awards last 3 to 5 years, and are available to individuals with clinical degrees (or those with Ph.D.s in a clinical field) who hold full-time employment at an academic institution. Recipients must be able to devote at least 75 percent of their time to research career development activities sponsored by the award. Figure 3-17 shows that K08 applications and awards have increased steadily over time.

Figure 3-17.—Number of new K08 applications and awards and success rate, by year



The K08 dates to the early 1970s and has been widely offered by NIH ICs. Earlier versions of the award were made under the title "Clinical Investigator Award," and although early K08s shared the same basic aims and structure as current awards, some differences in eligibility and levels of support exist. For example, K08 awards offered by the National Cancer Institute and National Heart, Lung, and Blood Institute in the early 1980s were available only to individuals with clinical degrees (e.g., M.D. or D.O. degrees). Salary and research support vary across ICs, and have increased over time. In 1985, K08 salary support was limited to $40,000 per year, and has since increased to at least $85,000 per year.[16] This increase is roughly consistent with the rate of inflation between 1985 and 2006. Recent K08 awards have offered up to $50,000 per year for research and training expenses, up substantially from some earlier

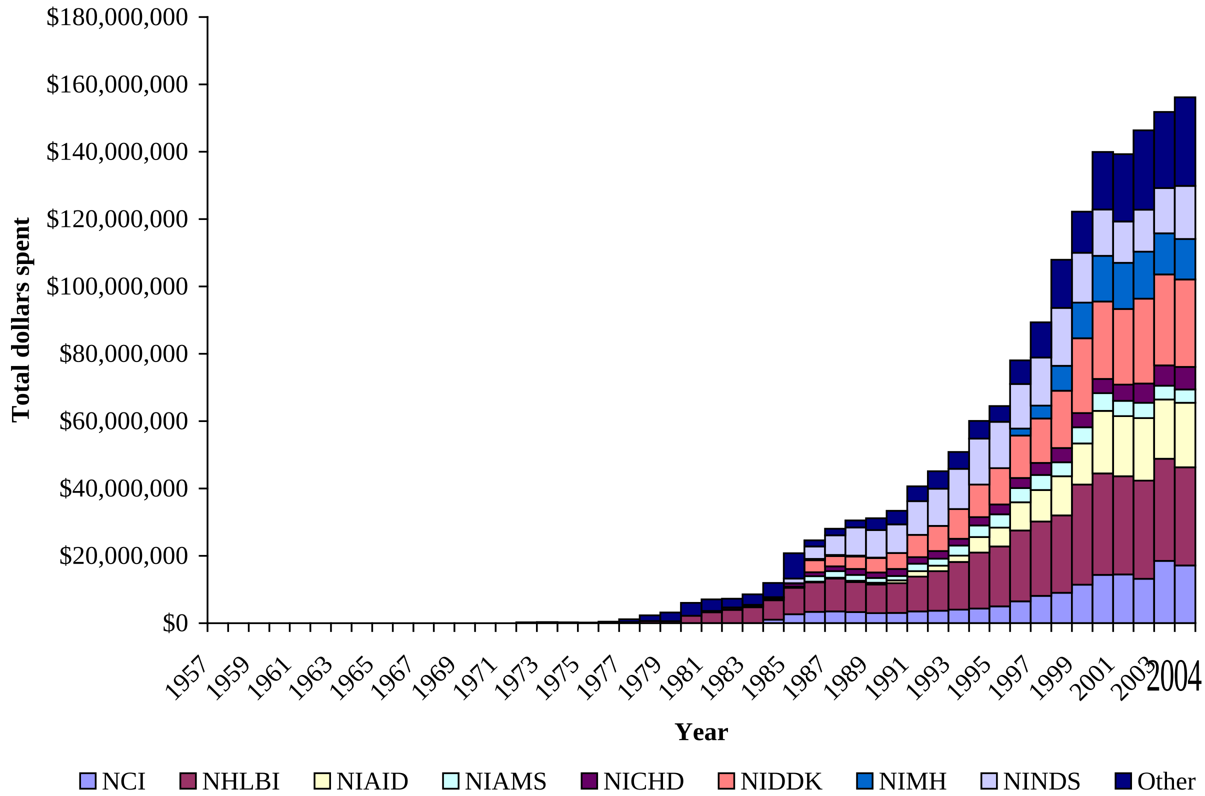
---

[15]In the mid-1990s, the K08 replaced the Clinical Investigator Award (old K08), individual Physician Scientist Award (K11), individual Dentist Scientist Award (K15), and Scientist Development Award for Clinicians (K20).

[16] See, for example, the 2005 NINDS K08 for Translational Research.

grants that limited research contributions to $10,000 per year.[17] As Figure 3-18 shows, spending on the K08 has risen steadily since the mid-1980s, with total spending reaching nearly $160 million in 2004.

Figure 3-18.—Total dollars spent on K08 award by IC, by year

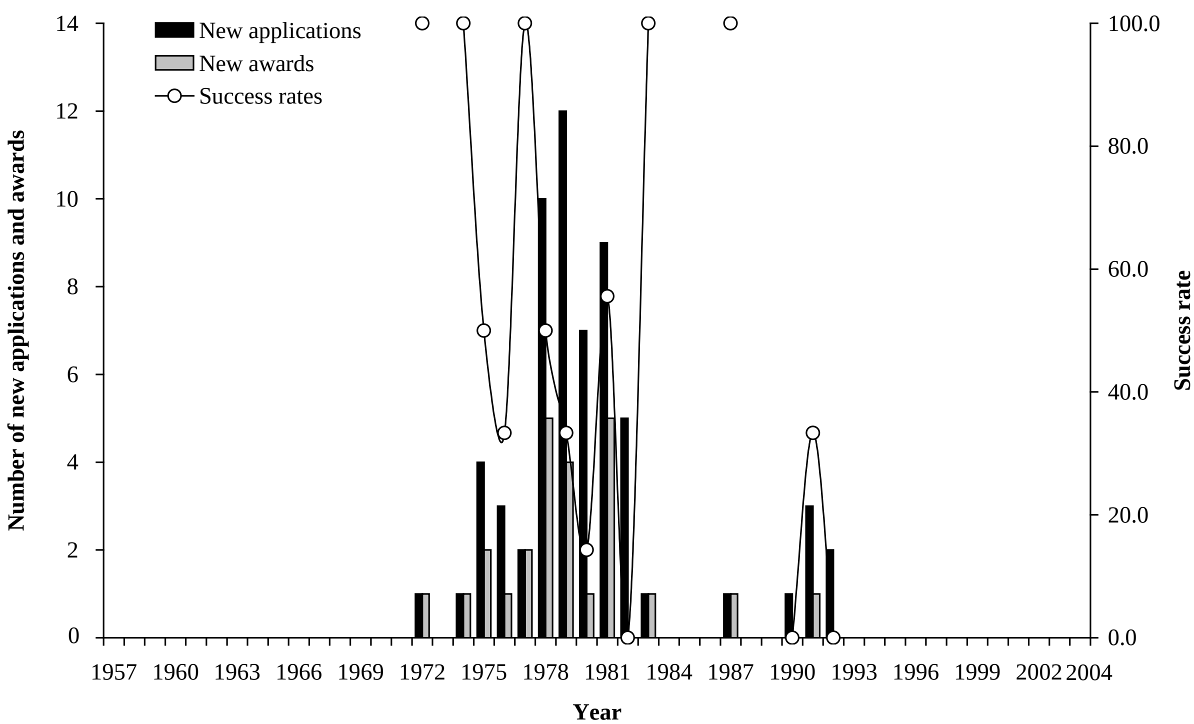


[17] For example, the 1983 NHLBI K08 offered $10,000 per year to cover research and training costs.

K10: Special Scientific Projects

The purpose of the K10 award was to facilitate the utilization and to increase the understanding of record information in fields related to health. Awards were offered by the National Library of Medicine between 1972 and 1992. As shown in Figure 3-19, the K10 attracted a fairly limited pool of applicants and made a small number of awards during its 20-year lifespan. The program enabled qualified individuals to devote a period of full-time effort to the scholarly documentation, evaluation, and analysis of social, cultural, or scientific advances in various fields, disciplines, or specialties of the health sciences. Additional historical information on this award is not available.

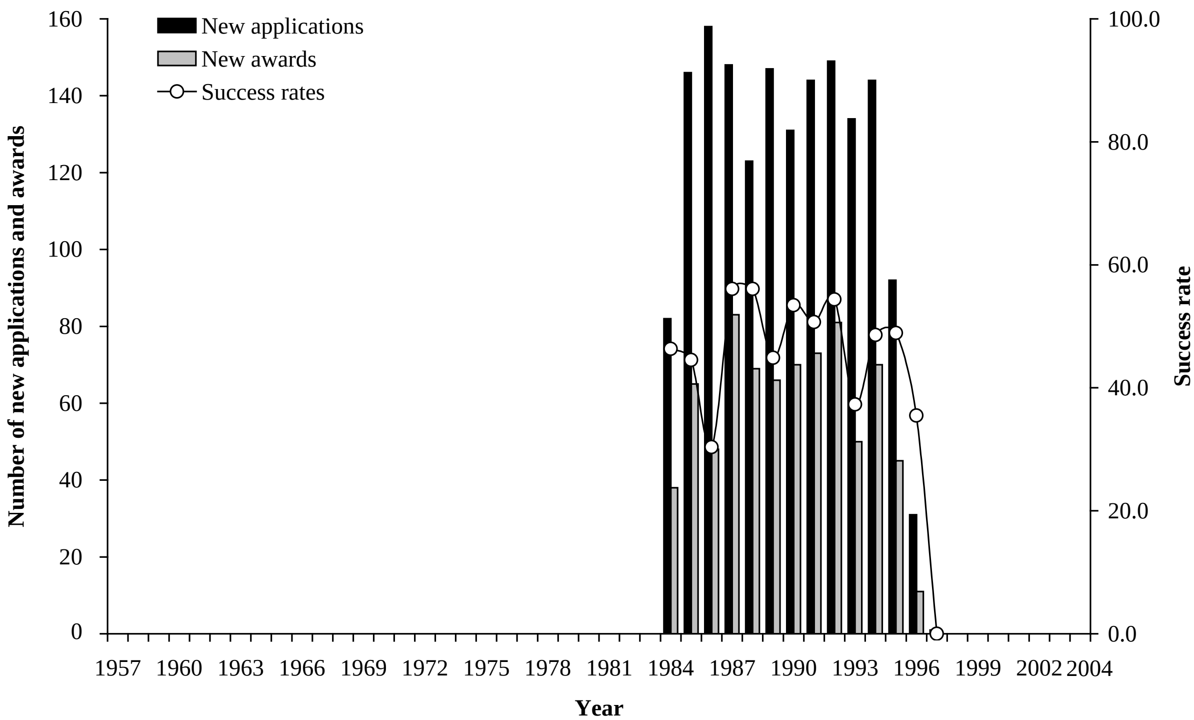
Figure 3-19.—Number of new K10 applications and awards and success rate, by year



K11: Physician Scientist Award (individual)

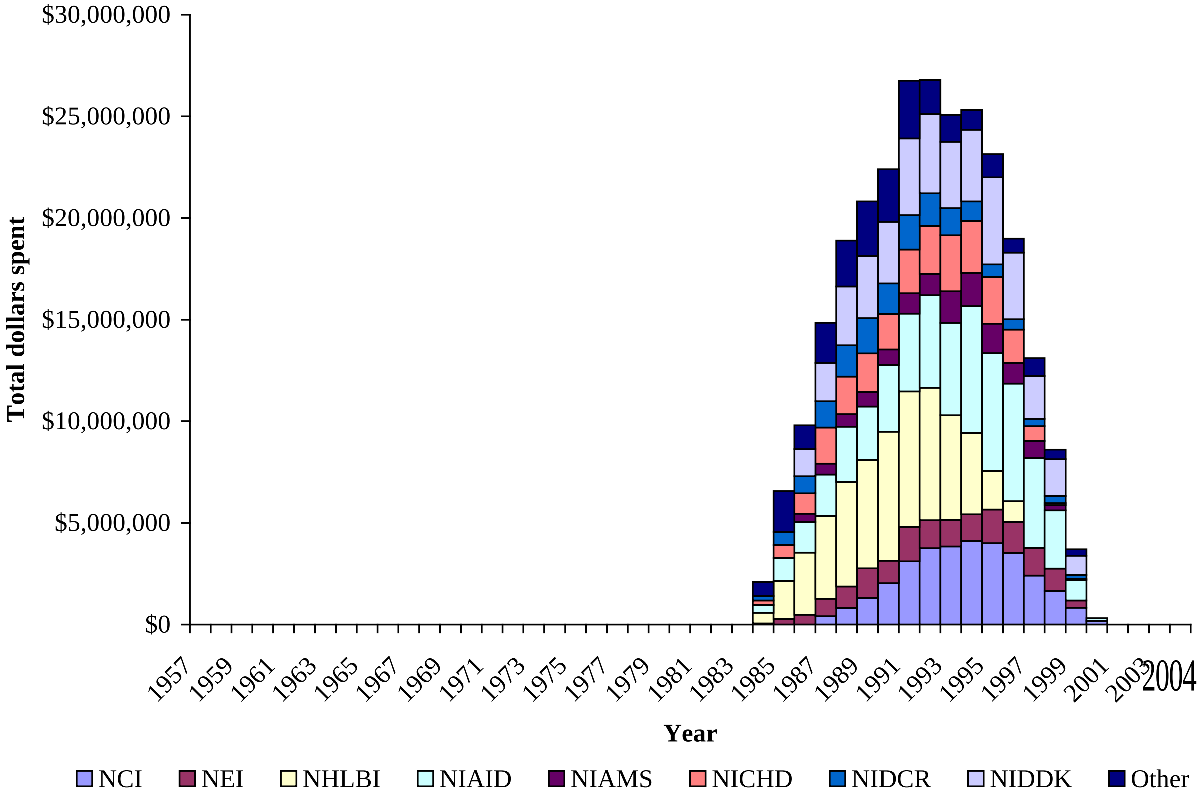
The K11 award, which was replaced in the mid-1990s by the K08 Mentored Clinical Scientist Development Award, offered clinically trained individuals an opportunity to gain research experience in a fundamental science through up to 5 years of intensive training. The award was offered in up to two separate stages, depending on the experience of the recipient. The first stage, targeted at clinicians with little research experience, consisted of 2 to 3 years of basic scientific research training, while the second stage involved recipients in an intensive supervised research experience for 2 to 3 years. Clinicians with stronger research backgrounds could forgo the first stage in favor of concentrating solely on research development. K11 awards provided salary and research support for both award stages. As shown in Figure 3-20, the K11 was offered between 1984 and 1996, with between 40 and 80 awards granted in most years.

Figure 3-20.—Number of new K11 applications and awards, by success rate and year



Between 1983 and 1990 (the only time period for which records exist), K11 awards did not change substantially in their goals or structure. Salary support, however, increased modestly from an annual maximum of $30,000 in 1983 to $40,000 in 1990. Research support remained unchanged during this time period, with up to $10,000 per year offered for stage one activities, and up to $20,000 per year for stage two. Additional historical information about this grant is unavailable. Figure 3-21 shows that total spending on the K11 topped $25 million in the early 1990s, with notable expenditures made in the 1990s by the National Heart, Lung, and Blood Institute, National Cancer Institute, and National Institute of Allergy and Infectious Diseases.

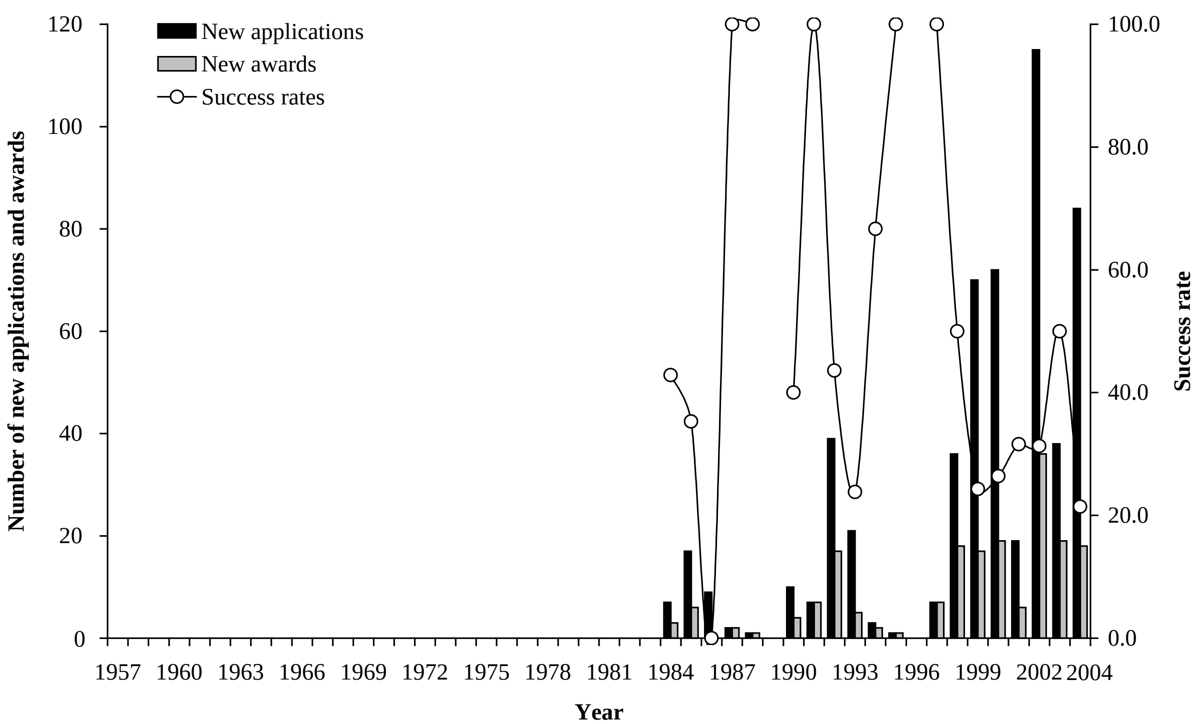
Figure 3-21.—Total dollars spent on K11 award by IC, by year



K12: Mentored Clinical Scientist Development Award (program)[18]

The K12 award is intended to support the development of didactic research training programs for individuals with doctorate-level clinical degrees. K12 awards are made to institutions that have the capacity to implement full-scale programs for 20-35 trainees per year. The overall goal of the K12 program is to increase the number of investigators with clinical research knowledge and skills, although the various NIH ICs that offer K12 awards often support programs tailored to their specific research area. Awards are typically made in 5-year, renewable periods. A variety of U.S. institutions and organizations are eligible, including for-profit, not-for-profit, public, and private organizations. Institutions must have adequate personnel capacity to provide program staff (including a program director and administrative staff), mentors for program participants, and highly trained faculty in clinical and basic scientific research. Figure 3-22 shows that K12 applications have increased substantially since the program's early days, while awards have risen to a more modest degree.

Figure 3-22.—Number of new K12 applications and awards and success rate, by year

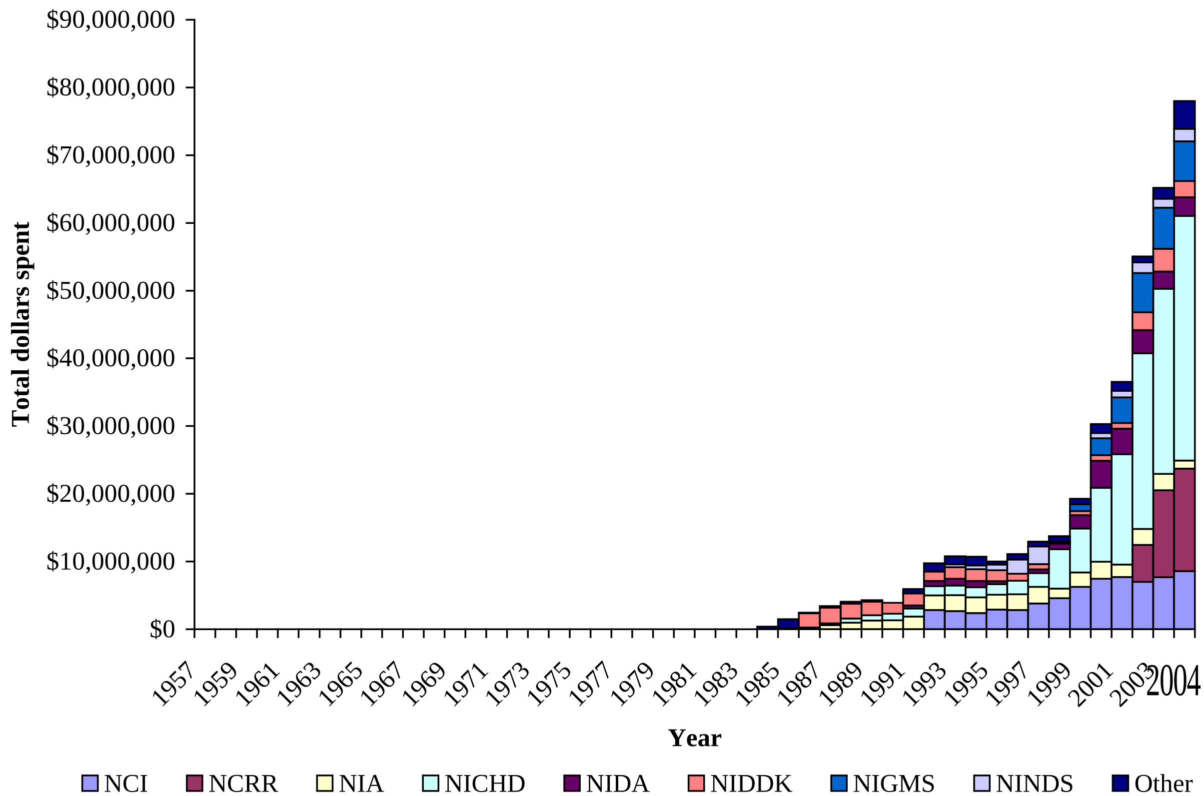


Participation in K12-sponsored programs is typically limited to health professionals with doctoral-level clinical training (e.g., M.D., D.D.S., etc.). Participants spend 3 to 5 years in the training programs, which are often split into two phases. The first phase, lasting 2 or 3 years, is a didactic training experience in a given field of clinical research. The second phase is designed as a practicum, where participants develop their research skills under the guidance of a mentor for an additional 2 to 3 years (although the maximum length of program participation is generally set at 5 years).

