IDA

INSTITUTE FOR DEFENSE ANALYSES

# An Outcome Evaluation of the National Institutes of Health (NIH) Director's Pioneer Award (NDPA) Program, FY 2004–2006

Bhavya Lal
Alyson Wilson
Seth Jonas
Elizabeth Lee
Amy Richards
Vanessa Peña

Draft Final
August 2012
IDA Paper P-4899
Log: H 12-000966/1
Copy

IDA SCIENCE & TECHNOLOGY
POLICY INSTITUTE
1899 Pennsylvania Ave., Suite 520
Washington, DC 20006-3602

*The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.*

**About this Publication**

This work was conducted by the IDA Science and Technology Policy Institute under contract OIA-0408601, Task OSTP-20-0004.87 (NH-20-6440), "Comparing the Outputs and Outcomes of the NIH Director's Pioneer Award with that of the NIH R01 Program," for the National Institutes of Health. The views, opinions, and findings should not be construed as representing the official position of the National Institutes of Health, the National Science Foundation, or the Office of Science and Technology Policy in the Executive Office of the President.

**Acknowledgements**

The authors would like to acknowledge the comments and observations of Susan Rose of IDA, Stephanie Shipp and Brian Zuckerman of the IDA Science and Technology Policy Institute, Liz Allen of the Wellcome Trust Foundation, and Juan Rogers from the Georgia Institute of Technology, who were the technical reviewers for this work. Thanks also go to Allison Riggieri, Jamie Doyle, Gina Walejko, and Rachel Parker of the IDA Science and Technology Policy Institute and Pierre Azoulay of the Massachusetts Institute of Technology for providing their subject matter expertise to our efforts.

**Copyright Notice**

© 2012 Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, Virginia 22311 ·1882 • (703) 845·2000.

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-4899

# An Outcome Evaluation of the National Institutes of Health (NIH) Director's Pioneer Award (NDPA) Program, FY 2004–2006

Bhavya Lal
Alyson Wilson
Seth Jonas
Elizabeth Lee
Amy Richards
Vanessa Peña

# Executive Summary

The National Institutes of Health (NIH) Director's Pioneer Award (NDPA) was initiated in fiscal year (FY) 2004 to support individual scientists of exceptional creativity who propose pioneering approaches to major contemporary challenges in biomedical research. The program is approaching its tenth year, which is likely to be the last year it will be fully supported through the NIH Common Fund. To help steer the future directions of the program, the NIH therefore requested the IDA Science and Technology Policy Institute (STPI) to evaluate the NDPA program.

## Goals and Parameters of the Evaluation

This evaluation addressed two primary study questions that followed from the program goals stated in the NDPA solicitations: To what extent does the research supported by the NDPA (or the "Pioneer") program produce unusually high impact, and to what extent are the research approaches used by the NDPA grantees (or the "Pioneers") highly innovative? The evaluation also addressed two secondary questions: To what extent is the Pioneers' research interdisciplinary, and to what extent are the Pioneers collaborative?

The evaluation focused on scientific publications between the date of award and the end of 2011. To ensure the entire 5-year NDPA grant period would be covered, STPI included only the first three cohorts of the Pioneers (35 in total), limiting the analysis to awardees in the period FY 2004–2006.

Because some of the study questions are expressed as statements of relative value (e.g., unusually high impact), and because NIH intends to use the evaluation to help decide the future of the program in light of its full portfolio, the study questions were addressed comparatively against the five comparison groups described in the following section.

## Comparison Groups

The first comparison group was a matched set of 35 new grants from the NIH Research Project Grant Program (R01) awarded from FY 2004 through FY 2006. The R01 set was constructed statistically using matching methods to ensure that the R01 grantees "looked" enough like the Pioneers that any differences in outcomes could be attributed to the NDPA rather than to characteristics of the investigators.

The second comparison group was the set of 35 recipients of the matched R01 grants. The first group allowed comparison with the NDPA grants (grant-level analysis), and the second group, with the Pioneers (researcher level analysis).

The third group comprised the 39 Howard Hughes Medical Institute (HHMI) investigators who received awards in FY 2005. This group allowed comparison of Pioneer outcomes with those of researchers from a similar research program that purports to fund similarly high-risk high-reward research.

We chose the 30 NDPA finalists from FY 2004–2006 as a fourth comparison group because finalists have demonstrated the most similarity to the Pioneers with respect to being exceptionally creative researchers having proposed high-risk ideas deemed worthy of consideration to be interviewed in the final round.

The fifth group comprised 30 sets or portfolios of randomly selected new grants issued from FY 2004 through FY 2006. Each set was selected to ensure that their direct costs added up to the same total direct costs as the NDPA portfolio. The portfolio level analysis allowed us to assess how much "impact" the same amount of direct funding would buy in the mainstream grant universe.

## Method

Analyses were performed both at the grant and researcher levels; grant-attributed publications represent the body of work produced as a result of a particular award, while researcher-attributed publications reflect the effect of multiple grants and funding sources on a researcher's published work. We analyzed over 20,000 publications and associated meta-data using Thomson Reuters' Web of Science database and the NIH's electronic Scientific Portfolio Assistant (eSPA), Scientific Publication Information Retrieval and Evaluation System (SPIRES), and SPIRES+ databases. The final set included 15,165 grant-attributed publications (518 Pioneer, 295 matched R01, and 14,352 from the random R01 portfolio) and 8,859 researcher-level publications (3,287 Pioneer, 3,274 matched R01, and 2,298 from the NDPA finalists).

We employed both bibliometric analyses and expert reviews to measure the scientific impact of researchers and their NDPA research and the innovativeness of their research approaches. Bibliometric analyses included computations of the numbers of citations of awardees, grants, and publications; SCImago Journal Rankings, a measure of journal impact; and h-indices, a measure of scientific productivity and impact based on numbers of publications and citations.

To complement the bibliometric analyses and their associated limitations, PI publications were also reviewed by experts to assess their impact and innovativeness. 94 experts were recruited, and reviewed 108 sets of "top five" papers from the NDPA award, the matched R01 grants, and the HHMI investigators. A total of 1,923reviews led to

assessments in the form of quantitative ratings and qualitative comments on the impact and innovativeness of the research approaches at a "body of work" and individual paper levels.

## Findings and Observations

On indicators of impact, we found that the Pioneers—both at the grant and researcher levels—scored as well as or higher than the matched R01 grants and PIs. The reasons for the differences are likely complex. By design, the two groups were well-matched on pre-award PI characteristics and research areas. However, the Pioneers received more funding than the matched R01 PIs, through a grant mechanism intended to provide more flexibility and to fund riskier ideas. The differences between the matched R01 grants and NDPAs are likely not attributable to PI differences or differences in research area, but may be due to differences in funding or programmatic differences.

