Outcome Evaluation of the NCI Career Development (K) Awards Program

Final Report

February, 2012

Prepared for:

The National Cancer Institute, Center for Cancer Training

Prepared by:

Discovery Logic, a Thomson Reuters business

National Cancer Institute / Center for Cancer Training Career Development (K) Awards Program Evaluation Working Group

Jonathan Wiest, NCI
Julie Mason, NCI
Ming Lei, NCI
Jessica Faupel-Badger, NCI
Erika Ginsburg, NCI
Melissa Stick, NIDCD
Joanne Tornow, NSF
Rodney Ulane, NIH OD
Lawrence Solomon, NCI
James Corrigan, NCI

Study Directors
Joshua Schnell and Laurel Haak
Discovery Logic, a Thomson Reuters business

Analysts
Leo DiJoseph, Faye Liu, Yvette Seger, John Seibel
Discovery Logic, a Thomson Reuters business

Partial support for this study came from the NIH Evaluation Set-Aside Program, administered by the NIH Evaluation Office, reference number 10-2008-NCI. We are grateful to have had permission to use data from the National Science Foundation Doctorate Records File and the Association of American Medical Colleges Faculty Roster. A portion of the data used in this report was made available by the International Cancer Research Partners. We are also grateful to the American Association for Cancer Research (AACR), the American Society of Clinical Oncology (ASCO), and the Federation of American Societies for Experimental Biology (FASEB) for membership directory data.

Contents

List of Tables 7
List of Figures 9
Executive Summary 10
1.0 Introduction 13
1.1 Overview of the NCI K Program 13
1.2 Evaluation Scope 14
1.2.1 NCI K Awards Targeted to Early Career Investigators 14
1.2.2 NCI K Award Targeted to Mid-Career Investigators 16
1.2.3 NCI K Institutional Training Award 16
1.3 Logic Model 16
1.4 Data Sources 17
1.4.1 Extant Data Sources 18
1.4.2 Manual Data Mining 19
1.4.3 K12 Scholar Analysis 19
2.0. Characteristics of Program Applicants and Awardees 20
2.1 Overview 20
2.2 Variables 21
2.3 Demographics of Applicants and Awardees 21
2.3.1 NCI K Applications, Applicants and Awardees by Mechanism 21
2.3.2 Qualifying Degrees of NCI K Applicants and Awardees 22
2.3.3 PhD Field of Study and Medical Specialty of NCI K Applicants and Awardees 23
2.3.4 Gender of NCI K Applicants and Awardees 24
2.3.5 Distribution of Race and Ethnicity among NCI K Applicants and Awardees 26
2.3.6 Age of NCI K Applicants and Awardees 27
2.3.7 Years Since Degree of NCI K Applicants and Awardees 28
2.3.8 Prior NIH Support of NCI K Applicants and Awardees 30
2.3.9 Applications and Awards by Institution Type 32
2.4 Summary of Applicant and Awardee Characteristics 36
3.0 Selected Outcomes of Program Applicants and Awardees 38
3.1 Overview 38

3.2 Data Sources 38
3.3 Methodology 39
3.3.1 Deriving Comparison Cohorts 39
3.3.2 Publication Matching 39
3.3.3 Name Matching to Outcomes Datasets 40
3.3.4 Subsequent Outcomes Timeline 41
3.3.5 Subsequent NIH or NCI Grant Activity 41
3.3.6 Time to R01 Award Analysis 42
3.3.7 Publication Productivity Analysis 42
3.3.8 Outcomes of Individuals without Subsequent Appointment Information in IMPAC II 43
3.4 Validation of the Comparison Cohort 43
3.5 Subsequent NIH Research Funding 45
3.5.1 Subsequent NIH Funding High Water Mark 45
3.5.2 Subsequent NCI Funding High Water Mark 46
3.5.3 Time to R01 48
3.6 Subsequent Research Funding from Non-NIH Sources 49
3.7 Subsequent Publications 50
3.8 K25 Abstracts Pre- and Post-Award 54
3.9 Subsequent Career Appointments 55
3.9.1 Faculty Appointments 55
3.9.2 Medical Practice 57
3.10 Subsequent Participation in Clinical Research 57
3.11 U.S. Patent Applications and Issued Patents 58
3.12 Service on NIH Review Groups and Other Federal Advisory Committees 58
3.13 Membership in Professional Societies 60
3.14 Outcomes of Individuals without Subsequent Appointment Information 61
3.15 Summary of Applicant and Awardee Outcomes 63
4.0 Models 64
4.1 Regression Modeling of Applicant Characteristics and Outcomes 64
4.2 What Applicant Characteristics are Correlated with Receiving a K Award? 64
4.2.1 Applicant Race/Ethnicity and Degree Type 64
4.2.2 Applicant Gender 65

4.2.3 Age at Application 65
4.2.4 Applicant Prior Support 66
4.2.5 Applicant Institutional Characteristics 67
4.2.6 Submission of Multiple Applications 67
4.3 What is the Impact of K Program Participation on Subsequent Career Outcomes? 67
4.3.1 Subsequent NIH funding 67
4.3.2 Subsequent Publication Productivity and Impact 68
4.3.3 Subsequent Federal Advisory Committee Service 69
4.3.4 Subsequent Funded Research Career 70
4.3.5 Subsequent Engagement in the Broader Research Enterprise 70
4.4 What are the Mechanism-Specific Impacts of K Program Participation on Career Outcomes? 71
4.4.1 Subsequent NIH funding 71
4.4.2 Subsequent Publication Activity 72
4.4.3 Subsequent Federal Advisory Committee Service 72
4.4.4 Subsequent Funded Research Career 73
4.4.5 Subsequent Engagement in the Biomedical Research Enterprise 74
4.5 Summary of Variables Important for K Success and the Impact of K Programs on Participant Careers 76
5.0 Policy Implications 77
6.0 Appendices 79
6.1 NCI K Mechanisms 79
6.2 Study Variables and Data Sources 81
6.3 Classification and Determination of Qualifying Degree 82
6.4 Resubmission of NCI K Applications 82
6.5 Determination of Age at Application and Years Since Degree 85
6.6 K07 Years Since Degree Analysis (Full Cohort) 86
6.7 Prior Support Categories 88
6.8 Construction of the Funding Bubble and Comparison Cohorts 89
6.9 U.S. Patent Applications and Issued Patents 91
6.10 Linear and Logit Models 92
6.10.1 Model Summary 92
6.10.2 Variable Correlations 94

6.10.3 Model Results 94
6.10.4 Model Data Dictionary 104
6.10.4.1 Outcome (Y) Variables 105
6.10.4.2 Funding Variable – Used as Both an Input and an Outcome Variable 108
6.10.4.3 Input Variables (Independent Variables, X Variables) 108
6.11 Statistical Methods Used 113
7.0 Supplement 115
7.1 Defining the NCI K Cohort 115
7.2 Freshness of Data Sources 116
7.3 PhD Fields of Study (FOS) of NCI K Applicants Matched to the DRF 117
7.3.1 Combined Top PhD FOS of NCI K Applicants Matched to DRF (in alphabetical order) 117
7.3.2 Top PhD FOS of Applicants, by NCI K Mechanism (in alphabetical order) 120
7.4 NIH High Water Mark Analysis by NCI K Mechanism 125
7.5 NCI High Water Mark Analysis, by NCI K Mechanism 129
7.6 Model Results for Publication Metrics including Applicants who did not Publish 133

List of Tables

Table 1. NCI K mechanisms and cohorts included in the study. 14
Table 2. General statistics of NCI K mechanisms included in this evaluation. 20
Table 3. Distinct new application and distinct individual counts for the mechanisms evaluated (top panel) and distinct individual counts (bottom panel) for other combinations of mechanisms used in the study. 21
Table 4. NCI K applicants and awardees, by primary mechanism. 22
Table 5. Top three Fields of Study for NCI K applicants with a PhD, by mechanism and award status. 24
Table 6. Overall gender distribution of NCI K applicants, excluding K04s and K12 scholars. 24
Table 7. Overall race and ethnicity distribution of NCI K applicants. 27
Table 8. Prior NIH support of NCI K applicants, by mechanism and award status. 31
Table 9. Distribution of K applications by NCI Cancer Center designation. 33
Table 10. Distribution of NCI K applications and awards, annual cumulative NCI funding level of institution. 35
Table 11. Subsequent grant activity high water mark categories. 42
Table 12. Demographic composition of the comparison and full cohorts. 45
Table 13. Subsequent non-NIH research funding, comparison cohort. 50
Table 14. Top 10 Web of Science Journal Subject Categories, awardees in the full cohort. 51
Table 15. Top 10 journals, awardees in the full cohort. 51
Table 16. Analysis of publications authored by NCI K applicants, by mechanism and funding status. 52
Table 17. Subsequent faculty appointments, full cohort. 56
Table 18. Subsequent private practice involvement, full cohort. 57
Table 19. NCI K08 and K23 applicant and K12 scholar participation in clinical trials, full cohort. 58
Table 20. Service of NCI K applicants on NIH review groups, full cohort. 58
Table 21. Participation of NCI K applicants on Federal advisory committees, full cohort. 59
Table 22. Participation of NCI K applicants as members of select professional societies, full cohort. 60
Table 23. Distribution of applicants from the full cohort included in the manual search, by mechanism and funding status. 61
Table 24. Detailed breakdown of current positions of applicants included in the manual search, grouped by broad position category, funding status, and refined position title. 62
Table 25. Contribution of applicant gender to K award success. 65
Table 26. Contribution of prior training and research support to K award success. 67
Table 27. Contribution of applicant institution characteristics to K award success. 67
Table 28. Contribution of K program participation on subsequent NIH and NCI grant activity. 68
Table 29. Correlation between participation and publication productivity and impact. 69
Table 30. Correlation between participation and subsequent service on a Federal advisory committee, full and comparison cohort. 69
Table 31. Correlation between participation and having a funded research career, full and comparison cohort. 70

Table 32. Correlation between participation and being engaged in the larger biomedical research enterprise, full and comparison cohort. 71
Table 33. Correlation between K participation and subsequent NIH and NCI awards by mechanism, full and comparison cohort. 72
Table 34. Correlation between K participation and subsequent Federal advisory committee service by mechanism, full and comparison cohort. 73
Table 35. Correlation between K participation and subsequent funded research career by mechanism, full and comparison cohort. 73
Table 36. Correlation between K funding and subsequent funded research career by qualifying degree, full and comparison cohort. 74
Table 37. Correlation between participation in particular K mechanisms and subsequent engagement in the larger biomedical research enterprise, full and comparison cohort. 75

List of Figures

Figure 1. Logic model for evaluation of the NCI K programs. 17
Figure 2. Schematic of the data sources used for the evaluation. 18
Figure 3. Qualifying degree of NCI K applicants, by mechanism and award status. 23
Figure 4. Gender distribution of NCI K applicants, by mechanism and award status. 26
Figure 5. Age distribution of NCI K applicants, by mechanism and award status. 28
Figure 6. Years since degree distribution of NCI K applicants, by mechanism and award status. 29
Figure 7. Comparison group methodology: The "funding bubble." 39
Figure 8. Attainment of NIH high water mark categories in the full and comparison cohorts, by funding status 46
Figure 9. Attainment of NCI high water mark categories in the full and comparison cohort, by funding status 47
Figure 10. Time to first NIH R01 following K application by mechanism and K funding status, full cohort 48
Figure 11. Publications per person per year for all individuals in the full cohort, for 2-year time slices. .. 53
Figure 12. Publication impact analysis per person in the full cohort, for 2-year time slices. 54
Figure 13. Estimated funding probability for applicants by age at application. 66

Executive Summary

The National Cancer Institute (NCI) career development (K) awards program includes a broad range of funding mechanisms, providing scientists with support for protected time to further develop their cancer research careers, transition to independence, expand their existing research programs, or mentor junior investigators. The NCI K awards program is administered by the Cancer Training Branch (CTB), the extramural branch of the Center for Cancer Training (CCT), and is a significant component of NCI's training effort. In fiscal year 2011, the CTB K program supported 365 awards at an approximate cost of $65M.

CTB evaluated the K program to determine program impact and whether any policy or programmatic changes were warranted. Nine of the NCI K mechanisms were included in the evaluation, spanning the timeframe of 1970 through 2008. The objectives of the evaluation were to:

(1) Determine the demographics and characteristics of NCI K award program awardees and comparison cohorts,
(2) Identify the post-award outcomes of NCI K awardees and comparison cohorts, and
(3) Determine the impact of NCI K awards on participant career outcomes.

Key Findings

Demographics and Characteristics of K Applicants

- During the time period of the evaluation, over 29% of NCI's K applications were awarded, and individual applicant funding success was just over 38%.
- Although men comprised 57.5% of the total K applicants, men and women were equally successful in receipt of K awards.
- An applicant's race/ethnicity did not influence K award rate or likelihood of resubmission.
- Degree field and clinical specialty of the K applicants matched program focus.
- Prior NRSA training support was common among NCI K applicants and increased an applicant's odds of K award receipt.
- More than 60% of the K awards went to 14% of the applicant institutions, representing the top bracket of NCI funding. These institutions tended to be affiliated with NCI-designated Cancer Centers.

Subsequent Career Outcomes and Impact

- Participation in the K awards program promoted increased likelihood of subsequent NCI and NIH funding with no effect on time to first R01.
- K awardees had increased subsequent publication productivity, and publication impact.
- K awardees had improved odds of having an independent, funded research career.

- For those K awardees that chose not to pursue funded research careers, participation in the K program positively influenced their odds of remaining engaged in the biomedical research enterprise.
- Among the various K mechanisms analyzed, there were slight differences in various outcome metrics, which are discussed in the body of the report. Overall, participation in any of the K mechanisms had measurable impact on the careers of awardees, not only as gauged by traditional factors such as research grant funding, clinical trials, and publications, but generally in terms of activities that signal engagement within the broader biomedical research enterprise such as Federal advisory committee service and membership in scientifically-oriented professional societies.

Conclusions and Recommendations

- To increase participation of women in cancer research at this stage, efforts could focus on attracting more women applicants.
- The NCI has a strong commitment of support for underrepresented minorities through the Center to Reduce Cancer Health Disparities and the Diversity Training Branch, whose K applications were not evaluated in the current report. A more complete picture of the contribution of race/ethnicity to K award success and future career outcomes should include analysis of the K awards made by NCI's Diversity Training Branch and deeper investigation into whether the gaps in the race/ethnicity data could be filled.
- The growing need for multi- and inter-disciplinary training may not be fully realized with NCI's multitude of specialized K mechanisms. Collapsing several mechanisms to break down the artificial barriers created by the discrete mechanisms could offer K applicants more freedom in designing their research and training experiences and would reflect NCI's message of encouraging interdisciplinary research. For instance, since K08 and K23 applicants come from similar disciplines but the K08s focus on basic research while K23s perform clinical research, using one mechanism that unifies the K08 and K23 applicant pools and invites applications across the cancer research disciplines could be the first step in facilitating seamless opportunities among basic and patient-oriented physician scientists.
- Prior NIH-supported NRSA training is an important factor in K award success and may be one of the first steps toward developing an independent cancer research career. Whether differences exist in future career success of trainees based on the type of prior training experience and institutional context is a question that merits further investigation.
- The majority of K awards were made to individuals at a small number of institutions with the highest amount of research funding, and these institutions were affiliated with NCI-designated Cancer Centers. This finding is also reflected in NCI's R01 pool. Programs such as the K22, which garners applicant interest from institutions outside of this sphere, should be maintained to attract and develop a workforce that can offer research resources to more communities.
- The length of this initial K training period should not deter applicants from pursuit of a career development award, given the similar time to R01 for both K awardees and non-awardees.

- A greater percentage of all K awardees published as authors compared to non-awardees, and awardees had higher publication productivity and impact. However, when only examining the subset of awardees and non-awardees who published, the numbers of publications per author per year were similar. A more in-depth examination of publication activity is required to fully understand the effect of K program participation on publications.
- In general, K awardees had comparable successful outcomes across the multiple K mechanisms explored in this evaluation. Most often, K awardees' outcomes showed similar trends (eg. higher median publications per person than non-awardees), and any differences lie in the magnitude of the change or in the ability to reach statistical significance, which was also influenced by individual mechanism size (applicant pool). Any other differences could be attributed to features of the target population each mechanism attracts (eg. applicants with MDs compared to PhDs).
- This study demonstrates the overall value of participation in NCI's K program and the positive impact of the program on awardees' future career successes both in the progression of funded research careers and in participation in careers to advance the biomedical research enterprise.

1.0 Introduction

1.1 Overview of the NCI K Program

The National Cancer Institute (NCI) career development (K) awards program aims to provide scientists from a variety of scientific backgrounds and at different stages of career development with support for protected time to further develop their cancer research careers, transition to independence, expand their research programs, or mentor junior investigators. Some of the K programs are designed to transition post-doctoral researchers from mentored research to independent investigator positions. Several programs provide clinical investigators with an opportunity to pursue mentored training in biomedical research, while others provide established/midcareer investigators with an opportunity to transition between research fields (e.g., from engineering to biomedical research), pursue patient-oriented translational research projects or pursue cancer prevention, control, behavioral, and population sciences research while mentoring junior investigators.

The K awards program at NCI is administered by the extramural branch of the Center for Cancer Training (CCT), the Cancer Training Branch (CTB), which currently sponsors thirteen 1-5 year training/career development award mechanisms for individuals working at universities and affiliated institutions, four 1-6 year National Service Research Award (NRSA) fellowships and four 1-3 year institutional training awards for fellows working at universities and affiliated institutions.[1] The NCI K program is a significant component of NCI's training effort, representing approximately 40% of the CTB training and career development budget, and close to 50% of the number of CTB awards. In fiscal year 2011, the NCI CTB K program supported 365 awards at an approximate cost of $65M.[2] There has not been a formal, systematic evaluation of the career outcomes of K grantees and scholars in the history of the program at NCI. A careful examination of program goals, the applicants the program attracts, and participants' subsequent outcomes is key to assessing program impact and value. In August 2010, the NCI CCT asked Discovery Logic to conduct an evaluation of the NCI K awards program based on the findings of a 2009 feasibility analysis[3]. The overall objectives of this evaluation were to:

1) Determine the characteristics and demographics of NCI K award program awardees and comparison cohorts,
2) Identify the post-award outcomes of NCI K awardees and comparison cohorts, and
3) Determine the impact of NCI K awards on participant career outcomes.

Across the K mechanisms included in this evaluation, the NCI CTB has received a total of 5,445 applications and made 1,609 awards since 1970. The 5,445 figure represents the total number of

[1] Information on NCI CTB awards available at: http://www.cancer.gov/researchandfunding/cancertraining/outsidenci/awardtype(last accessed October 21, 2011).
[2] Number of awards and approximate cost obtained using the NIH Information for Management, Planning, Analysis, and Coordination database (IMPAC II). Figures include three mechanisms (K05, K24, and K99) not included in the current evaluation. Total FY 2011 funding for mechanisms included in this evaluation was $51.7 million for 274 awards.
[3] National Cancer Institute Center for Cancer Training. (2009). "Assessing the Feasibility of Conducting an Evaluation of the NCI Career Development (K) Awards Program."

applications to all NCI CTB K programs since 1970. Individual applicants may have submitted multiple applications to one or more NCI K programs, but only one application per fiscal year is counted in this overall total. This total also includes applications that were withdrawn, as there is no code that distinguishes these applications from other unfunded applications. The 1,609 figure represents the total number of unique awards across all NCI K programs. A total of 188 K12 Scholars were included in this evaluation. **Table 1** lists the mechanisms and year ranges included in this evaluation. To support analysis of demographic and outcomes data, additional rules (described in Section 2.1) were applied to define the study cohort.

Mechanism	Years Offered	Years to be Analyzed
K01	1997 - 2011	1997 - 2007
K04	1970 - 1996	1970 - 1996
K07	1980 - present	1980 - 2008
K08	1984 - present	1984 - 2008
K11	1987 - 1996	1987 - 1996
K12	1992 - present	1992 - 2008
K22	1998 - present	1998 - 2008
K23	1999 - present	1999 - 2008
K25	2000 - present	2000 - 2008

Table 1. NCI K mechanisms and cohorts included in the study.

1.2 Evaluation Scope

The goal of this evaluation was to define the population of individuals who applied for NCI K awards and determine the impact that receipt of an award had on pursuit of a biomedical research career and more generally on contributions to the research enterprise. This evaluation explored six of the K award mechanisms currently offered by the NCI CTB: K07, K08, K12, K22, K23, and K25; as well as three CTB-administered K mechanisms that have been phased out (K04 and K11) or are in the process of being phased out (K01). Mechanisms targeted to more senior investigators, specifically K05, K18, K24; newer mechanisms including the K99; and those administered by NCI's Diversity Training Branch in the Center to Reduce Cancer Health Disparities to increase minority representation (a subset of K01, K08, K22, and K23 awards), were not included in this evaluation. Descriptions of the K mechanisms included are provided below and in **Appendix 6.1**.

1.2.1 NCI K Awards Targeted to Early Career Investigators

- **K01.** The goal of the NCI Howard Temin Award (K01) is to bridge the transition from a mentored research environment to an independent career in basic cancer research. The K01 provides awardees with up to five years of non-renewable support, allowing them the opportunity to gain additional skills and knowledge in human cancer research in a mentored environment culminating in a transition to an independent research/junior faculty position. In July 2006, NCI began to phase out its K01 program, replacing it with the K99/R00 Howard Temin Pathway to Independence Award mechanism. The K01 mechanism continues to be supported by the NCI Diversity Training Branch in the Center to Reduce Cancer Health Disparities.

- **K04.** The Research Career Development Award (K04) provided up to five years of non-renewable support and "protected time" for newly independent scientists to further develop their research programs. The NCI awarded K04 grants from 1970 through 1996. The K04 was phased out as part of a larger restructuring of the NIH career development awards program.

- **K07.** The Cancer Prevention, Control, Behavioral, and Population Sciences Career Development Award (K07) provides between three and five years of non-renewable support for early-career investigators who have made a commitment to focus their research on cancer prevention, control, behavioral, and population sciences. K07 candidates are typically post-doctoral fellows or non-tenured junior faculty, and the award provides an opportunity for specialized didactic study and mentored research to support the transition to independent research careers.

- **K08.** The Mentored Clinical Scientist Development Award (K08) provides up to five years of non-renewable support to individuals with a clinical doctoral degree to provide them with an opportunity to receive mentored training in laboratory-based biomedical or behavioral research. K08 support combines didactic study with methodological and theoretical laboratory training opportunities to develop the skills necessary to pursue independent clinical and/or translational research.

- **K11.** The Physician Scientist Award (K11) provided long-term basic, clinical, or behavioral research training to individuals with clinical science doctorates (MD, DDS, DVM, DO or equivalent) with two to seven years of clinical training at the postdoctoral level, allowing them to transition to independent biomedical investigator positions. The non-renewable award was administered in two phases: Phase I provided an opportunity for didactic study and laboratory experiences, while Phase II allowed recipients to pursue an intensive research project. The NCI awarded K11 grants from 1987 through 1996. This award was phased out to decrease the redundancy of awards targeted to clinician-scientists, with the K08 replacing the K11, K15 and K20 awards.

- **K22.** The NCI Transition Career Development Award (K22) provides up to three years of non-renewable support and "protected time" for newly independent investigators to develop and receive support for their initial cancer research programs, and to facilitate the transition from mentored to independent research. Applicants can be clinicians pursuing basic science careers; clinicians pursuing careers in patient-oriented research; or individuals pursuing careers in cancer prevention, control and population sciences. A unique feature of the K22 is that postdoctoral fellows at Federal agencies are eligible candidates. The K22 mechanism is also different from other K mechanisms in that applicants are not required to have a sponsoring institution/junior faculty appointment at the time the application is submitted.

- **K23.** The Mentored Patient-Oriented Research Career Development Award (K23) provides up to five years of non-renewable support for combined didactic study and mentored research to

individuals with a clinical doctoral degree, allowing awardees to acquire the methodological and theoretical research skills needed to pursue independent clinical and patient-oriented research.

1.2.2 NCI K Award Targeted to Mid-Career Investigators

- **K25.** The Mentored Quantitative Research Career Development Award (K25) provides three to five years of non-renewable career development support to investigators with backgrounds in quantitative and engineering sciences (e.g., outside classical biomedical research fields) to pursue basic or clinical research in the biomedical or behavioral sciences. Applicants can range from post-doctoral fellows to senior faculty members.

1.2.3 NCI K Institutional Training Award

- **K12.** The Paul Calabresi Award for Clinical Oncology (K12) is a multi- and trans-disciplinary institutional training award that supports the research career development of clinicians and basic science researchers to pursue patient-oriented research, translational research, and clinical studies focused on the development of cancer therapeutics. The K12 is awarded at the institutional level, rather than to an individual, and provides up to five years of support for clinical and research scholars.

1.3 Logic Model

A logic model was developed to identify, define and categorize critical components to be measured and analyzed in this evaluation (**Figure 1**). The logic model is divided into five categories:

- **INPUTS** include the features that define applicants to the NCI K programs, as well as the features of the individual mechanisms.
- **ACTIVITIES** include the actions that a funded researcher would take to further their research training and career plans.
- **CONTEXT** refers to specific features of the past and present environment in which the program participants are functioning.
- **OUTCOMES** include features that might be a direct result of participation in an NCI K program.
- **IMPACT** tracks the systemic effects of program participation at the individual level.

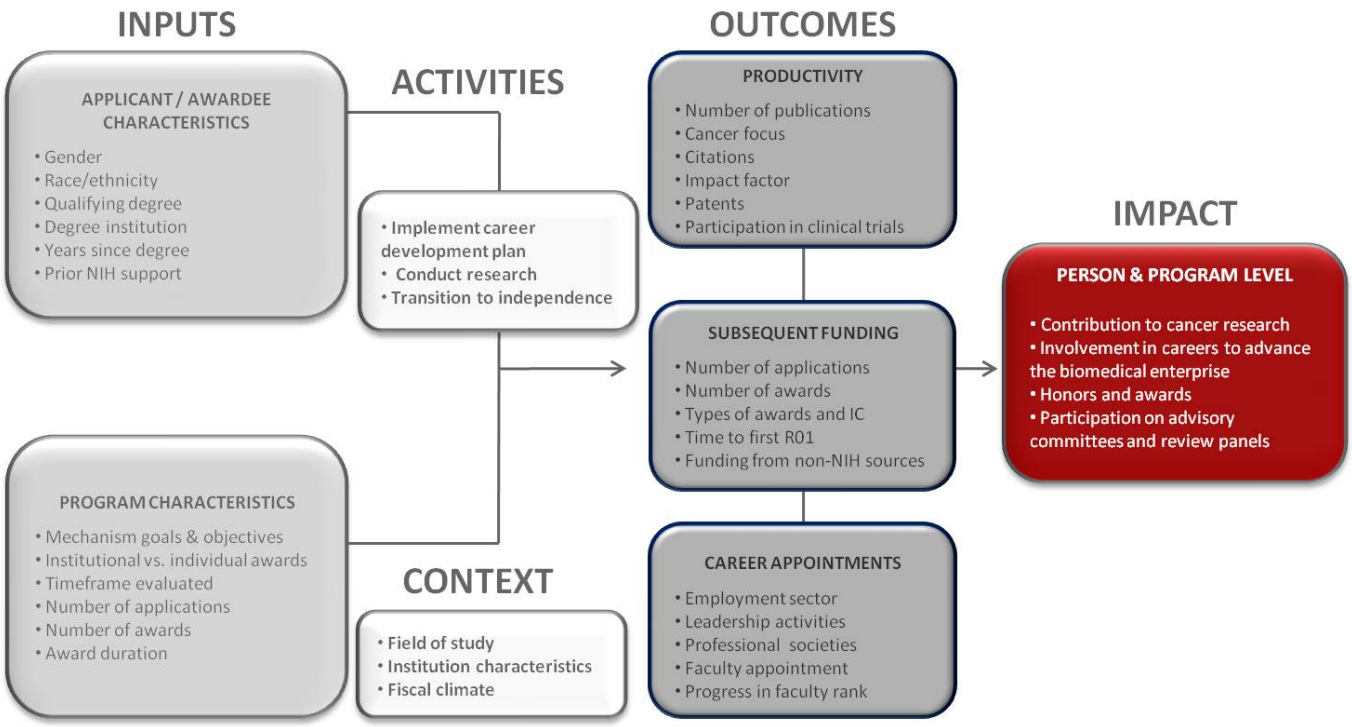


Figure 1. Logic model for evaluation of the NCI K programs.

1.4 Data Sources

Several databases were used to support this evaluation (**Figure 2**). The NIH grants database known as Information for Management, Planning, Analysis, and Coordination (IMPAC II) was the primary data source through which information about NCI K applicants and awardees was obtained, including basic demographic information and data regarding prior and subsequent NIH grant applications. The Association of American Medical Colleges Faculty Roster (AAMC Faculty Roster) and the National Science Foundation's Doctorate Records File (DRF) were used to supplement demographic information from IMPAC II, as well as provide additional fields to augment these analyses.

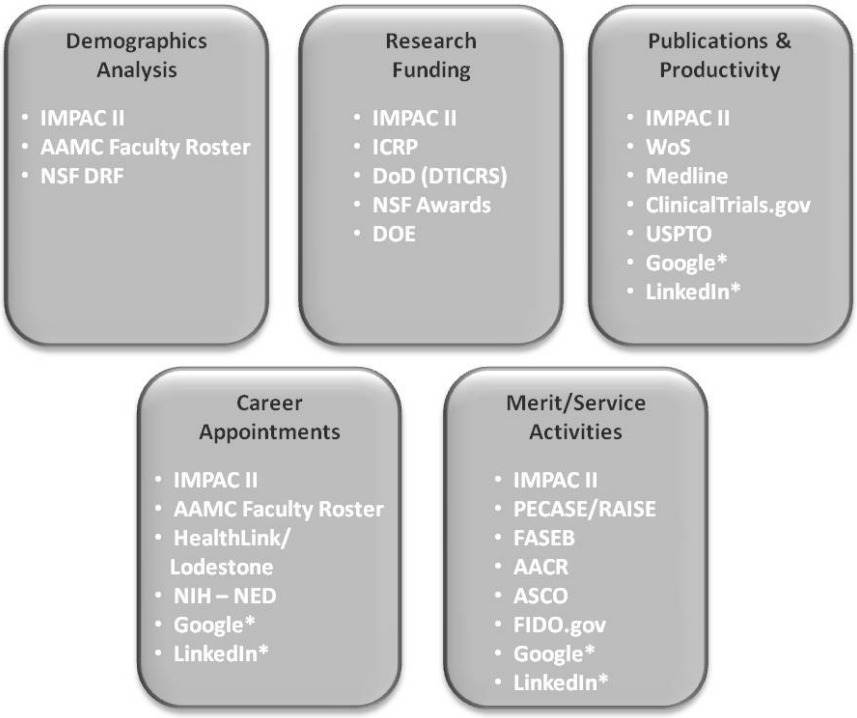


Figure 2. Schematic of the data sources used for the evaluation.
Items marked with an asterisk indicate data sources only used for obtaining outcomes information for a subset of NCI K applicants (~100 individuals) for whom no subsequent grant application or award information was available in IMPAC II. Acronyms: AAMC (Association of American Medical Colleges); NSF DRF (National Science Foundation Doctorate Records File); ICRP (International Cancer Research Portfolio); DoD-DTICRS (Department of Defense-Defense Technical Information Center Research Summaries; NSF (National Science Foundation); DOE (Department of Energy); WoS (Thomson Reuters Web of Science); USPTO (U.S. Patent and Trademark Office); NIH-NED (NIH Employee Directory); PECASE/RAISE (Presidential Early Career Award for Scientists and Engineers); FASEB (Federation of American Societies for Experimental Biology); AACR (American Association for Cancer Research); ASCO (American Society of Clinical Oncology); and FIDO.gov (Federal Interagency Databases Online).

1.4.1 Extant Data Sources

Other databases were used to obtain career outcome information for NCI K applicants and awardees. Non-NIH databases utilized as resources for research funding subsequent to an individual's K application or award included the International Cancer Research Portfolio (ICRP) database, National Science Foundation (NSF) FastLane database, and the Department of Energy (DOE) grants database. The National Library of Medicine (NLM) MEDLINE database and Thomson Reuters Web of Science (WoS) were used for matching publications to K awardees and applicants and collecting bibliometric data. The HealthLink/Lodestone database was used to track NCI K applicants and awardees in private medical practice, while clinicaltrials.gov was used to track the involvement of key personnel in clinical trials. The NIH Employee Directory (NED) was used to match NCI K applicants to NIH staff. Information on federal advisory committee service was obtained from the Government Accountability Office's (GAO) Federal Interagency Databases Online (FIDO.gov). Membership in select professional societies, American Association for Cancer Research (AACR), American Society of Clinical Oncology (ASCO), and the Federation of American Societies for Experimental Biology (FASEB), was determined through name matches to society member databases.

1.4.2 Manual Data Mining

To assess career outcomes of a selected subset of individuals for whom there were no subsequent records in IMPAC II, we conducted manual searches of the professional networking website LinkedIn and searched for online curricula vitae (CVs) using the Google search engine. This manual review was limited to a sample of 105 NCI K applicants (53 awardees and 52 non-awardees) across the K mechanisms, excluding K04 and K12. Web searches were conducted using the Google search engine and the applicant name as recorded in IMPAC II. Depending on the number of hits returned, the search string was modified to include/exclude middle name or initial, last known degree, or last known institution name. In some cases, the last known institution was used to help refine search results. An individual was considered "found" if information on a recent (i.e., 2009 and onward) website or version of an online document included a reference to the NCI K award. Individuals with common names or for whom information could not otherwise be retrieved were scored as "unknown."

1.4.3 K12 Scholar Analysis

Information regarding the Principal Investigators (PIs) of K12 grants is readily accessible through IMPAC II; these individuals represent the "mentors" for the K12 scholars. Data on the K12 scholars was provided by NCI CCT from roster lists maintained by program staff. Of the 373 K12 scholar names, 188 were matched to IMPAC II records, placed in a separate K12 scholar database, and included in several of the demographic and outcome analyses in this evaluation. K12 scholars were not matched to the DRF or AAMC Faculty Roster. Since the K12 mechanism is an institutional award, there is no "unfunded" comparison group for the scholars.