[18]In the mid-1990s, two award programs were merged into the K12, including the Physician Scientist program award (old K12), and Dentist Scientist program award (K16).

Prior to the mid-1990s, the K12 award was given as the Physician Scientist program grant. Early versions of this award, dating from the early 1980s, show strong similarity to the current K12. For example, a K12 program announcement from 1983—the earliest year information is available—outlined an opportunity for institutionally based research training programs of 3 to 5 years for clinicians. These grants also supported two-phase programs with both didactic and experiential components, and aimed to increase overall capacity in clinical research. In some ways, however, older K12 awards differ from new versions. For example, modern K12 awards tend to offer support to individuals with a more diverse range of training experiences (e.g., individuals with clinical Ph.D.s[19]), perhaps due to broadening conceptions of "clinical" experience. Levels of financial support for program participants have increased as well. In 1983, K12 trainee salary was set at a maximum of $35,000 per year for full-time effort, while current salaries are set at around $75,000 per year.[20] Early K12s offered program participants up to $10,000 per year for research support during their training period (phase one), and $20,000 per year for phase two. More recent K12 awards also offer research support to trainees, but the limit is typically set at $20,000 to $30,000 per year across each phase. A number of ICs have offered K12 awards, but, as shown in Figure 3-23, the majority of spending in recent years has come from a handful of ICs, including the National Institute of Child Health and Human Development and National Center for Research Resources.

Figure 3-23.—Total dollars spent on K12 award by IC, by year



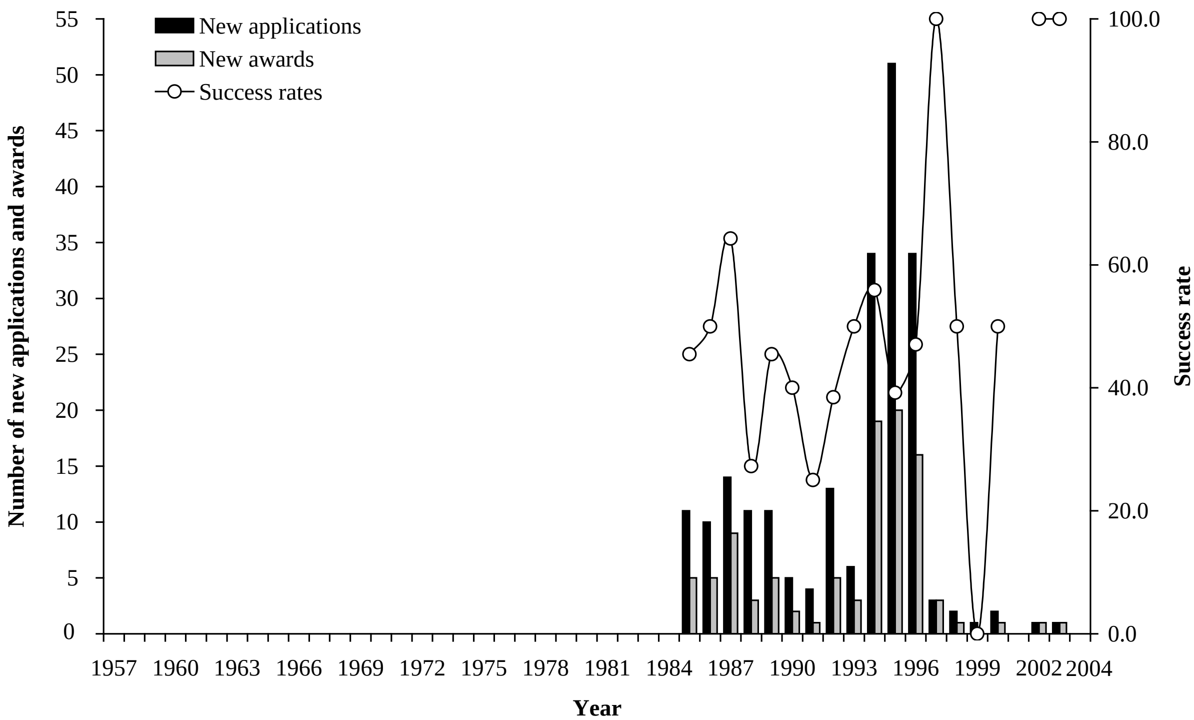
[19] See, for example, the 2005 Multidisciplinary Clinical Research K12 Request for Applications (RFA-RM-05-016).

[20] This increase outpaced the rate of inflation between 1983 and 2006.

K14: Minority School Faculty Development Award

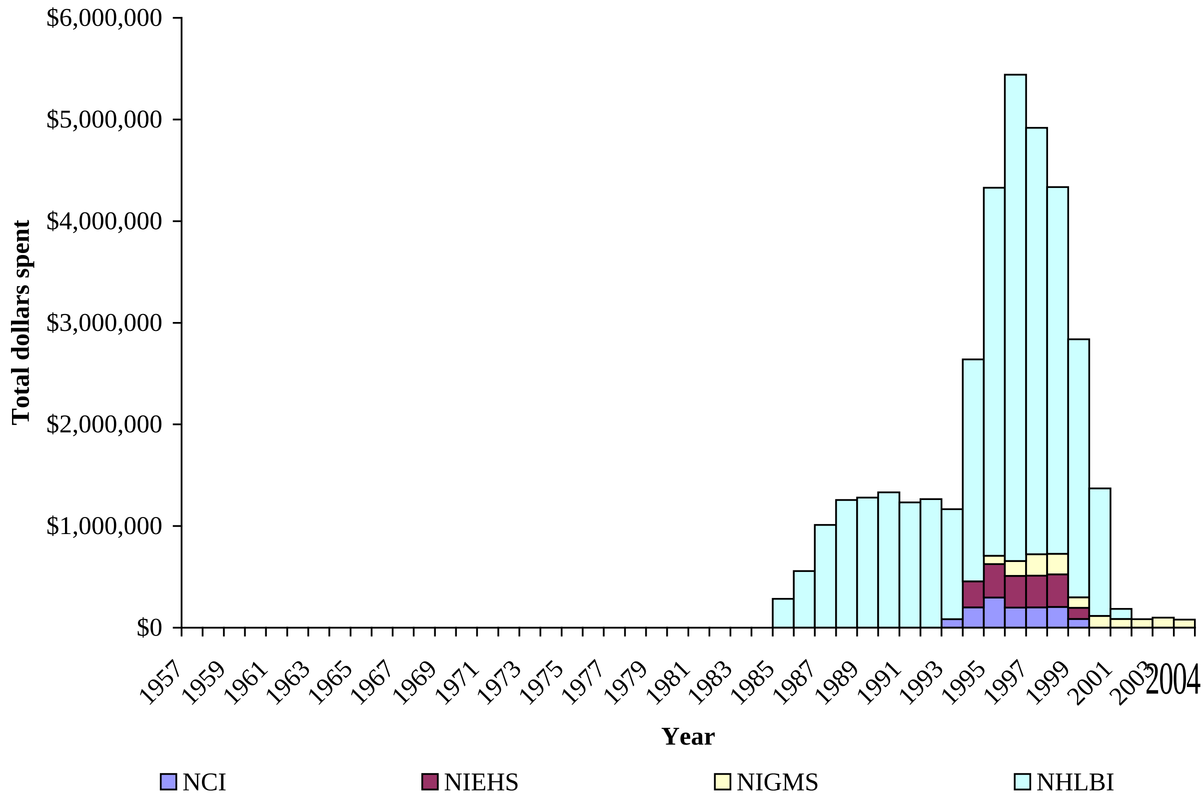
The K14 grant, replaced in the mid-1990s by the Mentored Research Scientist Development Award (K01), supported the development of faculty investigators at minority institutions. The award aimed to enhance the research capabilities of underrepresented scholars in specified areas of health research through up to 5 years of salary and research support. Recipients were expected to devote at least 25 percent of their effort to research development experiences during the academic year, and 100 percent of their time during summer periods, when recipients undertook intensive research training and development under the supervision of a mentor. As shown in Figure 3-24, the K14 was first offered in 1985, and only a few grants were made beyond 1996.

Figure 3-24.—Number of new K14 applications and awards and success rate, by year



Eligibility for this award was limited to faculty at predominantly minority institutions with doctoral-level training in a biomedical or behavioral science. As of the early 1990s, salary support for K14 recipients was limited to $50,000 per year, and research support was provided up to $20,000 per year. Additional historical information about this grant is not available. K14 expenditures exceeded $5 million in 1996, largely due to spending from the National Heart, Lung, and Blood Institute (see Figure 3-25).

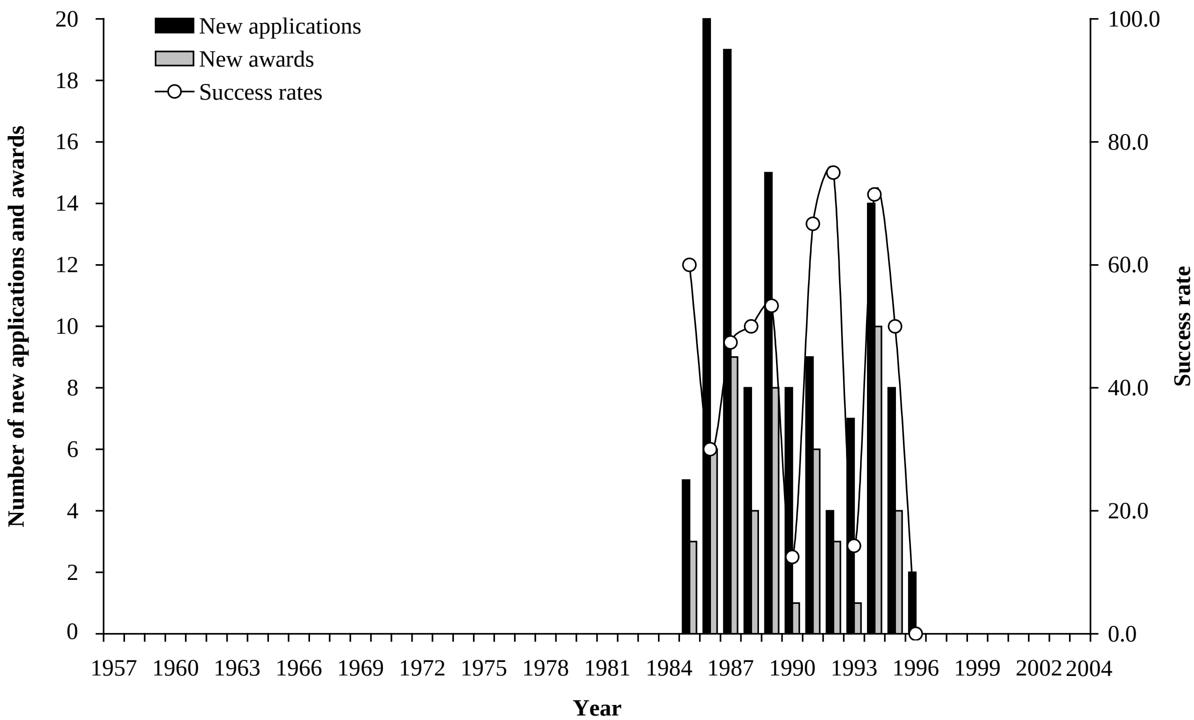
Figure 3-25.—Total dollars spent on K14 award by IC, by year



K15: Dentist Scientist Award (individual)

The individual Dentist Scientist Award, which was replaced in the mid-1990s by the Mentored Clinical Scientist Development Award (K08), was offered to support clinical research training for dentists through the development of basic scientific research skills, advanced clinical research development, and a supervised research experience leading to research independence. As of 1990, K15 recipients were offered salary support of up to $50,000 per year while pursuing grant-supported activities. Additional historical information about this grant is not available. As Figure 3-26 shows, the K15 was offered in modest numbers between 1985 and 1996. All awards were made by the National Institute of Dental and Craniofacial Research (figure not shown).

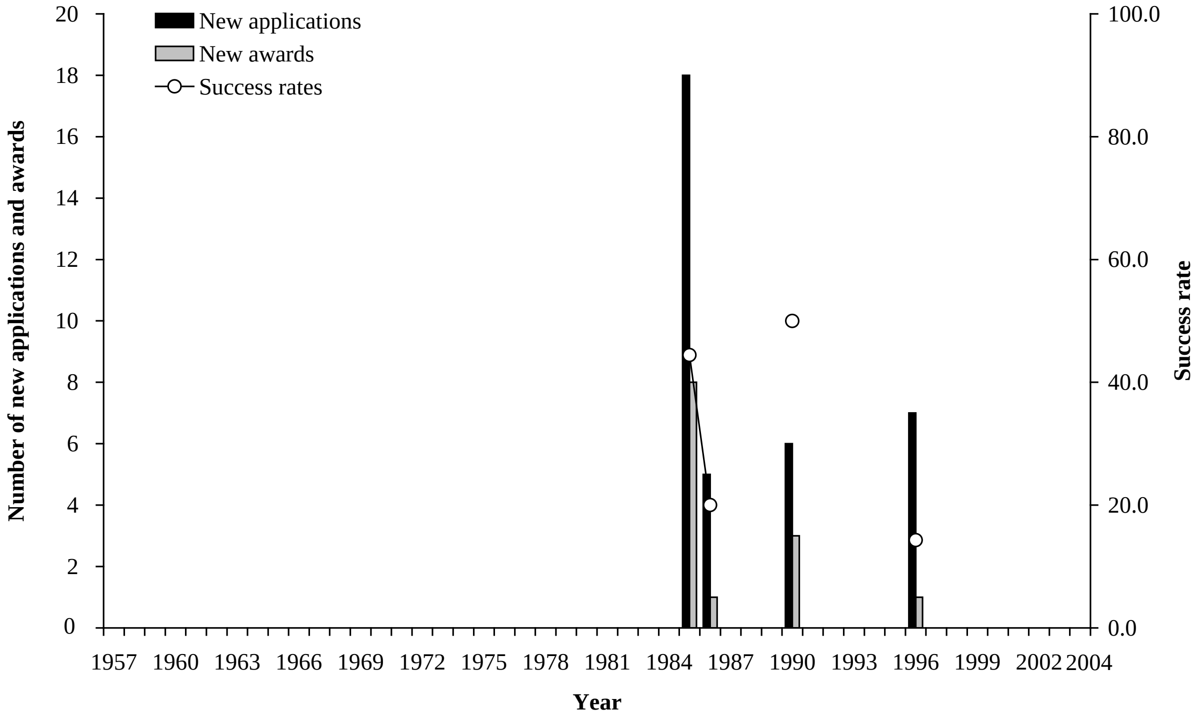
Figure 3-26.—Number of new K15 applications and awards and success rate, by year



K16: Dentist Scientist Award (program)

The program Dentist Scientist Award, replaced in the mid-1990s by the Mentored Clinical Scientist Development Program Award (K12), was intended to support the training and development of outstanding independent dental clinical scholars. The award, made to an institution capable of implementing a strong training program in dental research, offered participants up to 5 years of support for didactic training in basic scientific and clinical research methods, as well as supervised research experiences. As of 1990, K16 recipients were offered salary support of up to $50,000 per year while pursuing grant-supported activities. Additional historical information about this grant is not available. However, as shown in Figure 3-27, the award was given in modest numbers between 1985 and 1996. All awards were made by the National Institute of Dental and Craniofacial Research (figure not shown).

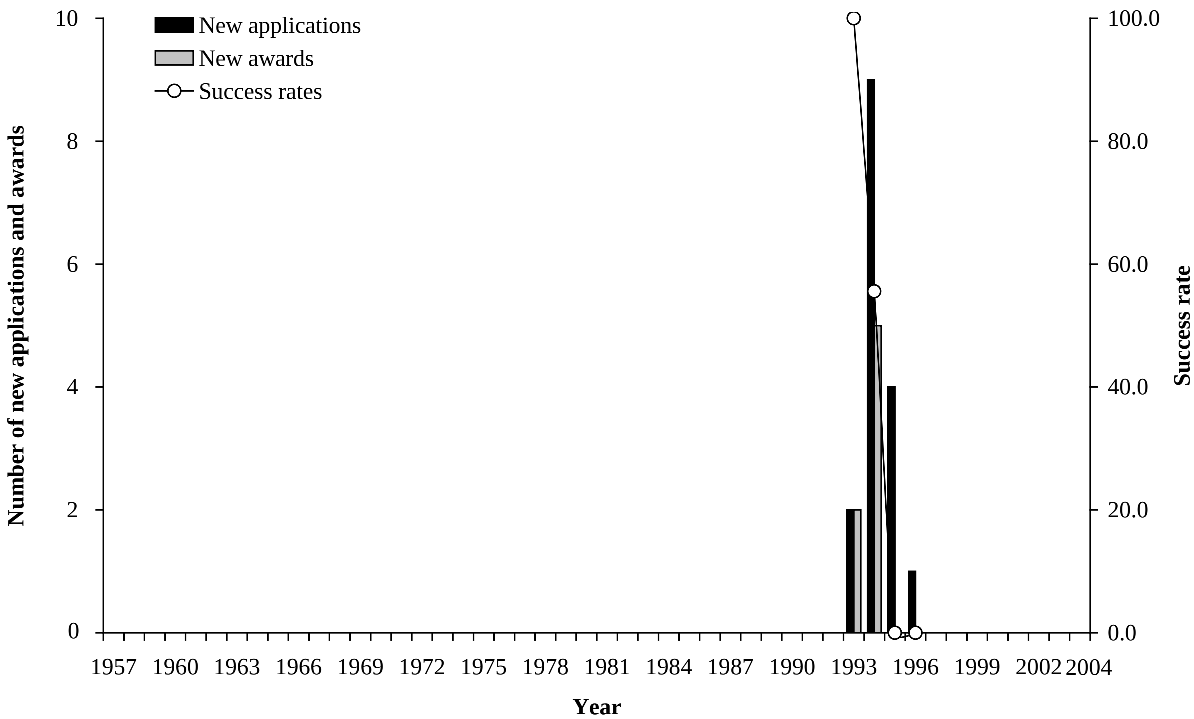
Figure 3-27.—Number of new K16 applications and awards and success rate, by year



K17: Research Career Re-entry Program

Replaced in the mid-1990s by the Mentored Research Scientist Development Award (K01), the K17 was intended to support basic or clinical scientists who planned to reenter their fields as active investigators after an absence. The award supported research development activities that would allow awardees of high potential to update their training and skills. As shown in Figure 3-28, the award was offered for a short period of time in the 1990s, during which time fewer than a dozen grants were awarded. No additional historical information about the K17 is available.