The random portfolio included 2-3 times as many grants as the Pioneer portfolio, and also produced more publications. However, on bibliometric measures of impact (number of citations, journal impact factors, and h-indices), the Pioneer portfolio scored as well as or higher than the 30 similarly sized random R01 portfolios. Since funding levels for the two portfolios were about the same, the differences between these groups are likely not attributable to funding, but may be due to differences in PI characteristics, research area, or program characteristics.

The Pioneers scored as well as or lower than HHMI investigators on all indicators of impact. While both programs provide flexibility and aim to fund riskier ideas, NDPA provides less funding than HHMI. The differences in outcomes are likely due to funding levels, differences in PIs, or differences in areas of science, or other programmatic characteristics.

Lastly, the Pioneers scored as well as or higher than the NDPA finalists on all measures except interdisciplinarity. The Pioneers and NDPA finalists were well-matched on pre-award characteristics, so these differences are possibly attributable to differences in subsequent funding or research area. The following tables summarize the differences across all groups visually.

**Productivity of NDPA Compared with Other Groups**

| | Matched R01 (Grant Level) | Random R01 Portfolios | HHMI Investigators | Matched R01 (Researcher Level) | NDPA Finalists |
|---|---|---|---|---|---|
| Number of publications | | | | | |
| Number of publications per grant funding amount | | | | N/A | N/A |

*Note:* Green indicates that the NDPA group rated higher than the comparison group, yellow indicates that the NDPA rated about the same as the comparison group, and red indicates that the NDPA rated lower. N/A indicates that it was not feasible to perform the requisite analysis.

**Bibliometric Impact of NDPA Compared with Other Groups**

| | Matched R01 (Grant Level) | Random R01 Portfolios | HHMI Investigators | Matched R01 (Researcher Level) | NDPA Finalists |
|---|---|---|---|---|---|
| Number of citations per awardee | | | | | |
| Number of citations per grant funding amount | | | | N/A | N/A |
| Number of citations per publication | | | | | |
| H-index | | N/A | | | |
| Journal ranking | | | | | |

**Expert Assessed Impact of NDPA Compared with Other Groups**

| | Matched R01 (Grant Level) | HHMI Investigators |
|---|---|---|
| Packets | | |
| Papers | | |

**Expert Assessed Innovativeness of NDPA Research Approaches**

| | Matched R01 (Grant Level) | HHMI Investigators |
|---|---|---|
| Packets | | |
| Papers | | |

**Bibliometric Interdisciplinarity and Collaborations of NDPAs**

| | Matched R01 (Grant Level) | HHMI Investigators | Matched R01 (Researcher Level) | NDPA Finalists |
|---|---|---|---|---|
| Integration score | | | | |
| Number of co-authors | | | | |
| Number of co-author affiliations | | | | |

*Note:* Green indicates that the NDPA group rated higher than the comparison group, yellow indicates that the NDPA rated about the same as the comparison group, and red indicates that the NDPA rated lower. N/A indicates that it was not feasible to perform the requisite analysis.

# Contents


1. Introduction ..... 1
   A. Overview of the Program ..... 1
   B. Purpose of the Outcome Evaluation ..... 1
   C. Study Questions ..... 2
   D. Scope of the Evaluation ..... 2
      1. Conceptualizing the Approach ..... 2
      2. Operationalizing Terms of Interest ..... 4
      3. Other Parameters to Bound the Evaluation ..... 5
   E. Overview of the Report ..... 5
2. Data ..... 7
   A. Description of Comparison Groups ..... 7
      1. Matched Set of R01 Grants and PIs ..... 7
      2. R01 Portfolios ..... 15
      3. HHMI Investigators ..... 17
      4. NDPA Finalists ..... 18
   B. Comparison Groups at Time of Award ..... 19
   C. Collection of Data ..... 25
      1. Publication Data (Matched R01 PIs, HHMI Investigators, and NDPA Finalists) ..... 25
      2. Publication Data (R01 Portfolios) ..... 26
      3. Publication Data (Text Mining) ..... 27
      4. Expert Review Data ..... 27
3. Methods ..... 29
   A. Bibliometric Methods ..... 29
      1. Background ..... 29
      2. Approach ..... 30
      3. Analyses Conducted ..... 32
   B. Expert Review ..... 37
      1. Expert Panel Experimental Design and Analysis ..... 37
      2. Analysis of Qualitative Data from the Expert Panel ..... 37
   C. Potential Limitations ..... 39
4. Impact and Innovation: Grant-Level Findings ..... 43
   A. Matched R01 Grants ..... 43
      1. Bibliometric Findings ..... 43
      2. Expert Review Findings ..... 51
   B. Random R01 Portfolios ..... 53
      1. Number of Publications ..... 53

2. Number of Publications by Portfolio Funding Amount .....53
3. Number of Citations .....54
4. Number of Citations per Portfolio Funding Amount .....55
5. Citations per Publication .....55
6. SCImago Journal Ranking.....58
7. Proportion of Publications by Journal Percentiles .....58
5. Impact and Innovation: Researcher-Level Findings.....61
A. Bibliometric Findings.....61
1. Number of Publications.....61
2. Number of Citations .....62
3. Number of Citations per Publication.....63
4. H-index.....63
5. SCImago Journal Ranking.....64
6. Proportion of Publications by Journal Percentiles .....65
7. Additional Bibliometric Analyses: Outcomes per Selected Direct Costs ...66
8. Additional Analyses: Text Mining .....69
B. Expert Review Findings .....70
C. Regression Analyses.....74
6. Impact and Innovation: Qualitative Findings from Expert Assessment.....77
A. Assessing Impact.....77
B. Assessing Innovation.....82
C. Summary of Qualitative Findings .....84
7. Interdisciplinarity and Collaborations Findings .....87
A. Interdisciplinarity .....87
1. Integration Score .....87
2. Additional Analyses: Proportion of Subject Area Categories.....88
B. Collaborations .....90
1. Number of Co-authors.....90
2. Number of Co-author Affiliations.....91
8. Summary of Findings .....93
A. Grant-Level Analyses of Impact and Innovativeness.....93
B. Researcher-Level Analyses of Impact and Innovativeness .....94
C. Qualitative Findings from Expert Assessment.....96
D. Regression Analysis .....96
E. Analyses on Interdisciplinarity and Collaborations .....97
F. Conclusion.....97
Appendix A. List of Illustrations ..... A-1
Appendix B. Potential Comparison Groups.....B-1
Appendix C. Researcher Profiles.....C-1
Appendix D. Expert Reviewer Characteristics ..... D-1
Appendix E. Expert Review Protocol.....E-1
Appendix F. Expert Review Data ..... F-1