2.0. Characteristics of Program Applicants and Awardees

2.1 Overview

To understand who is applying to the NCI K award program, we analyzed demographic characteristics using data from IMPAC II supplemented with information from the AAMC and DRF data files: qualifying degree type; field of study or medical specialty; gender; race and ethnicity; age; years since qualifying degree; and prior NIH support. We conducted a general analysis across all NCI K mechanisms, and also evaluated each mechanism individually by fiscal year.

Across the NCI CTB K mechanisms included in this evaluation, there were a total of 5,445 applications, of which 1,609 were awarded since 1970[4]. Three mechanisms (K04, K07, and K08) have been receiving applications for more than 25 years. Most programs have approximately 10 years of applicant data. Specific rules were applied to the K12 cohort (see Section 1.4.3). Initial analysis of the demographic and outcomes data of the K04 cohort indicated that they were further in their careers than applicants to the other mechanisms being evaluated, and therefore, they were excluded from many of the analyses in this study.

Table 2 provides a summary of applicant and award volume and award rates for the NCI K mechanisms in this study.

Mechanism	Years Analyzed	Total Applications	Total Awards	Percent Awarded
K01	1997 - 2007	624	153	24.52%
K04	1970 - 1996	1,249	341	27.30%
K07	1980 - 2008	825	274	33.21%
K08	1984 - 2008	1,638	515	31.44%
K11	1987 - 1996	216	86	39.81%
K12	1992 - 2008	115	44	38.26%
K22	1998 - 2008	324	73	22.53%
K23	1999 - 2008	366	98	26.78%
K25	2000 - 2008	88	25	28.41%
TOTAL	**1970 - 2008**	**5,445**	**1,609**	**29.55%**

Table 2. General statistics of NCI K mechanisms included in this evaluation.
Individual applicants might have submitted multiple applications to one or more NCI K programs, but only one application per fiscal year is counted in the total. Total awarded represents the number of unique awards across all NCI K programs.

Additional rules have been applied to the cohort. For instance, information collected from IMPAC II for K12s pertains to the mentor PIs on a grant, rather than the actual individuals (scholars) receiving training from the award. Therefore, for this evaluation, an NCI-maintained list of K12 scholars funded during the period of 1992 – 2008 was used. To learn more about the demographics and outcomes of the K12

[4] Individual applicants might have submitted multiple applications to one or more NCI K program, but only one application per fiscal year is counted in the overall total. This total also includes applications that were withdrawn, as there is no code that distinguishes these applications from other unfunded applications. The 1,609 figure represents the total number of unique awards across all NCI K programs.

scholar population, this list of names was compared to IMPAC II and 188 were matched. Unless otherwise noted, when available, demographics or outcomes are reported for the K12 subgroup.

Since different combinations of applicants and awardees are used throughout this evaluation, **Table 3** shows the distinct application and individual counts and totals for the most common combinations of mechanisms (top panel) and the distinct individual counts and totals for other mechanism combinations used (bottom panel).

Distinct New (Type 1) Application and Individual Counts and Totals				
Mechanisms Included	**Applications**	**Awards**	**Applicants**	**Awardees**
K01, K07, K08, K11, K22, K23, K25	4,081	1,224	2,893	1,206
K04	1,249	341	1,089	341
K12 (Mentor PIs)	115	44	87	43
K12 (Scholars) - List from CCT			373	373
K12 (Scholars) - Matched to IMPAC II [subset of 373 above]			188	188
Grand Total (Full Evaluation)	**5,445**	**1,609**		

Distinct Individual Counts for Other Mechanism Combinations	Applicants	Awardees
K01, K07, K08, K11, K22, K23, K25 + K04	3,982	1,547
K01, K07, K08, K11, K22, K23, K25 + K12 Scholars	3,266	1,579
K01, K07, K08, K11, K22, K23, K25 + K12 Matched Scholars	3,081	1,394
K01, K07, K08, K11, K22, K23, K25 + K04 + K12 Scholars	4,355	1,920
K01, K07, K08, K11, K22, K23, K25 + K04 + K12 Matched Scholars	4,170	1,735

Table 3. Distinct new application and distinct individual counts for the mechanisms evaluated (top panel) and distinct individual counts (bottom panel) for other combinations of mechanisms used in the study.
For application counts, only those classified as type 1 or "competing" were included. Non-competitive (e.g., type 5) and competitive renewals (e.g., type 2) were excluded.

2.2 Variables

The following parameters were used to analyze applicant and awardee characteristics: (1) qualifying degree type, (2) field of study or medical specialty, (3) gender, (4) race and ethnicity, (5) age, (6) years since qualifying degree, and (7) prior NIH research support. Single and cross-parameter analyses were performed for each mechanism. Representative data are shown in this report. Data sources used to derive variables are described in **Appendix 6.2**.

2.3 Demographics of Applicants and Awardees

2.3.1 NCI K Applications, Applicants and Awardees by Mechanism

Some individuals applied to multiple K mechanisms, and, under some circumstances, received more than one K award. The majority of applicants (3,863) applied to only one NCI K mechanism, but there were 119 applicants that submitted applications and/or received awards in multiple K mechanisms. To control for this, we implemented a "primary K mechanism rule" that places each applicant in a single K mechanism for analysis of demographics and career outcomes (**Table 4**). The "primary K mechanism"

was defined as the mechanism by which an applicant received their first award or, for non-awardees, the mechanism through which the last application was submitted.

A total of 18 individuals received awards in two K mechanisms, with the majority receiving a combination of a K07 or K08, followed by a K22 (14 individuals in total)[5]. For the purpose of the evaluation, the primary mechanism of these individuals was considered that of the first award, as the subsequent K award could be considered an outcome of receiving the first award.

For K12s, the 188 "awardees" represent the matched scholars (see Section 1.4.3).

Mechanism	Applicants (Primary Mechanism)	Awardees	Non-Awardees	% Awarded
K01	479	152	327	31.73%
K04	1,089	341	748	31.31%
K07	562	274	288	48.75%
K08	1,176	514	662	43.71%
K11	166	86	80	51.81%
K12	N/A	188	N/A	N/A
K22	200	57	143	28.50%
K23	254	98	156	38.58%
K25	56	25	31	44.64%
TOTALS	3,982	1,547 + 188 Scholars	2,435	38.85%

Table 4. NCI K applicants and awardees, by primary mechanism.
Primary mechanisms were utilized to ensure that information for an individual applicant was counted only once in the evaluation.

2.3.2 Qualifying Degrees of NCI K Applicants and Awardees

While all of the mechanisms included in this evaluation required applicants to have doctoral-level degrees, some mechanisms are targeted toward or were more appealing to researchers with specific doctoral degrees or training. For instance, mechanisms with a more clinical focus, such as the K08 or K23, are designed for applicants with a clinical degree. To examine trends based on the qualifying degrees of applicants and awardees, degree information was derived from IMPAC II and supplemented with data from the AAMC Faculty Roster or DRF as necessary (see **Appendix 6.3** for degree category descriptions).

As shown in **Figure 3**, the majority of applicants and awardees for the mechanisms focused primarily on the "non-clinical" sciences (e.g., K01, K04, and K25) held PhDs, while those with a more "clinical" focus (e.g., K08, K11, K23) included a larger proportion of awardees with MDs. Individuals with MD/PhDs were predominant in the K08 and K22 mechanisms. Most K12 scholars held MDs, although a significant proportion (~30%) of the trainees' degree types were not clearly documented. When comparing awardees and non-awardees, no significant differences were found based on degree type.

[5] Additional dual K award combinations were K23 then followed by K22, K11 then K22, K11 then K08, and K08 then K01.

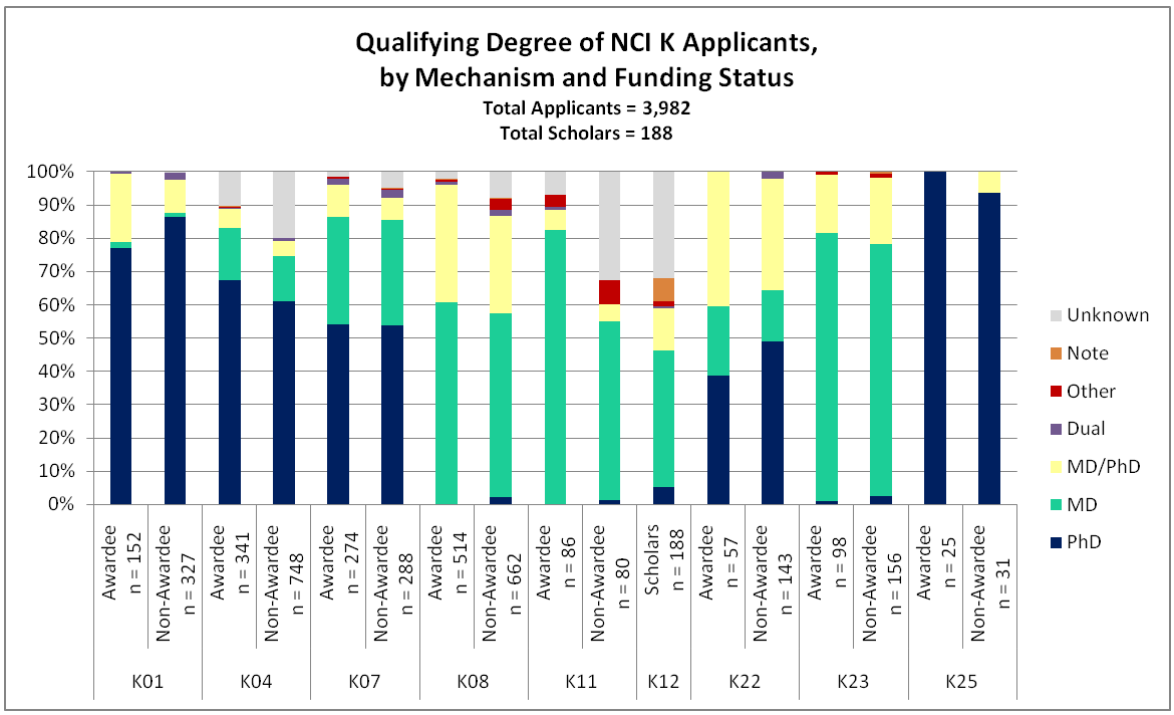


Figure 3. Qualifying degree of NCI K applicants, by mechanism and award status.
Qualifying degrees for awardees and non-awardees are shown side-by-side within each K mechanism, in addition to K12 scholars matched to IMPAC II. The number of unique awardees, non-awardees, or scholars is shown under each bar. The definition of Dual and Note can be found in Appendix 6.3. Data were obtained from IMPAC II, with incomplete records supplemented with data from NSF DRF and AAMC Faculty Roster.

2.3.3 PhD Field of Study and Medical Specialty of NCI K Applicants and Awardees

To determine whether the NCI K mechanisms attract applicants with specific educational backgrounds, we used field of study information from the DRF. We determined the most common PhD fields of study for each mechanism (**Table 5**).[6] Across mechanisms, applicants most commonly came from the following PhD fields of study: Molecular Biology, Biochemistry, and Immunology. The fields for the K07 mechanism – which is geared toward studies in cancer prevention control and the behavioral sciences – reflected the program focus, with the top fields represented being Epidemiology, Clinical Psychology, and Public Health. Similarly, the top PhD fields for applicants to the K25 mechanism, which supports investigators with quantitative scientific and engineering backgrounds, reflected the K25 program goal (data not shown)[7].

For applicants with medical degrees, AAMC Faculty Roster data were used to determine the top medical specialties of awardees across mechanisms. For all mechanisms, the top medical specialties were found to be Internal Medicine, Medical Oncology, and Hematology Oncology.

[6] The results for K11, K22, K23, and K25 mechanisms are not shown due to low numbers when analyzed by Tier 3 Field of Study category. K12 scholars were excluded because they were not matched to the DRF database.
[7] Field of Study was analyzed at both Tier 2 (general field, e.g., Biology, Chemistry, Engineering) and Tier 3 (specialties within a field, e.g., Molecular Biology, Biochemistry, Mechanical Engineering).

Mechanism	PhD Field of Study - Tier 3	Awardees (% Matched Awardees)	Total N Matched (% Awardees with PhDs Matched)	Non-Awardees (% Matched Non-Awardees)	Total N Matched (% Non-Awardees with PhDs Matched)
K01	Molecular Biology	17 (15.5%)	109 (73.6%)	31 (15.9%)	195 (61.9%)
	Biochemistry	14 (12.8%)		32 (16.4%)	
	Immunology	10 (9.2%)		20 (10.3%)	
K07	Epidemiology	38 (25%)	152 (87.4%)	19 (13.4%)	142 (81.6%)
	Clinical Psychology	24 (15.8%)		22 (15.5%)	
	Public Health	12 (7.9%)		11 (7.7%)	
K08	Molecular Biology	31 (20%)	155 (85.2%)	22 (15.3%)	144 (69.6%)
	Biochemistry	25 (16.1%)		19 (13.2%)	
	Immunology	21 (13.5%)		8 (5.6%)	

Table 5. Top three Fields of Study for NCI K applicants with a PhD, by mechanism and award status.
Only those applicants holding PhDs (PhDs or MD/PhDs) were included in this analysis. The percentage with PhD matched figures represent the percentage of PhDs matched to the DRF versus total number of PhD or MD/PhD awardees or non-awardees for each mechanism. Field of Study analysis was performed for all mechanisms included in this evaluation except K04 and K12 scholars; data are not shown due to low numbers. Data was obtained from NSF DRF.

2.3.4 Gender of NCI K Applicants and Awardees

To analyze the gender distribution across the NCI K mechanisms, we used IMPAC II data to determine the number of male and female applicants.[8] As reported in the National Science Foundation *Science & Engineering Indicators 2010*, on average, since 2000, men have received 52.6% of biological sciences PhDs and women received 47.4%.[9] During 2007, the most recent data, the proportion of men to women receiving PhDs was 50.6% to 49.4%, respectively.[10] Similarly, the AAMC reported in 2008 that 55.8% of MD degrees were earned by men and 44.2% by women[11]. Overall, the gender distribution seen across the NCI K mechanisms in this study echoed these trends (**Table 6**).

Gender	Applicants	Percent Applicants	Awardees	Percent Awardees	Non-Awardees	Percent Non-Awardees
Female	1,055	36.5%	441	36.6%	614	36.4%
Male	1,664	57.5%	734	60.9%	930	55.1%
Unknown	174	6.0%	31	2.6%	143	8.5%
TOTALS	2,893	100%	1,206	100%	1,687	100%

Table 6. Overall gender distribution of NCI K applicants, excluding K04s and K12 scholars.
Data were obtained from IMPAC II and supplemented by NSF DRF and AAMC Faculty Roster.

[8] Those applicants/awardees that did not specify gender on their applications were classified as "unknown".
[9] National Science Foundation. *Science and Engineering Indicators*. (2010). Available at: http://www.nsf.gov/statistics/seind10/c2/c2h.htm. Numbers reflect those for U.S. citizens/permanent residents. (Last accessed October 18, 2011).
[10] Ibid.
[11] Association of American Medical Colleges. *2008 Physician Specialty Data.* (2008). Available at: https://www.aamc.org/download/47352/data/specialtydata.pdf (Last accessed August 5, 2011).

However, several programs diverged from this pattern (**Figure 4**). The majority of K07 awardees (65%) and non-awardees (63%) were female. This is consistent with the K07 program's focus on PhDs and in areas of prevention, control, and behavioral and population sciences, which have higher percentages of female degree recipients.[12]

The majority of K08 (73%), K11 (72%), K23 (67%), and K25 awardees (73%) were male. These percentages, although lower than the MD degree distribution noted above, are consistent with the gender distribution of all active physicians, regardless of age, as reported in 2007: 71.7% were male, and 28.3% were female.[13]For K08 awardees, we considered that this data could be partially attributed to the fact that the first awards were made in 1984, when fewer women were in the biomedical workforce. However, upon further analysis, we found that award rates for K08 by gender were highly varied over time, but there was no overall significant difference between them during the period of this evaluation (data not shown).

We also examined resubmission of NCI K applications by gender. Using odds ratio tests, female applicants were determined to be 1.2x more likely to resubmit a K application than male applicants ($p=0.03$). However, further analysis found that this is driven primarily by the K07 applicant pool, as there is no significant difference in resubmission rates by gender when excluding K07 applicants. Resubmission analyses are presented in **Appendix 6.4**.

[12] Per the 2010 NSF Science & Engineering Indicators, over the period of 2000 through 2010, 58% of social/behavioral sciences PhDs, and 70% of psychology PhDs were granted to women. Numbers reflect percentage of total for each category for the years 2000 through 2007. *NSF Science and Engineering Indicators.* (2010). Available at: http://www.nsf.gov/statistics/seind10/c2/c2s3.htm#s5. (Last accessed November 10, 2011).
[13] Association of American Medical Colleges. *2008 Physician Specialty Data.* (2008). Available at: https://www.aamc.org/download/47352/data/specialtydata.pdf (Last accessed August 5, 2011).

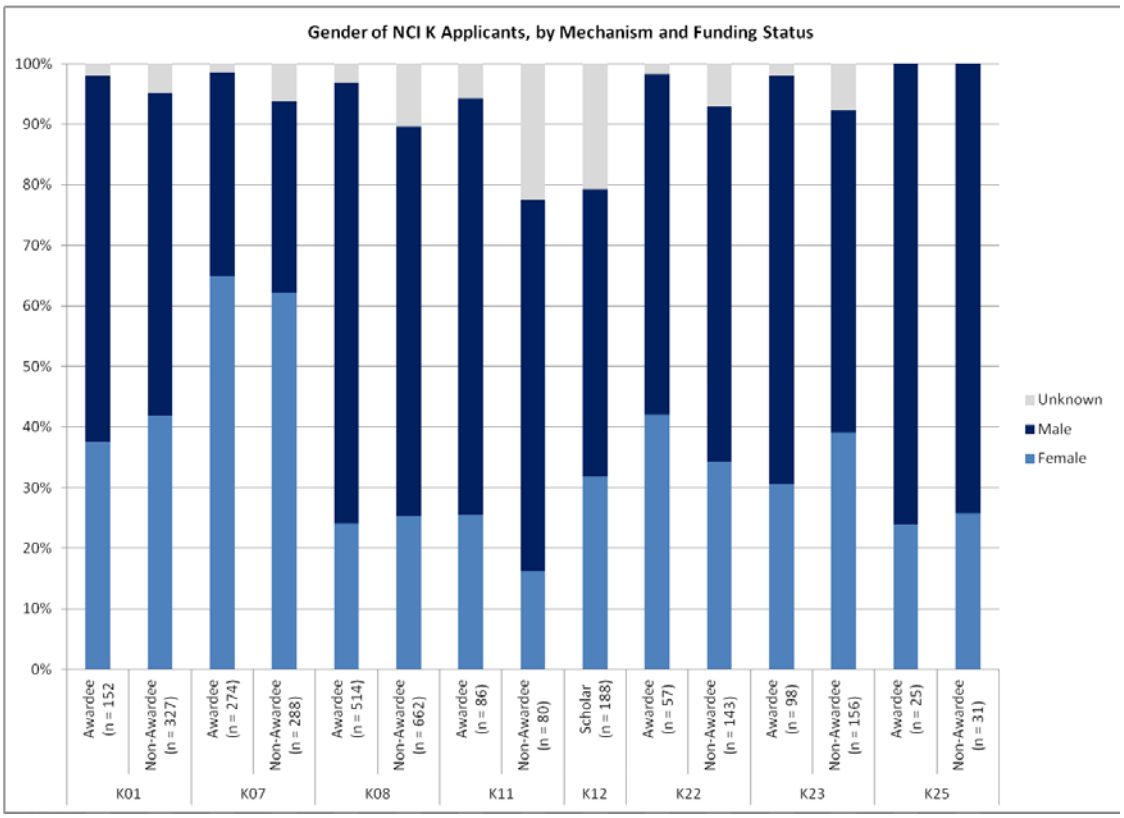


Figure 4. **Gender distribution of NCI K applicants, by mechanism and award status.**
Unknown category reflects gender not reported. Data are from IMPAC II, supplemented by NSF DRF and AAMC Faculty Roster.

2.3.5 Distribution of Race and Ethnicity among NCI K Applicants and Awardees

During the time period 2000 through 2007, Hispanics received 4.5% of biological sciences PhD degrees, Blacks 3.4%, Asians 11.6%, Native Americans 0.4%, and Whites 76.2%.[14] Because race and ethnicity are voluntarily reported and might not be consistently provided, a combination of data sources were used to compile race and ethnicity data for K program applicants. IMPAC II was the primary data source, supplemented as needed by the DRF and AAMC Faculty Roster (see **Appendix 6.2**). The category "Other" reflects applicants who listed more than one race or listed race(s) not included in the evaluation categories. The "Unknown" category was used for applicants who did not report race/ethnicity. Summary race and ethnicity data across the NCI K mechanisms is shown in **Table 7**. Clear trends regarding race and ethnicity of NCI K applicants and awardees are difficult to determine due to relatively large proportions of individuals with unknown or undeclared status (approximately 30% for all applicants, including 6% of awardees and 36% of non-awardees). A chi-squared test indicated that applicant and awardee race/ethnicity was independent of application resubmission (**see Appendix 6.4**).

[14] National Science Foundation. *Science and Engineering Indicators*. (2010). Available at: http://www.nsf.gov/statistics/seind10/c2/c2s3.htm (Last Accessed October 18, 2011).

Race / Ethnicity	Applicants	Awardees	Non-Awardees
Asian	450	195	255
Black	42	17	25
Hispanic	49	27	22
Native American	*	*	*
White	1,519	751	768
Other	6	3	3
Unknown	826	212	614

Table 7. Overall race and ethnicity distribution of NCI K applicants.
The Other category represents selection of a race/ethnicity not included in the list or selection of multiple race/ethnicities. The Unknown category indicates that no race/ethnicity was found in any of the data sources utilized. An asterisk indicates data that have been suppressed due to a low (<25) number of applicants. K04 and K12 scholars are excluded from this analysis due to a low match rate. The primary data source was IMPAC II supplemented by NSF DRF and AAMC Faculty Roster.

2.3.6 Age of NCI K Applicants and Awardees

The majority of K mechanisms included in this evaluation are specifically targeted to early-career researchers, and thus we explored whether the average age of applicants and awardees was commensurate with those of post-doctoral researchers and post-residency clinicians. Rules for determining applicant and awardee age are described in **Appendix 6.5**. For all mechanisms, the median age of awardees was 36 years, while the median age of non-awardees was 37 years (**Figure 5**). While the median age of the awardees for the majority of NCI K mechanisms fell within the range of mid-to-late thirties, the median age of the majority of K12 scholars fell within the early thirties, and 18% of scholars for which age data were available were 25-29 years old.[15]

[15] Age data were available for 142 of the 188 matched K12 scholars.

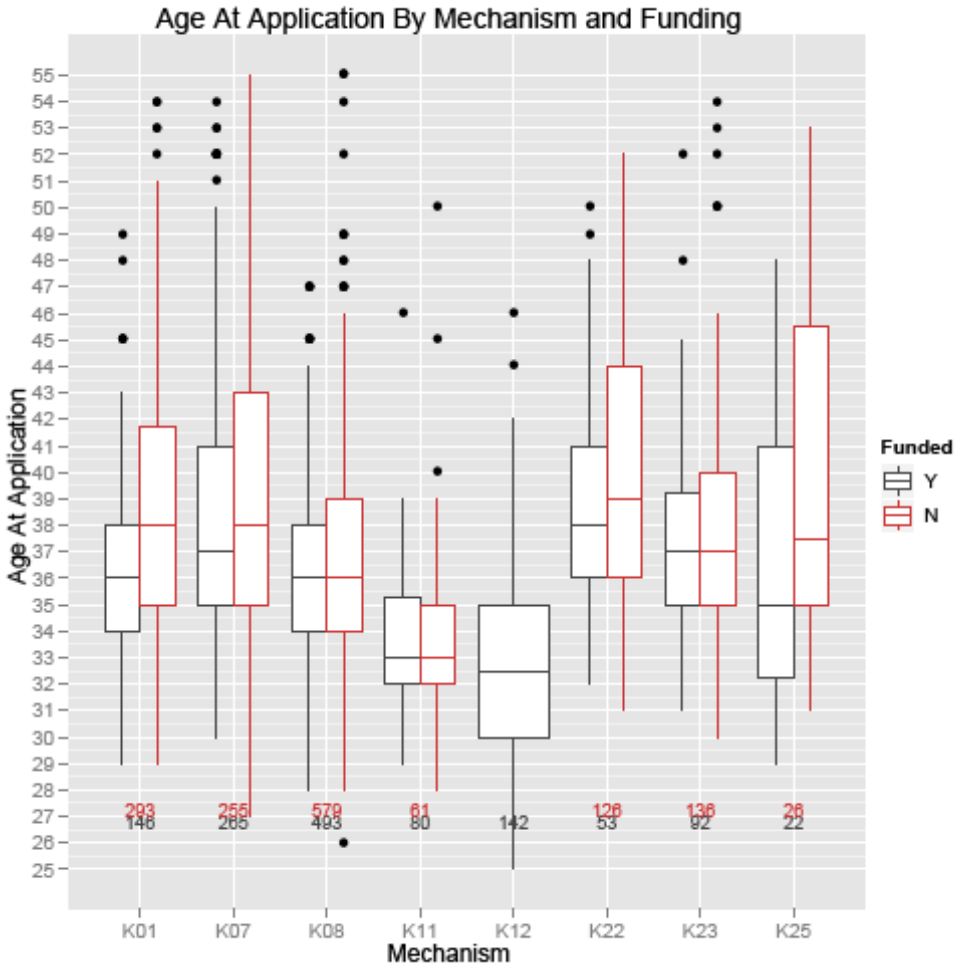


Figure 5. Age distribution of NCI K applicants, by mechanism and award status.[16]
In this representation, the box represents the interquartile range (IQR) with the middle horizontal line of the box representing the median, the lower horizontal line of the box representing the 1st quartile and the upper horizontal line of the box representing the third quartile. The lower line (whisker) represents the first quartile – 1.5x the interquartile range, and the upper line (whisker) represents the third quartile + 1.5x the interquartile range. Dots above or below the whiskers represent data points that are considered outliers. Numbers shown below the lower whisker indicate awardees (black) or non-awardees (red).

2.3.7 Years Since Degree of NCI K Applicants and Awardees

As discussed in Section 1.3, several K mechanisms are targeted toward investigators early in their careers, while others are intended to attract mid-career researchers or clinicians. An analysis of average years since qualifying degree (YSD) allows for the detection of trends within a particular mechanism. This is dependent on an individual's rate of progression through graduate and post-graduate studies, rather than biological age, which can vary depending on the age at which an individual chooses to pursue a career in biomedical research (**Figure 6**). Across mechanisms, the majority of NCI K applicants submit applications 6 to 10 years following the conferral of their qualifying degrees. This time frame is commensurate with completion of a post-doctoral research experience or medical residency. The K12 scholars have the shortest median time lag (4 years) between conferral of qualifying degree and first

[16] K12 data represent 142 of the 188 IMPAC II matched scholars for which age data was available.

K12 appointment. K07 applicants tend to have a shorter median lag time between conferral of their qualifying degrees and first application when compared to the other K mechanisms, but have a larger overall range. This could be attributed to the nature of the program, which unlike most programs in this evaluation, is not focused on career transition, but on bringing investigators into the field of cancer prevention and control research. It is also one of the longest running programs included in the evaluation. Additional analysis of the years since degree for the K07 mechanism is presented in **Appendix 6.6**.

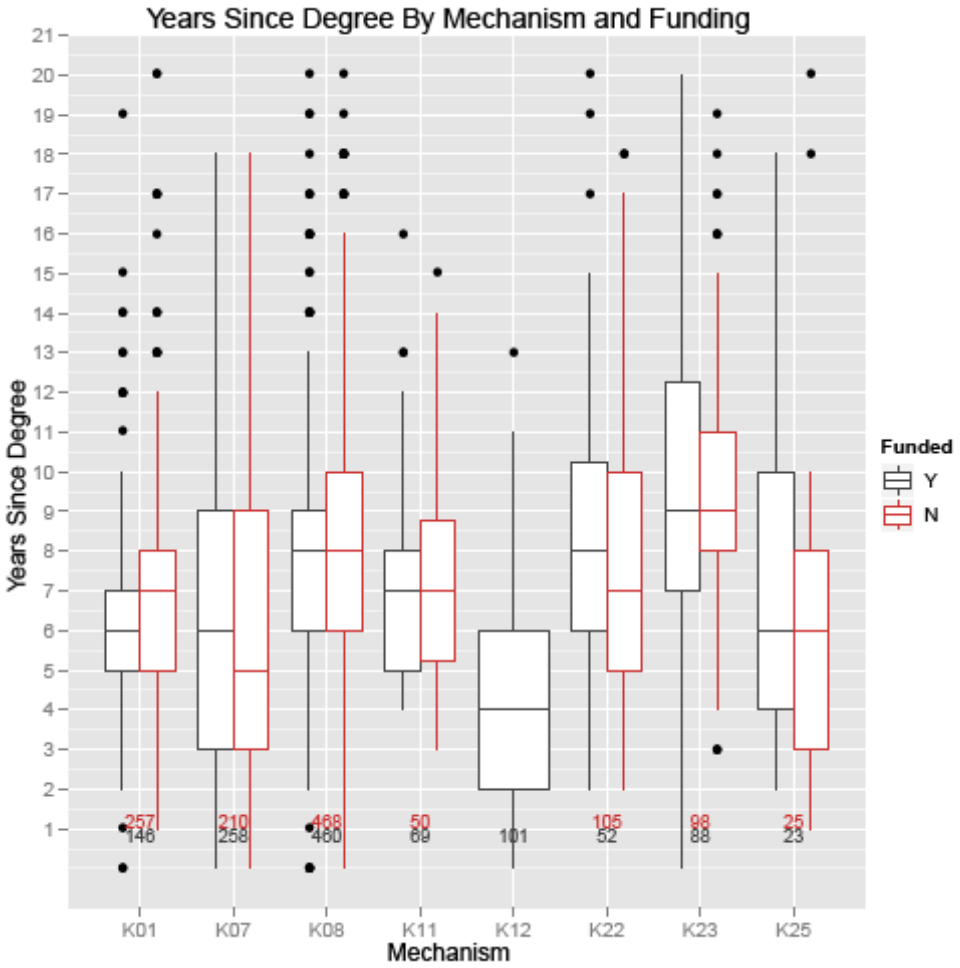


Figure 6. Years since degree distribution of NCI K applicants, by mechanism and award status.[17]
In this representation, the box represents the interquartile range (IQR) with the middle horizontal line of the box representing the median, the lower horizontal line of the box representing the 1st quartile and the upper horizontal line of the box representing the third quartile. The lower line (whisker) represents the first quartile – 1.5x the interquartile range, and the upper line (whisker) represents the third quartile + 1.5x the interquartile range. Dots above or below the whiskers represent data points that are considered outliers. Numbers shown below the lower whisker indicate awardees (black) or non-awardees (red).

[17] K12 data represent all 188 scholars matched to IMPAC II.

2.3.8 Prior NIH Support of NCI K Applicants and Awardees

Next, we examined prior NIH funding. We used IMPAC II to determine whether applicants had received previous NIH-supported training, an NIH Research Project Grant (RPG)[18], a combination of a training grant and RPG, or other support prior to applying for or being awarded a K award (**Table 8**). NIH-supported training included: 1) National Research Service Awards (NRSA), which are institutional (T) grants, or fellowships (F); and 2) loan repayment contracts (L). We were unable to examine whether K applicants received prior training from institutional R25 or K12 grants because trainees or scholars appointed to these grants are not captured in IMPAC II. Across mechanisms, if an applicant had prior support, the most common type was T, F, or L. Prior RPG support was most common among K04 applicants, reflecting the more senior status of these investigators. The currently existing mechanisms in this study are specifically targeted to early career investigators.[19] Modeling analyses described in **Section 4.0** describe the relationship between prior support and award status.

[18] Full list of grants included in this category is available in Appendix 6.7.

[19] Applicants with R03, R21, R29, and R55 awards are eligible for K awards. If a K application is received in which the applicant has had an R01, it is immediately withdrawn and not reviewed.

Prior Support	K01 Awardee n = 152	K01 Non-Awardee n = 327	K04 Awardee n = 341	K04 Non-Awardee n = 748	K07 Awardee n = 274	K07 Non-Awardee n = 288	K08 Awardee n = 514	K08 Non-Awardee n = 662	K11 Awardee n = 86	K11 Non-Awardee n = 80	K22 Awardee n = 57	K22 Non-Awardee n = 143	K23 Awardee n = 98	K23 Non-Awardee n = 156	K25 Awardee n = 25	K25 Non-Awardee n = 31
Had T Support	60 (39%)	130 (40%)	0 (0%)	5 (0.7%)	90 (33%)	73 (25%)	216 (42%)	227 (34%)	15 (17%)	12 (15%)	25 (44%)	40 (28%)	51 (52%)	71 (46%)	4 (16%)	6 (19%)
Had Only T Support	47 (31%)	92 (28%)	0 (0%)	0 (0%)	71 (26%)	45 (16%)	191 (37%)	191 (29%)	15 (17%)	12 (15%)	16 (28%)	27 (19%)	44 (45%)	46 (29%)	4 (16%)	5 (16%)
Had F Support	23 (15%)	53 (16%)	73 (21%)	166 (22%)	14 (5%)	10 (3%)	47 (9%)	32 (5%)	5 (6%)	3 (4%)	7 (12%)	11 (8%)	6 (6%)	5 (3%)	1 (4%)	0 (0%)
Had Only F Support	9 (6%)	18 (6%)	17 (5%)	65 (9%)	6 (2%)	3 (1%)	34 (7%)	22 (3%)	5 (6%)	3 (4%)	1 (2%)	3 (2%)	3 (3%)	4 (3%)	1 (4%)	0 (0%)
Had L Support	1 (0.7%)	3 (0.9%)	0 (0%)	0 (0%)	20 (7%)	31 (11%)	11 (2%)	37 (6%)	0 (0%)	0 (0%)	2 (4%)	4 (3%)	11 (11%)	23 (15%)	0 (0%)	1 (3%)
Had Only L Support	1 (0.7%)	0 (0%)	0 (0%)	0 (0%)	8 (3%)	17 (6%)	2 (0.4%)	13 (2%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)	5 (5%)	4 (3%)	0 (0%)	1 (3%)
Had RPG Support	1 (0.7%)	14 (4%)	232 (68%)	363 (49%)	34 (12%)	35 (12%)	6 (1%)	9 (1%)	0 (0%)	0 (0%)	1 (2%)	16 (11%)	1 (1%)	8 (5%)	0 (0%)	3 (10%)
Had Only RPG Support	0 (0%)	10 (3%)	169 (50%)	241 (32%)	21 (8%)	16 (6%)	1 (0.2%)	8 (1%)	0 (0%)	0 (0%)	1 (2%)	8 (6%)	1 (1%)	2 (1%)	0 (0%)	2 (6%)
Had Multiple T, F, or L Support	13 (9%)	36 (11%)	0 (0%)	1 (0.1%)	11 (4%)	14 (5%)	19 (4%)	30 (5%)	0 (0%)	0 (0%)	7 (12%)	8 (6%)	7 (7%)	20 (13%)	0 (0%)	0 (0%)
Had Multiple Support, including RPG	1 (0.7%)	4 (1%)	54 (16%)	100 (13%)	13 (5%)	16 (6%)	4 (0.8%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)	6 (4%)	0 (0%)	4 (3%)	0 (0%)	1 (3%)
Had Only Other Support	3 (2%)	1 (0.3%)	5 (1%)	12 (2%)	5 (2%)	10 (3%)	0 (0%)	6 (0.9%)	0 (0%)	0 (0%)	1 (2%)	2 (1%)	0 (0%)	5 (3%)	0 (0%)	0 (0%)
No Prior Support	78 (51%)	166 (51%)	85 (25%)	303 (41%)	135 (49%)	159 (55%)	260 (51%)	385 (58%)	66 (77%)	65 (81%)	28 (49%)	85 (59%)	36 (37%)	68 (44%)	20 (80%)	22 (71%)

Table 8. Prior NIH support of NCI K applicants, by mechanism and award status.
An individual may be in multiple categories except those labeled "Had Only (T, F, L, RPG, or Other) Support" or No Prior Support. Percentages are calculated within each funding group (e.g., K01 awardees). A table showing the award mechanisms included in each prior support group is available in Appendix 6.7. All data from IMPAC II.