Figure 3-28.—Number of new K17 applications and awards and success rate, by year



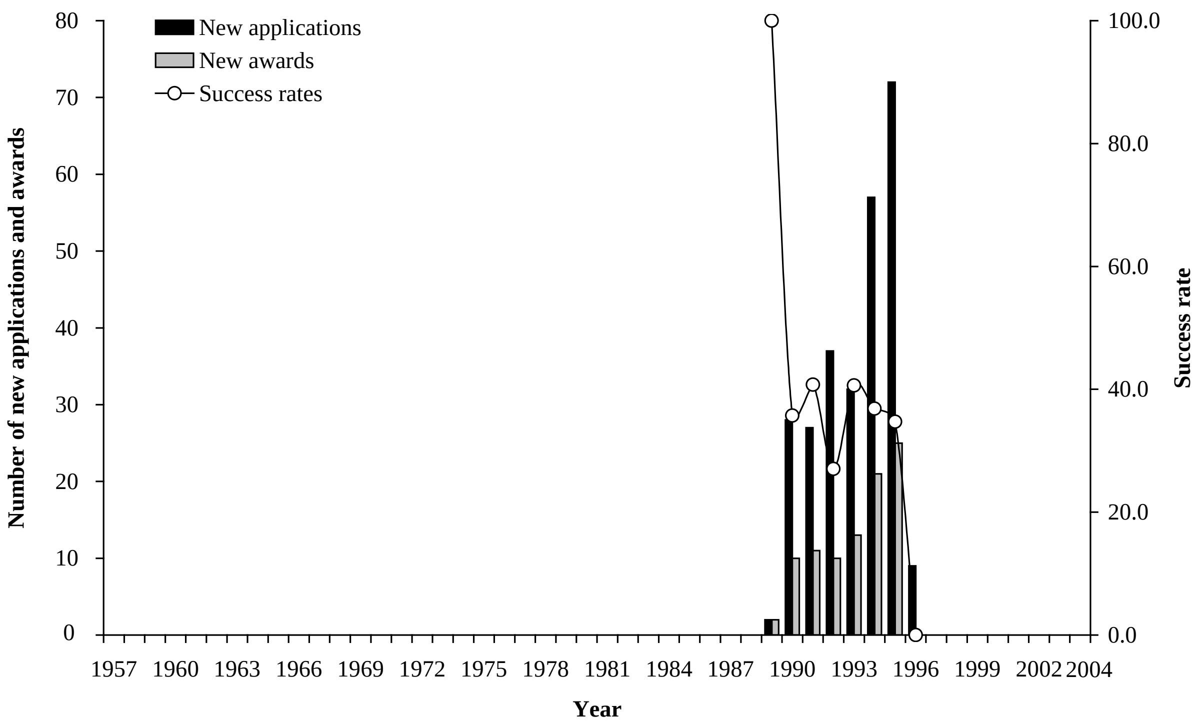
K18: Career Enhancement Award

The K18 award offers support for short-term research training experiences for new and established investigators. K18 recipients are provided with salary support and funds for research and training for a period of 6 months to 1 year. Salary support is currently limited to $180,000 per year, and yearly research support is capped at $50,000. The awards may be used to supplement an investigator's research capabilities or to provide new skills to a scientist pursuing new career directions. Regardless of their purpose, K18 awards require investigators to pursue their training with a qualified mentor who is an expert in the chosen field of study. Offered since 2003, the K18 has been given in support of stem cell research by the NIH, and for language disorder and autism research by the National Institute on Deafness and Other Communication Disorders (NIDCD). NIH application data show that four K18s were awarded in 2003 and 2004, with an application acceptance rate of slightly less than 40 percent each year (figures not shown). The NIDCD K18 is given to established investigators in academic positions for supplementary training, and is designed to be completed in two intensive 3-month training periods. The stem cell research K18 is targeted toward both junior and senior faculty who may wish to gain basic training or supplement their experience in stem cell research. As it is a new program, the K18 has not undergone any significant changes to its structure or goals.

K20: Scientist Development Award for Clinicians

The Scientist Development Award, which was replaced in the mid-1990s by the Mentored Clinical Scientist Development Award (K08), was designed to support clinically trained individuals with high potential and desire to become clinical researchers. The award provided salary and research support for up to 5 years, during which time recipients engaged in a supervised research development and training experience. Recipients were expected to devote at least 80 percent of their time to grant-supported activities, including career development, supervised research, and research skill training. Salary support as of 1992 was capped at $75,000 per year, and research support was provided up to $50,000 per year. Total spending on the award peaked at nearly $10 million in 1995, with expenditures made by the National Institute of Mental Health, National Institute on Drug Abuse, and National Institute on Alcohol Abuse and Alcoholism (figure not shown). Candidates for the award were clinicians with 2 to 3 years of postdoctoral research training or experience. Additional historical information about this grant is not available. As Figure 3-29 shows, applications and awards rose steadily following the program's introduction in 1989. No awards, however, were given after 1995.

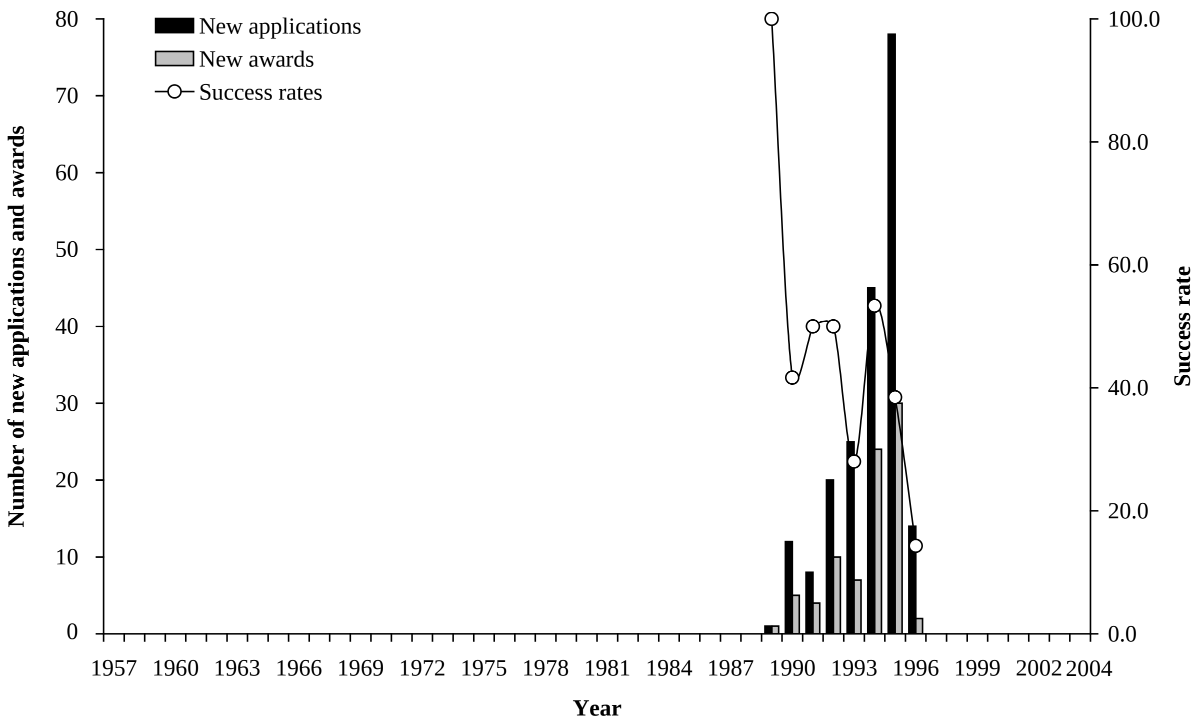
Figure 3-29.—Number of new K20 applications and awards and success rate year



K21: Scientist Development Award

The K21 award, which was replaced in the mid-1990s by the Mentored Research Scientist Development Award (K01), was intended to foster the development of outstanding biological or behavioral scientists pursuing alcoholism, drug abuse, or mental health research. The primary development mechanism offered by the K21 was a supervised research experience for awardees. Additional historical information about the award is unavailable. As Figure 3-30 shows, the award was offered between 1989 and 1996, with applications and awards peaking in 1995. Spending on the award followed a similar pattern, with expenditures reaching a high of over $7 million in 1995, and decreasing every year thereafter (figure not shown).

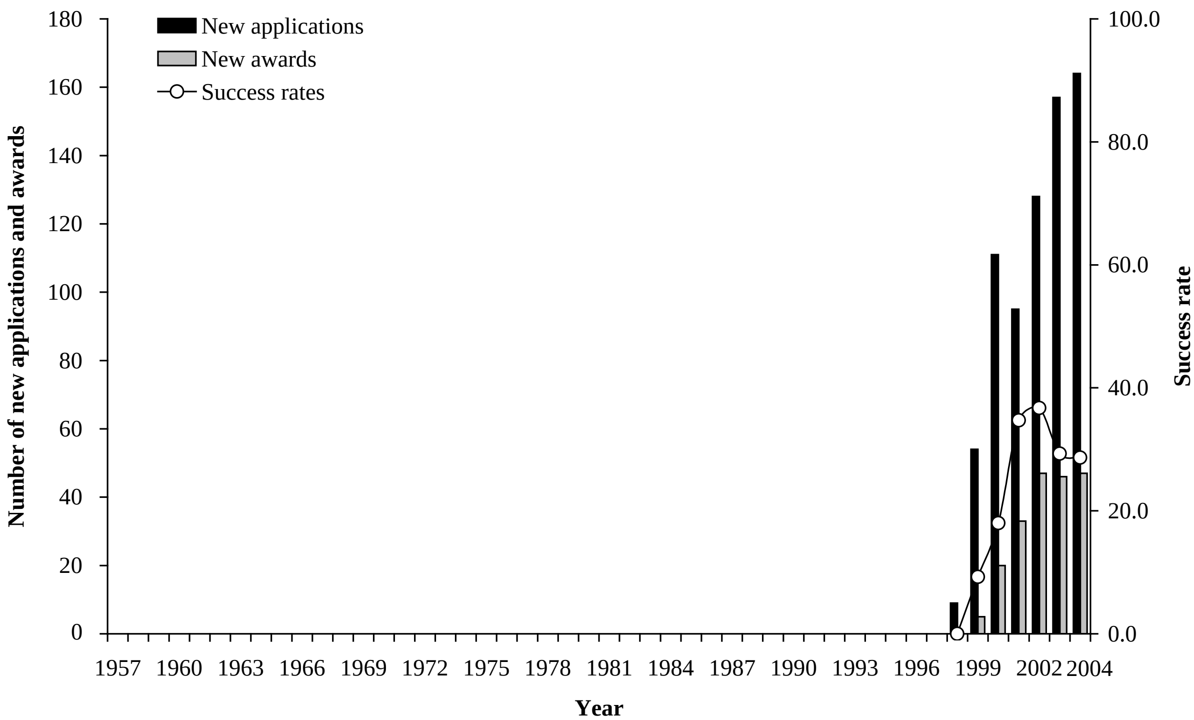
Figure 3-30.—Number of new K21 applications and awards and success rate, by year



K22: Transition Career Development Award

The K22 award is designed to ease the transition of new investigators from postdoctoral trainee to independent scientist. The award accomplishes this goal through two distinct approaches. The first provides K22 recipients with two periods of support—one focused on training at the postdoctoral level (typically within an NIH institute), and the other for research as they make the transition to faculty positions. The second approach focuses primarily on research and salary support for investigators who have already made or anticipate making the transition to a faculty position. As shown in Figure 3-31, K22 applications have risen nearly every year since the program's inception in 1998, while the number of awards offered remained fairly stable between 2002 and 2004.

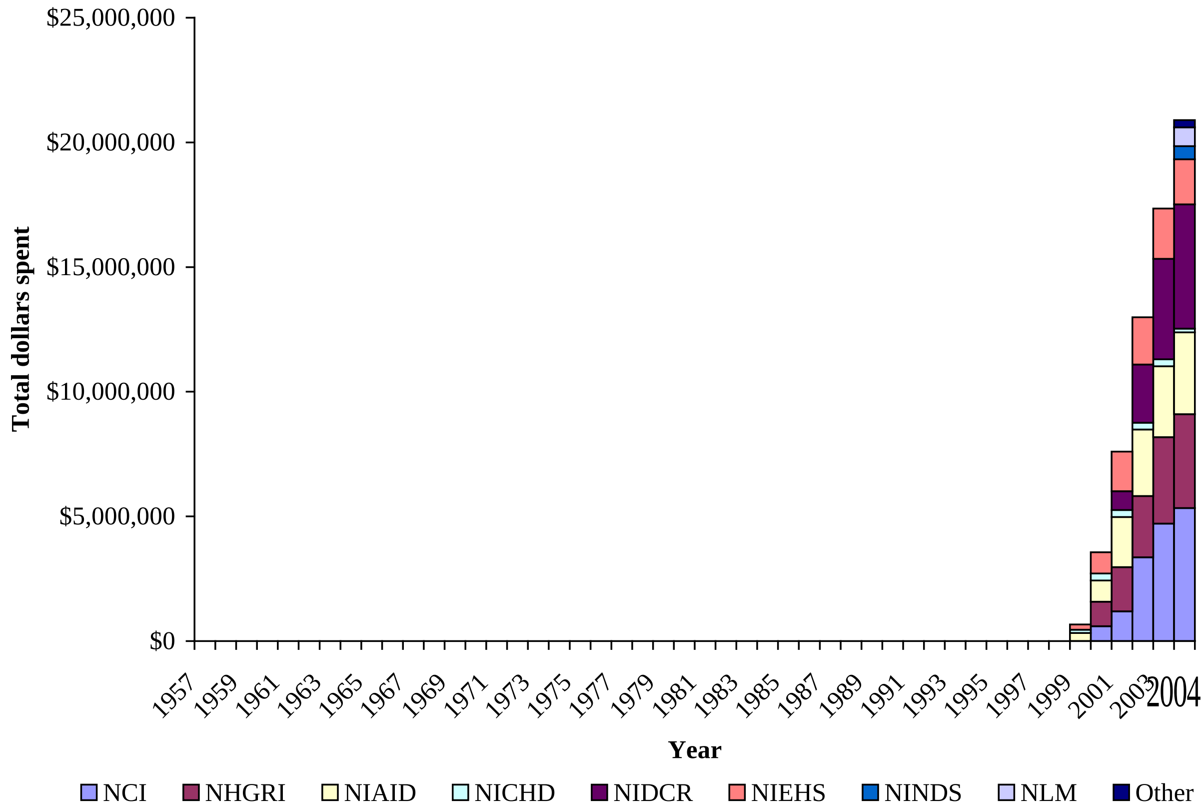
Figure 3-31.—Number of new K22 applications and awards and success rate, by year



A number of NIH institutes have offered the two-stage K22, including the National Institute of Nursing Research, the National Institute of Mental Health, the National Heart, Lung, and Blood Institute, the National Eye Institute, and the National Institute of Dental and Craniofacial Research. The NIH K22 in Women's Health Research also provides recipients with support for both the postdoctoral training and early faculty career stages. These awards typically provide 2 to 3 years of postdoctoral salary support for training within an NIH institute (intramural training), and 2 to 3 years of salary and research support for recipients once they have moved into junior faculty positions. Exceptions are the NIDCR and National Center for Complementary and Alternative Medicine K22 awards, which allow for both intramural and extramural postdoctoral training support. Salary, research support, and eligibility criteria vary across institutes, although two-stage K22 awards are typically directed at investigators with less than 5 or 6 years of postdoctoral training at the time of the award. Since 1999 (the earliest year data are available), salary support limits have ranged from $75,000 to $140,000 per year for postdocs, and from $50,000 to $125,000 for junior faculty. Research allotments for junior faculty range from $75,000 to $100,000 per year over the same time period.

K22 awards that support only new faculty members or postdocs with immediate plans to transition to faculty positions are offered by the National Cancer Institute, the National Institute of Allergy and Infectious Disease, and the National Library of Medicine. These awards typically offer 2 to 3 years of support to new faculty or experienced postdocs to help them solidify their independent research programs. The awards are normally given to postdoctoral researchers with several years of experience who are poised to make the transition to junior faculty positions, or to new faculty members who have held their positions for less than 2 years. Awards include a salary component to offset teaching and administrative responsibilities, which in recent years has ranged from $50,000 to $75,000 per year across institutes. The award also provides funds in support of research and training, running between $50,000 and $75,000 per year in recent grants. As shown in Figure 3-32, K22 spending increased sharply in the years following the program's inception, with significant expenditures made by several ICs.

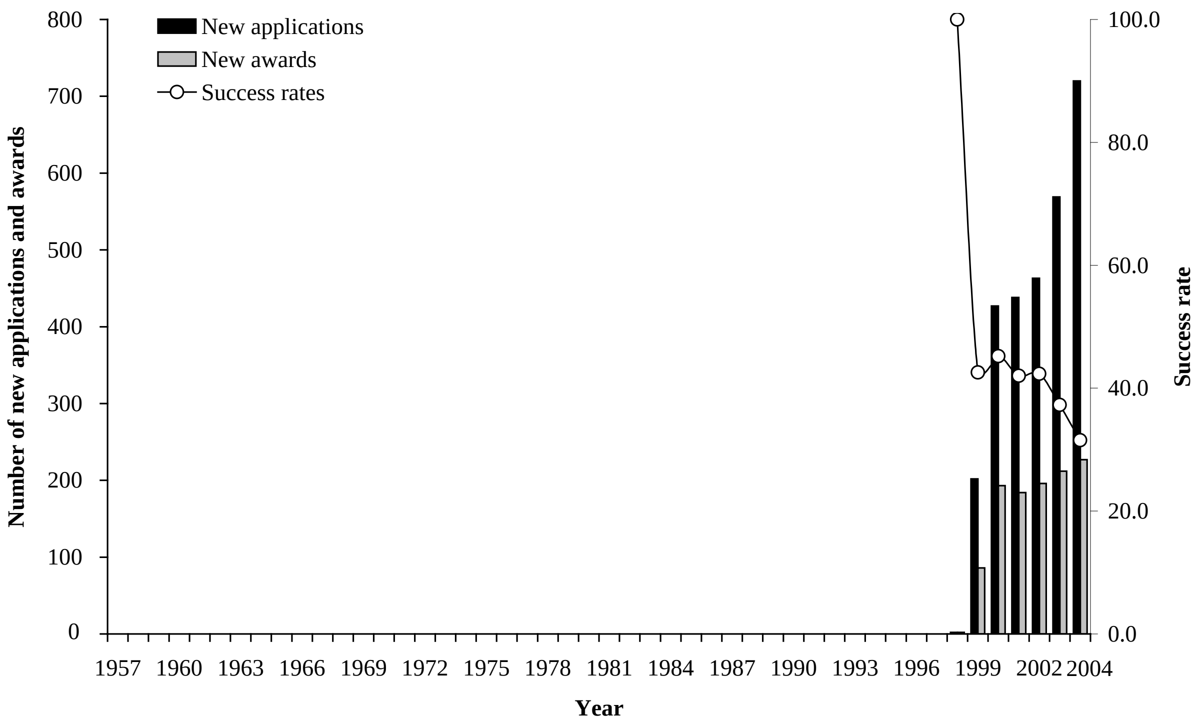
Figure 3-22.—Total dollars spent on K22 award by IC, by year



K23: Mentored Patient-Oriented Research Career Development Award

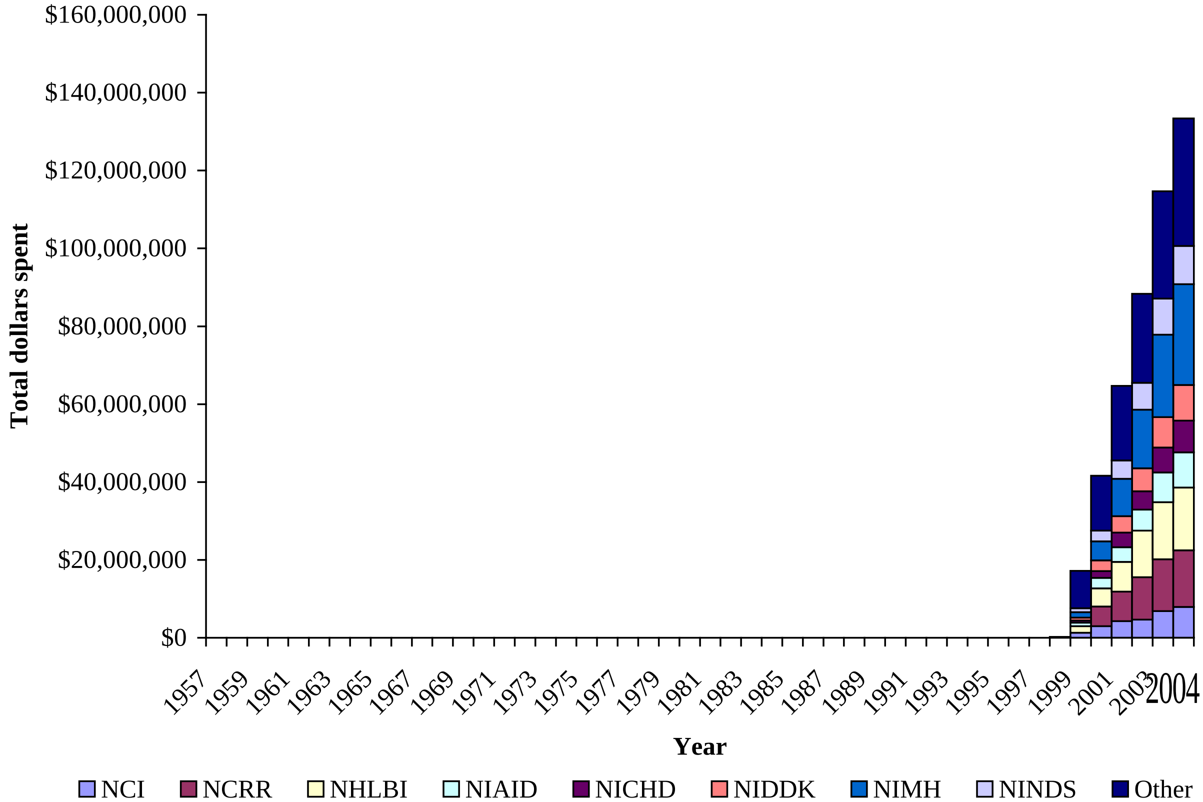
K23 awards support the career development of investigators pursuing patient-oriented research (POR), or research that is conducted with human subjects. The award aims to assist clinical researchers in gaining skills and experience related to POR, increase the pool of highly qualified POR clinical investigators, and support the career development of those who have already made a commitment to POR. To meet these goals, the K23 offers salary and research development support for a period of 3 to 5 years, during which time recipients are to devote at least 75 percent of their effort to mentored career development experiences in POR. The award is targeted toward clinically trained individuals (including those with a Ph.D. in a clinical field) who require an intensive, supervised research development experience in POR to reach independent investigator status. Current K23 salary support levels across NIH ICs range from $75,000 to $180,000 per year, while research development support ranges from $20,000 to $50,000 per year. Figure 3-33 shows that the K23 has been offered since the late 1990s, with over 700 applications resulting in 227 awards in 2004, the most recent year data are available.