Appendix G. Qualitative Analysis Codebook ..... G-1
Appendix H. Protocols for Downloading and Cleaning Publication Data for Bibliometric Analysis ..... H-1
Appendix I. Statistical Details ..... I-1
References ..... J-1
Abbreviations ..... K-1

# 1. Introduction

## A. Overview of the Program

The National Institutes of Health (NIH) Director's Pioneer Award (NDPA) was initiated in fiscal year (FY) 2004 "to support individual scientists of exceptional creativity who propose pioneering approaches to major contemporary challenges in biomedical research" (NIH 2004). It grew out of concerns that the traditional NIH peer review process had become overly conservative and the belief that the NIH required specific means to fund high-risk research (Brenner 1998, Mervis 2004).

The program provides individual researchers (called "Pioneers") with $500,000 in direct costs for each of the 5 years, for a total direct cost of $2.5 million. Compared to the average NIH Research Project Grant Program (R01) grant, the NDPA provides a greater amount of funding per year over a longer duration, and does not require a project budget specifying how the funds will be allocated.

The NDPA is currently administered through the NIH Office of the Director and is annually awarded to a small number of researchers across a range of scientific disciplines. Since 2004, the NDPA has been awarded to over a hundred Pioneers.

## B. Purpose of the Outcome Evaluation

FY 2013 will be the NDPA's tenth year and its last supported fully through the NIH Common Fund, which supports trans-NIH programs with participation of two or more NIH Institutes and Centers (ICs). After this year, the program will remain within the Office of the Director, but only the first year of the awards will be funded by the Common Fund; future years of the program are expected to be funded or co-funded by the individual ICs.[1]

To facilitate future IC participation, NIH Director Francis Collins requested that the IDA Science and Technology Policy Institute (STPI) compare NDPA to other grant programs in order to evaluate its effectiveness in producing high-impact research. Prior process and outcome evaluations of the NDPA program provide details on the program's inception, evolution, selection, outputs, and outcomes.[2]

---

[1] Email correspondence with an NIH program officer.

[2] The reports summarizing these analyses are available on the NIH Common Fund website, http://commonfund.nih.gov/pdf/Pioneer_Award_Outcome%20Evaluation_FY2004-2005.pdf and http://commonfund.nih.gov/pdf/PioneerAwardProcessEvaluation_2004-2008.pdf, accessed June 11, 2012.

## C. Study Questions

The NDPA solicitation provides the framework for the goals of this evaluation. The 2012 solicitation, which has not changed significantly since the program's inception, begins with the following statement:[3]

> Pioneer Awards are designed to support individual scientists of exceptional creativity who propose pioneering—and possibly transforming—approaches to major challenges in biomedical and behavioral research. The term "pioneering" is used to describe *highly innovative approaches* that have the potential to produce an *unusually high impact* on a broad area of biomedical or behavioral research…. (italics added)

This program goal established the evaluation's primary study questions, which are:

- To what extent does NDPA-funded research produce *unusually high impact*?
- To what extent are the *research approaches* used by the Pioneers highly innovative?

The evaluation also addresses two secondary questions that emerged from NIH leadership's interest in high-risk, high-reward research and are of special interest to the science and technology policy community. They also have strong policy implications. The secondary questions are:

- To what extent is the Pioneers' research interdisciplinary?
- To what extent are the Pioneers collaborative?

## D. Scope of the Evaluation

### 1. Conceptualizing the Approach

Science progresses through a combination of incremental and breakthrough research—both are essential for progress. In the current traditional research portfolio (e.g., R01), some percentage of the funded research outcomes are breakthrough—primarily through serendipity. If there is to be a focused effort on producing breakthrough research through a set-aside program (such as the NDPA), to justify the new program, either the percentage of breakthrough research must be greater for the set-aside program than the traditional program, or the amount of "breakthrough-ness" achieved per breakthrough must be greater.

---

[3] Department of Health and Human Services, Funding Opportunity Announcement RFA-RM-11-004, August 5, 2011, http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-11-004.html, accessed December 12, 2011.

The Pioneer program does several things to distinguish itself from traditional research programs. For example, it attracts highly meritorious researchers, whether recognized as such in the community or not; encourages them to propose high-risk ideas; gives them a larger sum of money and a longer time horizon; and gives them the flexibility to do what they need to do to pursue their high-risk ideas. The combination of exceptionally bright people pursuing high-risk ideas, in theory, increases either the probability of producing pioneering research outcomes—as distinct from incremental advances, or the degree of "pioneeringness" of the research outcomes.

Building on this rationale, STPI developed a conceptual approach to the evaluation. The evaluation examined how NPDA outcomes differentiate themselves from traditional research outcomes. There are two ways this differentiation could manifest itself. In one manifestation, the distribution of pioneeringness, however it is measured, shifts rightward, meaning that as a result of NDPA support, the overall pioneeringness of an entire group of researchers improves. This is represented in Figure 1 by the notional normal curve shifted below wholesale (dotted blue line). Alternatively, NDPA support could enable a larger number of instances of extreme pioneeringness. This could occur were NDPA support to increase the variance of the distribution by changing the shape of the distribution so that there was a longer tail in Figure 1 (dotted red line). In either case, NDPA could increase pioneeringness, although in the latter case, increased pioneeringness as measured by a longer tail comes at the cost of an increased percentage of supported awards that are unsuccessful.