2.3.9 Applications and Awards by Institution Type

The NCI Cancer Centers program supports the long-standing commitment of the NCI to support a system of integrated, multi-disciplinary cancer research centers across the United States[20]. We examined two characteristics of the institutions from which NCI K applications were received: (a) whether an institution has an NCI Cancer Center designation, and (b) the level of overall annual NCI funding the institution received during the study period.

To determine an institution's Cancer Center status, institutions with 10 or more K applications were compared with a list of NCI-designated Cancer Centers that included current Cancer Centers and institutions previously designated as a Cancer Center in *The NCI Annual Fact Book* (FY 1971 through FY 2010), or designated as a Comprehensive Cancer Center (**Table 9**). Applications to the NCI K01, K07, K08, K11, K22, K23, and K25 mechanisms were received from 363 total institutions, of which 46 (13%) were institutions that have NCI-designated Comprehensive Cancer Centers and 35 (9%) were institutions that have NCI-designated Cancer Centers. The majority of institutions (282, or 78%) did not have an NCI Cancer Center designation. However, across mechanisms, the majority of applications (71%) were received from and awards granted to institutions that have NCI-designated Cancer Centers, and about half of the applications and awards were from Comprehensive Cancer Centers (48%). The K22 mechanism was slightly different, garnering interest from applicants at institutions with and without NCI Cancer Centers. Half of the K22 applications came from institutions that have NCI-designated Cancer Centers (36% Comprehensive and 16% basic Cancer Centers), versus the nearly two-thirds seen for all other mechanisms. This is likely attributed to two related factors: 1) K22 applicants are not required to have a sponsoring institution at the time of application; and 2) the K22 is one of only two mechanisms (K22, K99) open to post-doctoral fellows at Federal laboratories. Whether this difference also reflects project proposals that do not necessarily require the support of NCI-designated Cancer Center resources requires further analysis

[20] National Cancer Institute Cancer Center Program website: http://cancercenters.cancer.gov/about/our-history.html (Accessed on July 7, 2011).

K Mechanism	NCI-designated Comprehensive Cancer Centers		NCI-designated Cancer Centers		Institutions that are not NCI-designated Cancer Centers		Total	
	46 Institutions* (13%)		35 Institutions* (9%)		282 Institutions (78%)		363 Institutions	
	Applications	Awards	Applications	Awards	Applications	Awards	Applications	Awards
K01	272	63	168	47	184	43	624	153
K04	467	139	264	72	518	130	1,249	341
K07	469	162	156	53	200	59	825	274
K08	796	235	463	163	379	117	1,638	515
K11	120	46	51	22	45	18	216	86
K22	118	36	52	19	154	18	324	73
K23	244	66	68	19	54	13	366	98
K25	53	18	9	1	26	6	88	25
Total	**2,539 (48%)**	**765 (49%)**	**1,231 (23%)**	**396 (25%)**	**1,560 (29%)**	**404 (26%)**	**5,330**	**1,565**

Table 9. Distribution of K applications by NCI Cancer Center designation.

Cancer Center status was obtained through manual review of *The NCI Annual Fact Book* collection (1971 through 2010) and comparison to the current list of NCI-designated Cancer Centers. Applicants with multiple applications were counted only once in this analysis. K12 scholars were not included. Institution names were obtained from IMPAC II. *In IMPAC II, each grant application has an institutional profile code identifying the institution submitting the application. The institutional profile code can and does change over time. Organizations will merge, separate, or become part of a state system and consequently, the institutional profile code changes. Thus, although 46 and 35 instititutions were identified from K applications in IMPAC II covering the period of analysis, there are currently only 41 NCI-designated Comprehensive Cancer Centers and 26 Cancer Centers.

We also classified applicant institutions by the average amount of annual NCI funding received between 1970 and 2011. Institutions were separated into three funding-level groups: those with $10 million to <$100 million; $1 million to <$10 million; and $0 to <$1 million in NCI funding.[21] The funding boundaries of each group are listed in the headers of **Table 10**.[22] Across mechanisms, but particularly in the K07, K08, K11, and K23 programs, the majority of applications originated from institutions receiving $10 million to <$100 million in NCI funding. Two mechanisms – K04 and K22 – had a slightly lower proportion of applications from institutions in the highest tier, but still, more than one-third of their application pool came from this group. The second-highest funding bracket, $1 million to <$10 million, represented approximately one-third of the applications for most mechanisms, except K11, K23, and K25, which had less than 25% from institutions in this funding level. Applications submitted from institutions in the highest funding bracket had a 32% funding success rate, while those submitted from institutions in the middle and lowest funding brackets had 28% and 18% funding success rates, respectively. In summary, regardless of mechanism, the majority of K applications and awards were made to 47 institutions with the highest average NCI funding within the study, even though applications spanned 352 total institutions. [23]

[21] These funding levels reflect the average annual amount of NCI funding received over the course of the study period.
[22] 11 institutions were excluded from this analysis because they did not receive any NCI funding during the evaluation period, which excluded 14 applications but 0 awards from this analysis.
[23] K12 not included.

K Mechanism	Funding Level: $10 Million to <$100 Million		Funding Level: $1 Million to <$10 Million		Funding Level: $0 to <$1 Million		Totals	
	47 Institutions (14%)		135 Institutions (38%)		170 Institutions (48%)		352 Institutions	
	Applications	Awards	Applications	Awards	Applications	Awards	Applications	Awards
K01	369	97	227	53	26	3	622	153
K04	481	140	531	153	236	48	1,248	341
K07	516	181	270	83	37	10	823	274
K08	1,097	378	497	129	43	8	1,637	515
K11	160	65	48	20	8	1	216	86
K22	132	40	97	28	88	5	317	73
K23	293	83	59	11	14	4	366	98
K25	56	18	16	4	15	3	87	25
Total (% of Total Applications or Total Awards)	**3,104 (58%)**	**1,002 (64%)**	**1,745 (33%)**	**481 (31%)**	**467 (9%)**	**82 (5%)**	**5,316**	**1,565**
Funding Success Rate	**32%**		**28%**		**18%**			

Table 10. Distribution of NCI K applications and awards, annual cumulative NCI funding level of institution.
Funding data collected from IMPAC II for FY 1971 through FY 2010. K12 scholars are excluded from this analysis.

2.4 Summary of Applicant and Awardee Characteristics

Overall, the K mechanisms are attracting applicants with appropriate training and specialization. There appears to be a somewhat skewed gender representation among applicants; however, award rates did not vary by gender. Many of the K awards are granted to a small proportion of applicant institutions, in particular, those which have received the NCI Cancer Center designation.

Awards and Award Rates

- The nine NCI K mechanisms, spanning the years 1970 through 2008, received 5,445 applications and made 1,609 awards. Combined, 29.55% of applications were awarded. There was no difference in award rate by gender or race/ethnicity.
- The majority of applicants received only one NCI K award. 3,863 applied to only one NCI K mechanism, 119 applied to multiple mechanisms, and 18 individuals received two NCI K awards. Individual applicant funding success was over 38%.
- Initial applications (A0) were awarded at a rate of 22.8%, while resubmissions were funded at a rate of 37.2%. The probability of being funded increased with the number of submissions.

Demographics

- Gender distribution varied by K program. Men comprised 57.5% of the total K applicants, and there was no difference in K funding success based on gender. Women made up 65% of applicants in the K07 cancer prevention program, consistent with its focus in the areas of prevention, control and behavioral and population sciences, where women have been majority participants. Men comprised 67-75% of the more clinically focused K08, K11, and K23 programs; this proportion was consistent across years suggesting it is independent of the gender distribution of MD graduates.
- Across NCI K mechanisms, the median age of awardees was 36 years.
- The majority of applicants to the NCI K mechanisms apply within 6 to 10 years of receiving their qualifying degree.
- Resubmission of an application was found to be independent of an applicant's race/ethnicity.

Degree and Specialty

- K mechanisms attracted the intended applicants. Those focused on "non-clinical" research (e.g., K01, K04, K25) attracted PhD applicants, and those with a "clinical" research focus (e.g., K08, K11, K23) attracted MD applicants.
- Degree field and clinical specialty of the applicants matched program focus. Of those NCI K applicants with medical degrees, the top three medical specialties seen among awardees were Internal Medicine, Medical Oncology, and Hematology oncology. This was true across mechanisms. Among K01 and K08 PhD applicants, the most common fields of study were found to be Molecular Biology, Biochemistry, and Immunology. Among K07 applicants, the most common PhD fields were Epidemiology, Clinical Psychology, and Public Health, reflecting the emphasis of the K07 mechanism on behavioral aspects of cancer prevention and control.

Prior Support

- The most common prior NIH support mechanism among NCI K applicants was determined to be T or F grants (NRSA training).

Institutional Characteristics

- Nearly half of all NCI K applications and awards were made to individuals at NCI-designated Comprehensive Cancer Centers, representing only 13% of all applicant institutions. Another 25% of awards were made to

individuals at NCI-designated basic Cancer Centers. Twenty-six percent of NCI K awards were made to individuals at the 282 institutions that did not have an NCI Cancer Center designation.

- More than 60% of awards went to 14% of the applicant institutions representing the top bracket of NCI funding.

3.0 Selected Outcomes of Program Applicants and Awardees

3.1 Overview

In this section, we describe our methods and present findings on the outcomes of NCI K program applicants and participants. The traditional measures of a research career, such as subsequent publications (publication number, citation counts, top journals and journal subject categories), grant applications and awards, time to receipt of first R01, and faculty rank progression, were explored in parallel with other measures of scientific engagement, such as subsequent funding from non-NIH sources, U.S. patent applications and issued patents, participation in clinical studies, membership in key professional societies, receipt of research awards, and participation on advisory and review panels.

3.2 Data Sources

IMPAC II served as the data source for subsequent NIH outcomes. Subsequent publications were collected from the NLM MEDLINE database and augmented with information such as citation count and journal subject category from Thomson Reuters Web of Science. The International Cancer Research Partners (ICRP) matched a list of NCI K applicant names to their database to provide data on subsequent awards from non-NIH sources, including the Department of Defense Congressionally Directed Medical Research Program (DOD-CDMRP), American Cancer Society, Prostate Cancer Foundation, Susan G. Komen Breast Cancer Foundation, California Breast Cancer Research Program, and the Canadian Institutes of Health Research. The Thomson Reuters ScienceWire grant catalog was used to match NCI K applicant names to grant awards sponsored by the U.S. Department of Energy (DOE) and National Science Foundation (NSF). The Thomson Reuters ScienceWire patent catalog was used to obtain U.S. Patent & Trademark Office (USPTO) patent applications and awards that cited NCI K awards included in this evaluation. The ClinicalTrials.gov database was used to match NCI K08 and K23 applicant names to those of key personnel for clinical trials registered through this public site. IMPAC II and data obtained from the Federal Interagency Databases Online (FIDO.gov) database were used to collect information about participation in NIH advisory groups and other federal advisory committees. The NIH-Presidential Early Career Award for Scientists and Engineers (PECASE) website[24] was used to match names of NCI K applicants to those listed as PECASE awardees. Matching of NCI K applicant names to the membership directories of the American Association for Cancer Research (AACR), the American Society of Clinical Oncology (ASCO) and the Federation of Societies for Experimental Biology (FASEB) was completed in cooperation with each society to obtain information regarding participation of K applicants in these organizations. Finally, the Google search engine and LinkedIn professional networking website were used to obtain subsequent career information on a subset of individuals for whom no additional career appointment data were available through IMPAC II.

[24] http://grants.nih.gov/grants/policy/pecase.htm. Last Accessed on October 25, 2011.

3.3 Methodology

3.3.1 Deriving Comparison Cohorts

To more accurately examine the effect of participation in the K program on subsequent career outcomes, we used application priority scores – an NIH-wide quantitative metric of application quality – to determine nearly identical groups of applicants to each NCI K mechanism[25] among whom there was an equal likelihood of the application being funded or not funded. We then restricted career outcome analyses to applicants within this "funding bubble" – the priority score range within which there was an equal chance of an application being funded or not funded (**Figure 7**). Additional details regarding the methodology used to construct this comparison cohort are presented in **Appendix 6.8**.

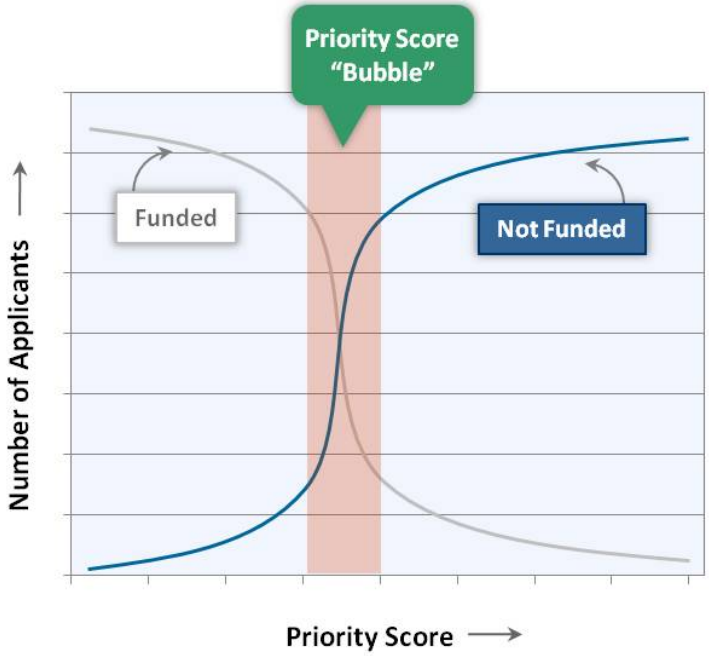


Figure 7. Comparison group methodology: The "funding bubble."

3.3.2 Publication Matching

Several independent but overlapping matching rules were used to identify MEDLINE publication records in which a study applicant appeared as an author. To be considered for matching, the publication date had to be at least one year after the application date of the last in-study K application for that applicant. The upper bound for the publication date was April 1, 2011. These rules are summarized below:

1. Match publications for which there was an exact match of the MEDLINE author email address and the IMPAC II PI email address, and a moderate-strength fuzzy name match between the MEDLINE author name and the IMPAC II PI name. "Fuzzy" matching accommodates for misspellings and other variations.

[25] K04 and K12 Scholars were excluded from this analysis.

2. Match publications for which there was an exact match of the MEDLINE author email address and the IMPAC II PI email address, and a name match between any of the other MEDLINE author names and the IMPAC II PI name.

3. Match publications for which there was an exact match of the Web of Science author email address (for MEDLINE publications that have been matched to Web of Science) and the IMPAC II PI email address, and a moderate-strength fuzzy name match between the MEDLINE author and the IMPAC II PI name.

4. Using the set of matches found using the first three rules and also publications found by funding acknowledgment (which are omitted from the overall match set if not also matched through one rules 1 – 3 to reduce a potential recall bias favoring awardees), find additional publications for which the MEDLINE author names have high name-frequency-corrected overlap and a fuzzy name match between the MEDLINE author name and the IMPAC II PI name.

The matching process was conservative, favoring precision over recall. This approach results in high-confidence in the papers that are assigned to individuals, but means that some publications for individuals are missed.

3.3.3 Name Matching to Outcomes Datasets

Names in the ICRP grant dataset were matched by ICRP personnel to a provided list of K program applicant and scholar names. ICRP reported the matches found using 6 different matching rules. We assessed the quality of the match data and selected the 2 highest precision matches (exact first and last name match/exact last name match and first name Levenshtein distance < 2).[26] Name matching to professional society membership lists was performed by ASCO, AACR and FASEB staff and these matches were accepted without change.

Data for most other non-IMPAC II data sources (with the major exception of MEDLINE/Web Of Science, discussed in 3.3.2) used a baseline name match, with corrections applied as necessary based on quality checks on the matching results. The baseline rule required an exact first and last name match, and if a middle name was present in both IMPAC II and the other data source, the first characters in each string had to match. Patent data, ClinicalTrials.gov data, PECASE data, and DOE grant data all used the baseline rule with no further restrictions. For FIDO data, there was an additional restriction that the IMPAC II name matched to a unique person in the FIDO database. For NSF grant data, the matches were first restricted to less common names and then further restricted by manual review. Common names were determined by a weighted average of the first and last name frequencies as measured in IMPAC II, with a threshold determined by examining a sample of names.

[26] The **Levenshtein distance** is a string metric for measuring the amount of difference between two sequences, and is defined as the minimum number of edits needed to transform one string into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character.

3.3.4 Subsequent Outcomes Timeline

For each individual in the study, all outcomes (e.g., publications, committee service, NIH grant activity, etc.) were measured starting with the fiscal year following the first NCI K award (for those with awards) or the last unsuccessful NCI K application (for those who had no NCI K award). Prior support was measured up to the fiscal year preceding the first award or last unsuccessful application.

When reporting applicant data by mechanism, unless otherwise specified, applicants are reported only once under the mechanism in which they obtained their first award or else had their last unsuccessful application (see Section 2.3.1). For six non-awardees who had unsuccessful applications in more than one mechanism in the same, latest year, additional rules were applied to select a unique mechanism: five were placed in the mechanism in which they had the best (numerically lowest) priority score, and in one case K22 was selected over K23 since there were fewer K22 applications overall in the study. With this rule, an awardee or non-awardee in the study overall is also an awardee or non-awardee, respectively, in the mechanism in which they are reported.

3.3.5 Subsequent NIH or NCI Grant Activity

Using IMPAC ll, subsequent NIH grant application activity was collected for all NCI K applicants. Grant activity was ordered into achievements, or "high water mark" levels, and each NCI K applicant was placed into the best (closest to 1) level possible based on application activity up to FY 2011 (**Table 11**). The high water mark was captured for NIH using application activity to all NIH institutes, including NCI, as well as for NCI-only, which used the same rules, but restricted the set of subsequent applications to only those submitted to NCI. In the NCI-only version of the high water mark analysis, individuals with post K application or awards from NIH ICs other than NCI were placed in Category 11 along with those with Type 3 or 5 grants from NCI (Other Future).

The first five categories in **Table 11** reflect the successful receipt of competitive research and training awards, the next five categories (6-10) reflect application to competitive research and training awards. Following that is category 11, grant application activity not meeting the specifications of levels 1-10. The remaining applicants are placed in Category 12 with no subsequent NIH or NCI funding application or awards found.

High Water Mark	Full Description	Abbreviated Description
1	Awarded a new (Type 1,2) P01 (primary PI) or R37 grant	Awarded P01 (primary PI) or R37
2	Awarded a new Institutional Training Grant (T32, K12, R25)	Awarded Institutional Training Grant
3	Awarded a new (Type 1,2) R01, U01, or P01 subproject grant	Awarded R01, U01, or P01 sub
4	Awarded a new (Type 1,2) RPG grant (other than above)	Awarded other RPG
5	Awarded a new grant other than an RPG or ITG	Awarded other grant
6	Applied for a new (Type 1,2) P01 primary PI or R37 grant	Applied for P01 primary PI or R37 (unfunded)
7	Applied for a new ITG (T32, K12, R25) that was not awarded	Applied for ITG (unfunded)
8	Applied for an R01, U01, or P01 subproject grant that was not awarded	Applied for R01, U01, or P01 sub (unfunded)
9	Applied for a new RPG grant (other than the above) that was not awarded	Applied for other RPG (unfunded)
10	Applied for a new grant other than an RPG or ITG that was not awarded	Applied for other grant (unfunded)
11	Some grant application activity not meeting the specifications of levels 1-10	Other Future
12	No grant activity found in the post-K period	None Found

Table 11. Subsequent grant activity high water mark categories.
For categories 1 and 6, the P01 primary PI represents the individual on the grant who is responsible for managing the entire P01 project. This individual may also serve as the PI on one of the subprojects comprising the P01 grant. Categories 2 and 7 indicate that the individual is applying to serve as the administrator of the institutional training grant rather than participating as a trainee.

3.3.6 Time to R01 Award Analysis

For individuals who received a subsequent R01 award as of FY 2011, the time to R01 value is the numerical difference between the fiscal year of that first R01 award and the fiscal year after their first NCI K award or their last unsuccessful NCI K application. According to this rule, individuals who received an R01 in the fiscal year immediately following their K award, or last application, will have a time to R01 value of zero.

3.3.7 Publication Productivity Analysis

To measure post-study publication productivity for each individual, the number of publications in each fiscal year after their first NCI K award or else their last unsuccessful NCI K application was recorded. A series of successive 2-year time periods up to and including the 12th year (counting the 1st year as Year 1) and a final time period for the 13th year to the present were constructed. The publications per person per year ratio for each individual was computed by dividing the number of publications for each person in a given time period by 2 for all but the last time period. In the last time period, the number of years from the start of the time period up to the last year with publication data was used as the denominator. In tabular or graphical summaries, the resulting publications per person per year in each time period were averaged over all individual cases represented by a given table cell or graph point. For example, a

graph showing a value of 1.6 for K22 awardees in year 3 means that 1.6 is the average of the publications/person/year for all K22 awardees in years 3 and 4 after their first award.

The Citation Benchmark is a standard against which to measure the actual citations received by a given publication. The Citation Benchmark is the median of the total number of citations at 24 months after publication of articles that share the following characteristics with those of a given publication:

1. Have the same article type (abstract, article, review, note, etc.),
2. Are published in the same journal,
3. Are published within six months (before or after) of the date of the study article,
4. Include the given article.

The count of actual citations received in the first 24 months is divided by the Citation Benchmark to obtain a ratio which can be used to compare citation performance among different types of publications.

The Actual/Benchmark ratios were analyzed in the same set of 2-year time ranges used for productivity, by taking the total of actual citations divided by the total benchmark citations for all publications by a given person in a given time range, and then averaging these ratios over the individuals in each group (e.g., K22 awardees).

3.3.8 Outcomes of Individuals without Subsequent Appointment Information in IMPAC II

To learn more about the career outcomes of individuals without subsequent appointment information in IMPAC II, Internet searches for a subset of NCI K applicants (both awardees and non-awardees) were conducted. This random subset contained 105 distinct individuals (excluding K04 and K12) who applied to the program, with distribution across mechanisms comparable to that in the full cohort. For additional information, see **Section 1.4.2**.

3.4 Validation of the Comparison Cohort

When using the comparison cohort to assess the impact of participation in the K program, it was important that the demographic characteristics of the comparison, or bubble, cohort reflected that of the population applying to the program as a whole. As shown in **Table 12**, across all parameters analyzed, the comparison cohort generally reflects the full cohort.

Parameter	Category	Full Cohort - Applicants (n = 2,893)	% Full Cohort	Comparison Cohort – Applicants (n = 586)	% Comparison Cohort
Primary K Mechanism	K01	479	16.1%	100	17.1%
	K07	562	18.8%	82	14.0%
	K08	1,176	39.4%	284	48.5%
	K11	166	5.6%	20	3.4%
	K22	200	6.7%	42	7.2%
	K23	254	8.5%	50	8.5%
	K25	56	1.9%	8	1.4%
Degree Type	PhD	869	29.1%	160	27.3%
	MD	1,209	40.5%	245	41.8%
	MD/PhD	613	20.5%	157	26.8%
	Dual	43	1.4%	6	1.0%
	Other	40	1.3%	9	1.5%
	Note	4	0.1%	0	0.0%
	Unknown	115	3.9%	9	1.5%
Gender	Male	1,664	57.5%	365	62.3%
	Female	1,055	36.5%	204	34.8%
	Unknown	174	6.0%	17	2.9%
Race/Ethnicity	White	1,519	52.5%	313	53.4%
	Hispanic	49	1.7%	13	2.2%
	Black	42	1.5%	10	1.7%
	Asian	450	15.6%	97	16.6%
	Native American	*	*	*	*
	Other	6	0.2%	4	0.7%
	Unknown	826	28.6%	149	25.4%
Prior Support	Had T Support	1,020	35.3%	247	42.2%
	Had Only T Support	806	27.9%	189	32.3%
	Had F Support	217	7.5%	47	8.0%
	Had Only F Support	112	3.9%	21	3.6%
	Had L Support	144	5.0%	30	5.1%
	Had Only L Support	53	1.8%	7	1.2%
	Had RPG Support	128	4.4%	25	4.3%
	Had Only RPG Support	70	2.4%	11	1.9%
	Had Multiple T, F, or L Support	165	5.7%	41	7.0%
	Had Multiple Support, including RPG	50	1.7%	12	2.0%
	Had Only Other Support	33	1.1%	8	1.4%
	No Prior Support	1,573	54.4%	286	48.8%

Parameter	Category	Full Cohort - Applicants (n = 2,893)	% Full Cohort	Comparison Cohort – Applicants (n = 586)	% Comparison Cohort
Institution Type (by applications)†	NCI-Designated Comprehensive Cancer Center	2,539	47.6%	343	48.4%
	NCI-Designated Cancer Center	1,231	23.1%	165	23.3%
	Not an NCI-Designated Cancer Center	1,560	29.3%	200	28.2%
Average Age at Application#		37.4	N/A	37	N/A
Average Years Since Degree^		7.7	N/A	7.9	N/A

Table 12. Demographic composition of the comparison and full cohorts.
Full cohort and comparison cohort include K01, K07, K08, K11, K22, K23, and K25 (K04 and K12 excluded). *: Indicates data that have been suppressed due to a low number of applicants, †: Institution Type data is for applications rather than individual applicants (n = 5,330 for full cohort; n = 708 for comparison cohort). #: Full cohort had age information available for 2,627 applicants; comparison cohort had age information for 551 applicants. ^: Full cohort had Years Since Degree information available for 2,330 applicants; comparison cohort had Years Since Degree information for 508 applicants. Definition of Dual and Note degree types may be found in Appendix 6.3.

3.5 Subsequent NIH Research Funding

3.5.1 Subsequent NIH Funding High Water Mark

One of the key program goals of the NCI K career development series is to prepare participants for careers in cancer research. One indication of a subsequent research career is an applicant's actions and success in obtaining subsequent research funding. As described in Section 3.3.5, we obtained information on subsequent NIH and NCI funding and categorized outcomes by high water mark.

Participation in NIH's mentored K programs was correlated with an increased likelihood of subsequent research activities across NIH for funded comparison cohort members[27]. This was also found to be the case for NCI's K awardees in both the full and comparison cohorts. Awardees in the comparison cohort had more awarded grants from all NIH ICs (including NCI) than non-awardees (56% in high water mark categories 1-5 for awardees, compared to 43% in non-awardees; Fisher's exact two-tailed test p=0.0038; **Figure 8**).

[27] See the NIH Mentored K evaluation report, Figure 15, available online at http://grants.nih.gov/training/K_Awards_Evaluation_FinalReport_20110901.pdf. Last accessed October 27, 2011.

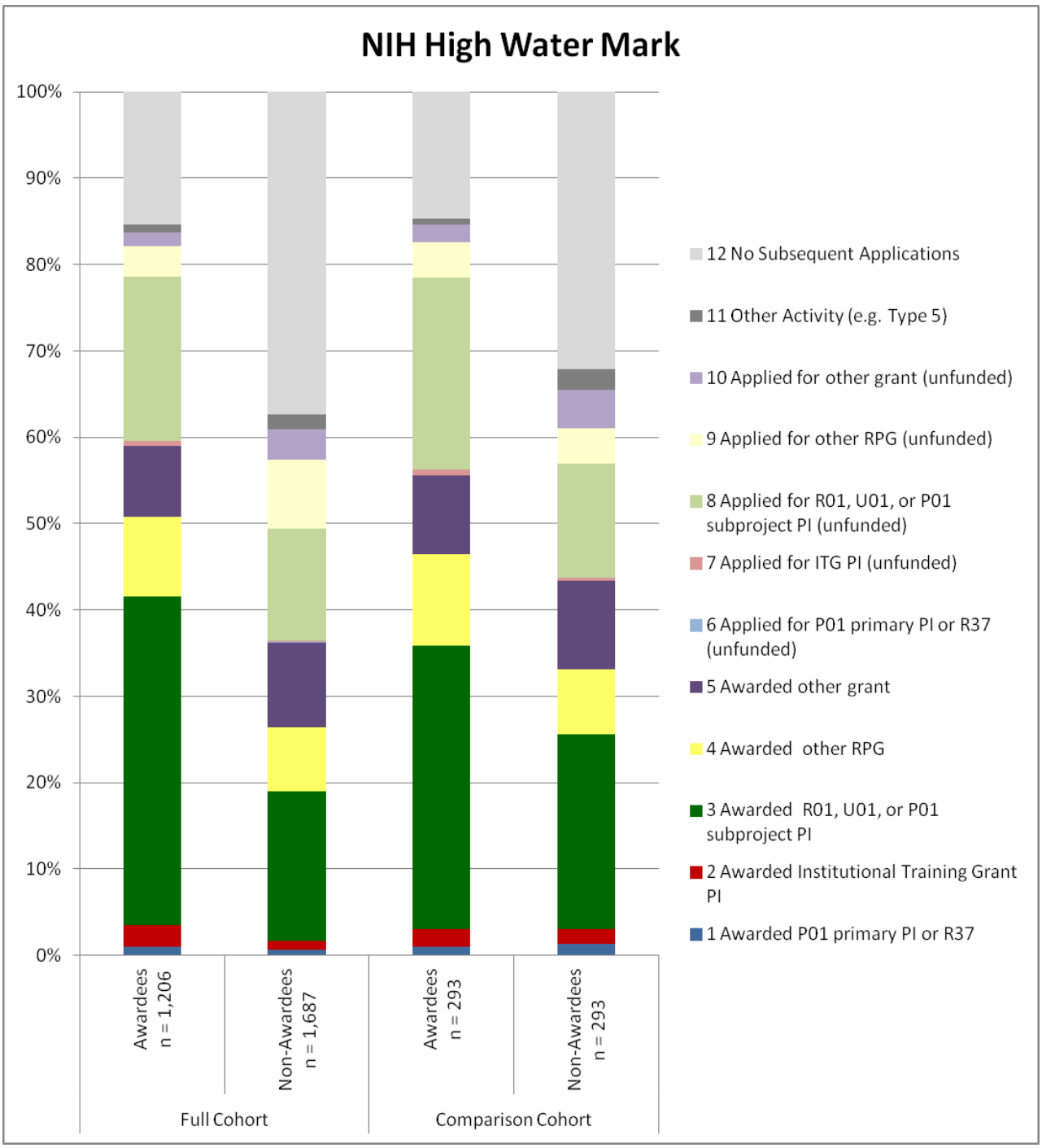


Figure 8. Attainment of NIH high water mark categories in the full and comparison cohorts, by funding status.
Individuals are represented only once in each bar in the highest category achieved. Analysis includes funding from all NIH ICs, including NCI.

3.5.2 Subsequent NCI Funding High Water Mark

K awardees in the comparison cohort had more awarded grants from NCI (high water mark categories 1-5) than non-awardees (43% compared to 29%; Fisher's exact two-tailed test p=0.0008; **Figure 9**). In addition, 76% of awardees in the comparison cohort had some level of NCI involvement (either awards

or unfunded applications, high water mark categories 1-10) compared to 53% of non-awardees (p=6.3E-9). Thus participants in the K program are significantly more likely to be involved in subsequent NCI research than similar unsuccessful applicants.