Figure 3-33.—Number of new K23 applications and awards and success rate, by year



NIH ICs offer a number of K23 variants. These include K23s focused on genomic or proteomic therapies, muscle disease research, aging research, AIDS research, and services research for homeless populations. The National Cancer Institute also offers a K23 to promote diversity in cancer research (formally known as the K23 for underrepresented minorities). Although disparate in research focus, these awards share the same basic aim of supporting clinical researchers who would benefit from a supervised research development experience in POR. As Figure 3-34 shows, K23 spending is spread across a wide range of ICs, which altogether spent over $130 million on the K23 in 2004.[21]

Figure 3-34.—Total dollars spent on K23 award by IC, by year



Prior to 2000, the NIH General Clinical Research Centers offered the Clinical Associate Physician (CAP) and Minority Clinical Associate Physician (MCAP) awards, which were the precursors to K23 grants. The CAP and MCAP awards shared many of the K23 characteristics, including support of mentored research development for clinically trained individuals. MCAP awards were targeted toward underrepresented minority applicants as a means of increasing diversity in clinical research fields. CAP awards were first issued in 1974, while MCAPs were introduced in 1990. CAP and MCAP awards dating from the early 1990s, while still intended to give clinically trained individuals an opportunity to pursue research, seemed to place less emphasis on patient-oriented research than modern K23s.[22] CAP and MCAP salaries from this time period were limited to $42,500 per year, while research support was offered up to $5,000 per year. When comparing these figures to current K23 support levels discussed above, it appears that both salary and research support increases have significantly outpaced the rate of inflation between 1990 and 2006.

---

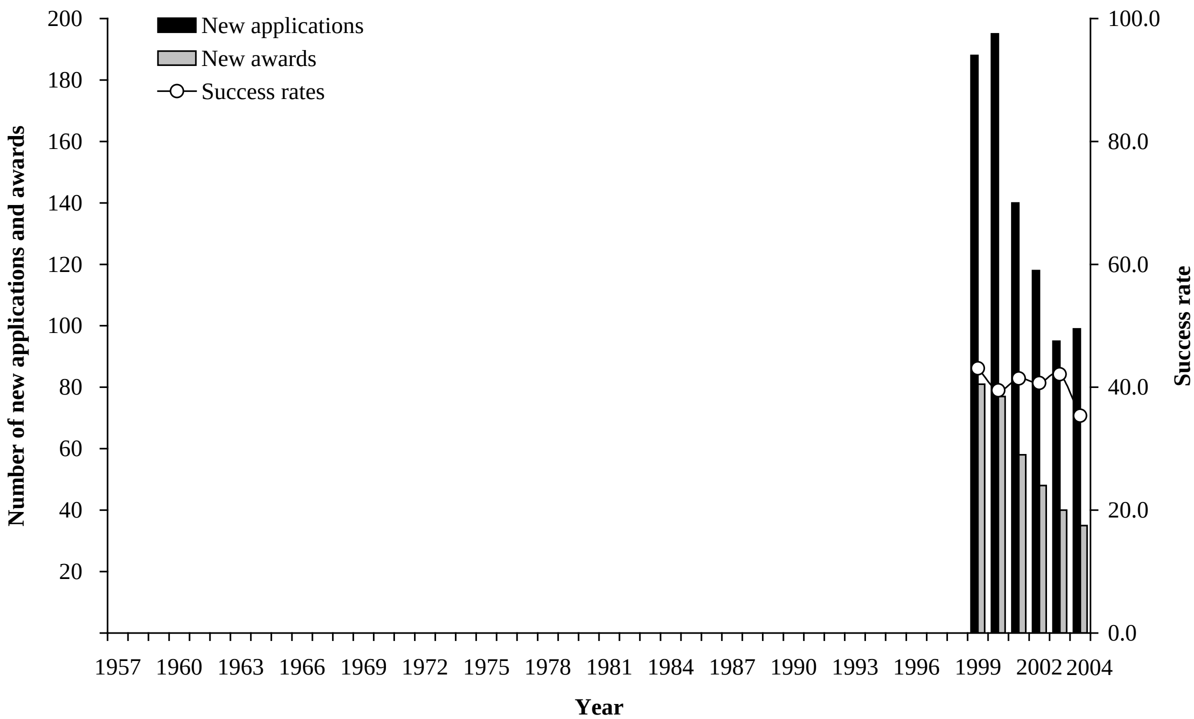
[21]It bears repeating that figures showing total spending disaggregate only the 8 top-spending ICs, resulting in the large “other” category representing spending by 12 additional ICs, for a total of 20.

[22]See, for example, the CAP/MCAP announcements published in the NIH Guide on August 24, 1990. Earlier information about the CAP/MCAP awards is unavailable.

K24: Midcareer Investigator Award in Patient-Oriented Research

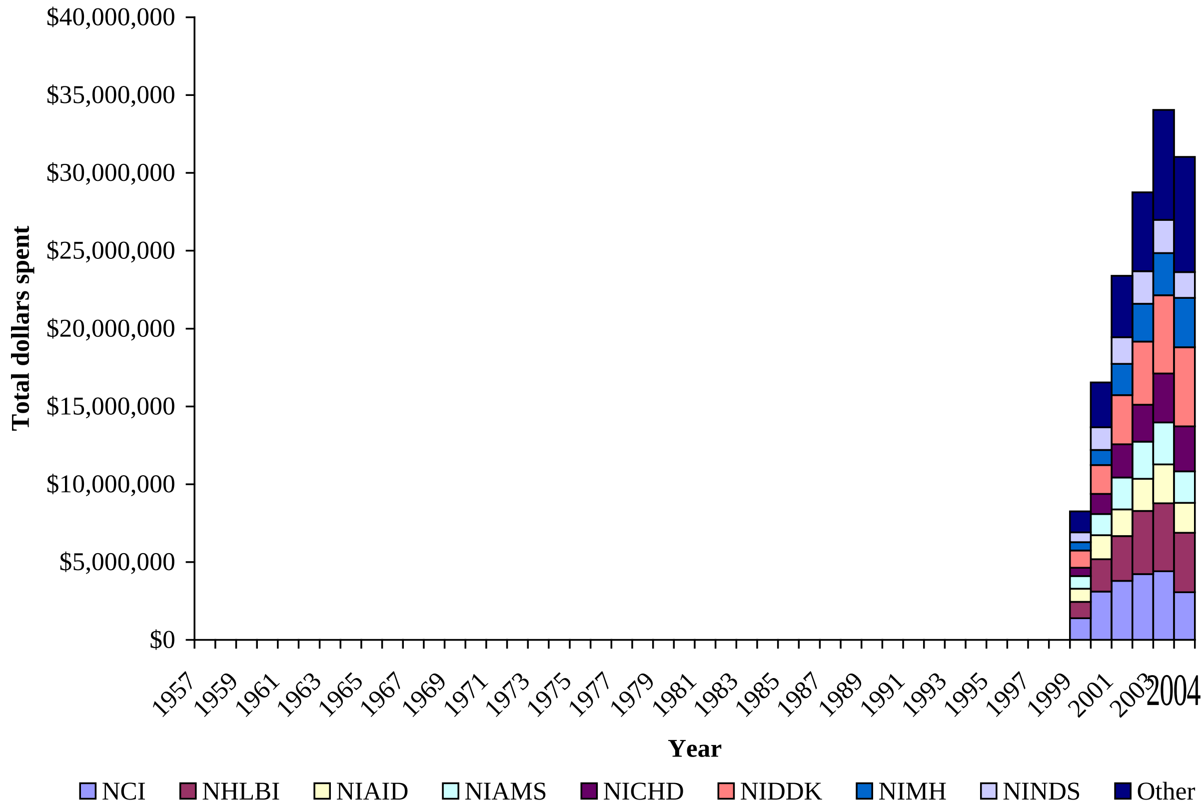
The K24 award supports established clinician-investigators to give them protected time free from administrative and teaching responsibilities. The award is intended to allow these investigators more time to devote to patient-oriented research and mentoring activities, both enhancing the recipient's research program and making them available as a resource for beginning scientists. As shown in Figure 3-35, the K24 has been offered since 1999, and both applications and awards have declined somewhat since the program's inception.

Figure 3-35.—Number of new K24 applications and awards and success rate, by year



K24 awards provide salary support commensurate with 25–50 percent effort by the investigator, and $50,000 per year for research expenses (up from $25,000 in previous competitions). Grants are made for a period of 3 to 5 years, and are available to scientists who have reached the associate professor level or equivalent designation in a nonacademic setting. Recipients must also hold independent research support at the time of the grant (e.g., R01 funding), and be able to devote at least 25 percent of their time to grant-supported research and mentoring activities. Although K24 awards fell throughout the early 2000s, Figure 3-36 shows that total expenditures rose until 2004, when expenditures declined slightly.

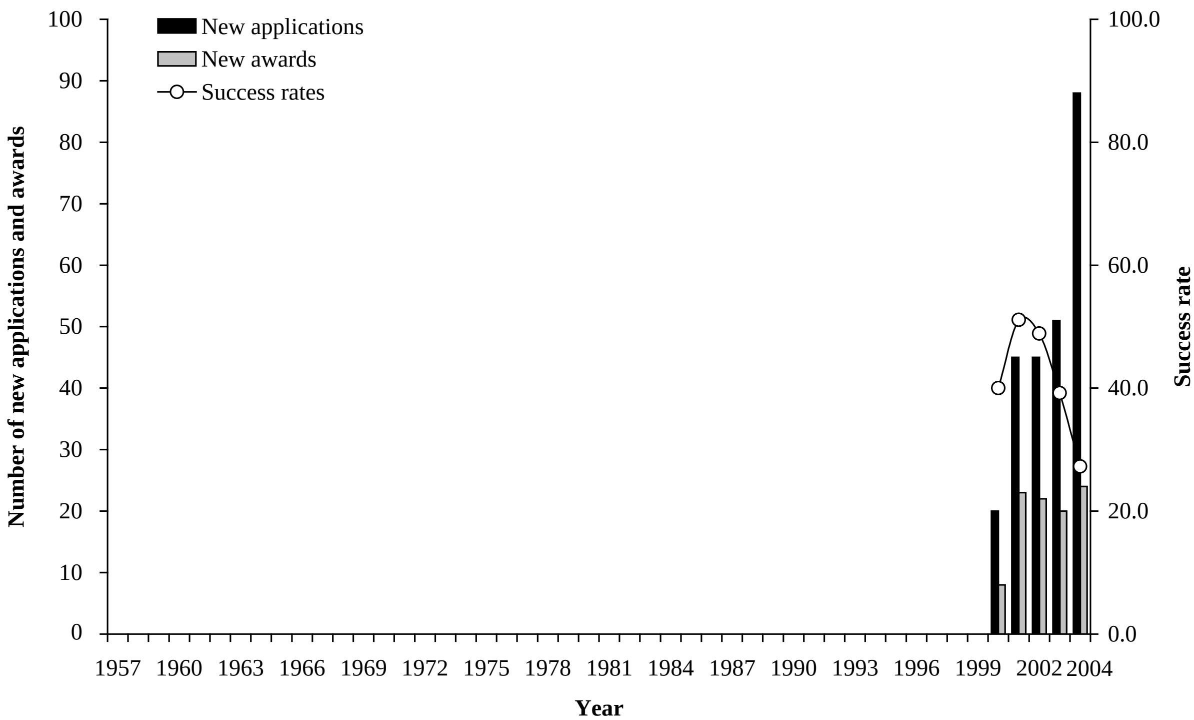
Figure 3-36.—Total dollars spent on K24 award by IC, by year



K25: Mentored Quantitative Research Development Award

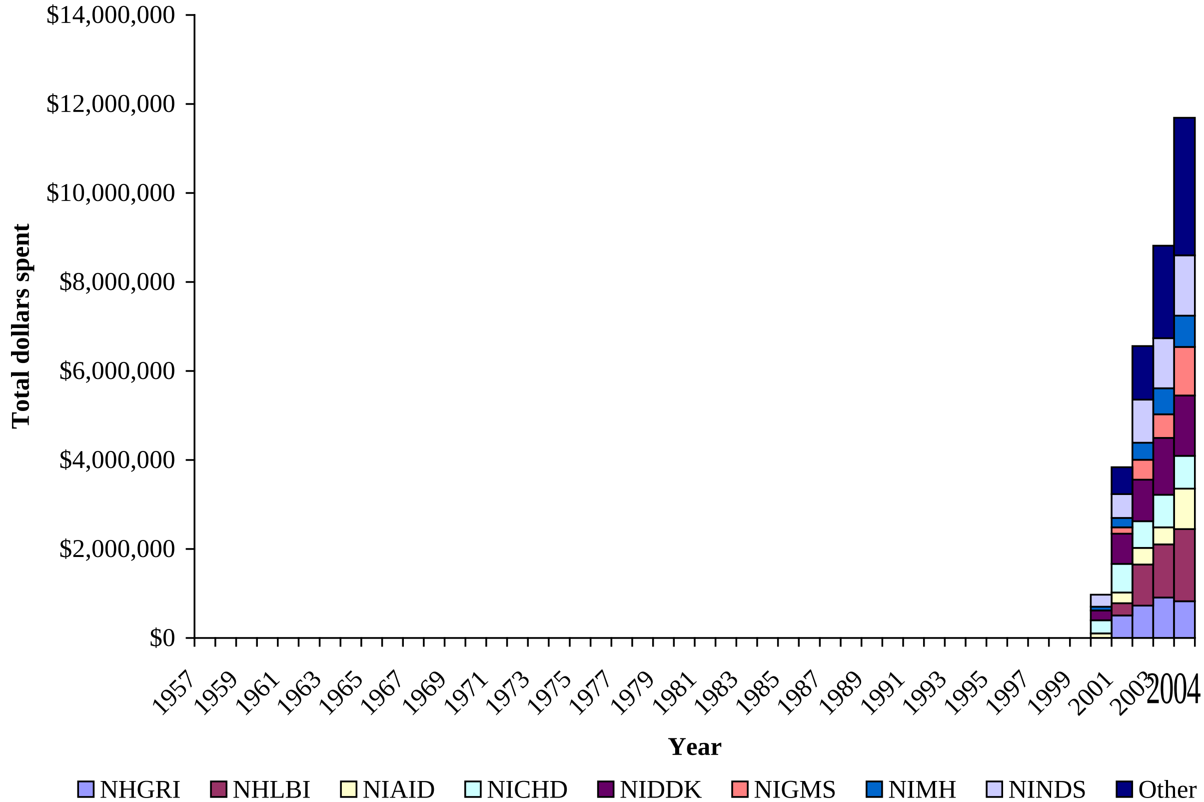
A relatively new award, the K25 seeks to attract investigators with strong quantitative backgrounds to health and disease research. The award is intended to increase the pool of quantitative researchers in fields such as biomedicine and bioengineering, and offers these researchers a unique opportunity to devote significant time to developing new fundamental knowledge and skills. The award provides salary support and funds for research and training experiences over a period of 3 to 5 years. Salary support ranges from $75,000 to $180,000 per year, while $20,000 to $50,000 in yearly research support are available. Investigators are expected to devote at least 75 percent of their time to activities supported by the award. Qualified applicants for the K25 include those with advanced degrees in quantitative science or engineering and a demonstrated record of productive research. As shown in Figure 3-37, about 20 K25 awards have been made per year in recent years, although applications have steadily risen, leading to a declining application success rate.

Figure 3-37.—Number of new K25 applications and awards and success rate, by year



The K25 has been offered by a wide range of ICs, with 15 ICs making expenditures on the award in 2004. As shown in Figure 3-38, expenditures have been distributed fairly evenly across the top-spending ICs in recent years.[23]

Figure 3-38.—Total dollars spent on K25 award by IC, by year

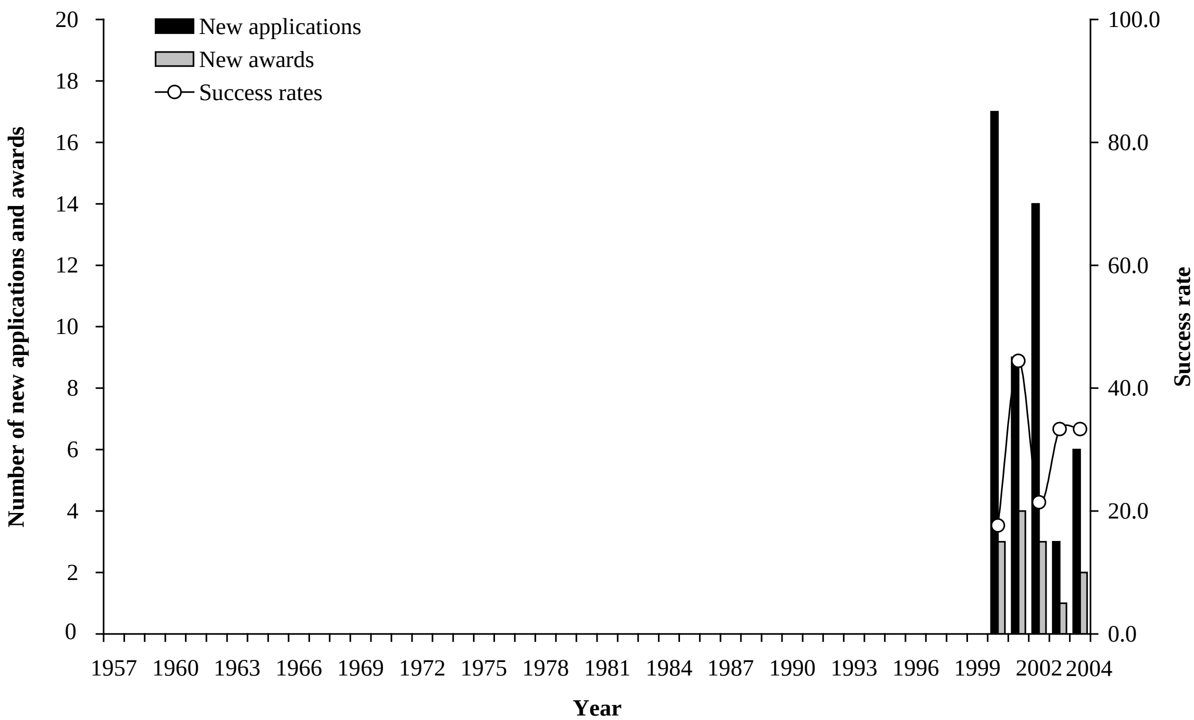


[23]It bears repeating that figures showing total spending disaggregate only the eight top-spending ICs, resulting in the large "other" category representing spending by the remaining seven ICs.

K26: Midcareer Investigator Award in Biomedical and Behavioral Research

The K26 grant, offered since 2000, supports investigators who are within 15 years of their formal training by offering protected time free from teaching and administrative responsibilities. The award provides both salary support (up to $92,000) and some support for research and training experiences (up to $25,000). The goal of the program is to allow established biomedical and behavioral scientists more time to devote to their research programs and mentoring of new investigators. The award is made for a period of 3 to 5 years, during which time recipients are expected to devote at least 50 percent effort to grant-supported activities. The K26 has been implemented by the National Center for Research Resources and the National Institute on Aging to support midcareer mouse pathobiologists in conducting research and mentoring activities, and has not undergone any significant changes to its goals or structure since its inception. Total spending on the award rose steadily following its inception, topping $1.3 million in 2004 (figure not shown). Figure 3-39 shows that K26 applications and awards declined somewhat between 2000 and 2004.