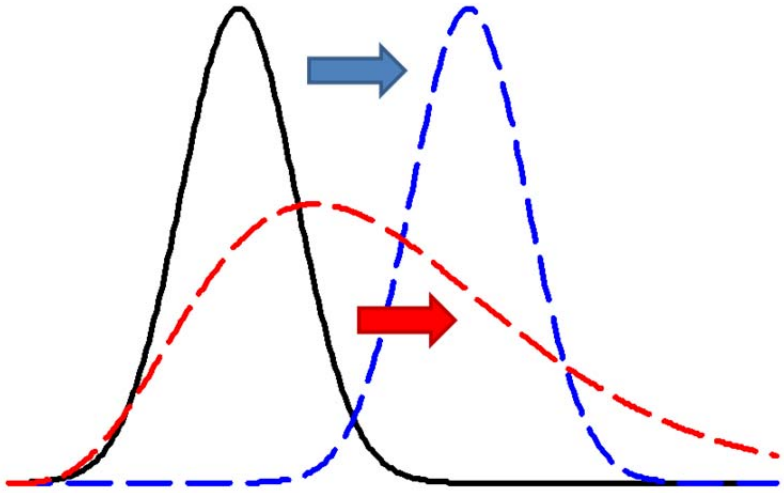


**Figure 1. Notional Manifestation of Two Types of Differences (Shifts) in Performance**

## 2. Operationalizing Terms of Interest

After conceptualizing the evaluation, STPI considered ways to operationalize the key metrics of interest. The first primary study question focused on the concept of unusually high impact. The term "impact" is not defined in the solicitations, and in general is a concept that is difficult to quantify or standardize. Unusually high impact is even more problematic to quantify. The solicitation attempted to clarify its expectations, and requested that, in their research strategy sections, applicants explicitly address the question: "What are the pioneering, and possibly high-risk, approaches that, if successful, might lead to groundbreaking or paradigm-shifting results?" However, terms such as "paradigm shifting" and "groundbreaking" are no easier to define and operationalize than "impact."

Godin and Dore (2004) propose a broad definition of scientific impact: the contributions to research through publications, including diffusion and appropriation of new knowledge, theories, methodologies, models, and facts; the formation and development of specialties and disciplines; the diversification of the type of research conducted (basic, applied, strategic), and the development of interdisciplinary, intersectoral, and international research. As a first step, in this evaluation, we took a narrow definition of scientific impact and examined impact only as expressed in scientific publications of researchers.

The second study question explores the concept of the innovativeness of research approaches. Again, while the term has many interpretations, it is difficult to standardize. Given these difficulties, STPI approached the evaluation from two directions.

First, we decided to follow a multi-method approach to include both objective and subjective assessments. The objective approach involved identifying bibliometric indicators of interest (Chapter 3 describes them in detail), and the subjective approach involved asking experts to provide a rating and discussion of impact and innovation with regard to the research in question. While the experts were provided high-level guidance on how to think about impact and innovation of research approaches, ultimately it was a subjective judgment on their part.

Second, we determined that any assessment of NDPAs or the Pioneers would be meaningful only in comparison. (Whether NDPAs yield "more" impact begs the question "compared to what.") This required a choice of comparison groups that were meaningful and had policy relevance. To fully isolate the effect of the Pioneer program would require randomizing principal investigators (PIs) to grant mechanisms; failing this, STPI chose comparison groups that overlap with NDPA in substantive ways.

### 3. Other Parameters to Bound the Evaluation

While the Pioneer program is ongoing, and eight cohorts of awards have been made since inception, this evaluation covers the 35 NDPAs from the first three cohorts (FYs 2004–2006). This was to ensure that the award period was complete and that (almost) all outputs of the awards had had a chance to be reviewed in the community.

As discussed previously, while there are many ways research creates impact, this evaluation focuses on impact as evident in scientific publications. Excluding materials like textbooks, websites, databases, and other outputs of research certainly limits the scope of the evaluation; however, it makes the data and findings more comparable.

Since researchers typically received multiple grants, to ensure appropriate comparison, analyses were performed both at the grant and the researcher levels. Grant-level analyses allow for the comparison of just the output of grants; researcher-level analyses permit inclusion of the publications of multiple grants and funding sources.

## E. Overview of the Report

Chapter 2 describes the methods we used to select and construct the comparison groups and summarizes our data collection approaches. Chapter 3 describes the analytic methods used both for the bibliometric analyses and for the expert review. Chapter 4 presents our findings on impact and innovation at the grant level. Chapter 5 presents these findings at the researcher level. Chapter 6 explains our qualitative findings from the expert assessment, and Chapter 7 provides findings related to interdisciplinarity and collaboration. Chapter 8 summarizes all our findings. Appendix A provides lists of figures and tables in the main report, and Appendix B through Appendix I provide supporting details.

# 2. Data

## A. Description of Comparison Groups

STPI used five different groups of awards/awardees for comparison with the NDPA and Pioneers. These groups, which are described in the following subsections, are: a matched set of 35 FY 2004–2006 NIH Research Project Grant Program (R01) grants, the 35 PIs on the same R01 grants, 30 FY 2004–2006 random R01 portfolios that received comparable amounts of funding to the NDPA portfolio, 39 FY 2005 Howard Hughes Medical Institute (HHMI) investigators, and 30 FY 2004–2006 NDPA finalists. Appendix B describes these groups and additional comparison groups that we considered before settling on these five groups. Appendix C provides profiles of the researchers that form the groups used for comparison purposes.

### 1. Matched Set of R01 Grants and PIs

The R01 is the original and historically oldest grant mechanism used by the NIH to support biomedical research and health-related research and development.[4] R01 applications may be initiated by the investigator or in response to a program announcement or request for applications.

Successful R01 applications are funded through one or more of the NIH's 27 Institutes and Centers, which have research agendas that often focus on specific diseases or body systems.[5] Grants are generally awarded for up to 5 years. In FY 2011, R01 PIs received an average of $440,000 for one year.[6]

From FYs 2004 to 2006, the NIH funded 12,007 new (Type 1) R01 applications. This is the population of R01 grants from which we derived our matched set of R01 grants. Since R01 grants are the most traditional research grant at the NIH, they provide a comparison group that helps clarify the effect of the mechanism on research impact and innovation. The scope of the R01 set was limited to Type 1 applications because NDPA projects support new research ideas.[7] We considered R01 PIs from FY 2004 through

---

[4] For more information on the R01 mechanism, visit: http://grants.nih.gov/grants/funding/r01.htm.

[5] For more information on the NIH's Institutes and Centers, visit: http://www.nih.gov/icd/.

[6] For more information on awards, visit NIH RePORT Funding Facts at http://report.nih.gov/fundingfacts/index.cfm.

[7] For more information on NIH application types, visit: http://grants.nih.gov/grants/glossary.htm#A27.

2006 to keep the bibliometric measures, which are time sensitive, as comparable as possible.