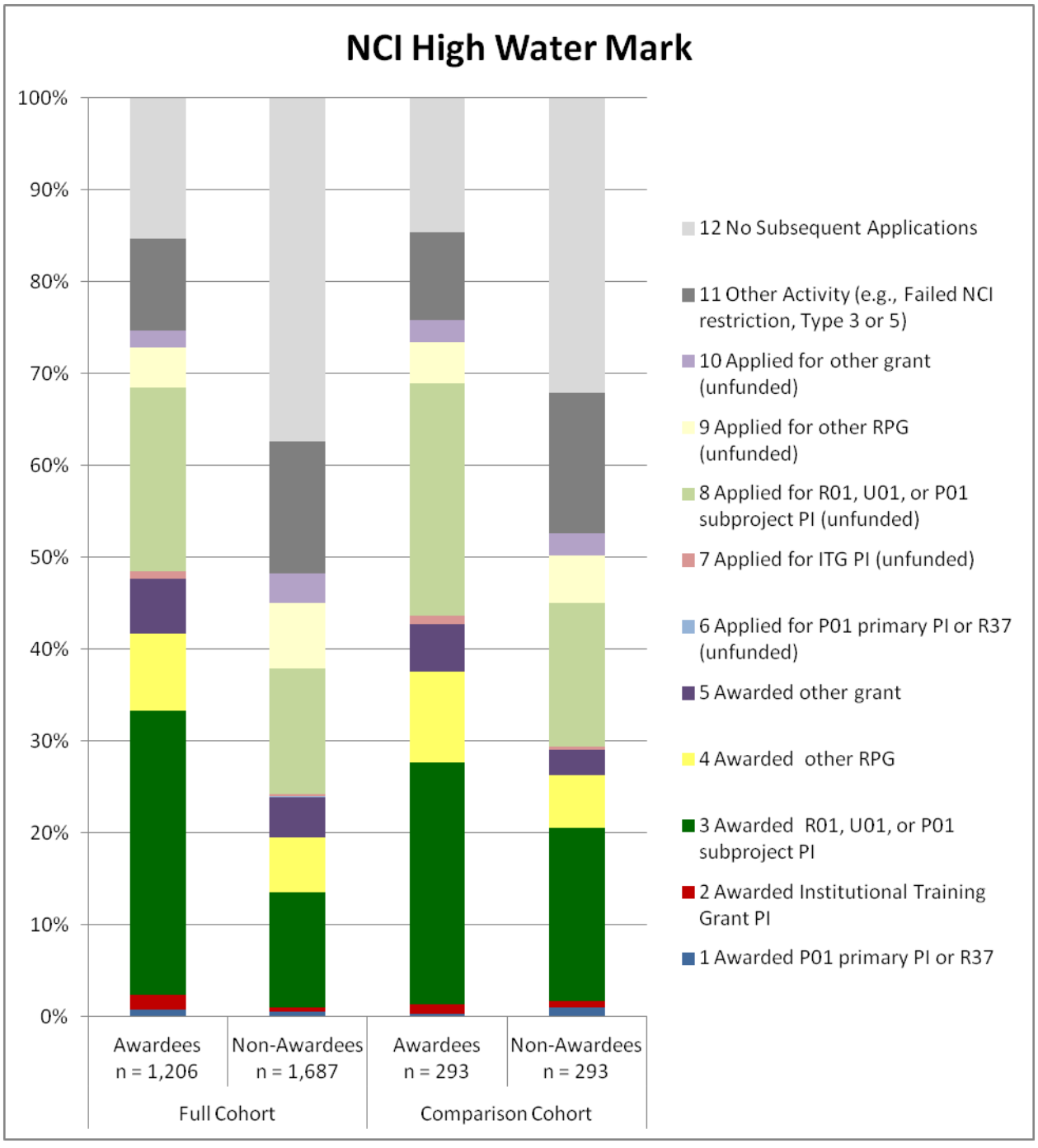


Figure 9. Attainment of NCI high water mark categories in the full and comparison cohort, by funding status.
Individuals are represented only once in each bar in the highest category achieved. For the NCI-specific analysis, category 11 (Other Activity) includes those individuals with applications to or awards from NIH ICs other than NCI (categories 1-10 in Figure 8), or type 3 or 5 NCI awards. Among all applicants placed in Category 11, the most common type of non-NCI activity was Category 3 (22% full cohort, 18% bubble cohort) followed by Category 5 (19% full cohort, 12% bubble cohort), Category 8 (19% full cohort, 25% bubble cohort) and Category 9 (14% full cohort, 8% bubble cohort).

3.5.3 Time to R01

To examine the effect that receipt of a K award had on an applicant's career progression, we determined the length of time between K application and subsequent R01 awards for individuals in the full cohort (**Figure 10**) and comparison cohort (data not shown due to small cohort sizes). Overall, the mean time to R01 for awardees and non-awardees was within 1 year of each other and the distributions have substantial overlap. A Mann-Whitney test of median differences was performed for each cohort, and the only significant difference was detected among the K01 awardees and non-awardees in the full cohort, in which non-awardees attained their first R01 in 3 years versus the awardees' 3.5 years (difference of 6 months) (data not shown). These findings reflect those seen in the NIH mentored K evaluation.[28]

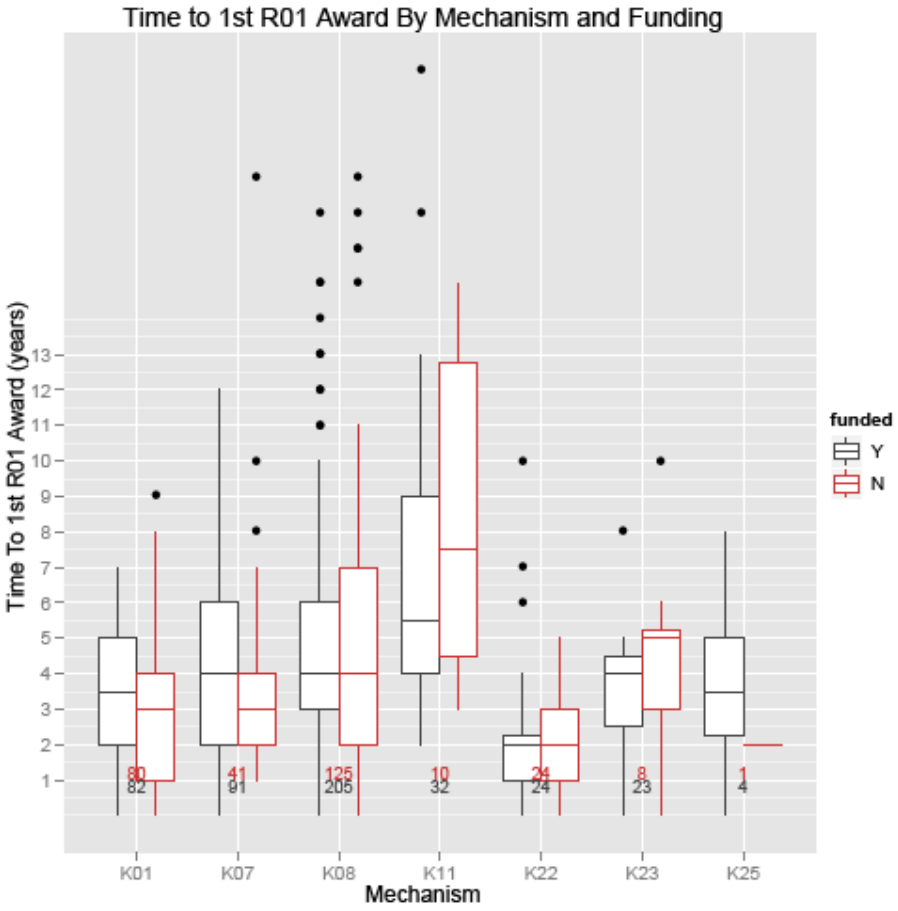


Figure 10. Time to first NIH R01 following K application by mechanism and K funding status, full cohort
In this representation, the box represents the interquartile range (IQR) with the middle horizontal line of the box representing the median, the lower horizontal line of the box represents the 1st quartile and the upper horizontal line of the box represents the third quartile. The lower line (whisker) represents the first quartile – 1.5x the interquartile range, and the upper line (whisker) represents the third quartile + 1.5x the interquartile range. Dots above the upper whisker represent data points that are considered outliers. The red line for unfunded K25s represents the single available datapoint. Numbers shown below the lower whisker indicate the awardees (black) or non-awardees (red).

[28] See the NIH Mentored K evaluation report, available online at http://grants.nih.gov/training/K_Awards_Evaluation_FinalReport_20110901.pdf. Last accessed October 27, 2011.

3.6 Subsequent Research Funding from Non-NIH Sources

Pursuit and receipt of subsequent funding from other funding agencies is another indicator of a research-oriented career (**Table 13**). Among US Federal agencies, the Department of Defense (DOD), Department of Energy (DOE), and National Science Foundation (NSF) offer funding for cancer-related research projects. The DOD's Congressionally Directed Medical Research Program (CDMRP) administered over $6.5 billion in research appropriations during the period of FY1992 through FY2011, covering a wide range of diseases, including various cancer types, neurological disorders, and deployment-related health issues[29]. DOE and NSF also offer research project grants to support biomedical discoveries, although on a smaller scale than NIH. Across mechanisms for both the full (data not shown) and comparison cohort, a handful of NCI K applicants were matched to DOD-CDMRP, DOE, and NSF awards. The K07 and K08 mechanisms had the greatest participation within the DOD-CDMRP program. The K25 mechanism had the largest number of applicants who pursued NSF support, likely reflecting the intent of the K25 program to encourage physical sciences researchers to collaborate with biomedical researchers.

To determine if NCI K applicants had received non-government research support, data from the International Cancer Research Portfolio (ICRP)[30], a database of information on cancer research awards sponsored by the International Cancer Research Partners, including the NCI and DOD-CDMRP, non-profit cancer research organizations, and other international groups were used (**Table 13**). NCI K applicants in both the full and comparison cohorts received funding from several of the ICRP organizations, namely the American Cancer Society, Susan G. Komen Breast Cancer Foundation, California Breast Cancer Research Foundation, and the Prostate Cancer Foundation.

[29] More information on DOD-CDMRP can be found at http://cdmrp.army.mil/default.shtml. Last Accessed October 17, 2011.
[30] http://www.cancerportfolio.org/faq.jsp. Last accessed October 17, 2011.

Mechanism	Organization	Number of K Awardees	Future Select Grants - K Awardees	Number of K Non-Awardees	Future Select Grants - K Non-Awardees
K01	American Cancer Society	2	2	0	0
	Susan G. Komen Breast Cancer Foundation	0	0	1	1
	U. S. Department of Defense, CDMRP	2	2	3	3
	National Science Foundation	0	0	2	2
	U.S. Department of Energy	0	0	3	3
K07	California Breast Cancer Research Program	1	2	0	0
	U. S. Department of Defense, CDMRP	2	3	1	1
K08	California Breast Cancer Research Program	1	1	0	0
	U. S. Department of Defense, CDMRP	3	6	4	6
	U.S. Department of Energy	3	6	0	0
K11	U. S. Department of Defense, CDMRP	1	2	0	0
K12 Scholar	U.S. Department of Defense, CDMRP	4	6	N/A	N/A
	U.S. Department of Energy	2	2	N/A	N/A
K22	Canadian Institutes of Health Research	0	0	1	1
	Prostate Cancer Foundation	0	0	1	1
	Susan G. Komen Breast Cancer Foundation	0	0	1	4
	U. S. Department of Defense, CDMRP	0	0	1	5
	National Science Foundation	1	1	1	1
K23	National Science Foundation	0	0	1	1
K25	National Science Foundation	4	4	2	2

Table 13. Subsequent non-NIH research funding, comparison cohort.[31]
Table shows the number of awardees and non-awardees in the comparison cohort who received subsequent grants from organizations within the ICRP, the NSF, or the DOE. The third and fifth columns show the number of individuals, and the fourth and sixth columns show the total number of awards received by those matched individuals.

3.7 Subsequent Publications

In addition to grants, peer-review publications are a relevant indicator of subsequent research activity. First, metadata from subsequent publications were mined to discern area of research for program participants. **Table 14** shows the most frequent journal subject categories represented in the full cohort awardee publication set, and **Table 15** the most frequent journals in which awardees in the full cohort were published. These data show that 11,602 of the more than 15,000 articles published by awardees are found in journals with an emphasis on cancer research and cancer-related fields. Non-awardees had similar trends for both journal subject category and journal (data not shown).

[31] NCI K applicant names were matched to the International Cancer Research Portfolio database (includes American Cancer Society, Susan G. Komen Breast Cancer Foundation, California Breast Cancer Research Program, Canadian Institutes of Health Research, Prostate Cancer Research Program and U.S. Department of Defense, CDMRP) and the ScienceWire Grants Catalog (National Science Foundation and U.S. Department of Energy).

WoS Journal Subject Category	Total Number of Articles
Oncology	5,636
Biochemistry & Molecular Biology	1,937
Hematology	1,489
Cell Biology	1,452
Public, Environmental & Occupational Health	1,125
Immunology	982
Genetics & Heredity	886
Medicine, General & Internal	676
Surgery	652
Medicine, Research & Experimental	608

Table 14. Top 10 Web of Science Journal Subject Categories, awardees in the full cohort.
Journals can be assigned to more than one Journal Subject Category, so the sum of the number of articles in each of the top 10 journal categories is greater than the number of distinct articles in the top 10 (n=11,602). The total number of articles published by awardees in journals linked to a WoS Journal Subject Category is 15,101. The K04 program is excluded from this analysis.

Journal Title	Total Number of Articles
Cancer Research	697
Blood	629
Journal of Clinical Oncology	463
Journal of Biological Chemistry	403
Clinical Cancer Research	402
Cancer Epidemiology Biomarkers & Prevention	393
Cancer	346
Proceedings of The National Academy of Sciences of The United States of America	297
Oncogene	235
Journal of Immunology	231

Table 15. Top 10 journals, awardees in the full cohort.
The K04 program is excluded from this analysis.

To assess whether participation in the K program affected subsequent publication activity, we examined publication activity of the full cohort, beginning the fiscal year after receipt of a K award (for awardees) or after the last K application (for non-awardees). **Table 16** shows the number and percent of NCI K awardees and non-awardees that authored subsequent publications by mechanism. Across all mechanisms, a larger proportion of awardees had subsequent research publications. In addition to awardees being more likely to publish, among those who publish, the average and median number of publications per awardee is higher than for non-awardees. Thus participation in the K program has a measurable and significant positive effect on subsequent publication productivity.

Mechanism	Funding Status	Total Individuals	Individuals with Subsequent Publications	% with Subsequent Publications	Total Publications	Mean Subsequent Publications per Author	Median Subsequent Publications per Author
K01	Awardee	152	140	**92.1%**	1,476	**10.6**	**7.5**
K01	Non-Awardee	327	191	58.4%	1,517	8.0	5.0
K04	Awardee	341	293	**85.9%**	14,133	**48.7**	**31.0**
K04	Non-Awardee	748	514	68.7%	18,942	37.4	23.0
K07	Awardee	274	256	**93.4%**	4,409	**18.6**	**11.0**
K07	Non-Awardee	288	209	72.6%	2,325	11.4	5.0
K08	Awardee	514	443	**86.2%**	7,638	17.6	**10.0**
K08	Non-Awardee	662	379	57.3%	6,727	**18.0**	8.0
K11	Awardee	86	71	**82.6%**	1,851	26.4	10.0
K11	Non-Awardee	80	27	33.8%	520	19.3	8.0
K22	Awardee	57	53	**93.0%**	462	8.7	6.0
K22	Non-Awardee	143	78	54.5%	544	7.0	4.0
K23	Awardee	98	92	**93.9%**	1,505	**16.5**	**11.0**
K23	Non-Awardee	156	100	64.1%	1,243	12.5	7.0
K25	Awardee	25	21	**84.0%**	159	7.6	6.0
K25	Non-Awardee	31	12	38.7%	104	8.7	2.5

Table 16. Analysis of publications authored by NCI K applicants, by mechanism and funding status.
Significant differences ($p<0.05$) between awardees and non-awardees within each mechanism are indicated by bolding the higher value within each mechanism.

We next tested how these differences in publication activity changed over time. **Figure 11** shows the publications per person for all awardees or non-awardees normalized over two-year time ranges for each mechanism. As shown here, K awardees published more papers per year compared to non-awardees, particularly in the years following the start of K program. When this same analysis was performed on only those awardees and non-awardees that published, the differences in productivity per year were reduced, although in later years (years six through thirteen), K01 and K11 awardees saw increases in productivity that were not matched by the non-awardee group (data not shown).

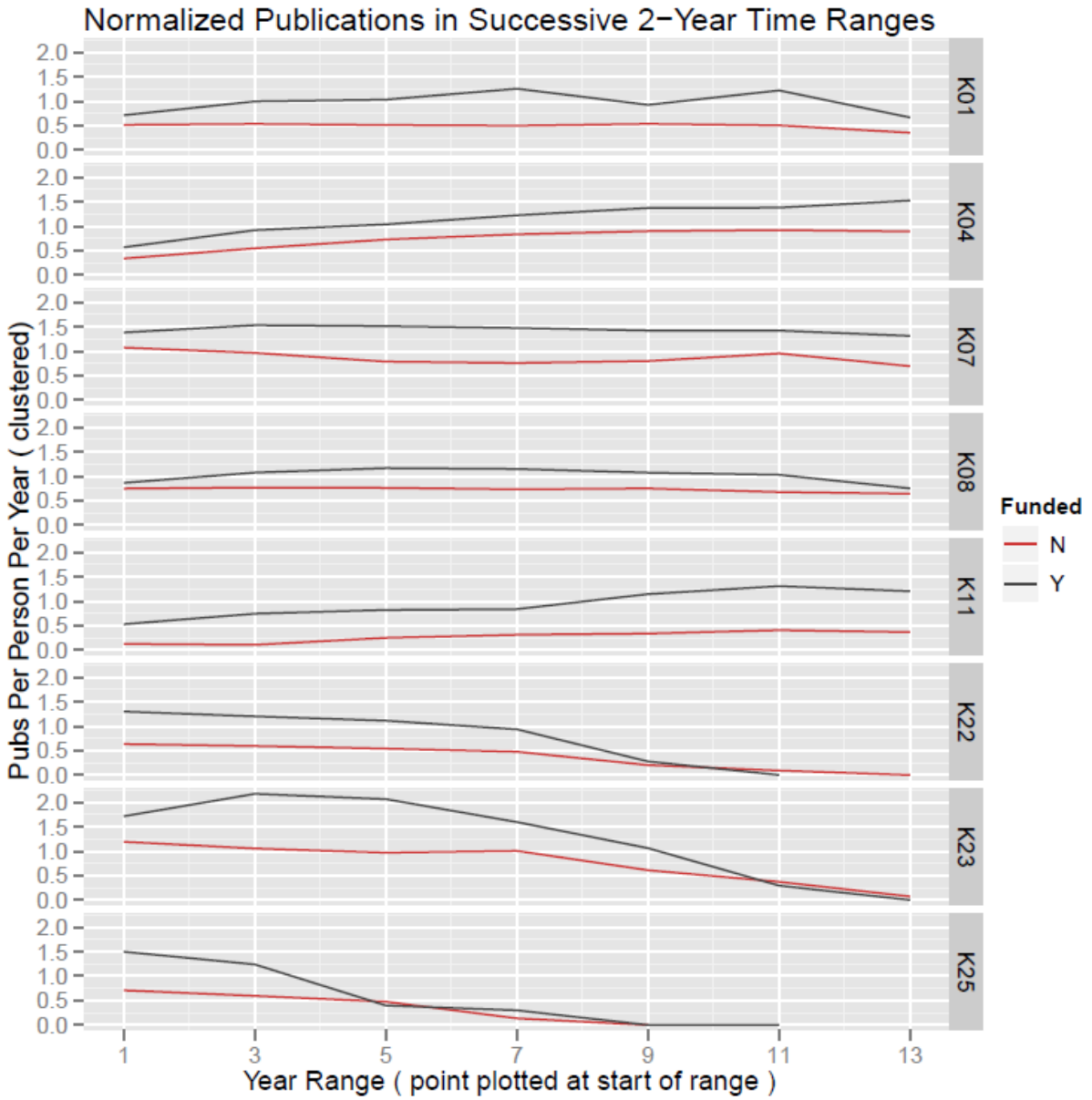


Figure 11. Publications per person per year for all individuals in the full cohort, for 2-year time slices.

Next we examined the impact of these publications by comparing the observed 2 year citation rates for the subsequent publications for the NCI K awardees or non-awardees to a benchmark citation rate for similar publications published during a similar time frame (see Section 3.3.7 for methodology). As shown in **Figure 12**, papers published by the K awardee group had a greater impact within the first two years following publication when compared to those papers published by non-awardees. This difference in impact was most notable in the first three years following the start of K awardee participation, consistent with the program having a direct effect on the quality of the participant's publications. K04, K07, K08 and K11 awardees were notable for their consistently higher citation ratio compared to their respective non-awardees. The overall effects of this analysis were reduced when restricted to only

those individuals who authored papers, but yielded similar results in the years immediately following the start of K program participation (years two through six) in all cases except K22s (data not shown).

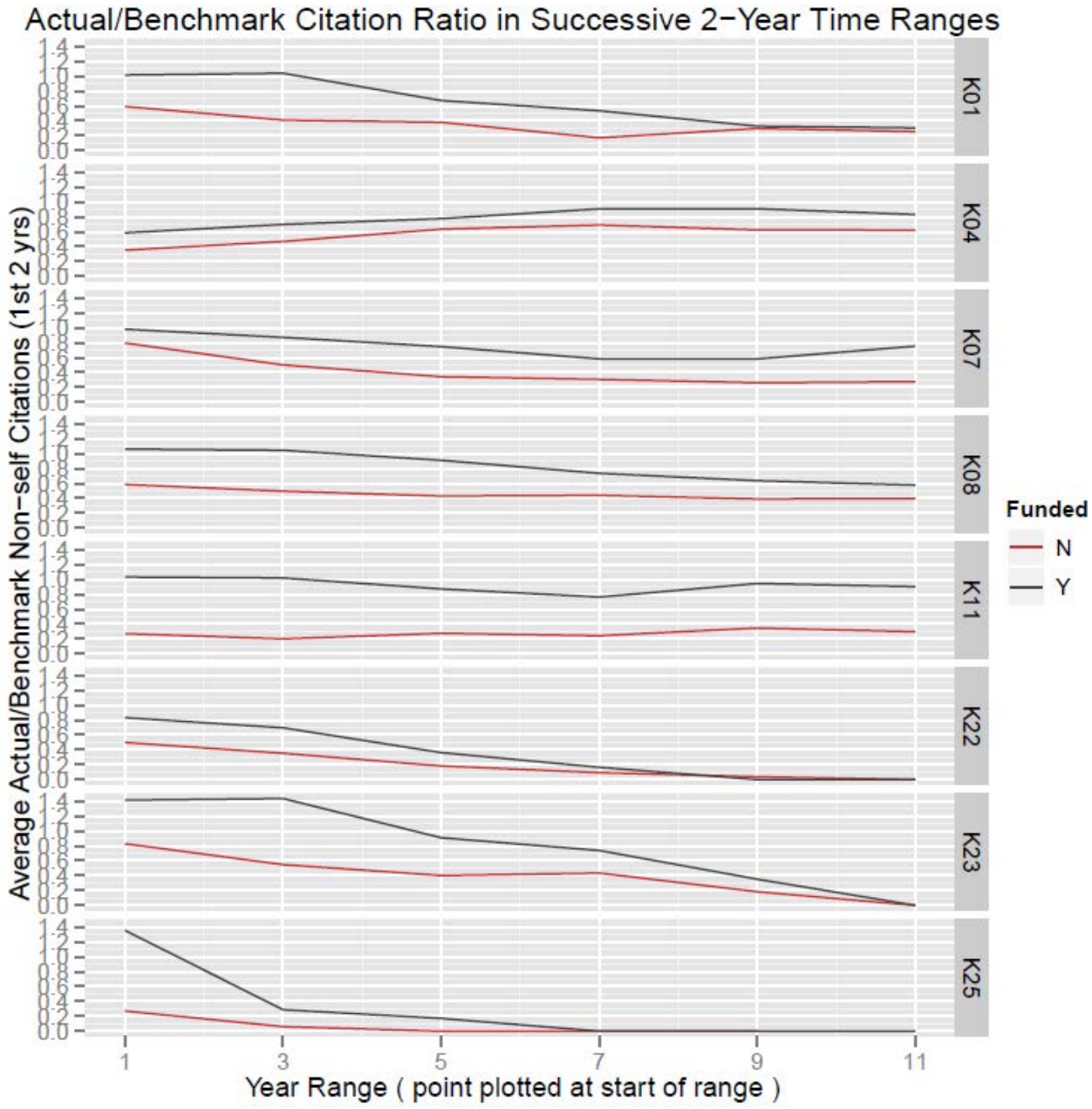


Figure 12. Publication impact analysis per person in the full cohort, for 2-year time slices.

3.8 K25 Abstracts Pre- and Post-Award

In addition to encouraging awardees to contribute to cancer research, the K25 mechanism has the goal of encouraging researchers with training in fields outside of biology and medicine to apply their expertise to studies in cancer research. To determine if receipt of a K25 award resulted in a subsequent

shift of an investigator's research, an analysis of publication subject areas pre- and post-award can be conducted by reviewing and coding abstracts. While this analysis is desirable, the K25 mechanism was only first awarded in 2001, with a total pool of 6 applicants in the comparison cohort. While initial findings suggested that some awardees continue or increase their publications in cancer-related fields, a larger sample population is necessary to confirm this analysis.

3.9 Subsequent Career Appointments

To understand how participation in the NCI K program affected employment outcomes, we analyzed data on faculty level appointments and physician listings as healthcare providers.

3.9.1 Faculty Appointments

Using data from IMPAC II and the AAMC Faculty Roster to determine the most recent faculty appointment, we found that across all of the K mechanisms analyzed, members of the K awardee cohort had proportionally more individuals attain a current rank of Associate Professor than non-awardees (**Table 17**). The most recent academic rank outcomes differ by mechanism, with K04s achieving predominantly Professor-level rank subsequent to the K program, commensurate with the longevity of the program and that participants were typically later-stage researchers at time of application (data not shown). The majority of K01, K07, K22, K23 and K25 applicants achieved current ranks at the Assistant or Associate Professor-level. More than half of the K08 awardees held a current rank of Associate Professor or Professor and in greater proportion than non-awardees. The K08 program has been administering awards for over 25 years, so it is to be expected that participants would achieve Associate- and Professor-level ranks compared to other mechanisms.

Appointment Rank	K01		K07		K08		K11		K22		K23		K25	
	Awardee n = 152	Non-Awardee n = 327	Awardee n =274	Non-Awardee n = 288	Awardee n = 514	Non-Awardee n = 662	Awardee n = 86	Non-Awardee n = 80	Awardee n = 57	Non-Awardee n = 143	Awardee n = 98	Non-Awardee n = 156	Awardee n = 25	Non-Awardee n = 31
Professor	7 (5%)	12 (4%)	43 (16%)	27 (9%)	117 (23%)	115 (17%)	25 (29%)	13 (16%)	1 (2%)	2 (1%)	9 (9%)	6 (4%)	1 (4%)	1 (3%)
Associate Professor	32 (21%)	36 (11%)	67 (24%)	39 (14%)	146 (28%)	107 (16%)	18 (21%)	11 (14%)	20 (35%)	16 (11%)	36 (37%)	33 (21%)	3 (12%)	1 (3%)
Assistant Professor	76 (50%)	116 (35%)	92 (34%)	109 (58%)	159 (31%)	179 (27%)	15 (17%)	11 (14%)	24 (42%)	54 (38%)	35 (36%)	67 (43%)	7 (28%)	7 (23%)
Instructor	4 (3%)	22 (7%)	7 (3%)	18 (6%)	21 (4%)	35 (5%)	6 (7%)	3 (4%)	1 (2%)	3 (2%)	7 (7%)	11 (7%)	1 (4%)	3 (10%)
Other	4 (3%)	7 (2%)	8 (3%)	7 (1%)	4 (1%)	13 (2%)	3 (3%)	0 (0%)	1 (2%)	5 (3%)	0 (0%)	0 (0%)	2 (8%)	0 (0%)

Table 17. Subsequent faculty appointments, full cohort.
Numbers indicate the count of individuals at the corresponding current academic rank; percentages show the proportion of total awardees or non-awardees within each mechanism. Excludes K04s and K12 scholars. Data from IMPAC II and AAMC Faculty Roster.

3.9.2 Medical Practice

Using the HealthLink database of physicians and healthcare providers, we determined that close to 40% of K applicants were listed as physicians following their application to the K program (**Table 18**). As expected, the K mechanisms geared toward clinical doctorates had the highest number of registered physicians. A greater percentage of K awardees (47%) were listed as physicians than non-awardees (33%).

Mechanism	Awardees in Mechanism	Number of Awardees Matched to Providers in HealthLink (% of total)	Non-Awardees in Mechanism	Number of Non-Awardees Matched to Providers in HealthLink (% of total)
K01	152	12 (7.9%)	327	9 (2.8%)
K07	274	76 (27.7%)	288	64 (22.2%)
K08	514	353 (68.7%)	662	342 (51.7%)
K11	86	50 (58.1%)	80	33 (41.3%)
K22	57	12 (21.1%)	143	25 (17.5%)
K23	98	69 (70.4%)	156	85 (54.5%)
K25	25	0 (0%)	31	0 (0%)
TOTAL	1,206	572 (47.4%)	1,687	558 (33.1%)

Table 18. Subsequent private practice[32] involvement, full cohort.
Results are based on a name match between the NCI K applicants and the physicians and healthcare providers in the HealthLink database. K04s and K12 scholars were excluded from this analysis.

3.10 Subsequent Participation in Clinical Research

One of the goals of the K08, K12, and K23 mechanisms is to recruit and retain early-stage investigators with clinical degrees in translational or patient-oriented research, an example of which is clinical trials. To determine the amount of subsequent engagement of K08, K12, and K23 applicants in clinical research activities, names of applicants were matched to those of key personnel on clinical trials registered in the NIH-sponsored ClinicalTrials.gov database (**Table 19**). We matched 34.6% of K08 awardees to 590 trials, 75.5% of K23 awardees to 450 trials, and 46.4% of K12 scholars to 623 trials. K08 and K23 non-awardees showed lower rates of participation than awardees as key personnel, but were matched to slightly more trials. However, the average number of trials per person was higher for K23 awardees than non-awardees.

Mechanism	Awardees Associated with Clinical Trials (% Awardees)	Number of Trials Matched to Awardees	Non-Awardees Associated with Clinical Trials (% Non-Awardees)	Number of Trials Matched to Non-Awardees
K08	178 (34.6%)	590	195 (29.5%)	747
K12 Scholar	173 (46.4%)	623	N/A	N/A
K23	74 (75.5%)	450	93 (59.6%)	497

[32] HealthLink is not explicitly restricted to physicians in private practice; they may have a joint appointment at a medical school/university which cannot be determined through this data source.

Table 19. NCI K08 and K23 applicant and K12 scholar participation in clinical trials, full cohort.
Applicants and scholars were matched to clinical trial personnel in the clinicaltrials.gov database by name. The full list (373) of K12 scholars was used for matching.

3.11 U.S. Patent Applications and Issued Patents

To determine whether research conducted while a researcher was supported by an NCI K mechanism contributed to a patent awarded by the U.S. Patent and Trademark Office (USPTO), the list of K award numbers included in this evaluation were matched to grant numbers appearing in the government interest section of USPTO patents awards (see **Appendix 6.9**) using the ScienceWire patent catalog. A total of 28 NCI K awards in the study sample (2.7%) were acknowledged in 50 USPTO patents, spanning the K01, K07, K08, and K11 mechanisms. The results from the name matching of NCI K applicants to USPTO patent assignees were not used, as there was not enough evidence to verify the match.

3.12 Service on NIH Review Groups and Other Federal Advisory Committees

Service on federal advisory committees is generally an indication that an individual is a notable contributor in their research area, and also suggests that the individual is an active participant in the research community. Furthermore, NIH review group service is correlated with likelihood of success on NIH R01 grant applications.[33] We investigated whether K applicants served on NIH review panels and other federal advisory committees, and compared service between funded and non-funded groups for the full cohort.

Using IMPAC II, we were able to identify K applicants who had served on several NIH committees. In all categories examined, K awardees were more highly engaged than the non-awardees (**Table 20**). This largely appears to be the case for federal advisory panel participation (**Table 21**), indicating that K awardees are valued members of the scientific community.

NIH Committee Category	NCI K Applicants	NCI K Awardees	Percentage of NCI K Awardees	NCI K Non-Awardees	Percentage of NCI K Non-Awardees	NCI K12 Scholars	Percentage of NCI K12 Scholars
National Advisory Council/Board	29	17	1.4%	12	0.7%	0	0%
Program Advisory Committee/Council and/or Board of Scientific Counselors	103	71	5.9%	32	1.9%	1	0.30%
Special Emphasis Panel and/or Initial Review Group	1,049	615	51.0%	434	25.7%	37	9.9%
No Service Found	1,838	587	48.7%	1,251	74.2%	336	90.1%

Table 20. Service of NCI K applicants on NIH review groups, full cohort.
Review group service data for the full study cohort (K01, K07, K08, K11, K12 scholars, K22, K23, and K25) was obtained from the IMPAC II cmte_type_code field, and committees were grouped into broad categories based on input from NCI. An individual may appear in multiple categories, with the exception of those in the category "No Service Found." The full list (373) of K12 scholars was used for matching.

[33] See, for example, Ginther, et al. (2011) **Science**, 333 (6045): 1015 -1019.

Mechanism	Agency	Matched Applicants (% within Primary Mechanism)	Awardees	Percent Awardees Within Primary Mechanism	Non-Awardees	Percent Non-Awardees Within Primary Mechanism
K01	Department of Health and Human Services	33 (6.8%)	13	8.6%	20	6.1%
	National Science Foundation	5 (1.0%)	2	1.3%	3	0.9%
K07	Department of Health and Human Services	60 (10.7%)	32	9.4%	28	9.7%
	National Endowment for the Arts	1 (0.2%)	0	0.0%	1	0.3%
	National Endowment for the Humanities	1 (0.2%)	0	0.0%	1	0.3%
	National Science Foundation	3 (0.5%)	1	0.3%	2	0.7%
K08	Department of Health and Human Services	80 (6.8%)	45	8.8%	35	5.3%
	Department of State	1 (0.9%)	1	0.2%	0	0.0%
	National Endowment for the Arts	1 (0.9%)	1	0.2%	0	0.0%
	National Science Foundation	6 (0.5%)	2	0.4%	4	0.6%
K12 Scholar	Department of Health and Human Services	10 (2.7%)	N/A	N/A	N/A	N/A
	National Science Foundation	3 (0.8%)	N/A	N/A	N/A	N/A
K11	Department of Health and Human Services	13 (9.6%)	6	7.0%	7	8.8%
	Department of State	1 (0.6%)	1	1.2%	0	0.0%
	Department of Veterans Affairs	1 (0.6%)	1	1.2%	0	0.0%
K22	Department of Defense	1 (0.5%)	0	0.0%	1	0.7%
	Department of Health and Human Services	14 (7.0%)	7	12.3%	7	4.9%
K23	Department of Health and Human Services	17 (6.7%)	6	6.1%	11	7.1%
	Department of the Interior	1 (0.4%)	0	0.0%	1	0.6%
	National Science Foundation	2 (0.8%)	1	1.0%	1	0.6%
K25	Department of Health and Human Services	3 (5.4%)	2	8.0%	1	3.2%

Table 21. Participation of NCI K applicants on Federal advisory committees, full cohort.