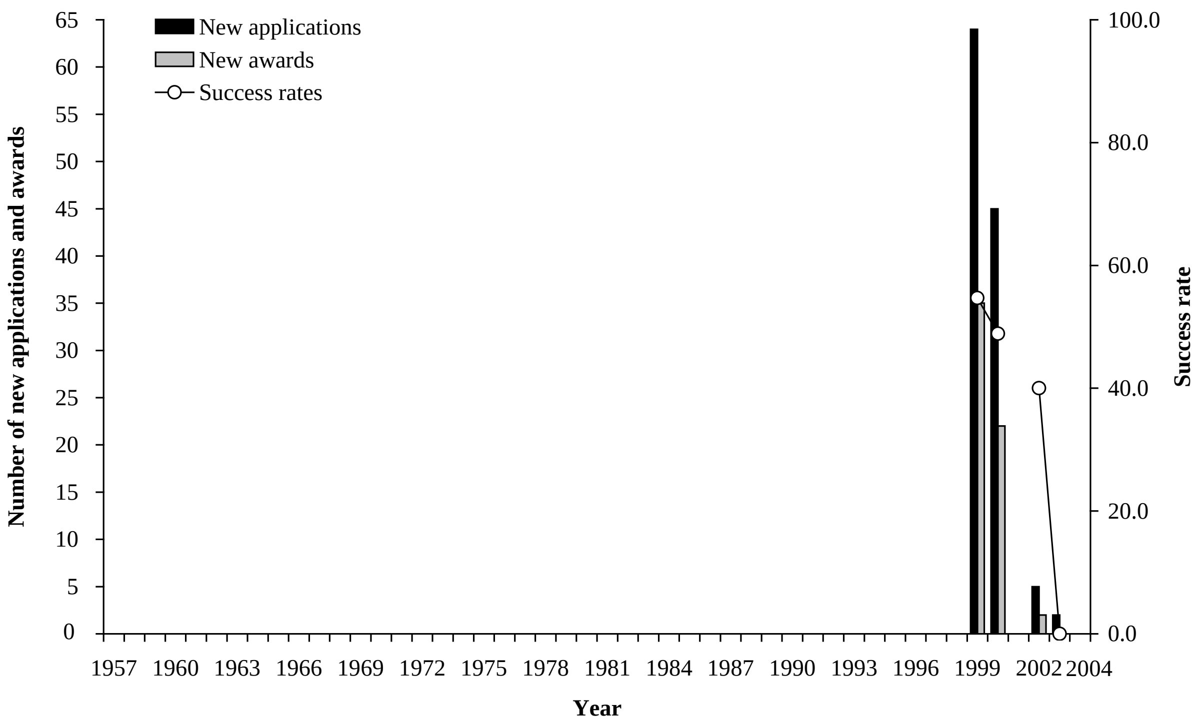
Figure 3-39.—Number of new K26 applications and awards and success rate, by year



K30: Clinical Research Curriculum Award

The K30 award is given to institutions to support the development of high-quality didactic training programs in clinical research. Programs are expected to provide in-depth instruction in the fundamental skills, methodology, and theory necessary for new investigators to reach independent scientist status and compete for prestigious research grants (e.g., R01 awards). The award has been offered since the late 1990s by the National Heart, Lung, and Blood Institute and the National Center for Complementary and Alternative Medicine, which together spent over $12 million on the K30 in 2004 (figure not shown). The K30 has not undergone significant changes to its structure or goals, except for a recent increase in total yearly costs supported—moving from $200,000 to $300,000 in the latest grant announcement. The award is offered for periods of 5 years. Eligible institutions must have strong, well-established clinical research and training programs, as well as a highly trained faculty in clinical research. The institution must also specify a well-qualified program director able to devote at least 20 percent of his or her effort to overseeing the program. Available announcements of K30 awards do not provide detailed information on eligible participants or participant salaries. As shown in Figure 3-40, the K30 was offered somewhat sporadically between 2001 and 2004 (the most recent year application and award data are available).

Figure 3-40.—Number of new K30 applications and awards and success rate, by year



Institutional Awards

NIH Has offered five types of institutional career development awards: K12, K16, K17, Clinical Associate Physician (CAP), and Minority Clinical Associate Physician (MCAP). As noted above, all K16 awards (Dentist Scientist Award) were made by the National Institute of Dental and Craniofacial Research between 1985 and 1996. All K17 awards (Research Career Re-entry Program) were made by the National Institute of Neurological Disorders and Stroke in 1993 and 1994. No information was obtained about individuals supported under the K16 or K17 awards.

The CAP and MCAP, precursors to K23 grants, were offered prior to 2000. All records for the CAP and MCAP are contained in hardcopy. Information obtained from the ICs regarding the K12 awards is shown in Exhibit 3-3, which shows that most of the ICs have maintained a spreadsheet or database, but some of these systems appear to be limited in scope and/or only complete for the past several years. Most of the available data consist of hardcopy records, either applications or progress report. Six of the ICs responding can provide the names of individuals supported, although one IC can provide this information for active awards only. The greatest number of individuals supported per year by the ICs providing a number was 13 (although this represented new awards only for this IC). At most, two of the ICs responding have information on unsuccessful applicants.

Thus, unlike the individual K awards for which data can be easily accessed and analyzed electronically using the CGAF, information on the institutional awards is scattered throughout the ICs, generally in hardcopy format. At this point the number of individuals supported under these grants is uncertain.

Exhibit 3-3.—Information on candidates supported by K12 awards

IC	Database or spreadsheet available?	Number of individuals supported by year?	Names of individuals supported?	Information on unsuccessful applicants?	Duration of appointments?	Hardcopy records?
NCI	?	?	?	?	?	?
NEI	?	?	?	?	?	?
NIA	Yes	Yes (12 per year)	Yes (active awards only)	No	Yes (only years supported thus far)	Yes (progress reports for each scholar)
NIAMS	Yes	Yes (need to review hard copies)	Maybe (need to review hard copies)	No	?	Yes
NICHD	Yes (Access database)	Yes (2 or 3 new candidates per year)	Yes	Yes (some information)	Yes (typically 3 years)	Yes (including applications and review scoring)
NCRR	Yes – Excel file with limited information starting in 2002	Yes (since 2002, but not broken out by year)	Yes	No	Yes (starting in 2002)	Yes
NIDA	No	Yes (need to review hard copies)	Yes (need to review hard copies)	Maybe (need to review hard copies)	Yes (need to review hard copies)	Yes (expired awards are kept at the Federal Records Center for 7 years)
NIDCR	?	?	?	?	?	?
NIDDK	Yes (spreadsheet)	Yes (up to three new appointees per K12 per year since 1997)	Yes	No	Yes (typically 2 years)	No response
NIEHS	?	?	?	?	?	?
NIGMS	No	Yes (need to review hard copies)	Yes (need to review hard copies)	?	?	Yes
NIMH	Yes (spreadsheet with limited information)	Yes (need to review hard copies)	?	?	?	Yes (from 2003 forward, which is 1 grant)
NINDS	No	No (not currently available)	Maybe (need to review hard copies)	No	Yes, if individuals are named (typically 3 years)	Yes
NINR	?	New awards per year (range:1-13)	?	?	?	?

4. ADDITIONAL FEASIBILITY STUDY FINDINGS

In this chapter, we present the recommended indicator variables and outcome measures for the evaluation, comparison groups considered, and sources of extant data that might be used in the evaluation.

Indicator Variables and Outcome Measures

Having documented the explicit and implicit goals of each K award, we turn to the identification of indicator variables that might be used to determine whether these goals have been achieved. Indicators provide a way of translating from more general constructs to specific behaviors, accomplishments, or activities that provide evidence that an activity has or has not taken place as planned and a goal has or has not been achieved.

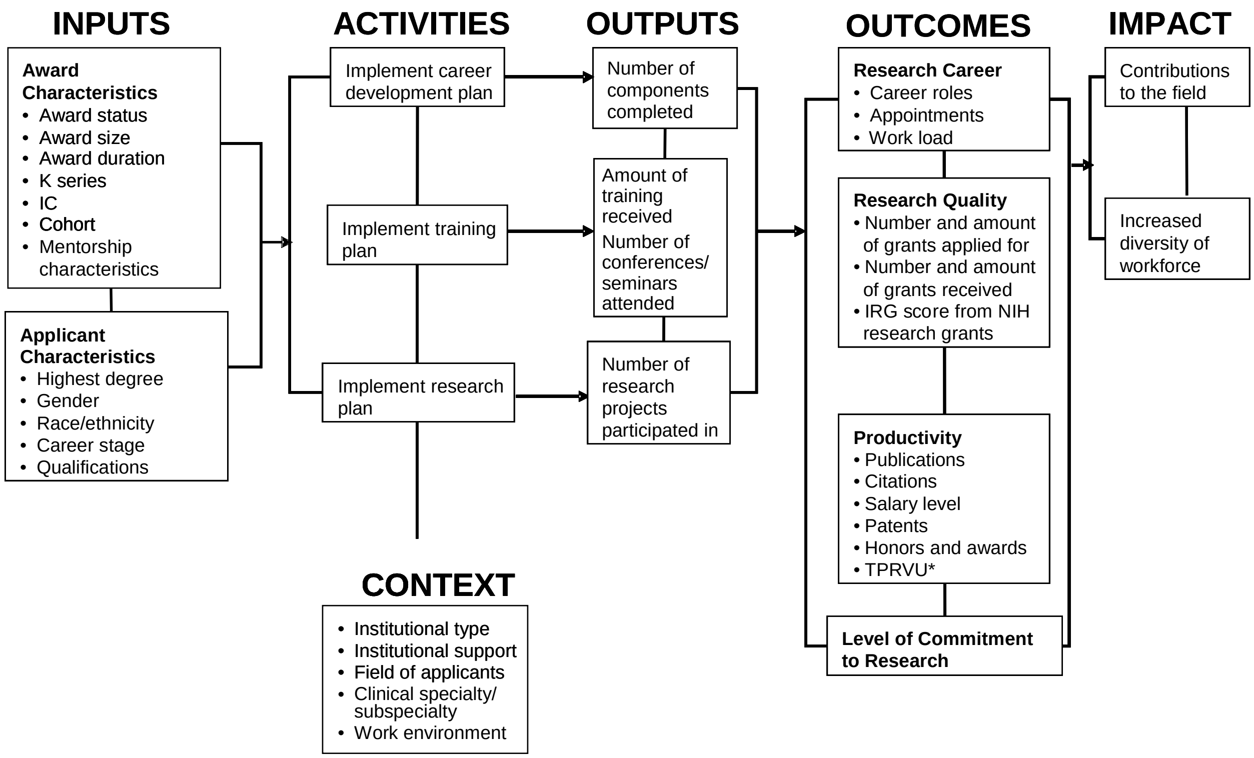
These indicator variables and outcome measures have been incorporated into a logic model for the Career Development Awards program. This logic model provides the scaffolding for identifying critical components to be measured and possible indicators that should be considered as part of the feasibility effort. In our previous evaluation work, we have found it very useful to develop a logic model and use it as initial consideration is given to a variety of design and measurement alternatives.

What is a logic model? A logic model is a visual depiction of the theory of change underlying the phenomena or processes to be studied. The logic model provides a way to identify the components and linkages believed to be critical. A logic model describes the following:

- "Context" describes the special features of the past and present environment in which the program participants are operating.
- "Inputs" describes the resources that support the program.
- "Activities" describes critical activities supported by the program.
- "Outputs" document the immediate products of the activities in quantitative terms.
- "Outcomes" are the program results and usually focus on individuals.
- "Impacts" documents the effects on the system and include impacts that are broader than those on individuals.

Exhibit 4-1 presents a logic model for the career development programs, showing the expected theory of change.

Exhibit 4-1.—Logic model



*Total professional relative value unit.

The logic model shows our thoughts regarding the components to consider in the development of indicators. In the following section, we provide a detailed description of each indicator.

Context

The context of both the institution where the awardees and their comparisons were at the time when the award was received and used, as well as the institution where they are currently working should be considered. We hypothesize that the environment of the first institution impacts how the award was utilized (outputs) and that of the second affects, at least in part, the outcomes. In addition, we posit that the field of the applicants and whether they work for academic environment or industry may affect the outcomes.

Five sets of related variables are recommended: institutional type, institutional support, field of the applicants, clinical specialty/subspecialty, and the work environment (academic vs. industry). Institutional type involves categorizing the institutions using for example the Carnegie system for academic institutions. Institutional support involves perception of factors such as quality of facilities, support from the leadership, and reward structures for research participation.

Inputs

Inputs are the awards (award characteristics) and the individuals (applicant characteristics).

Award characteristics document the inputs from NIH. They are designed to distinguish not only the presence or absence of the award, but also the potential differences within an award. They include:

- Presence or absence of the award;
- Award size in terms of the amount of individual award in constant dollars, as well as the total amount in cases in which multiple awards were received;
- Award duration;
- Type of K award, i.e., whether the award is for senior or junior applicants, and if there is a mentorship requirement;
- Institutions and centers (ICs);
- Cohort, i.e., the fiscal year in which the award is made; and
- Mentorship characteristics, such as amount of protected time, amount of face time with the mentor, and mentor qualities, for awards involving a mentor.

Applicant characteristics, which can be extracted from the CGAF, indicate what the applicant brings to fore at the time of application.

- Gender and race/ethnicity are common variables in this type of study.
- Highest degree received is the Ph.D., M.D. or both.
- Career stage suggests the professional age at the time of application.
- Application quality can be measured by the priority score.

Activities refer to areas where grant money was spent to enhance career development, such as the type of training received, career development attained, and type of research activities.

Outputs are numeric products of the activities such as the amount of training received, professional conferences/seminars attended and the number of research projects in which the individuals participated. While outputs are important in understanding the process, we do not recommend collecting these quantitative data because it is unlikely that the respondents will recall the information with accuracy, especially when the activities took place many years ago.

Outcomes

Outcome indicators are the major focus of the evaluation. We have grouped the outcome measures into four categories.

Research Career

- Career roles include type such as research, teaching, and administration, and whether it is science related.

- Employment is at universities and/or in industry and/or government. If at a university, recommended variables would include tenure status, research, and teaching load; if in industry/government, we recommend collecting information about the nature and level of the position, and the research and management load.

- Service on review committees, advisory groups and boards, and membership in research-oriented professional associations are recommended variables.

Research Quality

- The number and amount of grants applied for and received from NIH and other sources are suggested variables.

- NIH peer-review priority scores for the research grants is recommended as a measure of quality, because all of the NIH grant applications will have a score based on the panel review.

Productivity

- Salary level is a general measure of productivity frequently used by economists.

- Publications are major measures of productivity in academia. Criteria to consider are the number, prestige, and authorship contribution—a weighted measure to distinguish single vs. multiple authorship and account for difference in the order of authorship.

- Citations, including the count, rate, and prestige of the source are recommended.

- Patents can be used to measure productivity both in academia and industry.

- Total professional relative value units (TPRVU) billed annually[24] is a measure for clinical and non-research productivity (Eshelman et al. 2000). This is especially relevant for this study as a large number of awardees may be practitioners with M.D. degrees.

- Honors and awards include those given by the employers and by trade associations

Level of commitment to the research field can be measured both objectively by indicators such as time spent in the field and expected to be spent in the field, and subjectively including a series of questions such as how strongly an applicant is attached to the field and the his/her level of effort extended to advance the field.

---

[24] Relative value units have been used to evaluate physician clinical productivity. This method is considered a more reliable measure of clinical effort than hours worked, patients seen, or revenue generated because reimbursements can vary considerably. It is based on the resource-based relative value scale for payments under the Medicare system.

Impacts

Long-term effects include contribution to the research field and increased diversity of the work force at the national level. Because the focus of the evaluation is on individual outcomes, these impacts will not be addressed directly in the evaluation. One could assume that the aggregate of individual outcomes will shed light on the overall impact, although sometimes that sum may not equal the overall impact.

The evaluation will focus on outcomes but will also examine inputs, external factors, and development activities, which are captured in part by the type of K award received. Since outputs occur at the time the awards are active, which would be more than 25 years for some recommended participants, they will not be addressed in the evaluation. The impact of the NIH Career Development program on biomedical research also will not be examined because it is beyond the intended scope of the evaluation.

Comparison Groups

The purpose of this section is to explore different design options for constructing comparison groups for the Career evaluation. We begin with a general discussion of experimental and quasi-experimental designs and specific issues related to selection of comparison groups. Then, we examine different strategies used to create comparison groups in the evaluation of similar programs. Next, we analyze various options in the context of the Career evaluation from both methodological and practical perspectives. Finally, we discuss two related issues regarding the sampling plan and strategies for analyzing data.

Estimating Program Impacts

Perhaps one of the most challenging aspects of an impact evaluation is obtaining credible estimates of program impacts. By definition, program impacts are those outcomes that the program caused to happen and conversely would not have occurred without it. Therefore, to measure the impact of a program requires comparing its outcomes from a sample of participants with an estimate of what these outcomes would have been for the same group in the absence of the program (the counterfactual), an extremely difficult undertaking.

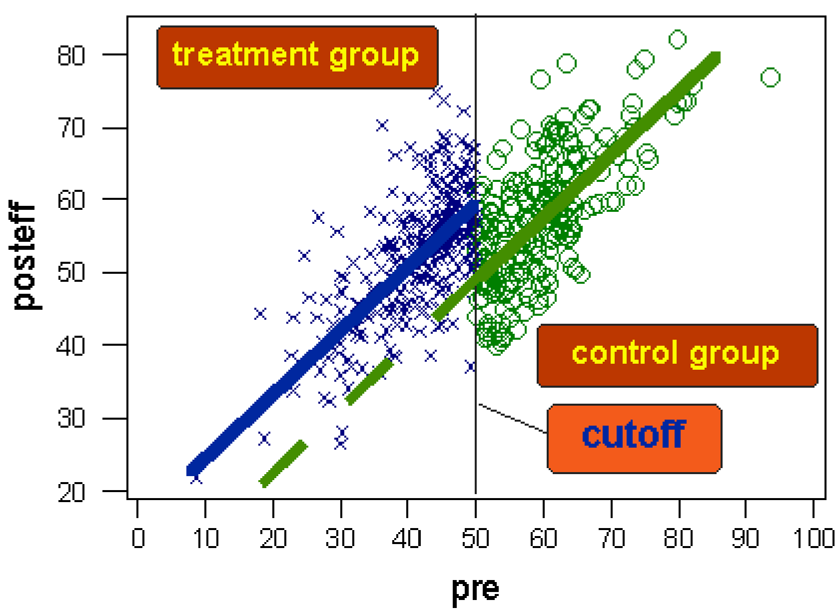
Program evaluation designs fall into three broad categories—experimental, quasi-experimental, and nonexperimental—based on how counterfactuals are constructed. Experimental designs randomly assign individuals to either treatment or control groups. Quasi-experimental designs identify comparison groups using a variety of techniques but not random assignment. Nonexperimental designs do not involve any comparison groups, and the impacts can only be estimated using statistical methods to analyze the natural variation among treatment subjects.

In principle, the best way to construct a comparison group is the experimental design. The laws of chance help to ensure that the two groups are initially similar in all ways. In practice, however, there are many situations in which it is not possible to use random assignment. For these situations, researchers have developed a broad array of alternative approaches using quasi-experimental comparison groups. Under existing conditions, it is infeasible, if not impossible, to conduct a randomized experiment for Career evaluation. A randomized experiment scenario would require half of the applicants to be awarded a grant and the other half to not receive a grant, determined through random assignment. Once assigned to treatment and comparison groups, the evaluation would then have to wait at least 10 years for the awardees to complete the program and for the programs outcomes to emerge. The merit-based nature of the program, as well as the desire for more immediate evidence about the program impacts, dictates the use of a quasi-experimental design for the study.

For most quasi-experiments, a fundamental problem is selection bias. Program and comparison groups may differ with respect to many factors that are related to their outcomes. These differences come about in ways that depend on how program and comparison members are selected. Hence, they might reflect how 1) individuals learn about, apply for, and decide whether to participate in a program (self-selection), 2) program staff recruit and screen potential participants (staff selection), or 3) researchers chose a comparison group (researcher selection).

Among various quasi-experiments, the major distinction concerns the ways in which comparison groups are developed in order to minimize the selection bias that results from the uncontrolled assignment of targets to one or more comparison groups (Rossi and Freeman, 1993). Of all the quasi-experimental designs, the regression-discontinuity (RD) approach is considered the strongest methodological alternative to randomized experiments (Trochim, 2006). Using this approach, individuals are assigned to a program or comparison group on the basis of cutoff scores on preprogram measures used to determine assignment. In other words, all persons on one side of the cutoff are assigned to one group; all persons on the other side of the cutoff are assigned to the other; a continuous quantitative preprogram measure is required. The logic is illustrated in Figure 4-1.

Figure 4-1.—Regression-discontinuity design with 10-point treatment effect



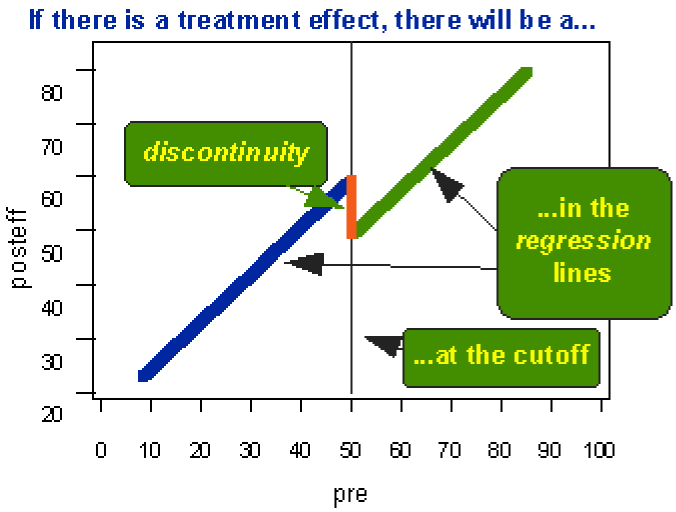
SOURCE: Trochim, 2006.

The figure shows a hypothetical bivariate distribution that might be found in a clinical trial. Each dot on the figure indicates a single person's pretest and posttest scores. The vertical line at the pretest score of 50 indicates the cutoff point, in this case representing an arbitrary threshold on a health index. The Xs to the left of the cutoff show the treatment cases, which are lower on both the pretest and posttest. The circles show the group that is comparative on both measures. The solid sloped lines in the bivariate distribution are the linear regression lines. In general, the distribution depicts a positive relationship, for both treatment and control groups, between the pretest and posttest—the higher a person scores at the pretest, the higher he/she tends to score on the posttest. If the treatment administered has a positive effect, in this case raising a treated person's score by 10 points, we will observe the distribution depicted in

Figure 4-1. The dashed line shows what we would expect the treated group's regression line to look like if the program had no effect.

Figure 4-2 presents a graphical depiction of the program effect when we observe a "jump" or "discontinuity" in the regression lines at the cutoff point.

Figure 4- 2.—How the regression-discontinuity design got its name



SOURCE: Trochim, 2006.