### a. Construction of the Matched R01 Set

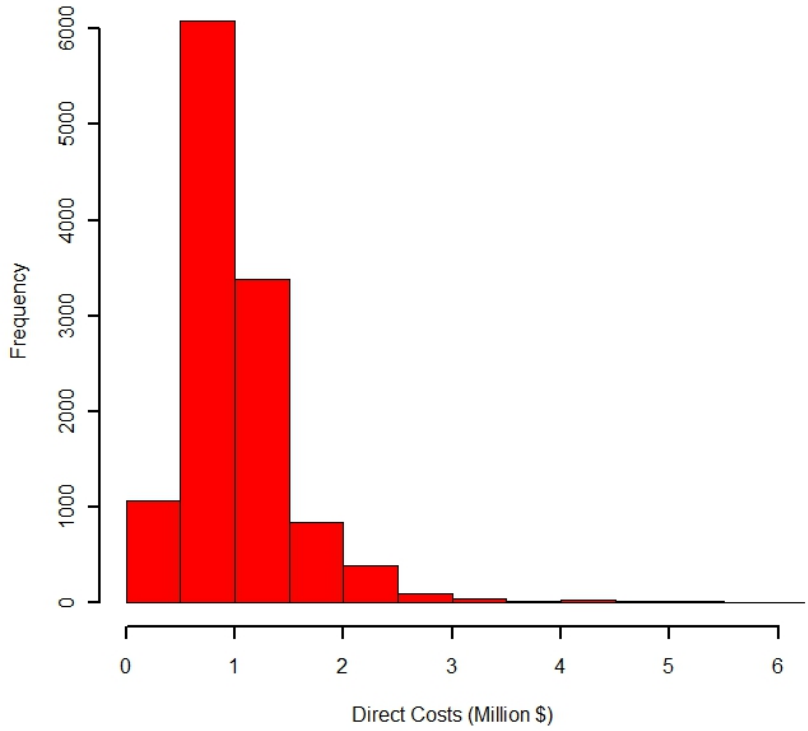
The matched R01 comparison groups were not selected randomly. This particular part of the evaluation used a quasi-experimental design. Quasi-experimental designs are typically developed when the evaluator must collect data after the program has been initiated or even completed. Because of these characteristics, evaluators compare the outcomes of the people who participated in the program with individuals that did not, and use statistical methods or research design methodologies to estimate the causal impact of the program. These methods are required because the causal impact can be difficult to disambiguate without them for several reasons, most notably non-random assignment to the program or treatment.

Non-random assignment happens because characteristics that predict which individuals self-select or are selected into a program or treatment are correlated with the outcome variable. An example would be nutritional classes for overweight people. If the participants self-select into a class (people who are motivated are more likely to sign up for these classes and therefore lose more weight), the measured treatment effect (for example, pounds lost), may be biased if the outcome is compared to people who did not participate, since the changes may be due to the motivation and not the actual treatment.

In this evaluation, we create a constructed control group that removes many of these differences between the control and comparison group. The construction of a control group can be done many different ways (Guo and Fraser 2010); this evaluation uses matching methods. Matching has been defined broadly as a method that aims to create a comparison group with equal or "balanced" distribution of specific covariates between the treated and the comparison group (Stuart 2010). Specifically, this evaluation used the rank-scale Mahalanobis distance approach, which uses a robust distance metric to account for nominal, ordinal, and continuous covariates (Rosenbaum 2005). This method was useful because the construction of the comparison group also included additional constraints, discussed below.

To construct a matched set of R01 grants, we started with the 12,007 Type 1 R01 grants awarded in FYs 2004–2006. One observed covariate is the total direct cost of the grant. For NDPA, this is $2.5 million (M) over 5 years. For the population of Type 1 R01 grants awarded in FYs 2004–2006, the distribution of direct costs is given in Figure 2. Grants of $2.5M are uncommon; only 2% of R01 grants we considered received $2.5M or more. (As a note, 84% of these were awarded for 4 or 5 years, with one awarded for 6 years and the remainder for 1 to 3 years.)

As a first step, we considered comparably sized R01 grants by restricting our set to the 183 R01 grants with total direct costs between $2.25M and $2.75M. We identified each of these as either basic or clinical, and found that approximately 85% of the R01 grants with these direct costs were clinical. NDPA, on the other hand, funds more basic research (Lal et al. 2010). We decided that a comparison of the outcomes of basic and clinical research would introduce an uncontrolled source of variation into the results and would be an inappropriate comparison group.




*Note:* Seventeen awards with total direct costs above $6M are not displayed here.

**Figure 2. Distribution of Direct Costs for Type 1 R01 Grants, FY 2004–2006 (*n* = 12,007)**

We next focused on identifying R01 grants that were topically similar to the NDPAs. In collaboration with Edmund Talley of the NIH, we used the NIH Topic Mapping Tool (Talley et al. 2011) to identify 821 R01 grants that were topically similar to the research being conducted by the Pioneers. This methodology uses Latent Dirichlet Allocation to categorize unstructured text. Some R01 grants were topically similar to more than one NDPA; overall there were 730 unique R01 grants identified. We use the term “research strata” to refer to the groups of R01s that are similar to the NDPAs.

We chose exactly one R01 from each research stratum under the constraint that the 35 R01s chosen must (1) contain the same distribution as for the NDPAs (9 awarded in

FY 2004, 13, in FY 2005, and 13, in FY 2006) and (2) be as close as possible to the NDPAs on a set of other characteristics. Choosing one R01 from each research stratum requires *fine balance*. "Closeness" is measured using the rank-scale Mahalanobis distance described by Rosenbaum (2005). The variables we used are year since degree, institutional prestige, prior NIH funding, terminal degree(s) received, receipt of early career awards, and receipt of R01 award within 5 years of most recent research or clinical doctorate. These variables are described in more detail in the next section.

A comparison of the funding received by the matched R01 awards and an NDPA award is shown in Figure 3. The specific R01 grants selected for the comparison group are listed in Table 1.

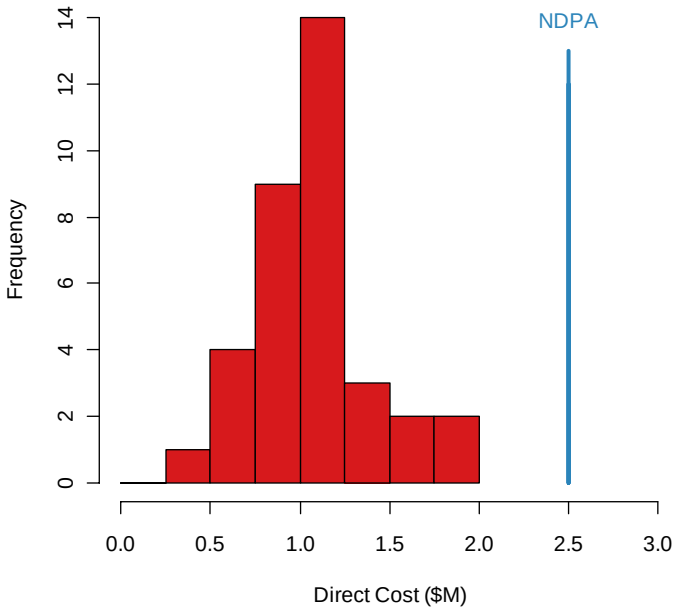



**Figure 3. Matched R01 Funding Compared to NDPA Funding**

This page has been redacted due to the personally identifiable information contained in this table.