A name match of the full NCI K cohort (including K12 scholars) was performed against data from the Federal Interagency Databases Online (www.fido.gov) Federal Advisory Committee Act (FACA) Database Committee Management Secretariat Website. Department of Health and Human Services committees may overlap with committees presented in Table 20.

3.13 Membership in Professional Societies

An additional measure of involvement in the research enterprise is membership in scientific professional societies. We engaged three professional societies to determine whether K applicants were present on active membership rosters: American Association for Cancer Research, American Society of Clinical Oncology, and the Federation of American Societies for Experimental Biology. As shown in **Table 22**, a greater percentage of K awardees from the full cohort for each mechanism are members of the three societies compared to the non-awardee group, consistent with K awardees being involved to a greater extent in the scientific community than unfunded applicants.

American Association for Cancer Research		
Mechanism	**Percent Awardee Members**	**Percent Non-Awardee Members**
K01	40.1%	35.2%
K07	37.6%	14.6%
K08	43.4%	33.1%
K11	38.4%	17.5%
K22	50.9%	39.2%
K23	43.9%	39.7%
K25	28.0%	22.6%
American Society of Clinical Oncology		
Mechanism	**Percent Awardee Members**	**Percent Non-Awardee Members**
K01	4.6%	4.9%
K07	21.2%	20.1%
K08	46.5%	42.7%
K11	43.0%	35.0%
K22	21.1%	17.5%
K23	76.5%	67.3%
K25	8.0%	0.0%
Federation of American Societies for Experimental Biology		
Mechanism	**Percent Awardee Members**	**Percent Non-Awardee Members**
K01	23.0%	17.1%
K07	12.4%	7.6%
K08	20.2%	14.2%
K11	25.6%	8.8%
K22	26.3%	12.6%
K23	10.2%	5.8%
K25	4.0%	3.2%

Table 22. Participation of NCI K applicants as members of select professional societies, full cohort.
A name match of the full NCI K applicant cohort was performed against the active membership databases of the American Association for Cancer Research, American Society of Clinical Oncology, and the Federation of American Societies for Experimental Biology.

3.14 Outcomes of Individuals without Subsequent Appointment Information

For a number of individuals in the study sample, there was no information in IMPAC II subsequent to the K application or award. To better understand outcomes for this group, we generated a random sample of 105 individuals, including awardees and non-awardees across mechanisms (excluding K04 and K12). The overall distribution by gender and degree matched that of the full cohort. **Table 23** shows the distribution of individuals across mechanism and by funding status within this group.

Mechanism	Awardees	Non-Awardees
K01	10	9
K07	14	7
K08	18	27
K11	8	3
K22	1	3
K23	0	3
K25	2	0
TOTALS	**53**	**52**

Table 23. Distribution of applicants from the full cohort included in the manual search, by mechanism and funding status.

We found that the majority of K awardees (72%) were currently employed in academia, with only a small proportion (9%) employed in private medical practice. Non-awardees, however, were fairly evenly split, with 40% holding academic positions, and another 42% working in private medical practice. Similar proportions of awardees and non-awardees held leadership positions within industry, Federal agencies, and non-profit organizations (13% and 14%, respectively) (**Table 24**). In this sample, NCI K awardees were more likely to pursue careers in academia, while non-awardees were evenly split between academia and private medical practice. An equal proportion of both groups pursued other careers in the scientific enterprise. Career outcomes were unknown for five individuals.

Current Position	Total Number (% of Total)	NCI K Awardees (% of Awardees)	NCI K Non-Awardees (% of Non-Awardees)
Academe	**59 (56%)**	**38 (72%)**	**21 (40%)**
Professor	15	9	6
Associate Professor	11	10	1
Assistant Professor	19	15	4
Instructor / Adjunct Professor	4	0	4
Research Associate / Fellow	4	0	4
Clinical Professor	2	1	1
Other University Position	3	2	1
Unknown[1]	1	1	0
Private Medical Practice	**27 (26%)**	**5 (9%)**	**22 (42%)**
Other Scientific Enterprise	**14 (13%)**	**7 (13%)**	**7 (14%)**
President & Chief Executive Officer	2	1	1
President	1	1	0
Chief Medical Officer	1	0	1
Chief Scientific Officer	1	0	1
Associate Director for Science [U.S. Federal Agency]	1	0	1
Senior Medical Director	1	1	0
Senior Vice President of Global Development & Chief Medical Officer	1	1	0
Group Director	1	1	0
Analyst / Programmer / Principal Scientist	2	1	1
Other Government Positions (State / Federal)	3	1	2
Unknown	**5 (5%)**	**3 (6%)**	**2 (4%)**
Total	**105**	**53**	**52**

Table 24. Detailed breakdown of current positions of applicants included in the manual search, grouped by broad position category, funding status, and refined position title.

[1]Unknown category indicates individuals listed on a university website that did not list specific appointment details.

In the next section we describe logit and linear models using program input variables described in Section 2 and the outcome variables described in Section 3, to determine which program features are most highly correlated with funding and outcomes.

3.15 Summary of Applicant and Awardee Outcomes

Receipt of an NCI K award was correlated with higher likelihood of applying for and receiving subsequent NCI and NIH grants, publication productivity and impact, and greater participation in the scientific community.

Subsequent NIH Activity

- NCI K programs are beneficial for awardees. K award receipt is correlated with an increased likelihood of subsequent NIH and NCI grant applications and awards (high water marks), compared to unfunded applicants, both in the full and comparison cohort.
- Participation in the K program did not affect career progress. Most K mechanisms studied showed no difference between awardees and non-awardees in mean time to first R01. K01s were found to have a significant difference of 6 months between awardees and non-awardees (3.0 years versus 3.5 years, respectively).

Subsequent Funding from non-NIH Sources

- Most K awardees remained in the NIH funding sphere. For the small numbers who pursued research funding from non-NIH federal sources, K07 and K08 applicants were more likely to receive funding from DoD-CDMRP, while K25 applicants were most likely to be matched to subsequent NSF funding. International Cancer Research Portfolio (ICRP) partner organizations, such as the American Cancer Society, the Susan G. Komen Breast Cancer Foundation, Prostate Cancer Foundation, and the California Breast Cancer Research Program, were also sources of subsequent funding for NCI K awardees and non-awardees.

Subsequent Publications

- Participation in the K program was associated with higher research publication productivity and impact. A substantially higher proportion of awardees authored subsequent publications, and awardee-authored publications tended to be more highly cited within the first two years following publication than those authored by non-awardees, compared to a benchmark citation rate.
- Participants pursue cancer research after program completion. Overall, NCI K awardees continue to publish in cancer or cancer-related fields.

Subsequent Careers

- Across mechanisms, K awardees were 10-20% more likely to be in faculty positions than non-awardees, and more likely to be at the Associate versus Assistant Professor rank.
- A greater proportion of K awardees than non-awardees were later listed as practicing physicians; this proportion was higher in the K mechanisms with a clinical focus.

Subsequent Participation in the Research Community

- Awardees in clinically-focused NCI K mechanisms (K08, K23) tended to be more likely to serve as PIs in clinical trials than non-awardees.
- Across all NIH committees examined, K awardees were about twice as likely to be engaged as non-awardees. K awardees were also more likely to serve on Federal advisory panels.
- K awardees were more likely to be active members of the major cancer-related professional societies than non-awardees.
- For those awardees with no subsequent NIH involvement, we found almost all of them to be involved in the scientific enterprise. More K awardees were in academic positions than non-awardees, and more non-awardees were in private medical practice.

4.0 Models

In Section 2.0 of this report, we presented descriptive findings of the characteristics of the applicants and awardees of the NCI K program mechanisms. In Section 3.0 we evaluated the career and funding outcomes of the applicants and awardees and showed that program participation was clearly beneficial for subsequent career success. To provide options for targeted program modifications, in this section we investigate more closely which aspects of the K program had the most impact on participant careers. We combine data on program participants and outcomes using linear and logit regression models to determine which variables are most highly correlated with successful outcomes.

This section is organized according to the following key evaluation questions:

- What applicant characteristics are correlated with receiving a K award?
- What is the impact of K program participation on subsequent career outcomes?
- What are the mechanism-specific impacts of K program participation on career outcomes?

4.1 Regression Modeling of Applicant Characteristics and Outcomes

With nine K programs and variation between programs and their applicants, we chose to use predictive modeling to tease apart the contribution of various applicant characteristics. A series of regression models was created for each outcome of interest with the contribution of all input variables being considered simultaneously. **Appendix 6.10** includes information on the model composition, the input and output variables, and the findings. Regression models served as a means to identify the variables important in program participant outcomes, and were used to further address the key evaluation questions. We confirmed model findings by applying statistical tests using program data. We verified that the statistical tests used were not affected by correlation between variables (see **Appendix 6.10**). In this section, we present the findings from these statistical tests, unless the predictive models were more informative.

4.2 What Applicant Characteristics are Correlated with Receiving a K Award?

4.2.1 Applicant Race/Ethnicity and Degree Type

We examined various key characteristics of K program applicants, as described in Section 2. Due to the large proportion of individuals with unknown or undeclared race/ethnicity, race/ethnicity was not considered for further analysis. Moreover, NCI's commitment to promoting workforce diversity is reflected in the K awards administered by NCI's Diversity Training Branch, which were excluded from the current study. The contribution of applicant degree to K award receipt was explored, and we determined that applicants holding an MD compared to those with a PhD had higher odds of K award receipt (see **Appendix 6.10**). However, given that a clinical degree is an eligibility requirement for many of the K mechanisms in this study, and since the awards target specific populations of clinical or non-clinical scientists in particular disciplines of cancer research, degree type was not further probed.

4.2.2 Applicant Gender

Slightly more than one-third of NCI K program applicants are female. We tested whether gender was an important factor in receiving a K award and found that applicant gender has no impact on the success of those who apply to the programs (**Table 25**). This suggests that any intervention to change the gender ratio of applicants should be focused on the application process.

Gender	Total N	Odds Ratio (95% CI)	p value
Is Male	2,716	1.10 (0.94,1.29)	0.2332

Table 25. Contribution of applicant gender to K award success.
Odds Ratio with 95% Confidence Interval and p-value presented for applicant gender in K award receipt. For this test, the N equaled the total number of observations with male or female gender. 173 observations had unknown gender. The odds ratio was found to be not significant. Reference group: female

4.2.3 Age at Application

We used the predictive logit model data (Model A) to examine the relationship between applicant age at the time of application and the probability of being granted a K award. As shown in **Figure 13**, funding probability peaked around age 33. To determine if age was an important variable for determining probability of receiving a K award, we carried out a 3-group Chi-square test of those applicants who were age 32 or younger (n=32), age 33 to 39 (n=2,303) or age 40 or older (n=554) at the time of application. The 33-39 age group had a funding probability of 35.1%, 1.8% higher than what was expected (p=3.01E-05). Applicants in the 40+ group were 1.7% less likely to receive a K award than expected and applicants 32 years or younger saw no statistical difference between actual and expected probabilities. These findings demonstrated that age at time of application was correlated with receipt of a K award.

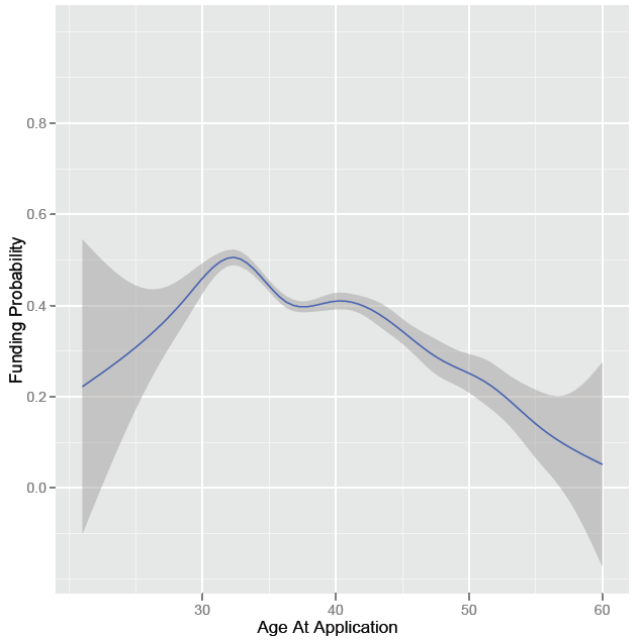


Figure 13. Estimated funding probability for applicants by age at application.
Logit Model A. 2,889 total observations. The average age of the sample was assigned to the 265 (9.2%) applicants for whom no age data were available, Grey area represents the margin of error for the model results.

4.2.4 Applicant Prior Support

Close to half of the applicants to the K programs had prior NIH research or training support. We found that applicants with prior Training (T) grant support ("Had T") had 1.25x higher odds of receiving a K award (p=0.0051) compared to applicants without prior T experience (**Table 26**). For applicants with a prior Fellowship (F) award, the odds ratio was 1.28 (p > 0.05) but this was not statistically significant, presumably due to a small number of F awards. Participation in the Loan Repayment program ("Had L") was correlated with reduced odds (0.62x) of receiving a K award (p=0.0092). The LRP program is a contract program intended to encourage new clinicians to engage in research. It is possible that many of the individuals who participated in the LRP did not receive K funding and therefore, directly sought RPG funding. Modeling data suggest that applicants who participated in the LRP have high odds of having subsequent funded research careers and obtaining high water mark NIH grant funding (see **Appendix 6.10).** The LRP interaction with the K program may merit further investigation[34]. We tested the impact of having participated in multiple prior training grants or fellowships using a three-group Chi-square (see **Appendix 6.11**), and found that having had more than one training grant or fellowship had no effect in increasing the probability of being funded beyond what would be expected if training and funding were completely independent (p = 3.01xE-5). We investigated whether having prior NIH research support in the form of a research project grant (RPG) was correlated with an applicant's odds of receiving a K award and found no significant effect ("Had RPG"), consistent with the overall program goals of supporting early career cancer researchers.

[34] For more information on the loan repayment program see: http://www.lrp.nih.gov/about_the_programs/index.aspx

Prior Support	Total N	Odds Ratio (95% CI)	p value
Had T	2,889	1.25 (1.07, 1.46)	0.0051
Had F	2,889	1.28 (0.96, 1.70)	0.0987
Had L	2,889	0.62 (0.42, 0.90)	0.0092
Had T, F, or L	2,889	1.3 (1.11, 1.51)	0.0007
Had RPG	2,889	0.70 (0.47, 1.03)	0.0664

Table 26. Contribution of prior training and research support to K award success.
Odds Ratio with 95% Confidence Interval and p-value presented for applicant prior support in K award receipt. Training grant (T), Fellowship (F), Loan Repayment (L) were considered training; Research Project Grants (RPG) were considered prior research support.

4.2.5 Applicant Institutional Characteristics

Most NCI K awards were made to a select group of highly funded institutions that had NCI Cancer Centers. Applicant institution funding level was strongly correlated with the probability of receiving a K award. As shown in **Table 27**, applicants from institutions in the top level of NCI funding (receiving $10 million or more annually) had 1.64x higher odds of receiving an award (p=4.83E-10) than applicants at any other institution. Affiliation of an applicant institution with an NCI-designated Cancer Center was also shown to be associated with improved odds for receiving a K award (1.36x, p=0.0005).

Applicant Institution Characteristics	Total N	Odds Ratio (95% CI)	p value
Institution received $10 million or more in NCI funding annually	2,881	1.64 (1.40, 1.94)	4.83E-10
Institution has an NCI-designated Cancer Center	2,889	1.36 (1.14, 1.62)	0.0005

Table 27. Contribution of applicant institution characteristics to K award success.
Odds Ratio with 95% Confidence Interval and p-value presented for applicant institutional characteristics in K award receipt. For the funding analysis, 8 observations did not have annual NCI funding information available. Reference groups: Not an NCI-designated Cancer Center or NCI funding of $0 to <$1 Million.

4.2.6 Submission of Multiple Applications

Finally, we investigated whether submission of multiple K applications was correlated with award success. Multiple applications includes those individuals who submitted more than one Type 1 NCI K application. Thus, this may refer to an individual who applied to multiple K programs, applied to the same K program with different projects, or submitted Type 1 amendments for the same project. We found that applicants with multiple K applications had 1.95x higher odds of K award receipt when compared to those who submitted just one application (95% Confidence Interval = 1.66 to 2.29; p = 1.26E-16).

4.3 What is the Impact of K Program Participation on Subsequent Career Outcomes?

4.3.1 Subsequent NIH funding

We tested whether participation in the K program had a positive impact on subsequent grant outcomes, both at NCI and more broadly at NIH (**Table 28**). Among all applicants and the comparison cohort, participation in the K program increased an applicant's odds of submitting applications to NIH or NCI for

funding, or an applicant's odds of being awarded an NIH or NCI grant. For example, in the full cohort, K awardees had 2.91x higher odds of being awarded a subsequent NCI grant (p=2.10E-40) or 2.53x higher odds of being awarded a subsequent NIH grant (p=6.90E-34). The effect was similar but diminished in the comparison cohort, likely due to the nearly identical characteristics (priority scores, etc.) of funded and unfunded individuals in the matched comparison cohort. Comparison cohort K awardees had 1.82x higher odds of being awarded a subsequent NCI grant (p=0.0008) or 1.64x higher odds of being awarded a subsequent NIH grant (p=0.0038). These findings are clear evidence of the value of this program in retaining participants in the funded NIH and NCI research workforce.

Subsequent NIH Grant Outcomes	Study Cohort	Total N	Odds Ratio (95% CI)	p value
All K applicants in the Study Cohort				
Had Subsequent NCI Application	**Full**	2,889	3.17 (2.69, 3.74)	1.80E-47
	Comparison	586	2.82 (1.96, 4.08)	6.30E-09
Had Subsequent NCI Award	**Full**	2,889	2.91 (2.47, 3.42)	2.10E-40
	Comparison	586	1.82 (1.27, 2.60)	0.0008
Had Subsequent NIH Application	**Full**	2,889	3.31 (2.76, 4.00)	5.20E-42
	Comparison	586	2.89 (1.91, 4.42)	1.10E-07
Had Subsequent NIH Award	**Full**	2,889	2.53 (2.17, 2.96)	6.90E-34
	Comparison	586	1.64 (1.17, 2.30)	0.0038
All K applicants in the study cohort that applied for subsequent NIH/NCI funding				
Had Subsequent NCI Award	**Full**	1,713	1.80 (1.48, 2.19)	2.40E-9
	Comparison	376	1.05 (0.68, 1.62)	0.8336
Had Subsequent NIH Award	**Full**	2,036	1.62 (1.35, 1.96)	2.50E-07
	Comparison	440	0.98 (0.65, 1.49)	1

Table 28. Contribution of K program participation on subsequent NIH and NCI grant activity.
Odds Ratio with 95% Confidence Interval and p-value for subsequent NIH and NCI grant outcomes for K participants, full and comparison cohort. The top panel includes all applicants in the full and comparison study cohorts. The bottom panel only includes those full or comparison cohort applicants that were found to have submitted a grant application (see Table 11, categories 1-10).

Among the full cohort applicants who applied for subsequent NIH or NCI funding (see **Table 11**, high water mark categories 1-10), participation in the K program increased their odds of receiving subsequent awards by 1.62x (p=2.50E-07) and 1.8x (2.40E-9), respectively. When selecting only those K applicants in the comparison cohort who subsequently applied for NIH or NCI funding, however, the impact of participation in the K program did not significantly increase their odds of being awarded a subsequent NIH grant (0.98; p=1) or NCI grant (1.05x, p=0.8336).

4.3.2 Subsequent Publication Productivity and Impact

As presented in **Section 3.7**, we found that a higher proportion of K awardees authored subsequent publications than non-awardees, and that the average and median number of publications per author were generally greater for awardees. Modeling results confirmed these findings for the full cohort of applicants: K awardees had higher average productivity as measured by the number of publications per

person per year (0.7 publications per person per year, p=5.4E-31), higher average 2 year citation rates (3.8 citations, p=8.0E-32), and a higher average 2 year citation to benchmark ratio (0.4 points, p=2.4E-35).

Next we analyzed only the subset of K awardees and non-awardees who published subsequent to their K award (or after their last unsuccessful K application). We did not find an observed difference in the average publication productivity between publishing awardees and non-awardees in either the full or comparison cohorts (**Table 29**). We next examined the impact of subsequent papers. For the full cohort, K awardees had higher average 2 year citation rates (2.2 citations, p=7.60E-08) and a higher average 2 year citation to benchmark ratio (0.15 points, p=0.0002). Thus, despite having similar numbers of publications as publishing non-awardees, K awardees published articles with higher citation rates and impact. We found similar trends for the comparison cohort, but the differences were not statistically significant.

Post K Participation Outcome	Study Cohort	Total N	Means Difference (Awardee – Non-Awardee)	p value
Average Publications Per Person Per Year	**Full**	2,070	0.04	0.548
	Comparison	449	-0.05	0.7249
Average Times Not Self-Cited in 2 Years	**Full**	2,070	2.20	7.60E-08
	Comparison	449	1.32	0.1663
Average Ratio of Actual to Benchmark Non-Self Citations in 2 Years	**Full**	2,070	0.15	0.0002
	Comparison	449	0.07	0.3615

Table 29. Correlation between participation and publication productivity and impact.
Means differences for subsequent publication outcomes, full and comparison cohort, with p values.

4.3.3 Subsequent Federal Advisory Committee Service

Participation on Federal advisory committees is an additional subsequent career outcome of interest. As shown in **Table 30**, we found that K participation increased an applicant's odds of subsequently serving on these advisory committees by 2.87x (p = 1.6E-41) for the full cohort and 1.8x (p=0.0007) for the comparison cohort.

Post K Participation Outcome	Study Cohort	Total N	Odds Ratio (95% CI)	p value
Served on a Federal Advisory Committee	**Full**	2,889	2.87 (2.45, 3.37)	1.6E-41
	Comparison	586	1.80 (1.27, 2.55)	0.0007

Table 30. Correlation between participation and subsequent service on a Federal advisory committee, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for subsequent membership on a federal advisory committee for K participants.

4.3.4 Subsequent Funded Research Career

One of the overall goals of the K program is to support participants as they establish an independent research career. To test whether the program achieved this during the study period, we created a composite outcome that served as a proxy for a funded research career. This composite outcome included the following subsequent activities[35,36]:

- Being awarded a competitive NIH grant, or
- Being awarded a Department of Energy grant, or
- Being awarded a grant from a member of the International Cancer Research Portfolio, or
- Being key personnel on a Clinical Trial.

Table 31 shows that K participants within the full analysis cohort were 2.39x more likely to be involved in a funded research career (p=1.0E-29). K participants in the comparison cohort had 1.58x higher odds for continuing funded research (p=0.0076). These findings demonstrate that the K program is achieving its goal of providing support for researchers to establish an independent research career.

Post K Participation Outcome	Study Cohort	Total N	Odds Ratio (95% CI)	p value
Is a Funded Researcher	**Full**	2,889	2.39 (2.05,2.80)	1.0E-29
	Comparison	586	1.58 (1.13, 2.23)	0.0076

Table 31. Correlation between participation and having a funded research career, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for the composite outcome "Is Researcher" for K participants. More information on this outcome is available in Appendix 6.10.

4.3.5 Subsequent Engagement in the Broader Research Enterprise

Research training such as that provided by the NCI K program is valuable for those individuals who, although they are not conducting funded research, continue to participate in the broader biomedical research enterprise. To determine the impact of K program participation on individuals who continued to work in broader roles in research, we developed a composite career outcome variable. This proxy includes individuals who:

- Were on the membership rolls of selected scientific professional societies, or
- Were registered health practitioners, or
- Served on Federal advisory committees (including NIH), or
- Published subsequent research papers, or
- Matched to the NIH Employee Directory (NED).

While engagement in the broader research enterprise could also include those in the "Is a Funded Researcher" (described previously in section 4.3.4), we specifically sought to test the relationship between K funding and subsequent engagement for the subset of applicants (46%) who were not classified in the "Is Researcher" group. Of the total applicants, 54% were classified in the "Is a Funded

[35] For more information on the Is a Funded Researcher composite outcome, see Appendix 6.10.4.1
[36] For a more complete description of the included funding agencies/groups, see Section 3.6.

Researcher" category, while 36% were placed in the "Is Engaged" category. Ten percent of the applicants had no subsequent involvement in activities that would result in classification in either of the above two categories. We found that K awardees in the full cohort had 5.33x higher odds (p=8.2E-21) of being engaged in the broader research enterprise after participating in the K program (**Table 32**). This was also true for the comparison cohort where K participants had 3.49x higher odds (p=0.0012) of achieving this broader career goal.

Subsequent Career Outcome	Study Cohort	Total N	Odds Ratio (95% CI)	p value
Is Engaged in Biomedical Research Enterprise, Not a Researcher	Full Cohort	1,334	5.33 (3.52, 8.34)	8.2E-21
	Comparison Cohort	255	3.49 (1.54, 8.69)	0.0012

Table 32. Correlation between participation and being engaged in the larger biomedical research enterprise, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for the composite outcome "Is Engaged" for K participants. More information on the "Is Engaged" outcome is available in Appendix 6.10.4.1

4.4 What are the Mechanism-Specific Impacts of K Program Participation on Career Outcomes?

4.4.1 Subsequent NIH funding

When considering the K program broadly, we observed that K awardees had increased odds of applying for and receiving NIH and NCI grants. We tested whether this differed among K mechanisms (**Table 33**). In this and most subsequent analyses we present only those findings in the tables for which p<0.05.

We found that K23 awardees had increased odds in receiving subsequent NIH grant awards, both for the full cohort (3.69x higher odds; p=1.44E-06) and the comparison cohort (5.42x higher odds; p=0.0095). When considering which program participants had the highest increase in their odds of receiving subsequent NCI grants, we found that K07 applicants in the full cohort experienced an increase in their odds by 3.75x (p=9.18E-14), and K23 applicants in both the full and comparison cohorts had increased odds, 4.34x (p=4.33E-07) and 4.91x (p=0.0186). K08 awardees in both the full and comparison cohorts also had significantly increased odds of subsequent NCI funding, 2.29x (p=2.5E-11) and 1.67x (p=0.0489).

Subsequent Grant Outcomes	Study Cohort	Mechanism	Total N	Odds Ratio (95% CI)	p value
Had Subsequent NIH Award	Full	K01	479	2.93 (1.93, 4.49)	9.20E-08
	Full	K07	561	2.69 (1.89, 3.85)	1.20E-08
	Full	K08	1,174	2.21 (1.74, 2.82)	3.20E-11
	Full	K11	166	2.55 (1.29, 5.11)	0.0045
	Full	K22	200	2.71 (1.38, 5.37)	0.0022
	Full	K23	253	3.69 (2.09, 6.57)	1.44E-06
	Comparison	K23	50	5.42 (1.43, 23.34)	0.0095
Had Subsequent NCI Award	Full	K01	479	3.18 (2.06, 4.93)	4.58E-08
	Full	K07	561	3.75 (2.59, 5.47)	9.18E-14
	Full	K08	1,174	2.29 (1.78, 2.94)	2.51E-11
	Full	K11	166	2.91 (1.41, 6.22)	0.0018
	Full	K22	200	2.70 (1.33, 5.47)	0.0032
	Full	K23	253	4.34 (2.37, 8.10)	4.33E-07
	Comparison	K08	284	1.67 (1.00, 2.81	0.0489
	Comparison	K23	50	4.91 (1.25, 22.46)	0.0186

Table 33. Correlation between K participation and subsequent NIH and NCI awards by mechanism, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for subsequent NIH and NCI awards for K participants by mechanism. Only findings for which p<0.05 are shown.

4.4.2 Subsequent Publication Activity

We investigated K program-specific effects on K awardee publication productivity in the full cohort including applicants without subsequent publications. For each mechanism, productivity as measured by the number of publications per person per year was higher for K awardees (p<0.05). When we analyzed only those awardees and non-awardees who had subsequent publications, the small number of individuals in each mechanism was too low to detect a significant positive effect, given the small differences between the awardee and non-awardee productivity rates and the variation in productivity within each group of individuals.

4.4.3 Subsequent Federal Advisory Committee Service

As shown in **Table 34**, the K07 and K01 program participants had the biggest increase in their odds of subsequent Federal advisory committee service, with a 3.34x increase in odds for K07 awardees (p=5.23E-12) and 3.06x increase in odds for K01 awardees (p=6.19E-08).

Subsequent Career Outcomes	Study Cohort	Mechanism	Total N	Odds Ratio (95% CI)	p value
Had Subsequent Federal Advisory Committee Service	Full	K01	479	3.06 (2.01, 4.67)	6.19E-08
	Full	K07	561	3.34 (2.33, 4.81)	5.23E-12
	Full	K08	1,174	2.48 (1.93, 3.18)	1.39E-13
	Full	K11	166	2.71 (1.35, 5.61)	0.0035
	Full	K22	200	2.78 (1.38, 5.62)	0.0032
	Full	K23	253	2.91 (1.64, 5.23)	0.0001
	Comparison	K08	284	1.81 (1.09, 3.02)	0.0209

Table 34. Correlation between K participation and subsequent Federal advisory committee service by mechanism, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for subsequent Federal advisory committee service for K participants by mechanism. Only findings for which p<0.05 are shown.

4.4.4 Subsequent Funded Research Career

When considering the impact of participation in specific K programs on a subsequent funded research career as described above, we found that K23 awardees in the comparison cohort had 8.08x higher odds (p=0.0014) for having a subsequent funded research career, and K23 awardees in the full cohort had 3.50x higher odds (p=1.54E-05; **Table 35**).

Subsequent Career Outcomes	Study Cohort	Mechanism	Total N	Odds Ratio (95% CI)	p value
Had Subsequent Funded Research Career	Full	K01	479	2.80 (1.84, 4.28)	3.03E-07
	Full	K07	561	2.49 (1.74, 3.58)	1.72E-07
	Full	K08	1,174	2.19 (1.72, 2.80)	6.90E-11
	Full	K22	200	2.29 (1.17, 4.53)	0.0115
	Full	K23	253	3.50 (1.88, 6.73)	1.54E-05
	Comparison	K23	50	8.08 (2.01, 38.57)	0.0014

Table 35. Correlation between K participation and subsequent funded research career by mechanism, full and comparison cohort.
Odds Ratio with 95% Confidence Interval and p-value for the composite outcome "Is Researcher" for K participants. More information on this outcome is available in Appendix 6.10. Only findings for which p<0.05 are shown.

We considered the impact of qualifying degree category on a subsequent funded research career (**Table 36**). Among K awardees in the full cohort, individuals who held PhDs were 2.53x (p = 4.09E-11) than PhD non-awardees to pursue a subsequent research career. Odds for MD awardees (2.12X, p = 2.42E-10) and MD/PhD awardees (2.11x, p=6.91E-6) in the full cohort were also significant. Among the comparison cohort, only PhD awardees had a significantly higher odds of pursuing a subsequent research career (1.97x, p=0.042885).

Subsequent Career Outcome	Study Cohort	Degree Category	Total N	Odds Ratio (95% CI)	p value
Had Subsequent Funded Research Career	Full	MD	1,212	2.12 (1.67, 2.71)	2.42E-10
		PhD	907	2.53 (1.90, 3.39)	4.09E-11
		MD/PhD	621	2.11 (1.50, 2.97)	6.91E-06
		Other	39	2.43 (0.29, 17.36)	0.354752
		Unknown	110	1.49 (0.41, 4.92)	0.559348
	Comparison	MD	244	1.20 (0.70, 2.07)	0.526081
		PhD	166	1.97 (1.02, 3.86)	0.042885
		MD/PhD	158	1.45 (0.72, 2.95)	0.318865
		Other	9	∞ (0.09, ∞)*	0.222222
		Unknown	9	2.24 (0.02, 234.24)	1

Table 36. Correlation between K funding and subsequent funded research career by qualifying degree, full and comparison cohort.

Odds Ratio with 95% Confidence Interval and p-value for the composite outcome "Is Researcher" for K awardees in the full and comparison cohorts. More information on this outcome is available in Appendix 6.10. All results shown, regardless of significance. *All awardees in the degree category Other in the comparison cohort were "funded researchers," thus odds ratio is infinite.

4.4.5 Subsequent Engagement in the Biomedical Research Enterprise

We also considered the impact of participation in a particular K mechanism on subsequent engagement in the biomedical research enterprise (described previously in Section 4.3.5). In the full cohort, awardees in all mechanisms except K22 and K23 were found to have significantly increased odds of being engaged in the biomedical research enterprise (**Table 37**). K07 awardees had the greatest advantage, with 8.15x odds of being engaged in the biomedical research enterprise than non-awardees. In the comparison cohort, only the K08s had significantly increased odds (5.6x) for the same outcome.

Subsequent Career Outcome	Study Cohort	Mechanism	Total N	Odds Ratio (95% CI)	p value
Is Engaged in Biomedical Research Enterprise, Not a Researcher	Full	K01	246	5.02 (2.00, 15.08)	0.0001
		K07	239	8.15 (2.44, 42.81)	3.3E-05
		K08	525	5.66 (2.75, 13.17)	2.2E-08
		K11	82	4.73 (1.16, 28.06)	0.0155
		K22	110	3.30 (0.87, 18.76)	0.0718
		K23	90	1.89 (0.37, 18.92)	0.7291
		K25	42	5.66 (0.97, 61.53)	0.0414
	Comparison	K01	45	1.99 (0.28, 23.42)	0.6808
		K07	30	∞ (0.62, ∞)*	0.1029
		K08	116	5.60 (1.47, 31.89)	0.005
		K11	0	No test possible*	N/A
		K22	23	2.59 (0.17, 157.06)	0.6036
		K23	22	1.63 (0.12, 98.21)	1
		K25	6	0.32 (0.002, 14.84)	1

Table 37. Correlation between participation in particular K mechanisms and subsequent engagement in the larger biomedical research enterprise, full and comparison cohort.

Odds Ratio with 95% Confidence Interval and p-value for the composite outcome "Is Engaged" for K participants. More information on the "Is Engaged" outcome is available in Appendix 6.10. *All awardees in the K07 comparison cohort are "engaged," and thus the odds ratio is ∞/3 = ∞ (not significant). All individuals in the K11 comparison cohort are "engaged," thus odds for both awardees are infinite/undefined and a test is not possible.