The RD design is distinguished from randomized experiments and from other quasi-experimental strategies by its unique method of assignment. The most common use has been in compensatory education evaluation. However, it has not been used frequently in social research for a number of reasons, not the least of which may be its deceptive name. Both terms have negative connotations to the lay audience. More to the point, though, is that in many situations, one or more key criteria are absent. For example, RD forces administrators to assign participants to conditions solely on the basis of quantitative indicators, thereby often restricting the degree to which judgment and discretion may be used. Perhaps the most telling reason for the lack of wider adoption is that at first glance, the design does not make sense. While other quasi-experimental designs aim to make program and comparison groups "similar," this approach appears to maximize the difference between two groups.

However, the strength of the design is that there is no selection bias, as both program and comparison groups chose to apply for the program, and the assignment strategy is perfectly known. The apparent difference between treatment and comparison groups at the preprogram stage can be "adjusted" statistically, thus yielding unbiased estimates of postprogram differences. Because of the lack of selection bias, some argue that regression-discontinuity is in a class of its own, but is as valid as randomized experiments in terms of providing causal evidence (What Works Clearinghouse, 2005).

The RD design is especially suitable to an evaluation of the NIH Career program because the program selects awardees on the basis of proposal merit, not random assignment. This consideration dictates the use of a quasi-experimental design for the study. The feasibility of the RD is greatly enhanced by the availability of priority scores. NIH research applications are reviewed by an Initial Review Group (IRG) who recommend approval or not of an application. In addition, the IRG members attach to approved applications a "priority score." This number, ranging from 100 (highest priority) to 500 (lowest priority), reflects their opinion concerning the scientific value of the proposed research relative to other research. The priority scores largely determine which approved applications are funded and which are not. IRGs are not uniform in the procedure by which they evaluate grant applications, nor are equal value of the priority scores necessarily equivalent across different IRGS or within the same IRG over time. Our initial analysis of the data shows that the average priority score for awardees is 160 with a standard deviation of 44, whereas the average score for unsuccessful applicants is 231 with a standard deviation of 80. To make priority scores comparable both across IRGs and within IRGs over time, we suggest employing a standardization procedure to create a standardized priority score using annual averages and standard deviations of the priority scores for all grant applications approved by each IRG. This procedure has been used by an earlier Career evaluation (Biddle et al., 1988).

Findings of Comparison Group Designs From a Literature Review and Evaluations of Similar Programs

Our review of evaluations from similar programs generated a few examples for comparison group design, which we will discuss in this section in detail. The majority of the evaluations, however, can be characterized as follow-up surveys of program participants involving no comparison group. The following provides four examples of evaluations that used comparison groups.

Example 1. The Rand evaluation (1988) of the NIH Career program (K01, K02, and K05) compared the career outcomes of the awardees (N=424) with a matched sample of R01 recipients who never applied for a K award (N=1,070). The comparison was based on the following: funding institute, three degree groups (M.D., Ph.D., and both degrees) with all Ph.D. groups further subdivided into whether or not their year of degree was missing, and all groups further subdivided in groups of 2 years each.

Example 2. Evaluation of the early career program of National Research Service Award (NRSA) predoctoral trainees and fellows (Pion, 2001) employed two comparison groups to NRSA trainees and fellows (N=12,441+2,440). The two groups differed in terms of the types of institutions where training occurred, with the first one said to be most similar to the NRSA recipients. Since the study involved analysis of records in existing databases and no additional data collection, the whole population was included in the study for certain outcomes (i.e., time to degree, immediate graduation plans, and NIH/NSF support). Other outcomes only used those sampled and responding to the SDR and the publication outcomes only included certain cohorts of the SDR respondents.

- NIH training institution group included individuals who graduated from departments that had NRSA predoctoral training grants, but unlike treatment groups, they were not supported by these training grants (N=15,992+6,684).

- Non-NIH training institution group included Ph.D.s who earned their degrees from departments that were not awarded NRSA predoctoral training support (N=15,037+14,338).

Example 3. Abt evaluation (Millsap et al., 2000) of the National Science Foundation (NSF) CAREER program was a quasi-experiment with the awardees group (N=1,037) and two comparison groups.

- NSF comparison group 1 included faculty members who applied for the CAREER award during the reference years and were not funded, and who subsequently received other NSF funding from regular research programs. There were 440 faculty in this group.

- NSF comparison group 2 consisted of faculty members who, between the reference years, met the eligibility requirements of the CAREER program but did not apply for the award. However, they applied for and received other NSF funding from the regular research programs. A random sample of 924 faculty was included in this group.

Example 4. In WestEd's evaluation of the NSF Graduate Research Fellowship Program (Goldsmith, Presley, and Cooley, 2002), in order to compare NSF disciplinary fellows, the evaluators employed one comparison group of "disciplinary peers" in four disciplines at 16 institutions who entered the same programs from the reference years (N=1,131).

A summary of comparison group strategies used in prior evaluations is provided in Exhibit 4-2.

Exhibit 4-2.—Programs and comparison group strategies

Program	**Evaluator**	**Comparison groups**
NIH Career	Rand	Eligible nonapplicants
NIH NRSA	NIH and Vanderbilt	Disciplinary peers
NSF CAREER	Abt	Unsuccessful applicants Eligible nonapplicants
NSF GRR	WestEd	Disciplinary peers

Evaluation of similar programs and literature suggested a number of options that the NIH Career evaluation may pursue. The next section provides a detailed analysis of these options in constructing comparison groups.

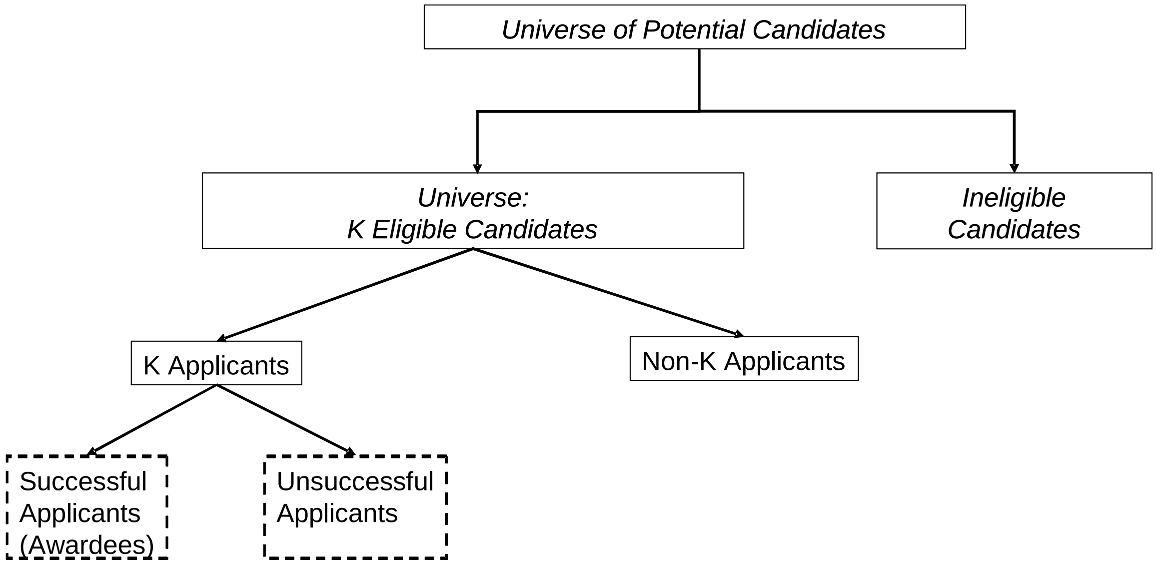
Analysis of Comparison Group Options

Selection of a comparison group for the Career evaluation is one of the most critical tasks for this planning effort. The choice affects which study questions can be addressed, how answers to study questions are interpreted, the types of generalizations that can be made, and the degree of confidence that can be placed in the findings.

The following is a discussion of a framework by which different comparison group strategies can be evaluated. We examine the appropriateness and feasibility of two types of comparison groups: macro-level and micro-level. The macro-level approach compares the awardees with those did not receive the award. The micro-level approach looks within the awardee group only and makes comparisons among awardees with different characteristics. Macro-level comparison is more critical, as it affects the overall design of the study.

1) Macro-level comparisons. The first step in identifying appropriate comparison groups for the survey was to define a universe that encompassed both Career awardees and appropriately defined groups of nonawardees. We pictured a universe of potential applicants. Such a universe is potentially very large. The size and indeterminate nature of this universe would make it both difficult and costly to identify individuals to construct the sampling frame. To create a more manageable sampling frame, we divided the universe into smaller, more definable subsets based on paths that these eligible individuals may take with regard to the Career program and other NIH grant programs. First, we distinguished between those who met the basic eligibility criteria for the Career program and those who did not. Among the eligibles, we further differentiated applicants and nonapplicants for Career grants. We then divided the applicant pool into successful (awardee) and unsuccessful groups. Through this process (illustrated in Figure 4-3), we arrived at the following possible groups for the study: eligible candidates, successful applicants, and unsuccessful applicants.

Figure 4-3.—Potential comparison groups



The next step in the process was to evaluate these groups and other alternatives such as disciplinary peers, as well as a retrospective comparison suggested by NIH staff. We evaluated these four options on two criteria: comparability to awardees and feasibility of data collection.

- **Comparability to awardees.** In this category, we considered three specific factors: (1) motivation/interest in the Career program, (2) eligibility under Career guidelines, and (3) participation in the same review process with known results.

- **Feasibility of data collection.** We used three factors to assess the feasibility and practicality of data collection: (1) ease of identifying potential group members for the sampling frame, (2) data availability, and (3) obtaining their cooperation with the primary data collection.

Exhibit 4-3 summarizes our analyses of each option based on these criteria, and the following sections present the details of our analyses. Areas identified as advantages were denoted with "+"; those regarded as "disadvantages" were marked with "-." In the section that follows the exhibit, we present a detailed discussion of the pros and cons of each option.

Exhibit 4-3.—Comparability and feasibility assessments of alternative comparisons groups for the Career evaluation

Comparability and feasibility	Option 1. Unsuccessful applicants	Option 2. Eligibles	Option 3. Departmental peers	Option 4. Retrospective
Comparability Criteria				
Motivation/interest in Career	(+) Career application suggests comparability to awardees.	(-) No basis to determine.	(-) No basis to determine.	(-) No basis to determine.
Eligibility under CAREER guidelines	(+) Applied under same guidelines as awardees in same cohort. (+) Met the eligibility criteria.	(-) Applied under different program guideline. (-) Met different program eligibility.	(-) Career eligibility unknown.	(-) Career eligibility unknown.
Consistent review process with known outcomes assessed	(+) Same as awardees. (+) Statistical adjustment possible based on reviews and panel ratings.	(-) Varies by NIH program. (-) No comparable measure available. (-) No statistical adjustment possible.	(-) No review process to compare. (-) No comparable measure available. (-) No statistical adjustment possible.	(-) No review process to compare. (-) No comparable measure available. (-) No statistical adjustment possible.
Feasibility Criteria				
Constructing the sampling frame	(+) Easily identified in NIH databases. (+) No additional screening required.	(+) Grantees easily identified in NIH databases. (-) Requires case-by-case screening to determine Career eligibility, comparability to awardees.	(-) Requires departments to supply current faculty list. (+) Current lists likely to be reliable. (-) Requires case-by-case screening to determine Career eligibility, comparability to awardees.	(-) The sampling frame needs to be known a priori.
Data availability	(+) Some extant data available from CGAF.	(+) Some extant data available from CGAF.	(-) New data collection required.	(-) New data collection required.
Obtaining cooperation (likely response rate)	(+) Relatively high, but may be lower than eligible group.	(+) Likely to be the highest.	(-) May be low if many members have no NIH affiliation.	(-) May be low if many members have no NIH affiliation.

Option 1. Unsuccessful Applicants. This option has the clear advantage of being the most rigorous in design as well as practical in data collection. It minimizes selection bias by using individuals who, like awardees, demonstrated an interest in the Career program's objectives and were motivated to apply for grants that emphasized both research and education. Awardees and unsuccessful applicants in the same cohort would have applied under and met the same eligibility guidelines and been evaluated by the same process and reviewers. The most critical difference between unsuccessful and awardees would be how their proposals fared in the competitions. Differences can be adjusted statistically using the proposal rating and/or panel score each application received.

Members of this group are easy to enumerate and describe. NIH databases contain their names, complete records of their NIH activity, and other key characteristics such as institution, year of highest degree, etc. Members of this group also are easy to locate. In one of the activities we carried out for the

NSF CAREER feasibility study, we located 92 percent of test cases (most of whom had changed institutions) through web searches that took, on average, less than 3 minutes. Other studies showed that response rates for similarly defined unsuccessful applicant groups were comparable to or lower than those of the eligible group: 59 percent and 56 percent, respectively, in an earlier evaluation of the CAREER program (Millsap et al., 2000) and 54 percent and 70 percent, respectively, in an evaluation of the Presidential Young Investigators program (NSF, 1990).

Option 2. Eligibles with Grants from Other NIH Programs. This group poses several threats to validity not encountered with unsuccessful applicants. Because the eligibles did not apply to the Career program**,** they may have had different professional interests than awardees. In addition, the winning proposals that "qualify" this group will have been submitted to NIH programs with different eligibility criteria that also placed no restrictions on the tenure or career stage of the applicant, and they were evaluated on different cycles, by different processes, and by different reviewers. For these reasons, it would not be possible to adjust for pre-award differences in the quality of winning proposals.

Assessments of the eligible group on feasibility criteria are mixed. NIH databases contain adequate information for identifying PIs who received other grants. However, to develop the sampling frame, eligibility would have to be established on a case-by-case basis using presurvey screening questionnaires. The likely response rate, as noted above, may be either comparable to or higher than that of the unsuccessful applicant group.

Option 3. Peers in the Current Departments of Awardees. Departmental peers can be used as comparison in order to examine awardees' accomplishments in the context of the departments in which they work. This approach is often used in evaluating awards that are made to institutions rather than individual PIs. This group may encompass noneligibles, eligible nonapplicants, and unsuccessful applicants. Clearly, this peer group could provide valuable perspectives on an awardee's prestige and contributions to the department. However, as a basis for comparison in determining the Career program's impact, this group would pose some of the same threats to validity as the eligible group. Both raise questions about comparability to awardees in interests and motivation, eligibility for the program, and participation in the same grant programs and review processes.

Using this peer group in the survey design also would require the following steps, which would be unnecessary if either the unsuccessful applicant or the eligible group is chosen for comparison:

- Departments with Career awardees would have to provide evaluators with a list of faculty members along with position titles, tenure status, and disciplines indicated.

- Evaluators would have to extract Career awardees from the faculty list to form a nonawardee group that may include unsuccessful applicants, eligibles, and others. These distinctions, however, will not be clear at this point in the process.

- Evaluators would have to narrow down the group using some predetermined criteria because the resulting pool is likely to be large.

While the first step is not difficult to implement, it represents an additional burden on departments and adds time to the evaluation. The second step could be quite tedious if names in the NIH databases and those from departmental lists are slightly different from one another (e.g., James Smith *vs.* James A. Smith *vs.* James Smith II). The third step could be accomplished through use of a presurvey screening questionnaire, which would be necessary for the eligible group. An alternative to screening would be to collect information on tenure, other status variables, and demographic factors in the main survey and to analyze the data using statistical models.

One of the few advantages of using departmental peers is that their locations would be known. The response rate from this peer group is likely to be lower than other comparison groups, however, because many members may have no affiliation with NIH.

Option 4. Retrospective Comparison (case-control). In constructing the comparison group for the previous three options, we tried to ensure that the comparison group was as similar as the Career awardees at the time the award was made. Unlike the three prospective options, NIH staff raised the possibility of a retrospective comparison group, that is, to look at a group of people that have achieved the goals for the Career program and examine whether they have received any Career award in a retrospective fashion. This is a very efficient design for studies of rare conditions (e.g., Nobel prizes) and has been frequently used in epidemiological studies to explore uncommon outcomes such as death (Khlat, 1997). However, such an approach requires that the sampling pool be perfectly known a priori. In other words, we need to begin with a given and exhaustive list of people who have achieved the Career goals. Given the diverse goals of different K awards, we would need multiple sample pools to satisfy that requirement.

The preceding analysis suggests that the unsuccessful applicants offer the best choice as a comparison group for Career evaluation. This group poses fewer threats to validity, as it lends itself nicely to the regression-discontinuity design, and also has greater practical advantages. The Career-eligible group and the other options we considered face validity challenges not easily overcome, and they also introduce additional problems for data collection.

2) Micro-level comparisons. Micro-level comparison groups are subgroups of Career awardees that can be distinguished, for example, by the differences in characteristics with regard to the awards, awardees, and institutions outlined in the logic model.

Comparisons based on these characteristics answer a somewhat different set of questions than the macro-level approach, focusing on the differential effect of the program under varying conditions, rather than the overall effectiveness. In addition, disaggregation of data based on micro-level criteria would also allow us to compare the profile of the Career awardees with those in the comparison groups. It may also allow, to some extent, comparisons of program effects among K-series and ICs.

We recommend considering both macro and micro approaches for creating comparison groups in order to optimize the information gained through the evaluation.

Analysis of Sampling Plan

Since 1957, the NIH has made about 16,000 competing career development awards, and it is impossible and unnecessary to sample the whole population. One of the first issues is to decide the time frame from which the sample will be drawn. We decided to focus on awards made after 1975 for the impact evaluation for two reasons. The CGFA data are more complete and well maintained. In addition, the subjects are more easily to be identified and located, as the majority are still active in work.

The second issue is about the sample size. As will be explained in a later section, the existing data sources do not contain information on all of the outcomes. In order to capture such information on the full array of outcomes, primary data collection such as survey will be necessary. Given the large size of the potential population (including comparisons), it is not necessary to survey the population to examine the Career award program as a whole. However, the design of an efficient sampling plan depends on an assessment of analytic needs (i.e., degree of precision desired, complexities of the analysis) and resources (i.e., time and money), and hence requires careful judgment.

Our selection of sample size is informed by power analysis (Kraemer and Thiemann, 1987). While researchers are often concerned with the significant level α (i.e., a reasonable level of remaining doubt), they sometimes overlook the power of analysis β (i.e., the convincing level beyond reasonable doubt). The latter directly affects the sample size. The power coefficient is related to the following factors: 1) the significant level (i.e., 1 percent, 5 percent, etc.), 2) how the hypothesis is formulated (i.e. one or two-tailed test), and 3) critical effect size. For example, under a standard 5 percent level two-tailed test, a power of 0.8 will require a sample size of 781 for an effect size of 0.1. However, the power analysis is based on a randomized experiment design. To attain the same power, an RD design will require a 2.75 times larger sample (N=2,148) (Trochim, 2006). Assuming a 50 percent overall response rate, the evaluation will need 4,296 completed surveys.

The last consideration is about sampling strategy. There are a number of possibilities, the pros and cons of which are summarized in Exhibit 4-4. One could draw a random or stratified sample from both treatment and comparison groups. This approach has the advantage of ascertaining that the sample is representative of the population. However, given the large number of K-series awards and long time frame involved, the approach is likely to introduce large variances from contextual effects, because various mechanisms used for K awards have changed, as has their target populations. For example, whereas an overwhelming majority of the awards were K04s in 1975-76 and most likely were awarded to Ph.D.s, the large majority in 1995-96 were K08s, which were explicitly targeted for M.D.s and other clinical doctorates. In addition, there have been noticeable changes in the resources available for research, the expectations for level of performance, and opportunities to achieve these outcomes. For example, the number of publication outlets has increased substantially, the competition for NIH funding has become increasingly difficult, and the research team versus the individual investigator approach has become more common in attaching several research problems. Again, these changes should affect the expected performance levels of awardees and their comparison groups. For example, it is known that publications that are authored by teams accrue more citations than those produced by small labs and individual investigators.

The second strategy is to sample all applicants from a short time frame (i.e., 1990–95). While this approach will minimize the contextual effects, the short time frame also decreases the utility of the findings. In other words, the evaluation will provide evidence of Career impacts on recent awards but not from the earlier periods.

A middle-of-the-road strategy is to sample all awardees and unsuccessful applicants from selected cohorts, assuming that the selected cohorts are representative of the cohorts from the same decade. Specifically, we recommend drawing all candidates from the 1975, and 1976 cohorts to represent candidates considered as “senior” in their career as of 2006 (i.e., 25 years after the end of the award, assuming 5 years as the average support time). Similarly, candidates from the 1985 and 1986 cohorts will form the “mid-careers” group (i.e., 15 years after the award); and those from the 1995 and 1996 cohorts will form the “young-career” group (i.e., 5 years after the award). Exhibit 4-4 presents the pros and cons of three options in terms of sampling strategies.

Exhibit 4-4.—Comparison of sample strategies

Criterion	Option 1. Random/stratified sampling	Option 2. Population sampling from a short time frame	Option 3. Population sampling from selected cohorts spanning a long time frame
Internal validity (degree to control for contextual effects)	Most difficult to control for contextual effects.	Sample is more homogenous and the contextual effects can be minimized.	Contextual effects can be somewhat minimized by using treatment and comparison groups from the same time period.
External validity (is the sample representative of the whole population?)	Sample is most representative of the population.	The sample only represents recent awardees and not earlier ones.	The sample is representative of the population and impacts from early to recent awards.