### a. Variables Used

A quasi-experimental design requires the construction of a comparison group that looks like the group of interest in ways that predict treatment and affect outcome. This section describes the covariates that we used to develop a sample of R01s. For both populations, we collected data from Query/View/Response (QVR), the NIH's internal interface for accessing information on funded and unfunded grant applications (NIH 2007). Additional data were collected from collaborations with an NIH topic modeling group, the National Science Foundation (NSF) website and award search database,[8] press releases, and program websites.

#### 1) Subject Area Distribution

Edmund Talley's topic modeling group at the NIH generated a graph-based layout algorithm of two-dimensional visualized output for all funded NIH grants whereby documents are clustered based on their topic- and word-based similarity over a visual map (Talley et al. 2011). Figure 4 is a visualization of NDPAs in the overall NIH grant landscape. We reasoned that subject area would have an effect on the bibliometric research outcomes because Cameron (2005) suggests that different research fields have different publishing patterns and citing conventions.

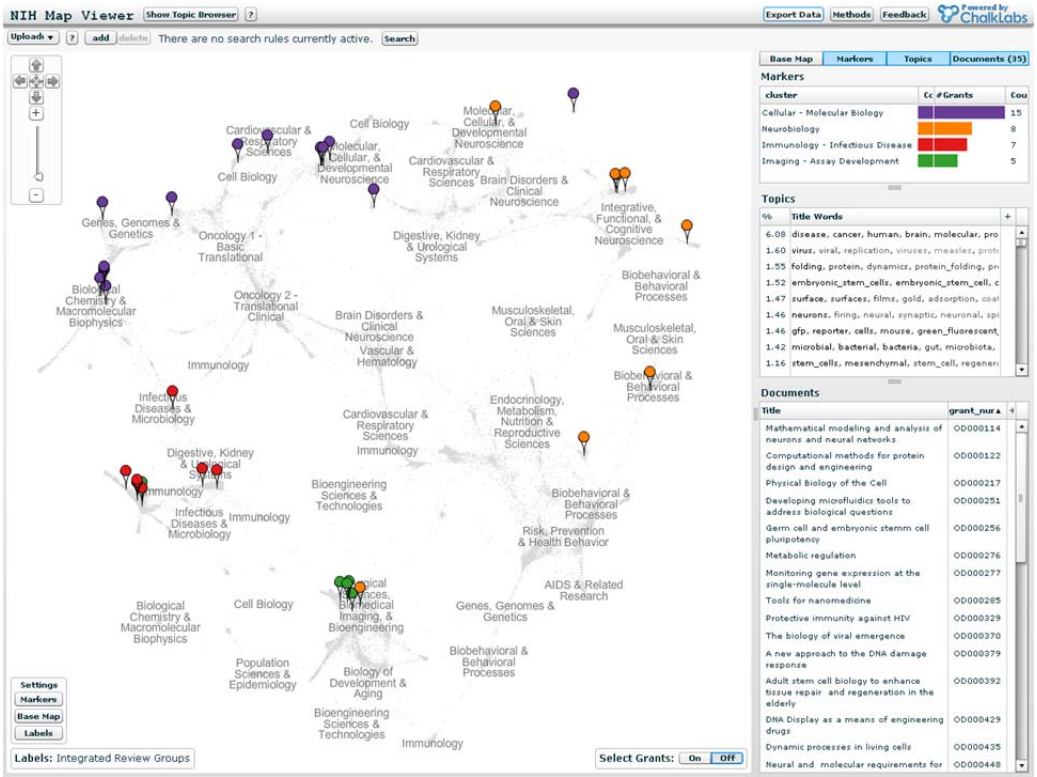



**Figure 4. Map of NDPA Research Areas within the NIH Landscape**

[8] NSF's award search database can be accessed at http://www.nsf.gov/awardsearch/.

### 2) Principal Investigator (PI) Characteristics

Using QVR, we gathered information about the personal characteristics of the PIs in the NDPA and R01 populations. We obtained data on institution at the time of award, year of doctoral degree conferral, and doctoral degree type(s). Using the SCImago Institutions Ranking (SIR), we gathered information on the rank of each PI's institution at time of award.[9] We also determined whether individuals in the NDPA and R01 populations received early career awards in the years prior to and including their year of grant receipt.

Using public sources such as NSF's award search page[10] and NSF and NIH program-related press releases,[11] we collected information on the following prestigious early career award programs that target their support towards early tenure-track faculty members and award scientists across different fields of research:

- NSF Faculty Early Career Development (CAREER)
- NSF Presidential Young Investigator (PYI)
- NSF Young Investigator (NYI)
- NSF Presidential Faculty Fellowship (PFF)
- NIH Presidential Early Career Awards for Scientists and Engineers (PECASE)[12]

Previous NIH funding data are available from QVR, the NIH's web-based tool on grant applications and awards. In a compiled database of all NIH funding prior to 2007, automated steps and human judgment were used to clean PI names, and CVs were compared to grant information to verify grant attribution. The prior NIH funding variable is a sum of the direct and indirect costs awarded through the R01 mechanism prior to and including FY 2004. Funding amounts were reported in current dollars each year, and there was no attempt to normalize the data for real dollar amounts. In addition to total R01 funding, another covariate available through QVR is the year of first R01 receipt. This variable denotes the year in which the PI of interest was awarded the first R01 as the lead PI.

Raw degree data for PIs were also obtained through QVR. We categorized the degrees obtained into "research" and "clinical" doctorates. As a starting point for identifying which doctoral degrees may be considered research doctorates, the NSF's

---

[9] For more information on SIR, visit: http://www.scimagoir.com/.

[10] NSF's award search page can be accessed here: http://www.nsf.gov/awardsearch/.

[11] Further information on the National Science Foundation early career awards chosen for these analyses can be found in National Science Foundation (2001), available at http://www.nsf.gov/pubs/2001/nsf01118/nsf01118.pdf.

[12] Further information on the NIH PECASE program is available at: http://grants.nih.gov/grants/policy/pecase.htm.