4.5 Summary of Variables Important for K Success and the Impact of K Programs on Participant Careers

Participation in NCI K programs was correlated with higher odds of subsequent research and career outcomes, as demonstrated using modeling studies that assigned values to specific variables associated with career outcomes.

Variables Important for K Success

- Applicants between 33 and 39 years of age at the time of application had increased odds of being granted a K award.
- Applicants with prior T training support had increased odds of being granted a K award.
- Applicants at institutions that received $10 million or more in NCI funding annually or were affiliated with an NCI-designated Cancer Center had increased odds of being granted a K award.
- Applicants who submitted multiple applications to the NCI K program had increased odds for receiving a K award.

K Program Participation Increased Odds of Subsequent Research Career Outcomes

- Participation in the K program increased applicants' odds for applying for NIH and NCI grant funding, and for receipt of NIH and NCI awards. Among those who applied for funding from NIH or NCI, K awardees had increased odds in receipt of NCI awards.
- K program awardees had increased odds for serving on Federal advisory committees.
- K program awardees had increased odds for having a subsequent funded research career.
- Among K awardees in the full cohort, PhD awardees had the highest odds of pursuing a subsequent funded research career versus PhD non-awardees. Awardees with MD and MD/PhD degrees also had increased odds for this outcome.
- Among K awardees in the comparison cohort, only those with a PhD had significantly higher odds of pursuing a subsequent research career.
- Participation in the K program increased applicants' odds for subsequently being involved in the broader biomedical research enterprise.

Mechanism-Specific Impact of K Program Participation

- K23 awardees appeared to have the greatest increase in odds for subsequent NIH and NCI grant funding and for attainment of a subsequent funded research career.
- K07 awardees had a sizeable increase in odds for subsequent NCI grant funding, subsequent participation on Federal advisory committees, and subsequent engagement in the broader biomedical research enterprise.
- K01 awardees had increased odds for subsequent participation on Federal advisory committees and subsequent engagement in the broader biomedical research enterprise.
- K08 awardees had increased odds for subsequent NCI grant funding and engagement in the broader research community.

5.0 Policy Implications

The NCI K program is attracting appropriate applicants and had a measurable impact on participants' subsequent research careers, not only in applying for and receiving NIH and NCI research grants and publishing high impact papers, but also in achieving other subsequent outcomes including the attainment of a funded research career and broader engagement in the biomedical research enterprise. Overall, the NCI K program is meeting its stated goals.

There are specific findings that point to opportunities for program policy adjustments:

- While gender did not play a role in an applicant's odds of receiving a K award, there were fewer than expected women applicants in several of the K mechanisms. The gender similarities in K award success mirror the finding in the total NIH pool of new competing R01s - success rates for men and women are equivalent and any gender disparities in receipt of NIH awards are attributable to differences in application rates rather than differences in success[37]. To increase participation of women, program recruitment could focus efforts to increase the number of women applicants at the K award stage or earlier in the pipeline.
- An applicant's race/ethnicity did not influence K award rate or likelihood of resubmission in this cohort. Although race/ethnicity appeared to influence success in NIH RPG funding in a recent report, that study was restricted to PhD investigators who submitted RPG applications between fiscal years 2000-2006[38]. A more complete picture of the contribution of race/ethnicity to NCI K award success and future career outcomes would need to include analysis of awards administered by NCI's Diversity Training Branch within the Center to Reduce Cancer Health Disparities.
- Prior NIH-supported NRSA training is an important factor in K award success and may be one of the first steps toward developing an independent cancer research career. Whether differences exist in future career success of trainees based on the type of prior training experience and institutional context is a question that merits further investigation.
- The majority of K awards were made to individuals at a small number of institutions with the highest amount of research funding, and these institutions tended to be affiliated with NCI-designated Cancer Centers. This finding is also reflected in NCI's R01 pool[39]. Maintenance of programs such as the K22, which attracts applicants from institutions outside of this limited sphere, may promote the development of a workforce that can offer research resources to more communities.
- Although degree field and clinical specialty of K applicants matched program focus, the growing need for multi- and inter-disciplinary training may not be fully realized with NCI's multitude of specialized K mechanisms. Consolidating several mechanisms to break down the artificial

[37] Pohlhaus, J., Jiang, H., and Sutton, J.S. (2010). Letter to the Editor, *Sex Differences in Career Development Awardees' Subsequent Grant Attainment*. Ann Intern Med (152): 616-617.
[38] Ginther, D.K., et. al. (2011) *Race, Ethnicity, and NIH Research Awards*. Science (333): 1015-1019.
[39] NCI Funded Research Portfolio. http://fundedresearch.cancer.gov/search/ResultManager?fy=PUB2010&mech=R01. (Last accessed February 9, 2012).

barriers created by the discrete mechanisms would offer K applicants more freedom in designing their research and training experiences and would reflect NCI's message of encouraging interdisciplinary research. Since K08 and K23 applicants come from similar disciplines but the K08s focus on basic research while K23s perform clinical research, a mechanism combining the two would be the first step in facilitating seamless collaborative opportunities among basic and patient-oriented physician scientists. Moreover, the existing K mechanism specificity in eligibility requirements and scientific discipline may preclude applicants from all fields from applying. A subset of applicants at the appropriate career stage who could potentially benefit from the K program may find themselves disqualified from all NCI K mechanisms if they do not have the exact combination of degree, scientific discipline, and experience.

- Although a larger percentage of all K awardees published compared to non-awardees and had higher productivity and impact, when examining only the individuals who published, the numbers of publications per author per year were similar between the awardees and non-awardees. A more in-depth examination of publications is required to fully understand the effect of K program participation on publications.
- In general, K awardees had comparable successful outcomes across the multiple K mechanisms explored in this evaluation. Most often, K awardees' outcomes showed similar trends (eg. higher median publications per person than non-awardees), and any differences lie in the magnitude of the change or in the ability to reach statistical significance, which was also influenced by individual mechanism size (applicant pool). Any other differences could be attributed to features of the target population each mechanism attracts (eg. applicants with MDs compared to PhDs).
- This study demonstrates the value of participation in NCI's K program as seen by the positive impact of the program on awardees' future career successes, both in the progression of funded research careers and in participation in careers to advance the biomedical research enterprise. The lack of major outcome differences between K mechanisms further suggests that mechanism consolidation may ameliorate the confusion caused by the currently complex organization of the K program. A previous report on biomedical workforce training by the National Research Council noted the complex nature of NIH's K program may discourage applicants[40]. A closer examination of potential redundancies and gaps in NCI's current K portfolio is justified.

[40] *Advancing the Nation's Health Needs: NIH Research Training Programs*. (2005). Committee for Monitoring the Nation's Changing Needs for Biomedical, Behavioral, and Clinical Personnel, Board on Higher Education and Workforce, National Research Council.

6.0 Appendices

6.1 NCI K Mechanisms

Mechanism	Description	Research Discipline	Support Years	Qualifying Degree(s)	Applicant Career Stage	Program Initiated	Program Terminated
K01	Award is intended to bridge the transition from mentored research environment to an independent research career in cancer research.	• Basic sciences	Up to 5 years	MD PhD MD/PhD	• Early career (post-doc to faculty transition)	1997	July 1, 2006 Program replaced by K99/R00 Award
K04*	Award is intended to provide "protected time" for a newly independent investigator to further develop their research.		Up to 5 years	PhD	• Newly independent faculty	1970	No new awards made after 1996
K07	Award provides support for early career investigators to conduct research in cancer prevention, control, behavioral and population sciences research	• Cancer prevention, control, behavioral, and population sciences	3 - 5 years	PhD Health Professional Doctoral degree (MD, DrPH, DDS, DO, DVM, PharmD or equivalent), Doctorally prepared oncology nurse	• Postdoctoral fellows • Non-tenured junior faculty	1980	
K08	Award provides support to individuals with a clinical doctoral degree to receive mentored training in laboratory-based biomedical, behavioral, or translational research targeted to the diagnosis, management, or prevention of cancer.	• Basic sciences • Translational research	Up to 5 years	MD (or equivalent) PhD in clinical discipline	• Postdoctoral and clinical fellows • Non-tenured junior faculty	1984	
K11*	Award is intended to provide long-term basic, clinical, or behavioral research training to MDs; Phase I provides didactic study and laboratory experiences, Phase II allows recipients to pursue an intensive research project.		Phase I, 2-3 yrs; Phase II, 2-3 yrs.	MD	• Early career	1987	NIH Notice (NIH Guide, Volume 24, Number 15, April 28, 1995) describes replacement programs, including K11.
K12	Award provides	• Clinical science:	Up to 5	MD (or	• Established	1992	

Mechanism	Description	Research Discipline	Support Years	Qualifying Degree(s)	Applicant Career Stage	Program Initiated	Program Terminated
	multi/trans-disciplinary institutional training to support development of research careers of clinicians or basic scientists in patient-oriented, therapeutics development research.	patient-oriented research • Translational research	years	equivalent) for clinicians; PhD (or equivalent) plus 2 years postdoctoral research for basic researchers	investigators		
K22	Award provides "protected time" for newly independent investigators to develop and receive support for their initial cancer research programs. Applicants can be clinicians pursuing basic science careers; clinicians pursuing careers in patient-oriented research; or individuals pursuing careers in cancer prevention, control, and population sciences.	• Cancer prevention, control, behavioral, and population sciences • Transdisciplinary research • Basic sciences (MDs only)	Up to 3 years	MD, PhD, DPH	• New faculty in first independent research positions	2000	
K23	Award provides support for the career development of clinical professionals to conduct mentored patient-oriented research projects.	• Clinical science: patient-oriented research • Translational research	Up to 5 years	Clinical doctorate (MD, oncology nurse) or equivalent. PhDs must be certified for clinical duties	• Non-tenured junior faculty with a clinical degree	1999	
K25	Award supports the career development of investigators with backgrounds in quantitative and engineering sciences who have chosen to focus their research on behavioral and biomedical research (basic or clinical).	• Cancer prevention, control, behavioral, and population sciences • Transdisciplinary research • Basic sciences (MDs only)	3 - 5 years	MSEE, PhD, DSc	• Postdoctoral to senior faculty quantitative scientists	2001	

* Terminated program for which limited descriptive information is available.

6.2 Study Variables and Data Sources

Variable	Data Source (s)
Gender	IMPAC II DRF AAMC Faculty Roster
Race/Ethnicity	IMPAC II DRF AAMC Faculty Roster
Age (derived from Date of Birth)	IMPAC II DRF AAMC Faculty Roster
Degree(s)	IMPAC II DRF AAMC Faculty Roster FASEB
Years Since Degree	IMPAC II DRF AAMC Faculty Roster FASEB
Prior NIH Support	IMPAC II
Subsequent NIH Support	IMPAC II
Non-NIH Federal Research Support	DoD DTIC DOE NSF FastLane
Non-Federal Research Support	International Cancer Research Portfolio
Faculty Appointment	IMPAC II AAMC Faculty Roster
Private Medical Practice	Lodestone / HealthLink Database
Non-Research Careers	LinkedIn, google.com
Professional Society Memberships	AACR ASCO FASEB
Publications	MEDLINE Thomson Reuters Web of Science
Scientific Awards	PECASE Award Website
Federal Advisory Committees and Grant Review Panels	IMPAC II FIDO.GOV

6.3 Classification and Determination of Qualifying Degree

Degree information (type of degree and year earned) was obtained from IMPACII, AAMC Faculty Roster, and DRF. The possible degree types were classified into the following main categories:

PhD	DMEDSC, DMSC, DNS, DNSC, DPH, DPHI, DPHIL, DPHL, DRPH, DRSC, DSC, DSW, EDD, PDFELLOW, PHD, POSTDOC, POSTDOCTRA, SCD, SD
MD	BAO, BCH, BDS, BDSC, BE, CHB, DO, MBBC, MBBCH, MBBCHB, MBBS, MBCHB, MD, MDCM, MRCOG, MSURGERY
MD/PhD	At least one PhD degree and at least one MD degree
Dual	At least one PhD OR at least one MD (but not both) and at least one Other degree (except for FAAN, RN, and OTH)
Other	APRN, BH, BVMS, BVSC, CRNP, DACVIM, DC, DCLINP, DCLINPSY, DDOT, DDS, DH, DMD, DNSCCNM, DOTH, DPHARM, DPM, DSN, DVM, FAAN, JD, JD1, LLD, MMED, ND, OD, OTH, PHAR, PHARMD, PHM, PHMD, PHRMD, PSYD, RN, RNP, VDOT, VMD
Note	AA, AACR, AAS, AB, AH, AM, AOCN, APRNBC, ARNP, AS, ASC, ATC, BA, BAMD, BAS, BBA, BC, BD, BM, BMATH, BMED, BN, BOTH, BPHA, BPHARM, BPHARMACY, BS, BSC, BSCH, BSCHONS, BSD, BSE, BSEE, BSN, BSPHAR, BSW, CCCA, CCCSLP, CE, CERT, CERTIF, CFNP, CLINRES, CM, CNM, CP, CPHIL, CRC, CS, DCH, DD, DDD, DGO, DIPACVS, DM, DMS, DRS, DTMH, EDM, EPI, FAAAAI, FAAP, FACC, FACEP, FACOG, FACP, FACS, FAHA, FCCM, FNP, FRACP, FRACPMHS, FRCA, FRCDC, FRCP, FRCPI, FRCS, GNP, HS, LCSW, LDNRD, LMT, LP, MA, MACP, MAPA, MAPP, MAS, MB, MBA, MBE, MCR, MDIV, MDOT, MDS, ME, MED, MEE, MGS, MH, MHA, MHS, MHSC, MHSE, MJ, MM, MMATH, MMS, MMSC, MOTH, MPA, MPE, MPH, MPHI, MPHIL, MPP, MPPM, MRCP, MRCPI, MS, MSBA, MSC, MSCE, MSCESCB, MSCI, MSCR, MSE, MSED, MSEE, MSHS, MSMBA, MSN, MSP, MSPH, MSSA, MSURG, MSW, MTR, MVSC, NULL, PAC, PD, PH, PHDMAB, PHDRESP, PNP, PT, RD, RNC, RPH, RVT, SB, SCB, SCM, SLP, SM, THM, WHCNP

Some synonymous degrees were standardized to a single display label for use in detailed reports: (DMD → DDS and VMD → DVM). Degree information from FASEB was used to make corrections to the degrees for 5 individuals in the study. All degrees were stored by application for each individual, with the degree selected for a given application being the closest in time before or equal to the application fiscal year. When reporting the years since terminal degree, the degree in the earliest year on record was used for each individual. For reports showing applicants by degree, the degree at the time of the first K application was used. For individuals with K applications in more than one fiscal year, only 10 individuals had a different degree at the start and end of their sequence of K applications.

6.4 Resubmission of NCI K Applications

The following rules were applied to the analysis of application resubmissions:

- K04 and K12 applications were excluded.
- Analysis was conducted on the combined group of applications (K01, K07, K08, K11, K22, K23, and K25) rather than for each mechanism separately.
- Only "pure resubmissions" (e.g., an application with the same serial number and suffix code that identifies it as an amended application [A1, A2, A3, or A4] with no examination of project title) were included. There were 840 individuals with pure resubmissions.
- An expanded version of the full cohort, which removed a subselection rule to only include one amendment per application per fiscal year, was used, allowing this analysis to include the full amendment history for each application.
- There were 42 individuals for whom there was no initial A0 submission, and these individuals and their applications were excluded from this analysis.
- Resubmission analysis did not track whether or not there were amendments missing within an application's sequence.

	Initial Submission (A0)	Resubmission (A1-A4)
Applications	3,067	1.349
Individuals	2,851	1,050

Table A6.4.1. Summary of resubmissions for NCI K programs.

Submission	Number of Applications	Number of Awards	Percent Awarded
Initial (A0)	3,067	700	22.8%
All Amendments (A1-A4)	1,349	502	37.2%
A1	1,072	353	32.9%
A2	276	148	53.6%
A3	1	1	100%

Table A6.4.2. Summary of submissions and awards by amendment number for NCI K programs.

We also looked at resubmission by gender.

Gender of Applicant	Unsuccessful A0 Applications	A1 Resubmissions	Percent of A0 applications that were Resubmitted
Female	874	415	47.5%
Male	1340	573	42.8%
Unknown	153	21	13.7%

Table A6.4.3. Resubmission of NCI K applications by gender

Using an odds ratio test (**Table A6.4.4**), we found that females were 1.2X more likely to resubmit NCI K applications than males (p=0.03)(CI [95%]: 1.02, 1.44).

Gender	No Resubmission	Resubmission
Male	767	573
Female	459	415

Table A6.4.4. Odds ratio test for resubmission of NCI K applications based on gender.

Because of the large proportion of female applicants to the K07 program we investigated whether this difference persisted for the applicant pool when K07 data were excluded. As presented in Table A6.4.5, this difference is not significant.

Gender	Unsuccessful A0 Applications	A1 Resubmissions	% of A0 applications that were resubmitted
Male	1,195	494	41.3%
Female	584	246	42.1%

Table A6.4.5. Resubmission of NCI K applications based on gender excluding K07 applicant data. Using the 2-sample test for equality of proportions with continuity correction found this difference to be non-significant (p=0.5784).

Finally, we also explored resubmissions by applicant race/ethnicity (**Table A6.4.6**). A chi-squared test indicated that race and application resubmission were independent variables ($\chi^2 = 0.1739$, $p = 0.9817$).

Race/Ethnicity of Applicant	No Resubmission	Resubmission
Asian	205	182
Black	19	15
Hispanic	17	14
White	630	540

Table A6.4.6. Resubmission of NCI K applications by race/ethnicity. Some races have been excluded from this analysis due to low total applicant number.

6.5 Determination of Age at Application and Years Since Degree

Age at Application

For each application for each individual, an age is computed if birth date information is available in any one of the following: IMPACII, AAMC Faculty Roster or DRF. Age is computed as the difference between the fiscal year of the application and the fiscal year of the individual's birth date. For all analyses in this report, we select the age associated with the first NCI K award or last unsuccessful NCI K application for each individual. Ages younger than 20, or older than 90, are treated as data errors and are marked as missing data cases. For model analysis, we used the overall average age to impute a value for all missing age cases.

Years since Degree

For each individual for which degree information was available along with a date or year of the conferral of the degree, we calculated the years since degree by subtracting the earliest year of any degree found from the fiscal year of the first NCI K award or last unsuccessful NCI K application.

6.6 K07 Years Since Degree Analysis (Full Cohort)

To better understand the results presented in Section 2.3.7, additional analyses exploring the years since qualifying degree of applicants to the NCI K07 mechanism were performed. To better discern the "career stage" of applicants, we first looked at the years since degree of applicants in increments of 5 fiscal years (**Figure A6.6.1**).

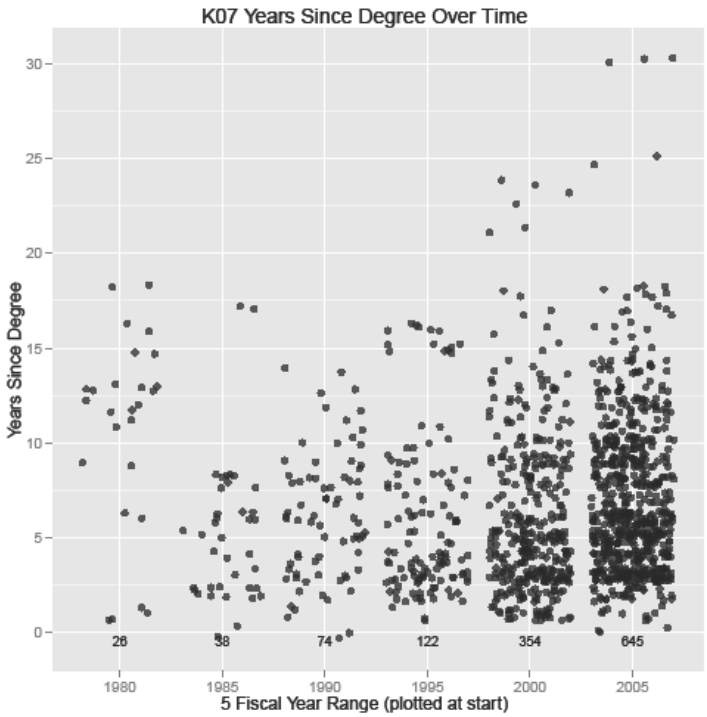


Figure A6.6.1. Years Since Degree for NCI K07 applicants in five fiscal year increments.
Scatter plot indicates that in the earliest years of the K07 mechanism (1980 – 1984), applicants were further along in their careers than more recent years (2000 – 2008), however, there is a broader range in years since degree over the life of the program.

Figure A6.6.2 shows the median years since degrees for these five-year increments, and confirms that the median of the first interval (1980-1984) was greater (~12 years) than subsequent years (range of 4-6 years), but with a greater range and more distinct outliers.

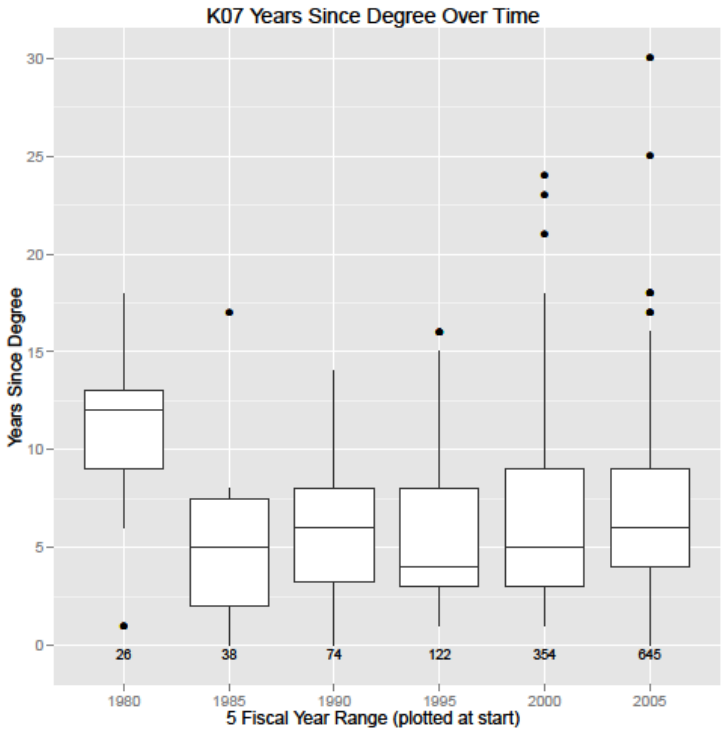


Figure A6.6.2. Median Years Since Degree for K07 applicants in five fiscal year increments.

6.7 Prior Support Categories

Prior support information was obtained from IMPAC II. The table below shows the prior support categories described in Section 2.3.8 (Table 8), and the award mechanisms contributing to each category. Prior support analyses include K04 and exclude K12 (3,982 total applications and 1,547 total awards).

Prior Support Category	Definition	Overlaps with other categories?	Prior Support Mechanisms for Individuals in this Category Mechansim (PI count)
Had T Support	At least one T mechanism w/wo other support	yes	T32 (999), F32 (81), L30 (56), T35 (35), R03 (19), L40 (16), R01 (12), R21 (10), F31 (9), M01 (9), P41 (5), Z01 (3), T01 (3), L60 (3), F30 (2), P01 (2), P20 (2), U01 (2), R37 (2), R23 (2), R25 (2), R29 (1), R36 (1), R55 (1), S15 (1), P50 (1), P60 (1), U54 (1), T12 (1), T15 (1), T34 (1), P30 (1), N01 (1), R10 (1), R13 (1), R15 (1), D15 (1), D33 (1), F02 (1), F34 (1), F35 (1)
Had Only T Support	Only T mechanisms	no	T32 (793), T35 (21), T15 (1)
Had F Support	At least one F mechanism w/wo other support	yes	F32 (287), R01 (151), T32 (94), F02 (65), F22 (37), F03 (31), F01 (26), F31 (14), R23 (9), R03 (7), L30 (7), R29 (4), P41 (3), L40 (3), F35 (3), F30 (3), M01 (2), Z01 (2), T35 (2), U01 (1), S10 (1), S15 (1), T34 (1), R21 (1), R26 (1), P20 (1), P30 (1), F37 (1), F38 (1), F34 (1), F06 (1), F11 (1)
Had Only F Support	Only F mechanisms	no	F32 (126), F22 (21), F02 (20), F03 (16), F01 (9), F31 (5), F35 (2), F37 (1), F38 (1), F30 (1)
Had L Support	At least one L mechanism w/wo other support	yes	L30 (103), T32 (67), L40 (21), L60 (19), T35 (13), F32 (10), M01 (7), R21 (6), R03 (4), L32 (1), P41 (1), U01 (1)
Had Only L Support	Only L mechanisms	no	L30 (38), L60 (11), L40 (3), L32 (1)
Had RPG Support	Had at least one of the RPG mechanisms (DP1, DP2, P01, P42, PN1, R01, R03, R15, R21, R29, R33, R34, R35, R36, R37, R55, R56, RL1, RL2, RL5, RL9, U01, U19, UC1, UC7) w/wo other support	yes	R01 (592), R03 (82), F32 (72), F02 (45), T32 (38), R29 (31), R21 (25), R23 (22), F01 (17), F22 (15), F03 (15), L30 (7), R26 (6), U01 (5), P01 (5), S15 (5), M01 (4), P41 (3), L60 (3), R55 (3), S10 (3), R25 (3), R37 (2), T01 (2), P20 (2), P30 (2), F34 (1), F06 (1), F11 (1), F31 (1), D33 (1), T12 (1), T35 (1), R41 (1), R36 (1), R10 (1), R13 (1), R15 (1), R18 (1), Z01 (1)
Had Only RPG Support	Only RPG mechanisms	no	R01 (406), R03 (48), R29 (22), R21 (11), R55 (2), U01 (2), P01 (2)
Had Multiple T, F, or L Support	Had (T and F) or (T and L) or (F and L) or (T and F and L) w/wo other support	yes	T32 (154), F32 (84), L30 (58), L40 (17), T35 (14), F31 (9), R03 (4), M01 (4), R21 (3), R01 (3), L60 (3), F30 (2), F02 (1), F34 (1), F35 (1), P20 (1), P30 (1), P41 (1), R23 (1), T34 (1), U01 (1), Z01 (1)
Had Multiple Support, including RPG	Had (R and T) or (R and L) or (R and F) or (R and T and F) or (R and L and F) or (R and T and L) or (R and T and F and L) w/wo other support	yes	R01 (160), F32 (72), F02 (45), T32 (38), R03 (26), F01 (17), F22 (15), F03 (15), R21 (14), L30 (7), R23 (5), R29 (5), U01 (3), L60 (3), P41 (2), R37 (2), S15 (2), T01 (2), P01 (2), R25 (2), R26 (1), P20 (1), P30 (1), R36 (1), T12 (1), R55 (1), S10 (1), T35 (1), M01 (1), F34 (1), R13 (1), R15 (1), F06 (1), F11 (1), F31 (1), D33 (1)
Had Only Other Support	Had Prior Support, but did not have any T, F, L or RPG Mechanism	no	M01 (7), R23 (7), R26 (4), N01 (3), P20 (3), R10 (3), R13 (3), R18 (3), P50 (2), U10 (2), D15 (1), R25 (1), R43 (1), S03 (1), S06 (1), P60 (1), P41 (1)
No Prior Support	No NIH awards prior to 1st K Award or last unsuccessful K application	no	N/A

6.8 Construction of the Funding Bubble and Comparison Cohorts

A method to identify candidates for the comparison cohorts for each NCI K mechanism based on priority score of the application was developed. The comparison cohort is constructed such that it includes applicants who have a similar quality application as determined by priority score and an equal likelihood of their application being funded or not funded. The priority score range in which there is an equal number of awarded and not awarded applications is referred to as the "funding bubble." This priority score range can change from year to year, and thus for each K mechanism evaluated, funding bubbles were determined for each fiscal year, and then combined into a single pool which was then used for evaluating career outcomes.

For each K mechanism and each fiscal year included in the evaluation, the priority score range (100 – 500) was divided into equal sized bins, and each bin was populated with the number of funded and not funded applicants, respectively. An ideal funding bubble bin contains an equal number of funded and not funded applicants; one would anticipate that bins at the low end of the priority score range to contain mostly funded applicants, while bins at the high end of the priority score range would contain mostly unfunded applicants. Bin width, or the range of scores included in a bin, is also an important factor. If the range of included scores is too wide, similarity of application quality is reduced; similarly, if the range of scores is too narrow, there is a risk of having a small and potentially unbalanced sample. The minimum bin width was a score range of 5 and the maximum allowed bin width was a score range of 50. Ten bin widths were considered for each FY-K mechanism combination, and the optimal bubble bin was selected using the following set of tie-breaker rules, applied in sequence:

- Highest "bubble usefulness score," calculated as 100% for any balanced bubble with an equal number of awardees and non-awardees and 0% for any unbalanced bubble multiplied by a factor measuring the "density" of applications (number of applications divided by the square root of the score range width)
- Highest score range upper endpoint (closest to 500)
- Smallest score range width (maximum allowed is 50)
- Largest application count (minimum allowed is 4 applications)
- Lowest score range low endpoint (closest to 100)

The optimal bins were selected for each K mechanism-FY combination and then pooled into a single comparison cohort to be used for the outcome analysis (**Figure A6.8.1**).

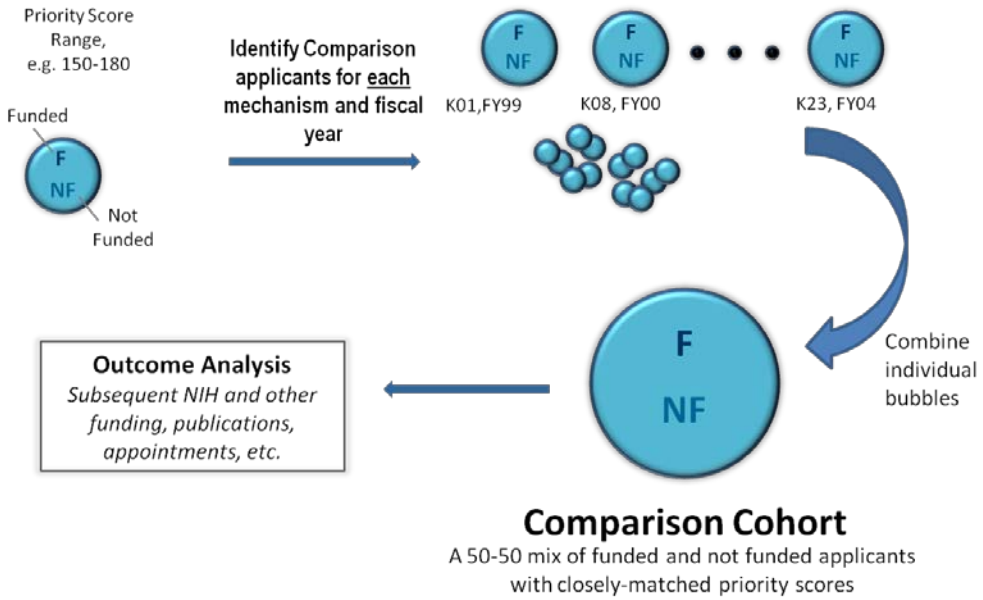


Figure A6.8.1. Identification of comparable NCI K applicants and composition of the comparison cohort.

6.9 U.S. Patent Applications and Issued Patents

The list of K award numbers included in this evaluation were matched to grant numbers appearing in the government interest section of U.S. Patent and Trademark Office patent applications and issued patents.

Mechanism	Total Awards	Number of K Awards Cited (% of total awards)	Number of Patents
K01	153	5 (3.3%)	8
K07	274	2 (0.7%)	3
K08	515	18 (3.5%)	32
K11	86	3 (3.5%)	7
TOTAL	1,028	28 (2.7%)	50

6.10 Linear and Logit Models

As discussed in the main report, a set of 26 regression models were constructed to assess the size, sense, and statistical significance of the contributions of applicant characteristics to various outcomes such as funding, publication productivity, and establishment of a research career.

Section 4 of the report presents a series of results based on direct tests (Fisher exact, chi-squared, Students' t) to measure the effect of single variables. The Section 4 tests and models were run using the same data used to run the linear and logit models described here. These linear and logit model results help to address concerns about confounding by marginalized variables in Section 4.

6.10.1 Model Summary

Table A6.10.1 summarizes the ordinary linear regression models and the binary logistic regression (logit) models used for this study. The symbols X and X' in **Table A6.10.1** are explained in **Table A6.10.2** and "Mech" is an abbreviation for the primary NCI K mechanism for each individual. Complete definitions of the input and output variables are given in the Data Dictionary in section 6.10.4.

Model Group	# of Models	Model			Type	Significance Threshold Used (α)
		Output (Y) variable	Y type	Input variables		
A	1	Funding	Yes = 1, No = 0	X+Mech	logit	0.05
B	6	K01, K07…K23 Funding	Yes = 1, No = 0	X, Mech set to constant	logit	0.01
C	1	Mech	K01-K25	X	ordered logit	0.05
E	1	NIH High Water	1-12 , 1 best	X'+Mech+Funding	linear	0.05
	1	NCI High Water	1-12 , 1 best	X'+Mech+Funding	linear	0.05
F	1	Publications Per Year Per Person	decimal >= 0	X'+Mech+Funding	linear	0.05
	1	Average/Benchmark Citation Ratio	decimal >= 0	X'+Mech+Funding	linear	0.05
H	1	Is Researcher	Yes = 1, No = 0	X'+Mech+Funding	logit	0.05
	1	Is Engaged	Yes = 1, No = 0	X'+Mech+Funding	logit	0.05
I	6	K01, K07…K23 Is Researcher	Yes = 1, No = 0	X'+Funding, Mech constant	logit	0.01
	6	K01, K07…K23 Is Engaged	Yes = 1, No = 0	X'+Funding, Mech constant	logit	0.01
Total Model Count	**26**					

Table A6.10.1 Linear and logit models

Input Variables	
X	Gender, Race, Age, Age^2 (= Age x Age), Degree, Institution Cancer Center Type, Institution Funding Level, Degree Field of Study, Fiscal Year of Award/Application, Age Imputation Flag, Multiple Application Flag, Multiple Mechanism Flag, Prior Support**
X'	X + Years Post-K + (Years Post-K) ^2
**Prior Support	7 separate Flag variables - Had T, Had L, Had F, Had RPG, Had Multiple T/F/L, Had Multiple including RPG, Had Only Other

Table A6.10.2 Model input variables

In each model, between 19 and 22 measurements were treated as independent input variables. Demographics variables such as gender, degree, institution funding level were included in all models. Age was included along with a term set equal to the square of the age to capture quadratic behavior often seen in social science modeling[41]. An individual's primary mechanism was included in all models, either as an input variable, or as a fixed value to study the effects of other variables strictly within a given mechanism. For post-K outcomes, a linear and quadratic term was added for the elapsed time from the K experience to the present (Years Post K).