We define treatment subjects as awardees who received the first competing award and comparisons as those who have applied but never received the award. Therefore, when selecting the awardees from the sampled cohorts, it is necessary to ensure that they are uniquely identified. If awardees are found in multiple selected cohorts, they will be identified as awardees from the earliest cohort group. Similarly, in identifying the unsuccessful applicants, those who were not awarded in the sampled cohorts but have received the awards either previously or afterwards must be removed. Although one could argue that doing so would increase selection bias and reduced the sample size, we maintain that the tradeoff of a small degree of selection bias is offset by introducing potential contamination from treating applicants who received later K awards as unsuccessful applicants. In addition, our preliminary analysis of unsuccessful applicants did not find evidence of serious sample reduction. Table 4-1 shows the total combined sample size from the selected cohorts is 4,965 (2,291+2,674). Assuming 50 percent response rate, the expected valid sample will be 2,482, which is slightly larger than our sample size derived from the power analysis.

Table 4-1.—Samples from the selected cohorts

Cohort	FY	Number of awardees		Number of unsuccessful applicants	
		All	No duplicate count	All	Never awarded
Cohort 1	1975	372	372	757	680
	1976	283	283	465	403
Cohort 2	1985	429	418	582	445
	1986	295	285	769	555
Cohort 3	1995	476	457	588	310
	1996	509	476	530	281
Total...		2,364	2,291	3,691	2,674

It is important to note that the above sampling strategies are designed to evaluate the NIH Career program as a whole, and not with respect to specific K awards or ICs. If the program is evaluated at the K awards or IC level, a much larger sample would be needed.

Table 4-2 shows the number of awards by K categories for all awards since 1975 and for awards in the selected cohorts. Table 4-3 shows the number of awards by IC for all awards since 1975 and for awards in selected cohorts. (Note that the numbers presented in these tables pertain to the awards rather than awardees, as in the previous table.) In general, we do not recommend any sublevel evaluation whose total number of valid sample is smaller than 100; this would double to 200 with a 50 percent response rate. This is a minimum sample size to provide sufficient power to show some differences between the awardees and the unsuccessful applicants. Therefore, in order to provide analyses by K or by IC, the recommended approach is to include the complete population since 1975. Even then, only 10 of the 19 K categories and 14 of the 23 ICs have sufficient sample size. The sample size will not be sufficient to examine K categories by ICs.

Table 4-2.—Number of K awards, by K categories

K category	All awards (1975–2004)	Awards from selected cohorts
K01	1,602	85
K02	1,162	167
K03	1,922	1
K04	3,412	747
K05	632	109
K06	595	0
K08	4,576	749
K09	14	1
K10	27	3
K11	783	174
K14	101	46
K15	57	13
K18	9	0
K20	94	26
K21	85	32
K23	1,200	0
K24	350	0
K25	98	0
K26	13	0
All K, total	16,732	2,153

NOTE: The number includes awards since 1975 and excludes K07, 12, 16 17, 22, and 30 because either they were not originally going to be part of the evaluation or they are institutional awards, which will be examined separately.

Table 4-3.—Number of K awards, by IC

IC	All awards (1975–2004)	Awards from selected cohorts
National Institute on Alcohol Abuse and Alcoholism (NIAAA)	279	29
National Institute on Aging (NIA)	535	79
National Institute of Allergy and Infectious Diseases (NIAID)	1,351	169
National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	1,036	147
National Institute of Allergy and Infectious Diseases Research Support (NIAID)	345	36
National Center for Complementary and Alternative Medicine (NCCAM)	34	0
National Cancer Institute (NCI)	1,412	216
National Institute on Drug Abuse (NIDA)	608	85
National Institute on Deafness & Other Communication Disorders (NIDCD)	162	18
National Institute of Dental and Craniofacial Research (NIDCR)	381	60
Formerly Division of Dentistry (DD-BHP)	2	0
National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK)	1,083	104
National Institute of Biomedical Imaging and Bioengineering (NBIBIB)	13	0
National Institute of Environmental Health Sciences (NIEHS)	122	21
National Eye Institute (NED)	272	44
National Institute of Child Health & Human Development (NICHD)	1,591	107
National Heart Institute	910	119
National Human Genome Research Institute (NHGRI)	140	0
National Heart, Lung, and Blood Institute (NHLBI)	40	4
Agency for Health Care Policy and Research (AHCPR)	2,452	425
National Library of Medicine (NLM)	43	0

Table 4-3.—Number of K awards, by IC—continued

IC	All awards (1975–2004)	Awards from selected cohorts
National Institute of Mental Health (NIMH)	37	6
Neurological Diseases and Blindness	2,118	252
National Institute of Nursing Research (NINR)	86	0
National Institute of Neurological Disorders and Stroke (NINDS)	95	3
National Institute for Occupational Safety and Health (NIOSH)	1,219	206
National Center for Research Resources (NCRR)	92	9
Fogarty International Center (FIC)	247	14
National Institute of Child Health & Human Development (NICHD)	27	0
All IC, total	16,732	2,153

NOTE: The number includes awards since 1975 and excludes K07, 12, 16 17, 22, 30 because either they were not originally going to be part of the evaluation or they are institutional awards which will be examined separately.

Given the strong desire from the ICs to compare the relative success of the K-series, we recommend a supplementary study that will examine the population of all awardees and unsuccessful applicants from 1975 to 2004. Only extant data in the CGAF and from NSF would be used in this study. The only outcomes that would be examined in this supplementary study would be success on later NIH research grants such as R01, R01 equivalents (i.e., R23, R29, R37) or other Research Project Grants (RPG) awards and success in obtaining NSF grants.

A separate sampling strategy is needed for the institutional awards (K12, K16, and K17). Records of individuals supported under these awards are housed within the individual ICs and are frequently in hard copy only. Evaluation of the outcomes of the institutional awards depends on the availability of names of individual recipients. If individual data are available, comparisons between recipients of institutional grants and individual K awards can be made.

Finally, to obtain a more detailed look at the publications, citations, and patents obtained by the awardees and the comparison sample, we recommend that a small bibliometric analysis be conducted. The sample would consist of 180 applicants (half awardees, half unsuccessful applicants)—60 from each of the three cohorts.

Analysis of Analytical Strategies

It is difficult to explain design without discussion about analysis because they are inherently related. Many of the studies reviewed incorporated bivariate comparisons between treatment and comparison groups using statistical tests such as *t*-test or chi-square tests. While these methods are easy to understand, for reasons explained earlier, estimating a program impact by a simple difference in mean outcomes may confound the true impact of the program with the effects of other factors. In quasi-experimental studies, there are three strategies in the analysis stage to further improve impact estimates (Bloom et al., 2002).

- **Use statistical balancing to eliminate differences in observed covariates (i.e., propensity score).** This procedure often entails developing an index based on propensity score, which is the probability of being in the program group instead of the comparison group as a function of observed covariates. It is used to deal with unknown non-random selection processes (Rosenbaum and Rubin, 1983). This method is limited in that it can only balance measured covariates.

- **Predict the counterfactual by modeling the outcome (i.e., ordinary least square model, hierarchical linear modeling).** The simplest and most widely used approach is regression analysis by modeling the systematic variation in the outcome. The ability to emulate the counterfactual for a program impact estimate depends on how well they account for the systematic determinants of the outcome.

- **Control for unobserved covariates.** This approach attempts to identify and measure variables that are related to selection but not to unobserved determinants of the outcome (i.e., the probability of being in the program group as a function of distance from the program). In practice, it is very difficult to identify the required exogenous selection correlates.

From the review of existing related evaluations, we found that the analysis is the weakest area. The majority of the evaluations present only descriptive data (i.e., count and percent) or examine the mean difference between treatment and comparison groups. While the information is important and informative for understanding the program, it is difficult to judge whether the differences observed between program and comparison groups are statistically significant and if they are, the extent of program effects.

In light of the RD strategy we recommended earlier, we suggest using 1) the propensity score approach (priority score) to model the selection process; and 2) both bivariate and regression approach for the impact analysis, which is the primary purpose of the analysis. The regression method is the most widely used and understood approach of the three options. For example, we may conceptualize a three-level hierarchical linear model (HLM). As a special case of regression, HLM models the hierarchical structure in a higher education institution, correcting for aggregation bias, misestimated precision, and the unit of analysis problem, thus producing more accurate results (Raudenbush and Bryk, 2002). At level 1, we predict individual outcomes as a function of program status (treatment vs comparison), priority scores, and other award and awardee characteristics identified at micro-level comparison, such as differences between mentored and non-mentored awards and between awardees with Ph.D. and M.D. degrees. At level 2, institutional characteristics would be controlled. Level 3 models area or field of research as contextual variables. However, the specification of model will also depend on the analysis of variance components at each level from the empirical data. Separate analyses will be conducted for the three cohorts. Different functional forms, including linear or quadratic, will be explored to model the relationship between the award status, priority scores, and the outcomes.

Qualitative techniques will be employed to analyze qualitative data from open-ended questions in the surveys. Content analysis will be used to code data according to themes, using the interview protocol as a framework. Descriptive codes, which simply classify data into thematic chunks, will later be replaced by pattern codes after subsequent rereadings. Pattern codes indicate emergent patterns in the data and are typically used in the last stages of analysis (Miles and Huberman, 1994).

Existing Data Sources

Several evaluations of NIH training programs have relied primarily upon existing data sources to assess outcomes of trainees and fellows. Summarized below are several data sources that include information of potential use in evaluating the Career Development Awards. These data sources include variables that contain information on (1) selected research-related outcomes, (2) variables that may prove useful in determining the extent to which awardees and relevant comparison groups differ at the time of application for an award on variables (other than award receipt) that can affect career progress, or (3) both. In contrast, these data sets seldom contain variables that measure such inputs and activities as quality of mentoring and the training experiences sponsored by the award.

Because the K awards include both Ph.D.s and M.D.s—two groups that may have somewhat different career paths as well as factors that affect career progress—the data sources are presented separately in terms of their relevance for these two groups. For each data source, a brief description of the variables and data collection procedures are provided. In addition, the usefulness of the data source is summarized in terms of its coverage of the relevant population, measurement of the key variables over time, and known aspects of the data quality.

Data Sources Relevant to K Awardees and Comparison Groups with Ph.D.s

Doctorate Records File (DRF)

Brief description and relevance. The Doctorate Records File is based on the Survey of Earned Doctorates (SED), which has been ongoing since 1920. This data collection effort is sponsored by several federal agencies (including the NIH) and is currently conducted by the National Opinion Research Center. The survey is a census of the research doctorates awarded by U.S. academic institutions. Included in the questionnaire, which is completed by graduates at the time that they complete all requirements for their doctorate, are questions on (1) demographic characteristics (e.g., gender and race/ethnicity); (2) educational history (e.g., undergraduate, master's, and Ph.D.-granting institutions); (3) degree characteristics (e.g., field and time to degree); (4) financial support for graduate study; and (5) immediate employment or postdoctoral study plans. A subset of these variables may be relevant to the Career Development Award evaluation inasmuch as they provide information on background characteristics that may distinguish K awardees and their comparison group counterparts at the time of award application. As such, they could serve as possible covariates (in addition to priority score) in the suggested propensity analyses. Of course, their usefulness depends on the extent to which such variables are both related to career outcomes and contribute over and above the priority score (e.g., information on their doctorate-granting institution and its research intensiveness).

Usefulness for the evaluation. The basic set of questions has remained relatively stable over time, and the data on K award recipients and comparison group members who earned Ph.D.s should be reasonably complete. Because the survey is a census of new research doctorate recipients and the response rates have remained extremely high (about 95 percent of new Ph.D.s), coverage and nonresponse bias for K awardees and comparison group members who earned Ph.D.s from U.S. institutions should be very low.[25] In addition, a very limited set of information on nonrespondents (the remaining 5 percent) is included, based on information extracted from commencement programs and other institutional sources. The quality of measurement is high, with a few exceptions. The most noticeable example involves the accuracy of self-report information on sources of financial support for graduate study, which has been found to be problematic (Ingram and Reis, 1994).[26]

Because the data are limited to predegree variables and immediate postgraduation plans, this data source contains no relevant data on actual outcomes for K awardees.

---

[25]The number of individuals with Ph.D.s from foreign institutions in both the K awardee and comparison groups is unknown, but given the citizenship requirements for award eligibility, it is expected to be reasonably small.

[26]With regard to predoctoral traineeships and fellowship, the Trainee-Fellow File (TFF) maintained by the NIH is a better source. At the same time, the largest source of NIH training support is provided by faculty research grants, and this information is not collected by the NIH. Here, the quality of self-report information has not been examined. This may or may not be information unless source of graduate support is somehow related to preexisting differences between K award recipients and the comparison groups.

Survey of Doctorate Recipients (SDR)

Brief description and relevance. The Survey of Doctorate Recipients is a biennial, panel survey currently conducted by the National Opinion Research Center. Initiated in 1973, this data collection effort is sponsored by the National Science Foundation, the National Institutes of Health, and other federal agencies. The survey target population consists of individuals who (1) had earned a research doctorate from a college or university in the U.S. in a science, engineering, or health science field; (2) were U.S. citizens or non-U.S. citizens who indicated in the SED plans to remain in the U. S. after receipt of their doctorate; (3) were under 76 years of age and not living in an institutionalized setting (National Science Foundation, 2006). For each survey wave, this target population is modified by adding individuals from the most recent two-year cohort of new doctorates (as identified by the SED) and eliminating those who have retired, are institutionalized or deceased, or exceed the age limit. There have been some changes to these eligibility rules over the course of the SDR, the most recent being that as of 2003, U.S. citizens who had been residing outside the U.S. for 4 or more years were included in the sampling frame.

The content covered by this survey includes variables relevant to the evaluation of K awardees. Of most interest are questions that pertain to career outcomes; examples include current employer, faculty position and rank, work activities, number of people supervised, federal contracts and grants received, current salary, papers presented, authored articles in refereed journals, published books and monographs, and patent applications and awards for the most recent 5-year period. A majority of these questions have been included in each survey wave. Surveys for a particular year also have frequently incorporated a “module” of items on special topics (e.g., the 1995 SDR covered current and previous postdoctoral training).

Usefulness for the evaluation. There have been several changes in the survey design since its inauguration in 1973, which may or may not affect its usefulness in evaluating the Career Development Award programs. In response to declining response rates experienced throughout the 1980s, the SDR underwent several changes in the sampling, questionnaire, and follow-up design, beginning in 1991. For example, the survey became a mixed-mode survey in 1993, which improved the response rate and thus reduced threats to data quality from nonresponse bias. In addition, the use of three main data collection modes (self-administered paper questionnaire, computer-assisted telephone interview, and self-administered online questionnaire) has not appeared to introduce factors that would affect the quality of data provided by respondents among the different modes of administration.

Since its inception in 1973, certain core survey questions have been reworded so as to improve data quality. At the same time, this action affects the ability to compare responses over time, and in a few cases, the utility of information for certain outcomes has been negatively affected. For example, the information on time spent in research as well as other activities has become less detailed. Whereas in earlier surveys respondents were asked to report the percentages of time spent in each activity, questionnaires administered since 1987 have only asked individuals to indicate either (a) their primary and secondary activities or (b) the activities on which they spent at least 10 percent of their time and, of these, which consumed the most hours. As such, these survey data on “time spent in research” provide cruder information on this outcome, which then affects its measurement for each group and the ability to identify potentially meaningful group differences. It also reduces the ability to compare different cohorts (e.g., changes in the time spent in research for the 1975 versus the 2000 cohorts).

The primary drawback in relying on the SDR for describing career progress for K award recipients and comparing it with individuals who did not receive the award is the limited sample size. Since its inception, the proportion of doctorates sampled has ranged between 5 and 8 percent of the population. Consequently, the number of K awardees and comparison group members included in the sample and who respond regularly is relatively small. For example, in the 2003 SDR, the total sampling frame for biological sciences Ph.D.s in the fields of interest to the NIH totaled 82,002; of this group, 4,476 (5.5 percent) were selected to be in the probability sample, and the final achieved sample (excluding nonrespondents and ineligibles) was 3,451 individuals (National Science Foundation, 2006). This suggests that among the total numbers of K grantees and unsuccessful applicants for 1975, 1985, 1995, and 2000, outcome data from the SDR would be available for less than 117 Career Development Award recipients and 159 unsuccessful applicants. Obviously, these small samples would yield reasonably imprecise estimates of career outcomes and impacts and preclude the ability to compare outcomes for the two groups for key subgroups (e.g., different cohorts, genders, or field of science).[27]

Data Sources Relevant to K Awardees and Comparison Groups with M.D.s

AAMC Medical School Graduation Questionnaire

Brief description and relevance. The Medical School Graduation Questionnaire (GQ) is a survey that is administered annually to U.S. graduating medical students. Sponsored and conducted by the American Association of Medical Schools (AAMC), this effort was initiated in 1978. Questions are included that ask about students' experiences in medical schools, educational and noneducational debt, and specialty choice, to name a few.

Usefulness for the evaluation. This survey of medical students has some relevance to the evaluation, primarily in the availability of information on preexisting differences between K awardees and comparison group members on variables that may affect performance on specific career outcomes. For example, questions are included as to whether individuals participated in a research project in medical school or authored a research publication during this time—characteristics that have been shown to be related to later involvement in research. Information on age at time of graduation from medical school, research intensity of the medical school, educational indebtedness, and intentions to pursue a research career also are collected by the survey, and these variables have been shown to affect specialty choice and later research involvement (e.g., Fang and Meyer, 2003; Rosenblatt and Andrilla, 2005). Given that students make these ratings at the time of graduation, they may possess more validity than retrospective accounts that are reported many years later of these experiences and plans.

Several of these questions have been included in each survey wave, suggesting that information should be available on K awardees and their comparison group counterparts who graduated from medical school during the past 25 or so years. However, detailed information on response rates is not readily available so as to better estimate their level of participation in the survey. Rough calculations of the number of respondents and the number of graduates indicate that response rates usually exceed 60 percent. Although expected to be a reasonably small number, Career Development awardees and comparison group members who earned their M.D.s at a foreign medical school would not be included in this data collection effort.

Similar to the SED, these data are useful only with regard to providing information on possible covariates that can be used in construction of propensity scores because they pertain primarily to individuals' experiences during medical school.

---

[27]These numbers are, in fact, serious overestimates, primarily because they are based on the total number of K award recipients and unsuccessful applicants for the respective cohorts; these individuals include those with degrees other than the Ph.D., who thus would not be included in the SDR's sampling frame. However, they help dramatically illustrate the problem that would result in relying on SDR data for this evaluation.

Data Sources Relevant to K Awardees and Comparison Groups with Ph.D.s and/or M.D.s

AAMC Faculty Roster

Brief description and relevance. This data initiative was begun in 1966 by the AAMC and includes information on the type of appointment, rank, and salary of paid faculty in accredited U.S. medical schools. Such information is reported to the AAMC by institutions that submit educational, employment, salary, and demographic data on new faculty members upon their first appointment to faculty positions in the institution and updates these data as necessary throughout the faculty member's tenure. Medical school participation in this data collection effort is voluntary, but all 125 medical schools participate and help to facilitate data collection by appointing a specific individual to coordinate reporting at their institution. Currently, the Faculty Roster contains records on approximately 119,000 full-time, active faculty; over 135,000 inactive faculty are retained for research purposes or in case of reactivation.

Usefulness for the evaluation. Relevant variables include those related to applicant characteristics (e.g., gender, race/ethnicity, educational history, and previous NIH training support) and outcome variables, including faculty appointments, tenure status, rank, involvement in research, service, and other activities (at least 10 percent). All changes in faculty status from initial employment to retirement are recorded, thus providing a career profile for individual faculty members. In 2002, approximately 90 percent of all full-time faculty in U.S. medical schools were included (Fang and Meyer, 2003). The Faculty Roster provides reasonably complete information on current and previous employment for both M.D. and Ph.D. faculty in medical schools. However, it does not capture participation in research grants and contracts, membership and leadership in professional associations, patents, or honors and awards, all outcomes of interest in the evaluation. In addition, faculty positions in colleges and universities without medical schools and nonacademic employment positions are not captured.

Institute for Scientific Information (ISI)

Brief description and relevance. The Institute for Scientific Information (now part of Thomson) not only compiles and creates the Web of Science, but also provides specialized data sets on publications and citations to these publications for research projects. For example, data on publication-related outcomes and impacts can be obtained for both K award recipients and their comparison group counterparts. For each publication that is authored or co-authored by these individuals, data can be obtained on the authors, order of authorship, and journal as well as other variables (whether the paper is an article, research note, letter to the editor, or review). Additional information such as expected citation rate also is available. ISI can also provide data on patents, such as number of patents registered, dates of patents, number of inventors, and position of inventors.

Usefulness for the evaluation. Publication and citation data are obviously relevant. These data are reasonably complete after 1982. Problems in using this database (e.g., common author names) can be minimized if supplementary information (curriculum vita) is available to accurately match specific publications to authors. It should again be noted that publication measures may be less suitable for evaluating research career outcomes for those working in business and industry, although this depends on the field and nature of the firm or corporation (Stephan and Levin, 1992). Obtaining information on patents as an outcome is also relevant. Patenting by academic institutions increased markedly between 1988 and 2003, rising from 800 to 3,200, and the growth in academic patents occurred primarily in the life sciences and biotechnology (National Science Board, 2006). Ding, Murray, and Stuart (2006) estimated that about 11.5 percent of life scientists in their sample (those in fields known for commercial applications and who were active publishers) were listed as inventors on one or more patents, although there were gender difference.