Survey of Earned Doctorates was consulted.[13] Additional judgment was used to categorize clinical degrees and doctorates not explicitly mentioned in the Survey of Earned Doctorates. Table 2 identifies which degrees were considered research and which were considered clinical. Number of years since degree was calculated from the most recent research or clinical doctorate.

**Table 2. Research and Clinical Degrees**

| Research Degrees | Clinical Degrees |
|---|---|
| doctor of philosophy | doctor of dental surgery |
| doctor of public health | doctor of dental medicine |
| doctor of education | other doctorate of medical dentistry, doctor of nursing science |
| doctor of social work | doctor of nursing practice |
| doctor of science | doctor of osteopathy |
| | doctor of podiatric medicine |
| | doctor of veterinary medicine |
| | other doctor of veterinary medicine |
| | bachelor of medicine |
| | bachelor of medicine and bachelor of surgery |
| | doctor of medicine |
| | doctor of optometry |
| | doctor of pharmacy |
| | doctor of psychology |

*Note:* Only degrees found among the Pioneer and R01 PI populations are listed in the table.

One of the NDPA review criteria in FYs 2004–2006 was "evidence of prior willingness to take risks or inclination towards creative and innovative research." Receipt of an early career award is an observable variable that has been linked to creativity and innovation (Azoulay et al. 2011).

After data on early career awardees were obtained, individuals were matched by their full names. Year of early career award receipt was also noted. Once the early career awardees were matched to the correct PIs in our compiled dataset, we derived the binary variable "Receipt of Early Career Award." This variable indicates whether the researcher received at least one early career award in a year prior to or including the year of grant receipt.

We chose to use the SIR to assess the prestige of the institution at which the PI was employed at the time of grant award (SCImago Research Group 2011a). The SCImago

[13] Data from NSF's Survey of Earned Doctorates is summarized in NSF (2010), http://www.nsf.gov/statistics/nsf11306/. Research degrees are identified in appendix table A-2: Research degrees included in the Survey of Earned Doctorates: 2005-09.

Research Group uses four indicators based on scientific output and citations to evaluate institutional research performance. The data for these rankings comes from Scopus. The indicators are output (number of scientific papers, pro-rated for co-authorship), percentage of international collaborations, normalized impact[14] ("the ratio between the average scientific impact of an institution and the world average impact of publications in the same time frame, document type, and subject area"), and the ratio of publications in first quartile journals.

We reasoned that these PI-level characteristics would have an effect both on the likelihood of being selected for an NDPA and the research outcomes of the award. Year of degree conferral is a proxy for research experience, SIR is a proxy for the research infrastructure supporting the PI, and early career award receipt is a proxy for the researcher's potential for conducting high-impact research.

We used the American rankings for institutions; institutions outside the United States were assigned the ranking of the American institution closest to but below it in the World Rankings.

#### 3) Research Grant Characteristics

Using QVR, we gathered information on the R01 the fiscal year of the award, grant's award length, and the sum of the direct costs of funded Type 1, Type 3, and Type 5 awards attributed to that grant number.

Since bibliometric measures of impact are time-sensitive, we decided that the fiscal year of the award was an important consideration in creating the matched R01 set. Two of the unique features of the NDPA mechanism are the 5-year award length and the $2.5M direct costs, so we reasoned that matching R01s on award length and direct costs would help us better understand the effects of the NDPA and R01 mechanisms as opposed to the effect of disparities in research funding support.

### 2. R01 Portfolios

Another comparison group of R01 grants was constructed to enable a comparison by direct cost. Using the same 12,007 base set of R01-funded grant applications we used in developing the matched R01 group, we selected 30 random samples of Type 1 R01 grants based on total direct cost. No matching was done for PI characteristics or for area of science. The direct costs of the random samples were stratified by year such that the total direct cost for grants in a given start year from FY 2004 through FY 2006) mirrored that

[14] Normalized impact is computed using a methodology established by the Karolinska Intitutet in Sweden, where it is named "item oriented field normalized citation score average."

of NDPA.[15] While it was not possible to get an exact match on direct costs, matches were made as close as possible. Figure 5 shows the distribution of total direct costs for the 30 R01 portfolios compared with that for NDPA.

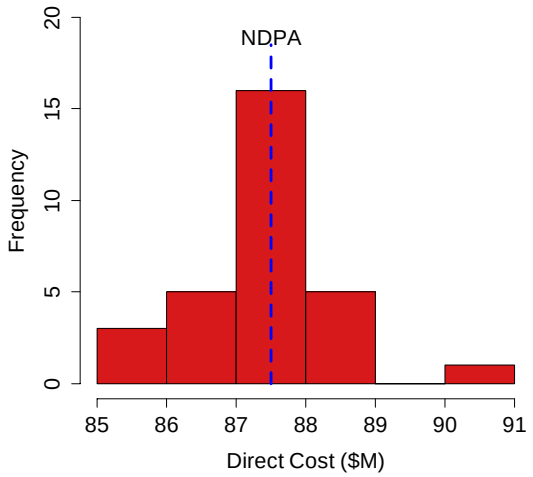



**Figure 5. Distribution of Total Direct Cost for the 30 R01 Portfolios**

A difference in funding levels between R01s and NDPAs (on a per grant basis) resulted in a larger number of grants per R01 portfolio when compared to the NDPA portfolio. Figure 6 compares the distribution of R01 grants per portfolio with that of NDPA.

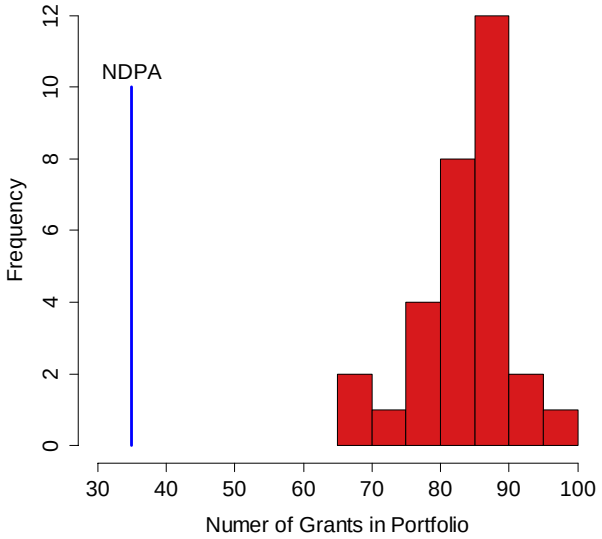



**Figure 6. Distribution of Number of R01 Grants per Portfolio Compared with NDPAs**

[15] NDPA issued 9 grants in FY 2004 (direct cost ~$22.5M), 13 grants in FY 2005 (direct cost ~$32.5M), and 13 grants in FY 2006 (direct cost ~$32.5M).