For variables such as Degree that take on descriptive (aka categorical) values, the model software automatically constructs a replacement set of "dummy" variables, each with a Yes or No values, for all but one pre-selected value of the original variable. The value for which no dummy variable exists is called the "reference level" and those individuals observed to have that value form the "reference group." For example, the Degree input variable is replaced by the following set of dummy variables: Had MD/PhD, Had MD, Had other, and Had Unknown. The PhD value was pre-selected as the reference level, and individuals with PhD degrees form the reference group for Degree.

The Data Dictionary in section 6.10.4 identifies the reference level for all categorical input variables in the model.

Whether an individual ultimately did or did not receive funding for one of their NCI K applications was included as a Yes/No variable first as an outcome variable in its own right in the model in groups A and B, and then as an input to the models in groups E, F, H, and I that address subsequent outcomes.

All models (except Group C[42]) assume that the outcome variable can be predicted from the input variables using the following formula:

$$\eta_i = B_0+B_1X_{i1}+B_2X_{i2}+\ldots+ B_nX_{in} + (\text{random error})_i$$

Where the X's are the input variables (Age, Had MD, etc.), and η is either the measured output (Y) variable for the models listed as "linear" in table A5.8.1, or else η is the log of the odds in favor of a Yes result for outcomes like funding , which are listed as "logit" in table A5.8.1. The B's are the coefficients that are estimated by the model and measure the contribution of each variable to the outcome. The subscript (i) indicates that this formula applies to each individual in the cohort.

[41] D. Ginther, personal communication, September 2011

[42] Group C models were an attempt to find predictors for primary mechanism based on the demographics and institutional characteristics. Overall, this modeling was not successful, and is not discussed in the report.

The coefficients are calculated so as to minimize the mean-square error for the linear models, or to maximize the likelihood of the observed pattern of Yes/No results for the logit models. Interpretation of the coefficients is discussed in section 6.10.3.

6.10.2 Variable Correlations

Before the models were run, we examined the degree of correlation between pairs of input variables. The Kendall tau correlations were calculated to estimate this correlation since most of the input variables are categorical.

The results are shown in **Table A6.10.3** below. Most correlations are very weak, with the strongest being those between degree and degree field of study and between having multiple NCI K applications and having applications for multiple mechanisms, followed by those between age and degree, degree and mechanism, gender and mechanism, and between degree field of study and mechanism. None of the correlations were so high as to make one input variable an effective proxy for another, so all were retained as inputs to the various models.

	Prior Support	Gender	Race	Age	Degree	Institution Cancer Center Type	Mechanism	Degree Field of Study	Multiple K applications	Multiple Mechanisms
Prior Support	1.00	-0.11	0.10	0.02	0.05	-0.01	-0.03	-0.21	0.08	0.04
Gender		1.00	-0.07	-0.02	0.08	0.01	0.14	0.06	-0.07	-0.01
Race			1.00	-0.08	-0.04	-0.02	0.00	-0.05	0.01	-0.01
Age				1.00	0.14	0.01	-0.04	-0.06	0.09	0.02
Degree					1.00	0.06	-0.14	-0.38	-0.02	0.00
Institution Cancer Center Type						1.00	-0.02	-0.03	-0.02	0.03
Mechanism							1.00	0.15	0.01	0.05
Degree Field of Study								1.00	0.02	0.02
Multiple K applications									1.00	0.30
Multiple Mechanisms										1.00

Table A6.10.3 Input variable correlations

6.10.3 Model Results

Table A6.10.4 below lists all coefficients whose p value was below the pre-determined threshold for α (= probability of a Type 1 error) listed in **Table A6.10.1**, using a significance test whose null hypothesis is that all coefficients are zero. For linear models, the coefficient is given directly and gives the predicted change in the outcome variable assuming a 1 unit change in the input variable, with all other variables held constant.

For logit models, the number shown is the exponentiated coefficient (e^{Bi}), which gives the estimated odds ratio in favor of a Yes outcome for a unit change in the input variable.

Note that for dichotomous input variables, the estimated odds ratio is different from the exact odds ratio reported in the 2x2 Fisher exact tests in Section 4. The exact odds ratios in Section 4 apply to 2 groups of actual observations in the data, those with input =Yes and input =No, and in each group the other variables take on their actual values but are not involved in calculating the odds ratio or p value. By contrast, the estimated odds ratios from the logit models apply to 2 hypothetical groups, again defined by having Yes or No value for the input variable, but with all other variables held constant.

Also note that for dummy variables, a coefficient is expressing the change in the outcome when the dummy variable changes from 0 (No) to 1 (Yes), which are its only possible values. But when the dummy variable "reaches" 1, all other related dummy variables must be 0 since only one out of a set of related dummy variables can be 1 for a given observation. For example when the Degree dummy variable for having an MD degree is 1, then the MD/PhD and Unknown degree dummy variables must both be 0. So to remain constant while the dummy of interest is changing, the remaining related dummy variables must all be 0 both before and after the change. This means the before state is all-0, which corresponds to the reference group. So a dummy variable coefficient expresses the change in the outcome variable as you "move" from the reference group to the Yes group for the dummy variable in question in a hypothetical movement that also keeps all other non-related variables constant.

These are the significant linear model coefficients and logit model exponentiated coefficients. The short mnemonic hint listed as "helps" or "hurts" should not be interpreted as a claim of a proven causal relationship. The hint indicates whether the input variable is positively correlated with outcome values that are considered to be more favorable or less favorable.

A final note - the publication productivity and benchmark citation ratio outcome variables in the table below represent modeling data on the subset of K awardees and non-awardees who published. For results of modeling when both publishers and non-publishers were included, see **Supplement 7.6**.

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
A	I(AA^2)	age squared	Funding	1.00	odds of funding	hurts	0.0422
A	NDQ2	applicant institution funding range $1 M - $10 M (2nd highest)	Funding	2.55	odds of funding	helps	0.0001
A	NDQ3	applicant institution funding range $10 M - $100 M (highest)	Funding	3.71	odds of funding	helps	2.8E-07
A	FY2004	applied in fiscal year 2004	Funding	0.27	odds of funding	hurts	0.0426
A	FY2007	applied in fiscal year 2007	Funding	0.26	odds of funding	hurts	0.0391
A	FY2008	applied in fiscal year 2008	Funding	0.27	odds of funding	hurts	0.0398
A	DUnknown	degree unknown	Funding	0.45	odds of funding	hurts	0.0318
A	FOSPsychology	field of study Psychology	Funding	0.58	odds of funding	hurts	0.0457
A	FOSSocial Science	field of study Social	Funding	0.26	odds of funding	hurts	0.0132

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
		Science					
A	Had_F	had F prior support	Funding	1.49	odds of funding	helps	0.0420
A	DPhD_MD	had MD/PhD degree	Funding	1.56	odds of funding	helps	0.0051
A	IsMulti	had multiple NCI K applications	Funding	2.39	odds of funding	helps	1.1E-18
A	MK07	K07 applicant	Funding	2.46	odds of funding	helps	1.5E-07
A	MK25	K25 applicant	Funding	4.06	odds of funding	helps	0.0002
A	RCU	race unknown	Funding	0.55	odds of funding	hurts	1.0E-05
B	DPhD_MD	had MD/PhD degree	Funding K01	2.90	odds of funding	helps	0.0045
B	IsMulti	had multiple NCI K applications	Funding K01	2.41	odds of funding	helps	0.0013
B	FOSSocial Science	field of study Social Science	Funding K07	0.16	odds of funding	hurts	0.0056
B	IsMulti	had multiple NCI K applications	Funding K07	4.13	odds of funding	helps	3.2E-09
B	FY1986	applied in fiscal year 1986	Funding K08	0.10	odds of funding	hurts	3.2E-05
B	FY1990	applied in fiscal year 1990	Funding K08	0.21	odds of funding	hurts	0.0061
B	FY1993	applied in fiscal year 1993	Funding K08	0.16	odds of funding	hurts	0.0022
B	FY1994	applied in fiscal year 1994	Funding K08	0.16	odds of funding	hurts	0.0015
B	FY1998	applied in fiscal year 1998	Funding K08	0.15	odds of funding	hurts	0.0002
B	FY2001	applied in fiscal year 2001	Funding K08	0.18	odds of funding	hurts	0.0010
B	FY2003	applied in fiscal year 2003	Funding K08	0.20	odds of funding	hurts	0.0027
B	FY2004	applied in fiscal year 2004	Funding K08	0.07	odds of funding	hurts	4.8E-06
B	FY2005	applied in fiscal year 2005	Funding K08	0.13	odds of funding	hurts	0.0002
B	FY2006	applied in fiscal year 2006	Funding K08	0.19	odds of funding	hurts	0.0031
B	FY2007	applied in fiscal year 2007	Funding K08	0.10	odds of funding	hurts	9.9E-05
B	FY2008	applied in fiscal year 2008	Funding K08	0.18	odds of funding	hurts	0.0018
B	IsMulti	had multiple NCI K applications	Funding K08	2.23	odds of funding	helps	8.0E-07
B	RCU	race unknown	Funding K08	0.35	odds of funding	hurts	7.8E-07
B	RCU	race unknown	Funding K11	0.06	odds of funding	hurts	4.1E-05
B	NDQ2	applicant institution funding range $1 M - $10 M (2nd highest)	Funding K22	25.36	odds of funding	helps	0.0004
B	NDQ3	applicant institution funding range $10 M - $100 M (	Funding K22	31.78	odds of funding	helps	0.0009

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
		highest)					
B	IsMulti	had multiple NCI K applications	Funding K23	3.03	odds of funding	helps	0.0026
E	AgeImpAgeImputed	age was imputed	NCI High Water	1.28	high water mark	hurts	0.0002
E	NDQ2	applicant institution funding range $1 M - $10 M (2nd highest)	NCI High Water	-0.95	high water mark	helps	0.0017
E	NDQ3	applicant institution funding range $10 M - $100 M (highest)	NCI High Water	-1.24	high water mark	helps	0.0002
E	NCCompCC	applicant institution is a comprehensive cancer center	NCI High Water	-0.38	high water mark	helps	0.0455
E	FY2008	applied in fiscal year 2008	NCI High Water	2.51	high water mark	hurts	0.0064
E	FOSPsychology	field of study Psychology	NCI High Water	-0.98	high water mark	helps	0.0140
E	FOSSocial Science	field of study Social Science	NCI High Water	1.58	high water mark	hurts	0.0311
E	Had_L	had L prior support	NCI High Water	-0.90	high water mark	helps	0.0118
E	DPhD_MD	had MD/PhD degree	NCI High Water	-0.83	high water mark	helps	0.0003
E	Had_RPG	had RPG prior support	NCI High Water	-1.07	high water mark	helps	0.0059
E	GM	is male	NCI High Water	-0.37	high water mark	helps	0.0064
E	MK08	K08 applicant	NCI High Water	1.06	high water mark	hurts	3.7E-05
E	MK11	K11 applicant	NCI High Water	2.18	high water mark	hurts	7.3E-08
E	MK23	K23 applicant	NCI High Water	1.00	high water mark	hurts	0.0016
E	RCU	race unknown	NCI High Water	0.79	high water mark	hurts	4.9E-05
E	Funded	was funded for NCI K	NCI High Water	-1.53	high water mark	helps	1.7E-29
E	AgeImpAgeImputed	age was imputed	NIH High Water	1.84	high water mark	hurts	1.2E-07
E	NDQ2	applicant institution funding range $1 M - $10 M (2nd highest)	NIH High Water	-1.11	high water mark	helps	0.0003
E	NDQ3	applicant institution funding range $10 M - $100 M (highest)	NIH High Water	-1.36	high water mark	helps	4.1E-05
E	FY2008	applied in fiscal year 2008	NIH High Water	2.54	high water mark	hurts	0.0063
E	FOSFOS_Unknown	field of study	NIH High	-0.46	high water	helps	0.0376

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
		unknown	Water		mark		
E	Had_L	had L prior support	NIH High Water	-1.68	high water mark	helps	3.0E-06
E	DPhD_MD	had MD/PhD degree	NIH High Water	-0.77	high water mark	helps	0.0008
E	IsMultiAC	had NCI K applications in multiple mechanisms	NIH High Water	-0.75	high water mark	helps	0.0238
E	DOther	had Other degree	NIH High Water	1.16	high water mark	hurts	0.0475
E	Had_Only_Other	had prior support but no T,F, L or RPG	NIH High Water	-1.33	high water mark	helps	0.0227
E	Had_RPG	had RPG prior support	NIH High Water	-1.68	high water mark	helps	2.0E-05
E	RCH	is Hispanic	NIH High Water	-0.97	high water mark	helps	0.0425
E	GM	is male	NIH High Water	-0.45	high water mark	helps	0.0012
E	MK08	K08 applicant	NIH High Water	1.35	high water mark	hurts	2.3E-07
E	MK11	K11 applicant	NIH High Water	2.62	high water mark	hurts	1.6E-10
E	MK23	K23 applicant	NIH High Water	1.40	high water mark	hurts	1.4E-05
E	RCU	race unknown	NIH High Water	0.90	high water mark	hurts	5.6E-06
E	Funded	was funded for NCI K	NIH High Water	-1.48	high water mark	helps	3.7E-27
F	FY1981	applied in fiscal year 1981	Actual to Benchmark Cite Ratio	2.12	cite ratio	helps	0.0011
F	FY2007	applied in fiscal year 2007	Actual to Benchmark Cite Ratio	-0.53	cite ratio	hurts	0.0500
F	FY2008	applied in fiscal year 2008	Actual to Benchmark Cite Ratio	-0.88	cite ratio	hurts	0.0012
F	MK23	K23 applicant	Actual to Benchmark Cite Ratio	0.20	cite ratio	helps	0.0355
F	AgeImpAgeImputed	age was imputed	Pub Productivity	-0.68	publications per person per year	hurts	0.0008
F	FY1984	applied in fiscal year 1984	Pub Productivity	1.10	publications per person per year	helps	0.0479
F	FY1986	applied in fiscal year 1986	Pub Productivity	1.29	publications per person per year	helps	0.0146
F	FY1988	applied in fiscal year 1988	Pub Productivity	1.68	publications per person per year	helps	0.0019

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
F	FY1990	applied in fiscal year 1990	Pub Productivity	1.35	publications per person per year	helps	0.0091
F	FY1991	applied in fiscal year 1991	Pub Productivity	1.32	publications per person per year	helps	0.0121
F	FY1992	applied in fiscal year 1992	Pub Productivity	1.14	publications per person per year	helps	0.0256
F	FY1993	applied in fiscal year 1993	Pub Productivity	1.30	publications per person per year	helps	0.0117
F	FY1994	applied in fiscal year 1994	Pub Productivity	1.21	publications per person per year	helps	0.0162
F	FY1995	applied in fiscal year 1995	Pub Productivity	1.40	publications per person per year	helps	0.0059
F	FY1996	applied in fiscal year 1996	Pub Productivity	1.25	publications per person per year	helps	0.0132
F	FY1997	applied in fiscal year 1997	Pub Productivity	1.32	publications per person per year	helps	0.0066
F	FY1998	applied in fiscal year 1998	Pub Productivity	1.27	publications per person per year	helps	0.0082
F	FY2000	applied in fiscal year 2000	Pub Productivity	1.43	publications per person per year	helps	0.0027
F	FY2001	applied in fiscal year 2001	Pub Productivity	1.13	publications per person per year	helps	0.0169
F	FY2002	applied in fiscal year 2002	Pub Productivity	1.43	publications per person per year	helps	0.0023
F	FY2003	applied in fiscal year 2003	Pub Productivity	1.33	publications per person per year	helps	0.0045
F	FY2004	applied in fiscal year 2004	Pub Productivity	1.29	publications per person per year	helps	0.0062
F	FY2005	applied in fiscal year 2005	Pub Productivity	1.24	publications per person per year	helps	0.0079
F	FY2006	applied in fiscal year 2006	Pub Productivity	1.03	publications per person per year	helps	0.0280
F	FY2008	applied in fiscal year 2008	Pub Productivity	1.16	publications per person per year	helps	0.0139
F	DPhD_MD	had MD/PhD degree	Pub Productivity	0.29	publications per person per year	helps	0.0136

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
F	Had_RPG	had RPG prior support	Pub Productivity	0.56	publications per person per year	helps	0.0030
F	Had_T	had T prior support	Pub Productivity	-0.24	publications per person per year	hurts	0.0034
F	GM	is male	Pub Productivity	0.46	publications per person per year	helps	2.8E-10
F	MK07	K07 applicant	Pub Productivity	0.80	publications per person per year	helps	3.1E-10
F	MK23	K23 applicant	Pub Productivity	0.71	publications per person per year	helps	1.6E-05
H	AA	age	Is Engaged	1.43	odds of being subsequently engaged	helps	0.0152
H	I(AA^2)	age squared	Is Engaged	1.00	odds of being subsequently engaged	hurts	0.0079
H	AgeImpAgeImputed	age was imputed	Is Engaged	0.41	odds of being subsequently engaged	hurts	0.0044
H	NCNonCompCC	applicant institution is a non-comprehensive cancer center	Is Engaged	1.59	odds of being subsequently engaged	helps	0.0435
H	FY1985	applied in fiscal year 1985	Is Engaged	13.86	odds of being subsequently engaged	helps	0.0267
H	FY1986	applied in fiscal year 1986	Is Engaged	9.28	odds of being subsequently engaged	helps	0.0432
H	FY1987	applied in fiscal year 1987	Is Engaged	17.40	odds of being subsequently engaged	helps	0.0173
H	FY1989	applied in fiscal year 1989	Is Engaged	22.24	odds of being subsequently engaged	helps	0.0105
H	FY1991	applied in fiscal year 1991	Is Engaged	13.48	odds of being subsequently engaged	helps	0.0341
H	FY1993	applied in fiscal year 1993	Is Engaged	15.52	odds of being subsequently engaged	helps	0.0197
H	FY1994	applied in fiscal year 1994	Is Engaged	15.01	odds of being subsequently engaged	helps	0.0194
H	FY1995	applied in fiscal year 1995	Is Engaged	29.82	odds of being subsequently engaged	helps	0.0091
H	FY1997	applied in fiscal year 1997	Is Engaged	8.99	odds of being subsequently engaged	helps	0.0466

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
H	FY2000	applied in fiscal year 2000	Is Engaged	8.93	odds of being subsequently engaged	helps	0.0457
H	FY2006	applied in fiscal year 2006	Is Engaged	9.68	odds of being subsequently engaged	helps	0.0296
H	DMD	had MD degree	Is Engaged	1.95	odds of being subsequently engaged	helps	0.0250
H	DPhD_MD	had MD/PhD degree	Is Engaged	2.83	odds of being subsequently engaged	helps	0.0003
H	Had_RPG	had RPG prior support	Is Engaged	4.38	odds of being subsequently engaged	helps	0.0268
H	GM	is male	Is Engaged	1.97	odds of being subsequently engaged	helps	2.5E-05
H	MK07	K07 applicant	Is Engaged	2.28	odds of being subsequently engaged	helps	0.0038
H	MK25	K25 applicant	Is Engaged	0.38	odds of being subsequently engaged	hurts	0.0406
H	RCU	race unknown	Is Engaged	0.61	odds of being subsequently engaged	hurts	0.0393
H	Funded	was funded for NCI K	Is Engaged	6.02	odds of being subsequently engaged	helps	1.4E-15
H	AgeImpAgeImputed	age was imputed	Is Researcher	0.41	odds of having a subsequent research career	hurts	0.0002
H	NDQ3	applicant institution funding range $10 M - $100 M (highest)	Is Researcher	1.86	odds of having a subsequent research career	helps	0.0062
H	Had_L	had L prior support	Is Researcher	3.55	odds of having a subsequent research career	helps	2.9E-07
H	DPhD_MD	had MD/PhD degree	Is Researcher	1.57	odds of having a subsequent research career	helps	0.0039
H	DOther	had Other degree	Is Researcher	0.29	odds of having a subsequent research career	hurts	0.0061
H	Had_RPG	had RPG prior support	Is Researcher	2.46	odds of having a subsequent research career	helps	0.0015
H	GM	is male	Is Researcher	1.29	odds of having a subsequent research career	helps	0.0056
H	MK11	K11 applicant	Is Researcher	0.47	odds of having a subsequent research career	hurts	0.0073

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
H	MK23	K23 applicant	Is Researcher	1.67	odds of having a subsequent research career	helps	0.0199
H	RCU	race unknown	Is Researcher	0.62	odds of having a subsequent research career	hurts	0.0003
H	Funded	was funded for NCI K	Is Researcher	1.73	odds of having a subsequent research career	helps	1.7E-09
I	Funded	was funded for NCI K	Is Engaged K01	5.52	odds of being subsequently engaged	helps	0.0002
I	AgeImpAgeImputed	age was imputed	Is Engaged K07	0.07	odds of being subsequently engaged	hurts	0.0025
I	Funded	was funded for NCI K	Is Engaged K07	10.64	odds of being subsequently engaged	helps	0.0012
I	NCNonCompCC	applicant institution is a non-comprehensive cancer center	Is Engaged K08	4.02	odds of being subsequently engaged	helps	0.0020
I	DMD	had MD degree	Is Engaged K08	5.20	odds of being subsequently engaged	helps	0.0081
I	DMD	had MD degree	Is Engaged K08	5.20	odds of being subsequently engaged	helps	0.0081
I	DMD	had MD degree	Is Engaged K08	5.20	odds of being subsequently engaged	helps	0.0081
I	DMD	had MD degree	Is Engaged K08	5.20	odds of being subsequently engaged	helps	0.0081
I	DPhD_MD	had MD/PhD degree	Is Engaged K08	7.64	odds of being subsequently engaged	helps	0.0008
I	DPhD_MD	had MD/PhD degree	Is Engaged K08	7.64	odds of being subsequently engaged	helps	0.0008
I	DPhD_MD	had MD/PhD degree	Is Engaged K08	7.64	odds of being subsequently engaged	helps	0.0008
I	DPhD_MD	had MD/PhD degree	Is Engaged K08	7.64	odds of being subsequently engaged	helps	0.0008
I	RCU	race unknown	Is Engaged K08	0.26	odds of being subsequently engaged	hurts	0.0006
I	RCU	race unknown	Is Engaged K08	0.26	odds of being subsequently engaged	hurts	0.0006
I	RCU	race unknown	Is Engaged K08	0.26	odds of being subsequently engaged	hurts	0.0006

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
I	RCU	race unknown	Is Engaged K08	0.26	odds of being subsequently engaged	hurts	0.0006
I	Funded	was funded for NCI K	Is Engaged K08	5.12	odds of being subsequently engaged	helps	8.0E-05
I	Funded	was funded for NCI K	Is Engaged K08	5.12	odds of being subsequently engaged	helps	8.0E-05
I	Funded	was funded for NCI K	Is Engaged K08	5.12	odds of being subsequently engaged	helps	8.0E-05
I	Funded	was funded for NCI K	Is Engaged K08	5.12	odds of being subsequently engaged	helps	8.0E-05
I	DPhD_MD	had MD/PhD degree	Is Researcher K01	2.95	odds of having a subsequent research career	helps	0.0050
I	Funded	was funded for NCI K	Is Researcher K01	2.09	odds of having a subsequent research career	helps	0.0044
I	Had_L	had L prior support	Is Researcher K07	3.29	odds of having a subsequent research career	helps	0.0054
I	Funded	was funded for NCI K	Is Researcher K07	1.82	odds of having a subsequent research career	helps	0.0077
I	FY2007	applied in fiscal year 2007	Is Researcher K08	0.20	odds of having a subsequent research career	hurts	0.0051
I	FY2008	applied in fiscal year 2008	Is Researcher K08	0.13	odds of having a subsequent research career	hurts	0.0003
I	Had_L	had L prior support	Is Researcher K08	6.87	odds of having a subsequent research career	helps	4.8E-05
I	DPhD_MD	had MD/PhD degree	Is Researcher K08	3.61	odds of having a subsequent research career	helps	0.0031
I	RCU	race unknown	Is Researcher K08	0.47	odds of having a subsequent research career	hurts	0.0001
I	Had_T	had T prior support	Is Researcher K11	0.09	odds of having a subsequent research career	hurts	0.0084
I	Had_T	had T prior support	Is Researcher K11	0.09	odds of having a subsequent research career	hurts	0.0084
I	RCU	race unknown	Is Researcher K11	0.05	odds of having a subsequent research career	hurts	0.0001
I	RCU	race unknown	Is Researcher K11	0.05	odds of having a subsequent research career	hurts	0.0001

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
I	Funded	was funded for NCI K	Is Researcher K23	3.57	odds of having a subsequent research career	helps	0.0009

Table A6.10.4 All significant model coefficients

6.10.4 Model Data Dictionary

Data used for the modeling was collected over the course of the study and collected into a single database table for export to the modeling software (R, version 2.13.1). This section describes the definition of the input and outcome variables.

Notes:

(1) For independent categorical variables (aka factors), i.e., variables with a small set of discrete values or levels, the notation [REF] indicates the reference value. Dummy variables are created by the modeling software corresponding to the other values, and the coefficients for each of those auto-generated dummy variables show the estimated change in moving from the reference value to a Yes value for the dummy variable. For example, for degree, the [REF] value is "had PhD", so a coefficient for "had MD" estimates the change in the outcome variable between those who have a PhD and those who have an MD.

Variable	Reference Level
Gender	Female
Race	White
Degree	PhD
Mechanism	K01
Institution Cancer Center Type	Not a Cancer Center
Institution Funding Level	1 (lowest of the 3 funding levels, 0 to < 1 million)
Degree Field of Study	Biological and agricultural sciences

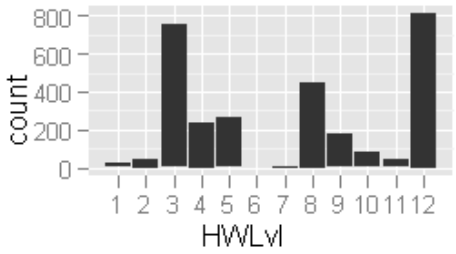
Fiscal Year	1980

(2) The maximum number of Individuals included in any model is 2,889 = 2,893 PIs (excluding K04 and K12) minus 3 PIs with Note degrees and 1 other case. Certain outcome variables have fewer individuals and the lower counts are listed under each variable below.

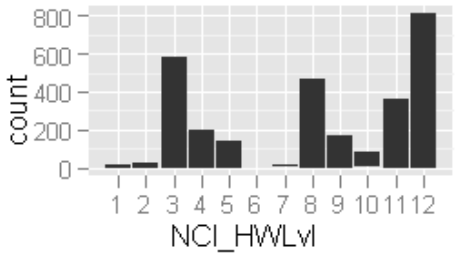
6.10.4.1 Outcome (Y) Variables

HW Lvl After their first NCI K award or their last unsuccessful NCI K application, the highest point achieved by each person on the ranked scale defined below (1-12, factor)

1 Awarded P01 prime or R37
2 Awarded Institutional Training Grant
3 Awarded R01, U01, or P01 sub
4 Awarded other RPG
5 Awarded other grant
6 Applied for P01 prime or R37 (unfunded)
7 Applied for ITG (unfunded)
8 Applied for R01, U01, or P01 sub (unfunded)
9 Applied for other RPG (unfunded)
10 Applied for other grant (unfunded)
11 Other Future
12 None Found

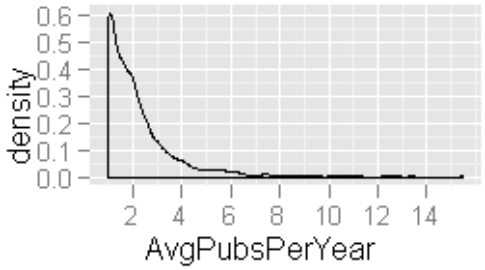


NCI_HWLvl - same as HWLvl but restricted to only NCI post-study applications. PIs with only non-NCI applications are counted in Level 11 (other future), along with those who only had Type 5 appls post-study , always ignoring post-study K continuations.



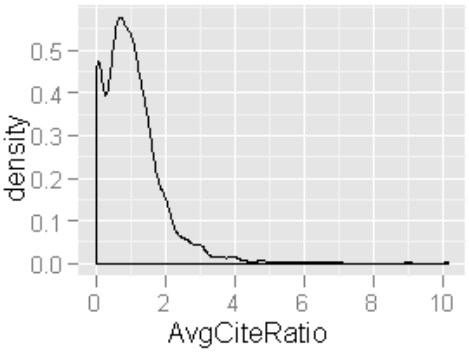
AvgPubsPerYear For those individuals with publications found after their first NCI K award or their last unsuccessful NCI K application, the number of publications in each post-study year were counted. This variable is the average of these per-year counts for each individual.

N= 2,070 individuals with publication data found



AvgCiteRatio For publications by an individual in the post-study period, this is the average of the ratio of actual citations received by a paper in the first 2 years following publication to the benchmark citation count for similar papers in the same time window (algorithm described in the main report).

N= 2,070 individuals with publication data found



Is Researcher Has post-K government funding. This is based on membership in "Group 1" defined below.

Yes 1,555
No 1,334

Is Engaged Is involved in the broader medical science enterprise in a variety of ways, including review committees and publishing and includes those in the "Is Researcher" group. This is based on membership in Groups 1, 2 or 3 defined below.

Note that within the main report, all results related to this variable were restricted to a subset of individuals for which membership in Group 1 (Is Researcher) = no. This allowed comparisons between the group of individuals for whom broader engagement was their highest subsequent career outcome and those who had no evidence of subsequent engagement.

Cohort	Is Engaged = Yes	Is Engaged = No
Full modeling cohort (includes Researchers)	2,599	290
Non-researchers only (used for tests)	1,044	290

For the coefficients related to "is Engaged" showing in **Table A6.10.4**, the full modeling cohort was used, with restrictions to specific K mechanisms as noted.

Below are the group definitions for Is Researcher and Is Engaged. The Boundary FY is the fiscal year of either an individual's first NCI K award or else their last unsuccessful NCI K application.

Group 1 – Carrying out a funded research program
IMPACII Application record (FY > Boundary FY)
- type 1 or 2, status 05, 06, or 99, any IC, any Mechanism, not a K-study continuation
DOE Project record (FY(StartDate) > Boundary FY)
ICRP record (FY(StartDate) > Boundary FY)
Clinical Trial record (FY(StartDate) > Boundary FY)
Group 2 – Involvement in the greater scientific enterprise
FASEB record (FY(MemberSince) > BoundaryFY)
AACR or ASCO record (no date restrictions)
Healthlink record (no date restrictions)
IMPACII Committee record (FY(appointment start date) > Boundary FY)
FIDO record (FY (OriginalEstablishmentDate)
NIH Employee Directory (as of Aug. 2008) – present in match from Feasibility study.
May also be in Group 1
Group 3 – Publishing
MEDLINE record (FY(Publication Year) > Boundary FY)
- does not have to be linked to Web Of Science

May also be in Groups 1 and 2

Group 4 – No other activity

Those individuals not found in Groups 1, 2, or 3.

6.10.4.2 Funding Variable – Used as Both an Input and an Outcome Variable

Funded 1,685 with value 0, 1,204 with value 1 (factor), mean= 0.41

1 = Person had at least 1 NCI K Award in the study period

0 = no NCI K Awards in the study period

6.10.4.3 Input Variables (Independent Variables, X Variables)

G Gender M=Male, F=Female [REF] , U=Unknown (factor) counts sensitive

RC Race (factor) counts sensitive

A	Asian
B	Black
H	Hispanic
N	Native
O	Other
U	Unknown
W	White [REF]

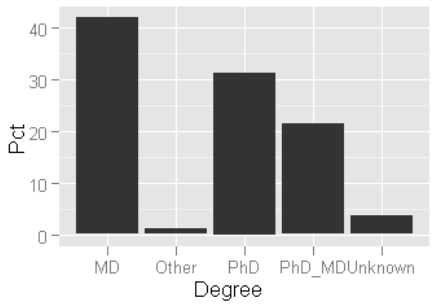
AA Age At First NCI K Award or last unsuccessful NCI K application (numeric)

Average of non-NULL cases was used to Impute NULL cases.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
21.00	34.00	37.00	37.39	39.00	60.00

D Chronologically last degree on record during the Study Period, with previously reported Dual category partitioned and recombined into MD or PhD. (factor)

PhD [REF], others shown below



Prior support Flags All Prior Support for each individual was examined and each of the flags below was set to either 1 if the person had prior support matching the description and 0 if they did not. An individual can have several flags set to 1. Note: For measuring Kendall correlations, a simple Yes(1)/No(0) for any prior support was used.

Had_T	Had any T mechanism (e.g., T32)
Had_F	Had any F mechanism
Had_L	Had any L mechanism
Had_RPG	Had any of the following: DP1, DP2, P01, P42, PN1, R01, R03, R15, R21, R29, R33, R34, R35, R36, R37, R55, R56, RL1, RL2, RL5, RL9, U01, U19, UC1, UC7
Had_Only_Other	Had Prior Support, but no T, F, L, or RPG
No_Prior_Support	No Awards prior to their boundary-year K application
Had_MultiTFL	Had T and F, or T and L, or F and L, or T and F and L, but no RPG
Had_Multi_incl_RPG	Had any of these combinations: RT, RL, RF, RTF, RLF, RTL, RTFL (with R= RPG).