NIH Consolidated Grant and Application File

Brief description and relevance. Based on administrative records, this file is compiled by the NIH and includes records for all applications for NIH research grants and contracts. Each record includes information on several outcome-related variables, including the name of the principal investigator, his/her institutional affiliation, the type of application (e.g., investigator-initiated or program project award), the status of the application (funded or not funded), the NIH study section and institute to which the application was assigned, the monies awarded, and the dates of council review and award.

Usefulness for the evaluation. Given that one major goal of NIH career development awards is to enhance the ability of early-career investigators to successfully compete for subsequent NIH funding, this information on NIH applications and awards addresses a key outcome of any evaluation of the K awards. Although not error free, the data are reasonably complete and thus should include the NIH grant-related activity for K award recipients and their comparison group counterparts. It is the case, however, that the data only pertain to the principal investigators who submitted the applications. As such, individuals who are co-principal investigators or serve in other key roles on the research project would not be listed. Previous research suggests that these individuals account for a healthy proportion of investigators. For example, a sample survey of FY 1994 applicants to the NIH found that an estimated 55 percent of applicants were PIs on NIH grants, and 47 percent were PIs on one or more projects funded by other sources; among those who did not have NIH support, 24 percent were principal investigators on grants awarded by sponsors other than the NIH (Pion et al., 1999). In addition, nearly one-third of all applicants were involved in NIH-funded research projects (but were not the principal investigator), and 21 percent were serving in similar roles on other externally supported research. More recent surveys of NIH applicants reaffirm this need to collect data on non-NIH research support. Malik and Pion (2003) found that slightly over 90 percent of neuroscience applicants reported being a principal investigator on one or more externally funded research projects, with 64 percent serving as principal investigator on an NIH grant. Among behavioral and social sciences applicants, 90 percent reported that they were serving as a principal investigator (PI) or in another key role on one or more sponsored research projects. Nearly half (49 percent) indicated that they currently were a PI on one or more NIH research grants, and another 25 percent were working on NIH research projects in another capacity (Center for Scientific Review and Pion, in press).

NSF Master Database of Proposals and Awards

Brief description and relevance. This file is quite similar to the Consolidated Grant Applicant File maintained by the NIH in structure and contents. Records cover all applications for NSF research grants and contracts and include information on the name of the principal investigator, the type of application, and its status. NSF awards also can be found for specific principal investigators on the searchable web database that is available to the public on the NSF website.

Usefulness for the evaluation. Data on application to the NSF and receipt of NSF grants are relevant outcome variables. These data are reasonably complete and should potentially be available for all individuals selected to participate in the evaluation, particularly those with Ph.D.s. In the Pion (2001) evaluation of NRSA predoctoral training, which included Ph.D.s and M.D./Ph.D.s, approximately 10 percent applied for NSF grants. However, it is likely that very few M.D.s will have applied to the NSF, and thus, this data source is less useful for tracking grant-related outcomes for this group. Further, similar to the Consolidated Grant and Application File, the data are only collected on the principal investigators. Finally, matching names across the two databases presents some challenges, such as when two individuals have the same name or when different forms of a name are used.

5. SUGGESTED EVALUATION DESIGN

In this chapter, we present the recommended evaluation questions, the suggested target population, data collection approach, clearance requirements, timeline, and resource and cost estimates.

Evaluation Questions

Based on the logic model and the information obtained in the feasibility study, we recommend the following evaluation questions:

- What were the characteristics of the K-series awardees at the time they received their award?
- How do the characteristics of K-series awardees compare to applicants who never received a K-series award at the time of application?
- What types of activities were supported under the K-series awards?
- How do the research careers of K-series awardees compare to those of applicants who never received a K-series award?
- How does the ability to obtain additional grant support for research for K-series awardees compare to applicants who never received a K-series? Are there any differences between the various K awards?
- How does the productivity of K-series awardees compare to applicants who never received a K-series award?
- How does the level of commitment to the research field of K-series awardees compare to applicants who never received a K-series award?

Target Population

One activity of the feasibility study was to document the Career Development Awards from their inception in 1957. However, the evaluation will focus on awards made from 1975 to the present because we would expect most of the oldest participants to be senior researchers who have not yet retired and therefore more easily located.

The unit of analysis is the individual participating in the various K-series awards, although individuals who received certain awards will be excluded. Individuals receiving the Career Transition Award (K22) and Curriculum Development Award (K30) will not be included because separate evaluations are underway on these activities. Similarly, the Mentored Clinical Scientist Development Awards (K12) being evaluated by the National Institute of Child Health and Human Development, that is, the Building Interdisciplinary Research Careers in Women’s Health (BIRWCH), Women’s Reproductive Health Research (WRHR), and the Roadmap K12, will not be included in the overall K evaluation. The matched comparison group will consist of applicants for K awards who never received one.

The target population and the matched comparison group will generally be identified through NIH's CGAF. However, individuals who were supported under institutional awards will be identified by reviewing the records of individual ICs as described in a section below.

Recommended Data Collection Approach

In this section, we describe our recommended data collection approach, which is based on considerations of data quality and completeness, efficiency of the data collection approach, data collection cost, and burden to the data provider. The recommended approach is represented by a core study with two supplementary studies of areas that are not adequately covered in the core study. One supplementary study will enable some analyses to be conducted at the K award and IC levels, and the other will provide information about individuals supported by institutional awards.

Core Study

The suggested core study will examine the three cohorts discussed in chapter 4. It will look at all applicants in 1975 and 1976, 1985 and 1986, and 1995 and 1996. The total sample size will be 4,965, and a 50 percent response rate is anticipated. The core study involves the following three components:

- Extant data in the Consolidated Grant Application File (CGAF),
- New data collection in the form of a survey, and
- Bibliometric analysis.

Exhibit 5-1 summarizes the source of data for each variable in the core study. These variables were discussed in chapter 4. The data collection approach for each type of data is discussed below.

Exhibit 5-1.—Data sources for the indicator variables

Variable	Data source		
	CGAF (specific variable)	Survey	Bibliometrics
Context			
Institution type		X	
Institutional support		X	
Field of applicants		X	
Clinical specialty/subspecialty		X	
Inputs			
Award			
CAREER award	AWARD		
Award size (dollars)	TOTDOL		
Award duration	SUPPYR		
K-series	ACTIVITY		
ICs	IC		
Cohort	Derived (AWARD, FY)		
Mentorship characteristics		X	
Applicants			
Highest degree	DEGLVE1		
Gender	MFSEX		
Application quality	IRGSCORE		
Race/ethnicity	MFRAE		
Career stage (professional age)	PHDYR		
Activities			
Training		X	
Career development		X	
Research		X	
Outcomes			
Research career			
Employment		X	
Committee/board service		X	
Career role		X	
Quality			
Number of grants applied for	Derived	X	
Number of grants applied for and amount of money received	Derived	X	
Priority score	IRGSCORE for Rs		
Productivity			
Salary level		X	
Publications (count, authorship contribution)		X	X
Citations (count, rate)		X	X
Patents		X	
Honors/awards		X	
Professional relative value unit billed		X	
Commitment			
Time spend in field		X	
Commitment index		X	

Extant data in the CGAF. NIH's CGAF contains records for all applications for NIH research grants and contracts. Details about the file and its limitations were provided in previous chapters. The CGAF contains both indicator and outcome variables recommended for the evaluation. It will be used to examine demographic characteristics of the applicants and to determine what additional NIH awards were obtained by the K award applicants.

For the feasibility study, only CGAF data on the Career Development Awards were examined. However, in the full evaluation, additional files containing applications for research awards will be needed, such as those for the R01, R01 equivalents, and RPG grants.

New data collection in the form of a survey. Many of the outcomes of interest are not available in any extant data source. Therefore, a survey of all members of the three cohorts is recommended.

The variables indicated in Exhibit 5-1 will be included on the survey that will be developed with input from NIH. After the survey is developed, it will be pilot tested with nine individuals, both awardees and unsuccessful applicants. The survey will also need to go through the OMB clearance process as described below. While the survey is going through this process, searches will be conducted to find the current contact information and to determine that the correct person has been located.

We suggest the use of multiple approaches for obtaining the survey data, including web-based supplemented by hard copy. We anticipate that extensive follow-up will be needed and suggest an 8-month data collection period. We suggest that respondents also be asked to provide a copy of their curriculum vita along with their completed survey as a mechanism for verifying other data in the evaluation. For the population involved, we suggest a $100 incentive for survey completion.

Bibliometric analysis. We recommend contracting with Thomson's Web of Science to conduct a bibliometric search of publications and citations as well as patents of K-series applicants. For each cohort in the core study, we recommend a sample of 60 individuals (30 awardees and 30 unsuccessful applicants). They will be randomly sampled from the individuals who have completed the survey. The data elements recommended in this analysis are:

- Publication counts and position of authorship
  - Number of publications authored
  - Number of papers presented
  - Dates of publications
  - Authors' or editors' names
  - Name of each article, paper, or book
  - Type of publication
  - Number of authors
  - Position of authorship
- Citation analysis

- Total number of publications of applicants cited in other publications

- Patentometrics analysis
  - Number of patents registered
  - Dates of patents
  - Name of patent and number
  - Number of inventors
  - Position of inventor

Supplementary Study of all Career Development Award Applicants from 1975 to the Present Using the CGAF and NSF's Master Database of Proposals and Awards

While the recommended core study will enable comparisons between K awardees and unsuccessful applicants for the Career Development Awards program as a whole, the sample size will not be sufficient to conduct separate analyses of individual K awards or the awards made by a particular IC. However, NIH would like some analyses by K award and by IC.

The approach that is recommended to provide information by K-series and IC is the analysis of extant data in the CGAF for the all K-series applicants, both successful and unsuccessful, since 1975. The variables that will be examined are shown in the second column of Exhibit 5-1 and include demographic variables concerning the award and the applicants. Outcomes will consist of NIH grants applied for and obtained and the priority scores for the applications. This approach is the same as the component of the core study involving analysis of CGAF data. However, the difference is that this study will include the full population of K applicants from 1975 to the present, while the core study is limited to the three cohorts. The total number of awardees included in these analyses is about 15,000; the total number of applicants who never received a K award is about 13,500.

In addition, we recommend that a similar analysis be performed using NSF's Master Database of Proposals and Awards to determine which K-series applicants, both successful and unsuccessful, since 1975 had applied for and obtained NSF awards.

Supplementary Study of Institutional awards

In the feasibility study, we were unable to obtain complete information about individuals supported under an institutional award. Therefore, this component of the evaluation is somewhat more exploratory.

Each IC having an institutional award and make arrangements for collecting the names of individuals support under the grants and whatever demographic information is available about them would need to be contacted. In only a few instances do we anticipate that this information will be available electronically. Generally, we expect that review of hardcopy documents such as grant applications and progress reports for the names of individuals will be necessary. As complete a list as possible should be developed from the records available. We do not have a good understanding of the number of people involved. One IC has supported 134 individuals under institutional grants, but this number is much higher than the estimates provided by other ICs. With a few exceptions, the records provided to NIH are not thought to include the names of individuals who applied for support under an individual grant but were not selected. Therefore, the supplementary study of institutional awards cannot use the same design as that recommended for the core study.

The suggested approach for the study of individuals supported by institutional awards is to perform an analysis using the CGAF to determine if they have received other K awards or research grants from NIH. In addition, up to 400 will receive the survey used in the core sample. Responses of those receiving institutional awards will then be compared to individual awardees in the core study who received awards with a similar purpose during similar time frames to examine whether, on average, there is any difference between recipients of individual and institutional awards. For example, K12 awards have only been given since the mid-1980s, and K17s were only given in the mid-1990s. Therefore, comparisons of individuals supported by these institutional awards would not be made with the oldest cohort.

Clearance and Data File Requirements

Several types of clearances will be needed in the conduct of this study. They are described in this section.

At Westat and presumably at most research companies, each project goes through an initial review to determine if it is subject to an Institutional Review Board (IRB) review. We would anticipate a full committee review for this study because the data will include income and career information, so there is some risk to the professional reputations of the participants. A full IRB review generally takes 4-6 weeks.

For the feasibility study, all staff using the database needed to have NIH security clearances for Level 5: Public Trust – Moderate Risk. We assume the same security level would be needed in the full evaluation.

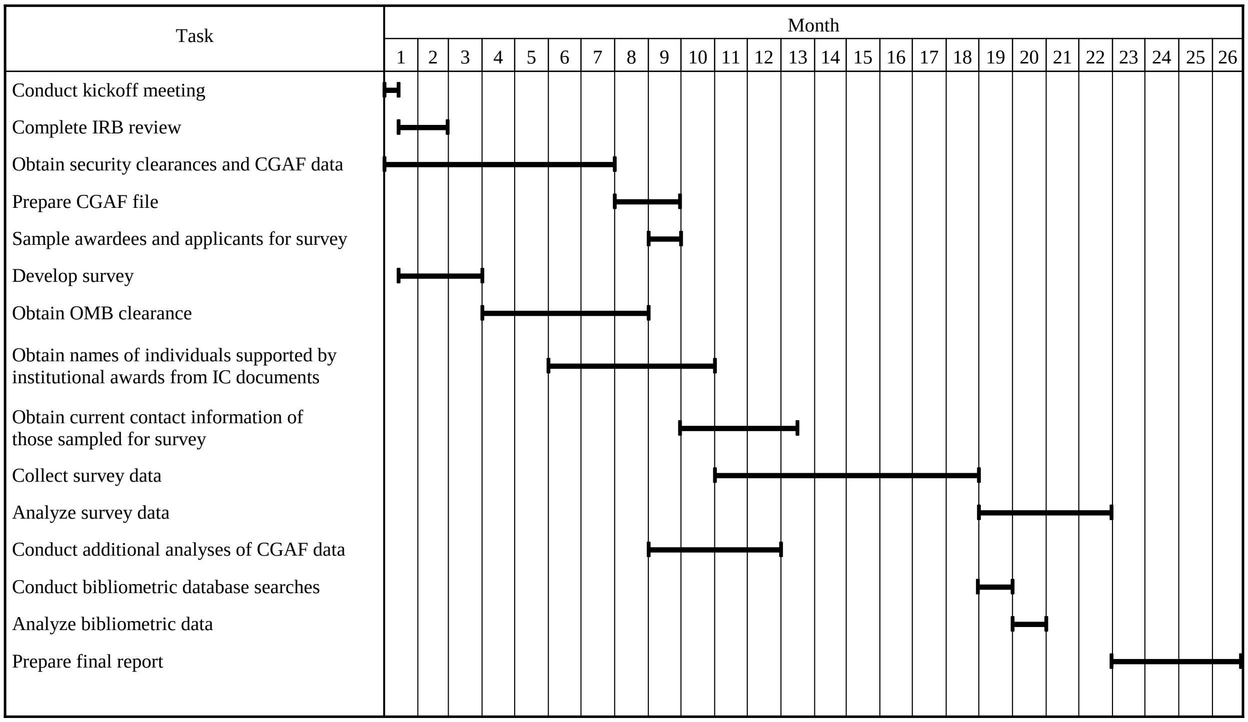
Once the security clearances have been obtained, NIH needs to prepare a data file and transmit it to the contractor. The mechanism used to provide the data for the feasibility study was a Virtual Private Network (VPN) account. In the feasibility study, it took a full 7 months to complete expedited security clearances and obtain the data file from NIH.

Since the evaluation will be a federal study and more than nine individuals will be asked the same questions in the survey, OMB clearance will be required. This involves the preparation of a clearance package and review by NIH and OMB. We recommend incorporating 5 months in the evaluation schedule for the clearance process.

Timeline

A recommended timeline for implementing the major study components is shown in Exhibit 5-2.

Exhibit 5-2.—Timeline for the evaluation of the NIH Career Development Awards



Level of Effort and Cost Estimates

A $100 incentive is suggested for both the awardees and the unsuccessful applicants who complete the survey. The awardees in the recommended sample received their grants at least 10 years ago, and some received them 30 years ago. This is quite a lot of time for individuals to still feel obligated to provide information. Unsuccessful applicants would not have any such obligation, but it would generally not be a good idea to provide incentives only to them.

Our estimated costs are as follows:

1. Core study plus the supplementary study of all applicants from 1975 to the present using the CGAF: $1,194,000 (excluding incentives);

2. Incentives for the core study: $250,000 (assuming a 50 percent response rate);

3. Supplementary study of institutional awards: $185,000 additional cost excluding incentives;

4. Incentives for the supplementary study of institutional awards: $20,000 (assuming a 50 percent response rate);

5. Additional cost to survey all applicants from 1975 to the present, who are not in the core sample: $2,404,000.

These costs do not include the analysis of the NSF Master Database of Proposals and Awards, which we would assume would be conducted under a separate contract.

REFERENCES

Arora, A., and Gambardella, A. (1998). *The impact of NSF support for basic research in economics*. Unpublished document.

Biddle, A.K., Carter, G.M., Uebersax, J.S., and Winker, J.D. (1988). *Research careers of recipients of the Research Scientist and the Research Scientist Development Awards*. Santa Monica, CA: Rand Corporation.

Bloom, H. S., Michalopoulos, C., Hill, C.J. and Lei, Y. (2002). *Can nonexperimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* MDRC working papers on research methodology. New York: MDRC.

Butler, P., Zhang, X., Wells, J., Havalchak, A., Liu, B., and Miyaoka, A. (2004). *Career Evaluation Studies: Final report of the feasibility study.* Rockville, MD: Westat.

Center for Scientific Review, and Pion, G. (in press). *Reorganization of peer review for the behavioral and social sciences at the Center for Scientific Review: A retrospective evaluation.* Bethesda, MD: National Institutes of Health.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cook, T.D., and Campbell, D.T. (1979). *Quasi-experiment: Design and analysis issues for field settings.* Chicago: Rand McNally.

Ding, W.W., Murray, F., and Stuart, T.E. (2006). Gender differences in patenting in the academic life sciences. *Science*, 313:665-667.

Eshelman, D.J., Sullivan, K.L., Parker, L., and Levin, D.C. (2000). The relationship of clinical and academic productivity in a university hospital radiology department. *American Journal of Radiology*, 174:27-31.

Fang, D., and Meyer, R.E. (2003). Effect of two Howard Hughes Medical Institute Research Training Programs for medical students on the likelihood of pursuing research careers. *Academic Medicine*, 78 (12): 1271-1280.

Goldsmith, S.S., Presley, J.B., and Cooley, E.A. (2002). *National Science Foundation Graduate Research Fellowship Program: Final evaluation report*. San Francisco, CA: WestEd.

Ingram, L., and Reis, P. (1994). *Validation study of the Survey of Earned Doctorates*. Washington, DC: Office of Scientific and Engineering Personnel, National Research Council.

Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards*. Thousand Oaks, CA: Corwin Press.

Khlat, M. (1997). Contribution of the case-control method to health program evaluation. In *Demographic Evaluation of Health Programs,* edited by M. Khlat, 99-109. Paris: CIRRED.

Kraemer, H.C., and Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage Publications.

Malik, K.F., and Pion, G. (2003). *Reorganization of neuroscience review at the Center for Scientific Review (CSR): A retrospective evaluation.* Bethesda, MD: National Institutes of Health.

Miles, M.B., and Huberman, A.M. (1994). *Qualitative data analysis,* 2nd Ed. Thousand Oaks, CA: Sage Publications.

Millsap, M.A., et al. (2000). *Faculty Early Career Development (CAREER) program external evaluation summary report.* Cambridge, MA: Abt Associates.

National Research Council. (2002). Scientific research in education. In *Committee on Scientific Principles for Education Research,* edited by R.J. Shavelson and L. Towne. Washington, DC: National Academy Press.

National Science Board. (2006). *Science and engineering indicators 2006*. Arlington, VA: NSF.

National Science Foundation. (1990). *NSF's Presidential Young Investigators program: A study of the first two "classes."* (Report 90-150). Arlington, VA: NSF.

National Science Foundation (2006). *Characteristics of doctoral scientists and engineers in the United States: 2003*. (NSF 06-320). Arlington, VA: NSF.

Pion, G. (2001). *The early career progress of NRSA predoctoral trainees and fellows*. Bethesda, MD: National Institutes of Health.

Pion, G.M., Schaffer, W., Seder, P., Marks, E., and Bouchard, J. (1999). *Customer satisfaction among applicants for NIH R01 and R29 grants.* Bethesda, MD: National Institutes of Health.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical linear models: Applicants and data analysis methods,* 2nd Ed. Thousand Oaks, CA: Sage Publications.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effect." *Biometrika*, 70(1):41-55.

Rosenblatt, R.A., and Andrilla, C.H. (2005). The impact of U.S. medical students' debt on their choice of primary careers: An analysis of data from the 2002 Medical School Graduation Questionnaire. *Academic Medicine*, 80(9): 815-9.

Rossi, P.H., and Freeman, H.E. (1993). *Evaluation: A systematic approach,* 5th Ed. Newbury Park, CA: Sage Publications.

Stephan, P., and Levin, S. (1992). *Striking the mother lode in science: The importance of age, place, and time*. New York: Oxford University Press.

Trochim, W.M.K. (2006). *The regression-discontinuity design*. Accessed at http://www.socialresearchmethods.net/kb/quasird.htm.

Trochim, W.M.K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage Publications.

What Works Clearinghouse. (2005). http://www.whatworks.ed.gov/. U.S. Department of Education, Institute of Education Sciences.

Yasumura, S. (1984). *The Research Career Award (K06): A 20-year perspective and an analysis of research productivity*. Unpublished report submitted to NIH.