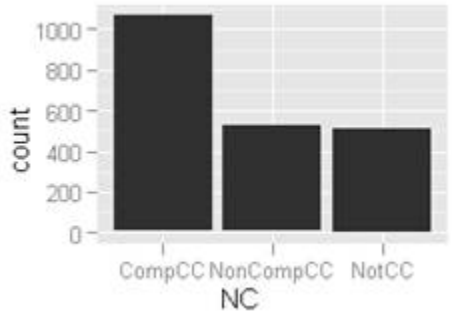
Prior Support	Applicants
Had_T	1,020
Had_F	216
Had_L	144
Had_RPG	128
Had_MultiTFL	165
Had_Multi_incl_RPG	50
Had_Only_Other	33
No_Prior_Support	1,570
	3,326

Note that the total of 3,326 is larger than the model cohort size of 2,889 since these categories are NOT mutually exclusive and Individuals are counted in multiple categories. The cardinality table is listed below: Note that due to the existence of the "Multi" flags, no person can be in exactly 2 of the above categories.

Individuals	# of Prior Support Categories	Total Individual-Category Events
2,682	1	2,682
192	3	576
7	4	28
8	5	40
2,889		3,326

NC Applicant's Institution's Cancer Center Type (factor)

CompCC = Comprehensive Cancer Center
NonCompCC = Non-Comprehensive Cancer Center
NotCC = Not a Cancer Center or Unknown [REF]

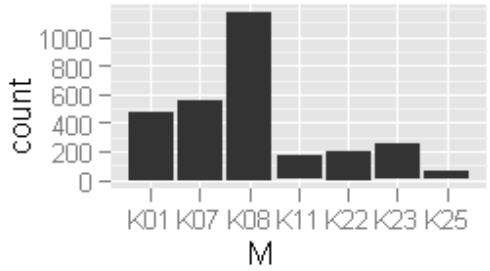


NDQ Applicant's Institution's average annual NCI Funding level (the NDQ symbol was created for an early version of the model that measured funding by quartile rank). For each institution associated with in-study applications and each FY, take the sum of each awarded non-subproject NCI application's calculated total cost amount (a field in the IMPACII database), including all years of funding and all NCI applications of any type or mechanism. Average over FY's per institution. Use the index of the funding range from the table below that includes each institution's average annual NCI funding.

Funding Range (left-inclusive)	NDQ=Funding Bin	Applicants
$0 to < $1 Million	1 [REF]	169
$1 Million to < $10 Million	2	870
$10 Million to < $100 Million	3	1,842
non-awardees whose institutions never received NCI funding	NULL	8
		2,889

M NCI K Primary Mechanism (factor). The primary mechanism is defined in Section 2.3.1 of the report.

K01 [REF], others shown below

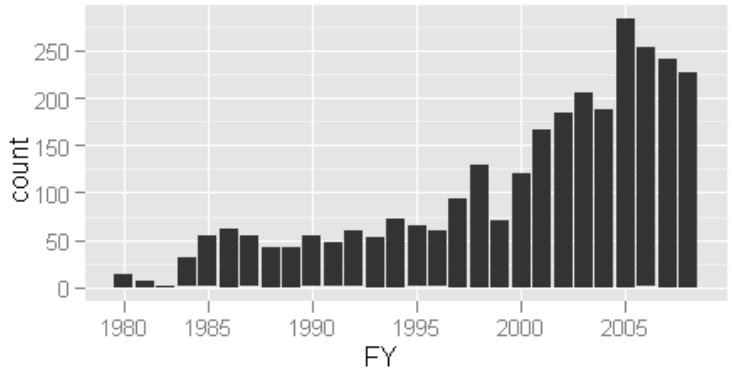


FOS Level 2 Field of Study from DRF (factor)

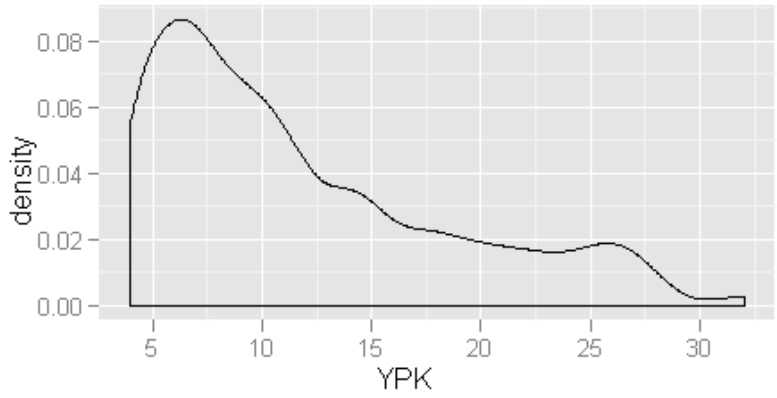
[REF] is Biological and agricultural sciences

FY The fiscal year of an individual's first NCI K award or their last unsuccessful NCI K application (numeric, but treated as factor)

[REF] is 1980



YPK Number of Years between the present (2011) and the fiscal year of an individual's first NCI K award or their last unsuccessful NCI K application (numeric)



Age Imp Age Is Imputed (factor)

AgeFromData	= Age is from a data record (IMPACII, DRF, or AAMC)	count = 2,624 [REF]
AgeImputed	= Age is Imputed to be the average of non-null Ages	count = 265

Is Multi Applicant had multiple NCI K applications (any mix of funded and non-funded) in the original study cohort, ignoring K04 and K12 (factor)

IsMulti	Applicants
0	1,947 [REF]

1	942

Is Multi AC Applicant had NCI K applications in more than one mechanism (any mix of funded and non-funded) in the original study cohort, ignoring K04 and K12 (factor)

IsMultiAC	Applicants
0	2,771 [REF]
1	118

6.11 Statistical Methods Used

To address the questions posed in this section, we employed statistical tests and models that directly addressed each question, based on the type of variables involved. The variable categories and the corresponding statistical tests are summarized in **Table A6.11.1** below. Note that this is not a complete list of all variable type combinations that could be analyzed with the data from this study – it is the set of combinations that were derived from the questions of interest.

Variable Types	Example	Statistical Tools	Effect Measurement	Comments
input: yes/no **outcome: yes/no**	input: K Funded outcome: Is Engaged	2x2 contingency table, Fisher exact test, 2 tailed	Odds ratio	Some tests were stratified (e.g. by mechanism) so the 2x2 table was built for each stratum and a Fisher test was run on each table.
input: 3 categorical values **outcome: yes/no**	Input: prior support 0, 1, or 2 trainings outcome: K Funded	χ^2 test of independence	χ^2 value and cell-wise difference between expected and observed frequencies	Also used to test age quadratic behavior by using 3 age groups. For age, graphical analysis was also used.
input: yes/no **outcome: numeric value**	input: K Funded outcome: Avg Pubs Per person per year	t test, 2-tailed	Difference in the yes/no group averages	Variances were not assumed to be equal. (Welch's t test)

Table A6.11.1. Variable categories and statistical tests used in Section 4.0.

All of the tests were performed using R, version 2.13.1.

The Fisher exact test is a conditional test that computes the exact probability of obtaining cell counts as far or further from a balanced table with an odds ratio of 1 (in either direction) given the values of the row and column totals. The probabilities are computed using the hypergeometric distribution, the cell values, and the range limits for the "yes/yes" cell. When this probability is below the threshold (0.05 was used for all tests in Section 4), we report the observed odds ratio as a significant result.

The χ^2 test measures the probability that the cell counts in a 3x2 table (n x m in general) match what would be expected if the variables that determine the overall row and column totals were statistically independent (so that the frequencies of each cell would equal the product of the corresponding row and column frequencies). The χ^2 value is the sum of the squared differences between the observed and expected frequencies divided by the expected frequencies. This value has a χ^2 distribution with 2 degrees of freedom, and when the probability of obtaining a value as large or larger is below 0.05, we report the overall set of deviations from the expected frequencies as a significant result and show the differences from the expected frequencies for each cell.

The t test directly measures the probability of observing a mean difference as large or larger (in either direction) between 2 independent groups. We used the Welch form of the test in which the variances of the groups are assumed to be different. The standard error is computed as a weighted average of the 2 group standard deviations, and there is an adjustment to the degrees of freedom (df) to yield a value less than the value of N-2 used when the variances are equal. The t value is computed as the observed difference in means divided by the standard error. When the probability of having an absolute value of t that large or larger is less than 0.05, we report the mean difference in the 2 groups as a statistically significant result; however, the observed difference can be small in practical significance.

Comparison	Report Section	χ2 value	p value
Age and NCI K Funding Probability	4.2.3	20.8	3.01E-05
Multiple Prior Support and NCI K Funding Probability	4.2.4	20.8	3.01E-05
Race/Ethnicity and Resubmission of Applications	6.4	0.1739	0.9817

Table A6.11.2 Comparisons and Report Sections in which Chi-Squared Tests Utilized.

7.0 Supplement

7.1 Defining the NCI K Cohort

To define the Demographics Cohorts used in the study, we started with a spreadsheet provided by NCI staff that identified all NCI K applications and primary investigators in the per-mechanism time frame of interest, excluding applications from the NCI's Diversity Training Branch. This set was further reduced using the rules listed below:

1. Restrict to only the following activity codes: K01, K04, K07, K08, K11, K12, K22, K23, and K25.
2. Remove 1 application (APPL ID: 2010808) and 1 individual (ID: 1860688) due to data errors detected by manual QA.
3. Restrict to Individuals with a PI role (excluding other roles like MPI to have a unique individual investigator associated with a given application).
4. Restrict to applications with a non-null status code and exclude status code 03 (so that it is unambiguous whether the application was funded or not).
5. Restrict to primary applications (excluding subprojects).
6. Restrict to Type 1 applications
7. Restrict the cardinality of applications over time so that there is a unique application for any given triplet of Individual, Fiscal Year, and K Mechanism by ranking multiple applications (if any) within each triplet by the following criteria, applied in order, and then picking the unique top-ranked application:
    a) rank Awarded applications over non-Awarded applications
    b) rank by decreasing suffix code, e.g., A(n),...A4, then A3, then A2, A1, and null (A0), initial submission
    c) rank by most recent council meeting date
8. Separate the K12 Cohort into a separate database (with linkage to the NCI-provided scholar list) since demographics and outcomes were studied almost exclusively for the scholars and not the K12 PIs.
9. The K04 applications were separated from the other mechanisms to form a new Cohort about midway through the study. At that time, 11 applications and 10 individuals were deleted to remove cases where the same person had both a K04 application and a non-K04 application.

7.2 Freshness of Data Sources

Data source	Approximate Data As-Of Date
IMPAC II	4/28/2011
DRF	Max PhD year is 2006
AAMC Faculty Roster	4/1/2010
DoD DTIC	7/31/2008
DOE	10/28/2010
NSF FastLane	3/15/2011
MEDLINE	5/13/2011
Thomson Reuters Web of Science	5/1/2011
Lodestone / HealthLink	4/11/2011
NIH - NED	8/5/2008
International Cancer Research Portfolio	7/31/2008
FASEB	10/7/2010
AACR*	2/14/2011
ASCO*	3/21/2011
FIDO.gov	9/30/2008
PECASE	5/24/2011
LinkedIn, google.com^	Searches conducted 8/1/2011 - 9/30/2011

* Indicates the date data received by Discovery Logic.
^ These resources were used only for the analysis of a random subset of NCI K program applicants, with data collected during the noted date range.

7.3 PhD Fields of Study (FOS) of NCI K Applicants Matched to the DRF

7.3.1 Combined Top PhD FOS of NCI K Applicants Matched to DRF (in alphabetical order)

Acoustics
Analysis and Functional Analysis
Analytical Chemistry
Anatomy
Animal Nutrition
Animal Science, Other
Anthropology
Applied Mathematics
Biochemistry
Bioengineering and Biomedical Engineering
Biology/Biomedical Sciences, General
Biology/Biomedical Sciences, Other
Biomedical Sciences
Biometrics and Biostatistics
Biophysics
Biotechnology
Botany/Plant Biology
Business/Managerial Economics
Cell/Cellular Biology and Histology
Chemistry, General
Chemistry, Other
Civil Engineering
Clinical Psychology
Cognitive Psychology and Psycholinguistics
Communication Research
Communication Theory
Communication, General
Computer Engineering
Computer Science
Condensed Matter/Low Temperature Physics
Counseling
Curriculum and Instruction
Demography/Population Studies
Developmental and Child Psychology
Developmental Biology/Embryology
Ecology
Educational Administration and Supervision
Educational Psychology
Educational Statistics/Research Methods

Educational/Instructional Media Design
Electrical, Electronics and Communications
Endocrinology
Engineering Mechanics
Engineering Physics
Engineering, Other
Environmental Health
Epidemiology
Experimental Psychology
Genetics
Genetics/Genomics, Human and Animal
Health Education
Health Sciences, General
Health Sciences, Other
Health Systems/Services Administration
Horticulture Science
Human Development and Family Studies
Humanities, General
Immunology
Inorganic Chemistry
Kinesiology/Exercise Science
Management Information Systems/Business Statistics
Mass Communication/Media Studies
Materials Science
Mathematics/Statistics, General
Mechanical Engineering
Medicinal/Pharmaceutical Chemistry
Medicinal/Pharmaceutical Sciences
Microbiology
Microbiology and Bacteriology
Molecular Biology
Neurosciences
Nuclear Engineering
Nuclear Physics
Nursing Science
Nutritional Sciences
Operations Research
Organic Chemistry
Parasitology
Parks/Sports/Rec./Leisure/Fitness
Particle (Elementary) Physics
Pathology, Human and Animal

Pharmacology, Human and Animal
Philosophy
Physical Chemistry
Physics, General
Physics, Other
Physiology, Human and Animal
Plant Genetics
Plasma/Fusion Physics
Polymer Chemistry
Psychology, General
Psychology, Other
Public Administration
Public Health
Public Health and Epidemiology
Public Policy Analysis
School Psychology
Science Education
Social Psychology
Social Sciences, General
Social Sciences, Other
Social Work
Social/Philosophical Foundations of Education
Sociology
Statistics
Toxicology
Veterinary Sciences
Zoology

7.3.2 Top PhD FOS of Applicants, by NCI K Mechanism (in alphabetical order)

K01 FOS

Acoustics
Analytical Chemistry
Anatomy
Animal Nutrition
Animal Science, Other
Applied Mathematics
Biochemistry
Bioengineering and Biomedical Engineering
Biology/Biomedical Sciences, General
Biology/Biomedical Sciences, Other
Biomedical Sciences
Biophysics
Botany/Plant Biology
Business/Managerial Economics
Cell/Cellular Biology and Histology
Chemistry, General
Chemistry, Other
Computer Science
Developmental Biology/Embryology
Endocrinology
Environmental Health
Genetics/Genomics, Human and Animal
Health Sciences, Other
Horticulture Science
Immunology
Inorganic Chemistry
Medicinal/Pharmaceutical Chemistry
Medicinal/Pharmaceutical Sciences
Microbiology
Molecular Biology
Neurosciences
Nutritional Sciences
Organic Chemistry
Parasitology
Pathology, Human and Animal
Pharmacology, Human and Animal
Physics, Other
Physiology, Human and Animal
Plant Genetics
Science Education
Toxicology
Zoology

K07 FOS
Anthropology
Biochemistry
Biology/Biomedical Sciences, General
Biometrics and Biostatistics
Cell/Cellular Biology and Histology
Clinical Psychology
Cognitive Psychology and Psycholinguistics
Communication Research
Communication Theory
Communication, General
Computer Science
Counseling
Curriculum and Instruction
Developmental and Child Psychology
Ecology
Educational Psychology
Educational Statistics/Research Methods
Educational/Instructional Media Design
Engineering, Other
Environmental Health
Epidemiology
Experimental Psychology
Genetics/Genomics, Human and Animal
Health Education
Health Sciences, General
Health Sciences, Other
Health Systems/Services Administration
Human Development and Family Studies
Inorganic Chemistry
Kinesiology/Exercise Science
Mass Communication/Media Studies
Medicinal/Pharmaceutical Chemistry
Medicinal/Pharmaceutical Sciences
Microbiology
Molecular Biology
Nursing Science
Nutritional Sciences
Organic Chemistry
Parks/Sports/Rec./Leisure/Fitness
Pathology, Human and Animal
Pharmacology, Human and Animal
Physiology, Human and Animal
Plasma/Fusion Physics
Psychology, General

Psychology, Other
Public Administration
Public Health
Public Health and Epidemiology
Public Policy Analysis
School Psychology
Social Psychology
Social Sciences, General
Social Sciences, Other
Social Work
Social/Philosophical Foundations of Education
Sociology
Statistics
Toxicology

K08 FOS
Anatomy
Biochemistry
Bioengineering and Biomedical
Biology/Biomedical Sciences, General
Biology/Biomedical Sciences, Other
Biomedical Sciences
Biophysics
Cell/Cellular Biology and Histology
Chemistry, General
Condensed Matter/Low Temperature Physics
Developmental Biology/Embryology
Ecology
Electrical, Electronics and Communications
Engineering
Endocrinology
Environmental Health
Epidemiology
Genetics
Genetics/Genomics, Human and Animal
Health Sciences, General
Health Sciences, Other
Humanities, General
Immunology
Medicinal/Pharmaceutical Sciences
Microbiology
Microbiology and Bacteriology
Molecular Biology
Neurosciences
Nutritional Sciences
Organic Chemistry
Particle (Elementary) Physics
Pathology, Human and Animal

Pharmacology, Human and Animal
Philosophy
Physical Chemistry
Physiology, Human and Animal
Toxicology

K23 FOS
Anthropology
Biochemistry
Biology/Biomedical Sciences, General
Biology/Biomedical Sciences, Other
Biomedical Sciences
Biophysics
Cell/Cellular Biology and Histology
Clinical Psychology
Epidemiology
Genetics/Genomics, Human and Animal
Health Sciences, Other
Immunology
Microbiology
Molecular Biology
Neurosciences
Nutritional Sciences
Pathology, Human and Animal
Pharmacology, Human and Animal
Public Health
Public Policy Analysis

K25 FOS
Analysis and Functional Analysis
Analytical Chemistry
Biochemistry
Bioengineering and Biomedical
Biometrics and Biostatistics
Biophysics
Biotechnology
Civil Engineering
Computer Engineering
Computer Science
Demography/Population Studies
Electrical, Electronics and Communications
Engineering Mechanics
Engineering Physics
Management Information Systems/Business Statistics
Materials Science
Mechanical Engineering
Nuclear Engineering

Nuclear Physics
Operations Research
Organic Chemistry
Physics, General
Physics, Other
Polymer Chemistry
Statistics

7.4 NIH High Water Mark Analysis by NCI K Mechanism

Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K01	1	Awarded P01 primary PI or R37	0	0	0	0
K01	2	Awarded Institutional Training Grant	0	0	0	0
K01	3	Awarded R01, U01, or P01 subproject PI	83	84	22	19
K01	4	Awarded other RPG	8	28	3	2
K01	5	Awarded Other Grant	8	15	5	2
K01	6	Applied for P01 primary PI or R37 (unfunded)	0	0	0	0
K01	7	Applied for Institutional Training Grant (unfunded)	1	0	1	0
K01	8	Applied for R01, U01, or P01 subproject PI (unfunded)	34	61	15	10
K01	9	Applied for other RPG (unfunded)	2	29	0	2
K01	10	Applied for Other Grant (unfunded)	0	7	0	1
K01	11	Other Activity (e.g., Type 5)	1	4	0	1
K01	12	No Subsequent Applications	15	99	4	13
Mechanism	**High Water Mark Category**	**High Water Mark Category Description**	**Full Cohort**		**Comparison Cohort**	
			Awardees	**Non-Awardees**	**Awardees**	**Non-Awardees**
K07	1	Awarded P01 primary PI or R37	0	2	0	1
K07	2	Awarded Institutional Training Grant	10	2	3	1
K07	3	Awarded R01, U01, or P01 subproject PI	91	41	9	8
K07	4	Awarded other RPG	45	32	7	4
K07	5	Awarded Other Grant	32	40	7	9
K07	6	Applied for P01 primary PI or R37 (unfunded)	0	0	0	0
K07	7	Applied for Institutional Training Grant (unfunded)	4	1	1	0
K07	8	Applied for R01, U01, or P01 subproject PI (unfunded)	48	31	8	5
K07	9	Applied for other RPG (unfunded)	6	30	2	2
K07	10	Applied for Other Grant (unfunded)	5	11	0	1
K07	11	Other Activity (e.g., Type 5)	2	3	0	0
K07	12	No Subsequent Applications	31	95	4	10

Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K08	1	Awarded P01 primary PI or R37	11	8	3	3
K08	2	Awarded Institutional Training Grant	11	13	1	4
K08	3	Awarded R01, U01, or P01 subproject PI	200	123	51	31
K08	4	Awarded other RPG	37	38	14	12
K08	5	Awarded Other Grant	34	66	8	15
K08	6	Applied for P01 primary PI or R37 (unfunded)	1	1	0	0
K08	7	Applied for Institutional Training Grant (unfunded)	1	1	0	1
K08	8	Applied for R01, U01, or P01 subproject PI (unfunded)	91	82	26	16
K08	9	Applied for other RPG (unfunded)	20	42	7	5
K08	10	Applied for Other Grant (unfunded)	13	21	5	7
K08	11	Other Activity (e.g., Type 5)	6	11	2	1
K08	12	No Subsequent Applications	89	256	25	47

Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K11	1	Awarded P01 primary PI or R37	1	1	0	0
K11	2	Awarded Institutional Training Grant	6	1	1	0
K11	3	Awarded R01, U01, or P01 subproject PI	29	8	1	2
K11	4	Awarded other RPG	7	4	0	0
K11	5	Awarded Other Grant	2	10	0	1
K11	6	Applied for P01 primary PI or R37 (unfunded)	0	0	0	0
K11	7	Applied for Institutional Training Grant (unfunded)	0	0	0	0
K11	8	Applied for R01, U01, or P01 subproject PI (unfunded)	11	2	4	0
K11	9	Applied for other RPG (unfunded)	1	2	0	1
K11	10	Applied for Other Grant (unfunded)	0	4	0	1
K11	11	Other Activity (e.g., Type 5)	1	2	0	1
K11	12	No Subsequent Applications	28	46	4	4

Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K22	1	Awarded P01 primary PI or R37	0	0	0	0
K22	2	Awarded Institutional Training Grant	1	0	1	0
K22	3	Awarded R01, U01, or P01 subproject PI	25	25	7	2
K22	4	Awarded other RPG	5	13	3	2
K22	5	Awarded Other Grant	2	10	0	2
K22	6	Applied for P01 primary PI or R37 (unfunded)	0	0	0	0
K22	7	Applied for Institutional Training Grant (unfunded)	0	0	0	0
K22	8	Applied for R01, U01, or P01 subproject PI (unfunded)	22	21	9	5
K22	9	Applied for other RPG (unfunded)	0	17	0	2
K22	10	Applied for Other Grant (unfunded)	0	5	0	0
K22	11	Other Activity (e.g., Type 5)	0	2	0	0
K22	12	No Subsequent Applications	2	50	1	8
Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K23	1	Awarded P01 primary PI or R37	0	0	0	0
K23	2	Awarded Institutional Training Grant	2	1	0	0
K23	3	Awarded R01, U01, or P01 subproject PI	27	9	5	4
K23	4	Awarded other RPG	8	8	4	1
K23	5	Awarded Other Grant	19	23	7	1
K23	6	Applied for P01 primary PI or R37 (unfunded)	0	1	0	0
K23	7	Applied for Institutional Training Grant (unfunded)	0	0	0	0
K23	8	Applied for R01, U01, or P01 subproject PI (unfunded)	17	17	2	2
K23	9	Applied for other RPG (unfunded)	9	14	3	0
K23	10	Applied for Other Grant (unfunded)	1	9	1	2
K23	11	Other Activity (e.g., Type 5)	1	7	0	4
K23	12	No Subsequent Applications	14	67	3	11

Mechanism	High Water Mark Category	High Water Mark Category Description	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K25	1	Awarded P01 primary PI or R37	0	0	0	0
K25	2	Awarded Institutional Training Grant	0	1	0	0
K25	3	Awarded R01, U01, or P01 subproject PI	4	1	1	0
K25	4	Awarded other RPG	2	3	0	1
K25	5	Awarded Other Grant	1	0	0	0
K25	6	Applied for P01 primary PI or R37 (unfunded)	0	0	0	0
K25	7	Applied for Institutional Training Grant (unfunded)	0	0	0	0
K25	8	Applied for R01, U01, or P01 subproject PI (unfunded)	7	5	1	1
K25	9	Applied for other RPG (unfunded)	4	1	0	0
K25	10	Applied for Other Grant (unfunded)	1	3	0	1
K25	11	Other Activity (e.g., Type 5)	0	0	0	0
K25	12	No Subsequent Applications	6	17	2	1

7.5 NCI High Water Mark Analysis, by NCI K Mechanism

Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K01	1	Awarded NCI P01 primary PI or R37	0	0	0	0
K01	2	Awarded NCI Institutional Training Grant	0	0	0	0
K01	3	Awarded NCI R01, U01, or P01 subproject PI	61	48	15	15
K01	4	Awarded other NCI RPG	5	17	3	1
K01	5	Awarded Other NCI Grant	4	4	2	1
K01	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	0	0	0
K01	7	Applied for NCI Institutional Training Grant (unfunded)	1	0	1	0
K01	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	38	67	18	7
K01	9	Applied for other NCI RPG (unfunded)	5	23	2	1
K01	10	Applied for Other NCI Grant (unfunded)	0	7	0	1
K01	11	Other NIH Activity, but not NCI (e.g., Type 5)	23	62	5	11
K01	12	No Subsequent Applications	15	99	4	13
Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K07	1	Awarded NCI P01 primary PI or R37	0	2	0	1
K07	2	Awarded NCI Institutional Training Grant	9	0	3	0
K07	3	Awarded NCI R01, U01, or P01 subproject PI	79	29	6	8
K07	4	Awarded other NCI RPG	41	29	7	4
K07	5	Awarded Other NCI Grant	25	13	2	2
K07	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	0	0	0
K07	7	Applied for NCI Institutional Training Grant (unfunded)	5	2	1	0
K07	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	47	35	11	8
K07	9	Applied for other NCI RPG (unfunded)	9	24	1	4
K07	10	Applied for Other NCI Grant (unfunded)	5	10	1	0
K07	11	Other NIH Activity, but not NCI (e.g., Type 5)	23	49	5	4
K07	12	No Subsequent Applications	31	95	4	10

Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K08	1	Awarded NCI P01 primary PI or R37	8	5	1	2
K08	2	Awarded NCI Institutional Training Grant	5	7	0	2
K08	3	Awarded NCI R01, U01, or P01 subproject PI	160	100	43	26
K08	4	Awarded other NCI RPG	36	33	11	9
K08	5	Awarded Other NCI Grant	28	35	6	5
K08	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	1	0	0
K08	7	Applied for NCI Institutional Training Grant (unfunded)	3	1	1	1
K08	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	98	89	30	21
K08	9	Applied for other NCI RPG (unfunded)	25	37	7	7
K08	10	Applied for Other NCI Grant (unfunded)	15	23	6	4
K08	11	Other NIH Activity, but not NCI (e.g., Type 5)	47	75	12	18
K08	12	No Subsequent Applications	89	256	25	47
Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K11	1	Awarded NCI P01 primary PI or R37	1	1	0	0
K11	2	Awarded NCI Institutional Training Grant	4	0	0	0
K11	3	Awarded NCI R01, U01, or P01 subproject PI	26	7	2	1
K11	4	Awarded other NCI RPG	6	4	0	0
K11	5	Awarded Other NCI Grant	1	5	0	0
K11	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	0	0	0
K11	7	Applied for NCI Institutional Training Grant (unfunded)	0	0	0	0
K11	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	10	3	4	2
K11	9	Applied for other NCI RPG (unfunded)	1	2	0	1
K11	10	Applied for Other NCI Grant (unfunded)	0	1	0	0
K11	11	Other NIH Activity, but not NCI (e.g., Type 5)	9	11	0	2
K11	12	No Subsequent Applications	28	46	4	4

Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K22	1	Awarded NCI P01 primary PI or R37	0	0	0	0
K22	2	Awarded NCI Institutional Training Grant	0	0	0	0
K22	3	Awarded NCI R01, U01, or P01 subproject PI	19	18	5	1
K22	4	Awarded other NCI RPG	6	8	4	2
K22	5	Awarded Other NCI Grant	0	6	0	0
K22	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	0	0	0
K22	7	Applied for NCI Institutional Training Grant (unfunded)	0	1	0	0
K22	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	20	18	7	5
K22	9	Applied for other NCI RPG (unfunded)	1	16	0	1
K22	10	Applied for Other NCI Grant (unfunded)	0	5	0	1
K22	11	Other NIH Activity, but not NCI (e.g., Type 5)	9	21	4	3
K22	12	No Subsequent Applications	2	50	1	8
Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K23	1	Awarded NCI P01 primary PI or R37	0	0	0	0
K23	2	Awarded NCI Institutional Training Grant	2	1	0	0
K23	3	Awarded NCI R01, U01, or P01 subproject PI	25	9	5	4
K23	4	Awarded other NCI RPG	6	7	4	1
K23	5	Awarded Other NCI Grant	14	10	5	0
K23	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	1	0	0
K23	7	Applied for NCI Institutional Training Grant (unfunded)	0	0	0	0
K23	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	21	17	3	2
K23	9	Applied for other NCI RPG (unfunded)	8	18	3	1
K23	10	Applied for Other NCI Grant (unfunded)	1	7	0	0
K23	11	Other NIH Activity, but not NCI (e.g., Type 5)	7	19	2	6
K23	12	No Subsequent Applications	14	67	3	11

Mechanism	High Water Mark Category	High Water Mark Category Description - NCI	Full Cohort		Comparison Cohort	
			Awardees	Non-Awardees	Awardees	Non-Awardees
K25	1	Awarded NCI P01 primary PI or R37	0	0	0	0
K25	2	Awarded NCI Institutional Training Grant	0	0	0	0
K25	3	Awarded NCI R01, U01, or P01 subproject PI	2	1	1	0
K25	4	Awarded other NCI RPG	2	3	0	0
K25	5	Awarded Other NCI Grant	0	0	0	0
K25	6	Applied for NCI P01 primary PI or R37 (unfunded)	0	0	0	0
K25	7	Applied for NCI Institutional Training Grant (unfunded)	0	0	0	0
K25	8	Applied for NCI R01, U01, or P01 subproject PI (unfunded)	7	2	1	1
K25	9	Applied for other NCI RPG (unfunded)	4	0	0	0
K25	10	Applied for Other NCI Grant (unfunded)	2	2	0	1
K25	11	Other NIH Activity, but not NCI (e.g., Type 5)	2	6	0	1
K25	12	No Subsequent Applications	6	17	2	1

7.6 Model Results for Publication Metrics including Applicants who did not Publish

The linear models in Model Group F (described in **Appendix 6.10.1**) were run against a data set that included non-publishers, with zero values for the Publications Per Person Per Year and Average Actual to Benchmark Citation Ratio outcome variables, but otherwise identical to that used for the results listed in **Appendix 6.10.3**. The table below lists the statistically significant coefficients from this model (new Model Group J) run using the same format and semantics as **Table A.6.10.4**.

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
J	Funded	was funded for NCI K	Actual to Benchmark Cite Ratio	0.28	cite ratio	helps	1.95E-15
J	FY2008	applied in fiscal year 2008	Actual to Benchmark Cite Ratio	-0.63	cite ratio	hurts	0.0073
J	GM	is male	Actual to Benchmark Cite Ratio	0.08	cite ratio	helps	0.0189
J	Had_RPG	had RPG prior support	Actual to Benchmark Cite Ratio	0.20	cite ratio	helps	0.0412
J	RCU	race unknown	Actual to Benchmark Cite Ratio	-0.16	cite ratio	hurts	0.0017
J	AgeImpAgeImputed	age was imputed	Pub Productivity	-0.61	publications per person per year	hurts	0.0001
J	DPhD_MD	had MD/PhD degree	Pub Productivity	0.38	publications per person per year	helps	0.0003
J	Funded	was funded for NCI K	Pub Productivity	0.46	publications per person per year	helps	2.19E-13
J	FY1988	applied in fiscal year 1988	Pub Productivity	1.07	publications per person per year	helps	0.0227
J	FY1990	applied in fiscal year 1990	Pub Productivity	0.99	publications per person per year	helps	0.0309
J	FY1991	applied in fiscal year 1991	Pub Productivity	0.91	publications per person per year	helps	0.0498
J	FY1993	applied in fiscal year 1993	Pub Productivity	1.05	publications per person per year	helps	0.0222
J	FY1997	applied in fiscal year 1997	Pub Productivity	1.15	publications per person per year	helps	0.0085

Model Group	Input Variable Symbol	Input Variable Description	Outcome Variable	Coeffi-cient	Coefficient Units (per x unit)	Mnemonic	p Value
J	FY1998	applied in fiscal year 1998	Pub Productivity	0.96	publications per person per year	helps	0.0265
J	FY2000	applied in fiscal year 2000	Pub Productivity	1.26	publications per person per year	helps	0.0035
J	FY2001	applied in fiscal year 2001	Pub Productivity	1.00	publications per person per year	helps	0.0189
J	FY2002	applied in fiscal year 2002	Pub Productivity	1.27	publications per person per year	helps	0.0029
J	FY2003	applied in fiscal year 2003	Pub Productivity	1.24	publications per person per year	helps	0.0036
J	FY2004	applied in fiscal year 2004	Pub Productivity	1.21	publications per person per year	helps	0.0046
J	FY2005	applied in fiscal year 2005	Pub Productivity	1.17	publications per person per year	helps	0.0056
J	FY2006	applied in fiscal year 2006	Pub Productivity	0.96	publications per person per year	helps	0.0232
J	FY2007	applied in fiscal year 2007	Pub Productivity	1.01	publications per person per year	helps	0.0175
J	FY2008	applied in fiscal year 2008	Pub Productivity	0.99	publications per person per year	helps	0.0199
J	GM	is male	Pub Productivity	0.44	publications per person per year	helps	2.46E-12
J	Had_RPG	had RPG prior support	Pub Productivity	0.79	publications per person per year	helps	9.22E-06
J	Had_T	had T prior support	Pub Productivity	-0.22	publications per person per year	hurts	0.0032
J	MK07	K07 applicant	Pub Productivity	0.74	publications per person per year	helps	7.89E-11
J	MK23	K23 applicant	Pub Productivity	0.50	publications per person per year	helps	0.0006
J	NDQ3	applicant institution funding range 10 M - 100 M (highest)	Pub Productivity	0.34	publications per person per year	helps	0.0228
J	RCU	race unknown	Pub Productivity	-0.28	publications per person per year	hurts	0.0018