### 3. HHMI Investigators

HHMI investigators were included as a comparison group because, like the Pioneers, they are also high-performing researchers funded at significantly higher levels than the traditional R01 awards. The Howard Hughes Medical Institute is one of the largest nonprofit biomedical research institutes in the world. There are approximately 330 current HHMI investigators. This is a prestigious group of researchers, and includes 13 Nobel laureates and 147 members of the National Academy of Sciences The HHMI website describes the program purpose as follows:[16]

> By appointing scientists as Hughes investigators, rather than awarding them grants for specific research projects, the investigators are provided with long-term flexible funding that gives them the freedom to explore and, if necessary, to change direction in their research. Moreover, they have support to follow their ideas through to fruition—even if that process takes a very long time.

In support of this purpose, HHMI has had a variety of selection processes over the years. In 1995, the first formal competition took place when HHMI solicited the presidents and deans of the top 200 NIH-funded universities for researcher nominations.[17] Our comparison group of 39 FY 2005 HHMI investigators was selected using this process. The May 2004 solicitation asked for "candidates from the full range of biological and biomedical inquiry who demonstrate exceptional promise early in their careers as independent researchers"[18] It also required that the investigators be 4 to 10 years from their first faculty appointment.

Starting in 2009, HHMI changed the selection process from a nomination-based process to an open competition. The frequency of the competitions varies and is at the discretion of the HHMI staff.

Additional characteristics of HHMI investigator support underscore the "person focus" as opposed to a traditional "project focus" of the award. When HHMI investigators are selected, they and some of their staff become employers of HHMI while the university where they work becomes the host institution. Total direct costs are approximately $650,000 per year.[19] Funding amounts differ between investigators because investigators can apply for equipment funds, which are awarded as needed. They are also required to spend at least 75% of their time conducting biomedical research.

---

[16] About HHMI Investigators: http://www.hhmi.org/research/investigators/.

[17] HHMI Institute News, "HHMI Taps 43 of the Nation's Most Promising Scientists," http://www.hhmi.org/news/032105.html, March 21, 2005.

[18] HHMI Research News, "Howard Hughes Medical Institute Seeks Up to 50 New Scientists," May 13, 2004, http://www.hhmi.org/news/pdf/051304.pdf.

[19] Figure comes from discussions with HHMI staff and program leadership.

Finally, in the belief that truly transformative research is accomplished over a long period of time, reappointment at the end of the 5-year award period is possible. A review panel evaluates the past achievements of the investigator; reappointment rate is approximately 80%, and investigators usually stay with HHMI for an average of 15 years.[20]

### 4. NDPA Finalists

In each year of the NDPA, the NDPA Oversight Committee co-chairs selected between 22 and 25 of the most promising applicants to present their ideas and interview with a panel of external experts at the NIH campus. These application finalists were selected using a combination of criteria, which included average overall application scores, "top four" scoring designations, other potential funding sources (e.g., interviews were not given to candidates who received HHMI fellowships while in consideration for the NDPA), and factors related to demographic and scientific diversity.[21]

Each finalist gave a 15-minute presentation and answered questions from the external review panel. Panelists then placed the finalists into three tiers. Interviewees in the top tier were recommended by the panel for funding, those in the middle tier were suggested for funding if money was available, and those in the bottom tier were not recommended for funding.

There is little documentation of the final phase of the NDPA selection process, though it is known that the NIH Director and the Advisory Committee to the Director, with input from NDPA program leadership, made final decisions on the award winners in all years. The likelihood of receiving an award was not based solely on the tiered decisions made by the interview panelists; program leadership also considered more subjective factors (i.e., existing funding, other potential funding sources, likelihood that the research could be funded with R01s, etc.) in the final phase of the selection process.[22]

---

[20] Conversation with Philip Perlman, VP of Research, HHMI, December 19, 2011.

[21] Finalists cannot be predicted solely from average overall scores and top four designations. In each year, individuals who had higher scores and more top four designations than some of the interviewees (and awardees) were not asked to interview. Conversely, some individuals asked to interview had lower scores and only one top four designation. Talks with NDPA program leadership revealed that other funding sources, demographic diversity, and scientific diversity were additional factors in the finalist selection process.

[22] For further information on the NDPA selection process, see Lal et al. (2010), https://commonfund.nih.gov/pdf/PioneerAwardProcessEvaluation_2004-2008.pdf.

Our NDPA finalist comparison group of 30 NDPA applicants who did not receive NDPA funding in FYs 2004–2006[23] had several limitations. First, we do not know if the finalists' NDPA-proposed ideas were subsequently pursued and funded. We also do not have complete information regarding the amount of research funding finalist received subsequent to their application to the NDPA. Nevertheless, we made the assumption that finalists continue to be productive biomedical researchers after their unsuccessful NDPA application. Due to the quality of the researchers and their potential for pioneering research, NDPA finalists were useful in comparisons to Pioneers at the researcher level.

## B. Comparison Groups at Time of Award

To assess the appropriateness of the comparison groups we selected, we examined several other variables that might influence post-award outcomes at baseline, that is, at time of award. As Figures 7 through 12 indicate, the Pioneers look similar to the matched R01 PIs, HHMI investigators, and NDPA finalists on most of these accounts. Although the covariates shown in Figures 7 through 12 were not used to select any of the comparison groups, overall, the groups appear to be quite comparable at time of award.

Figure 7 shows box plots for years since degree for the comparison groups. Results of a Kolmogorov-Smirnov test (K-S) indicates the HHMI group is statistically younger than the NDPA Pioneer group (K-S, $p = 0.0007$), which is not surprising since the HHMI investigators are chosen to be 4 to 10 years from first faculty appointment. The PIs from the grants selected in the random R01 portfolios ($n = 2297$) are also statistically different from the NDPA Pioneers (K-S, $p = 0.004$).

Figure 8 shows box plots for the SIRs for researchers' home institutions at time of award. The rankings for the NDPA Pioneers were similar to those for the matched R01 PIs (K-S, $p = 0.97$), HHMI investigators (K-S, $p = 0.91$), and NDPA finalists (K-S, $p = 0.99$). The rankings for the institutions of the random R01 portfolios are statistically different from those of the NDPA Pioneers (K-S, $p < 0.0001$).

Figure 9 shows box plots for the number of publications per researcher at time of award. NDPA Pioneers had similar citation distributions as the matched R01 PIs (K-S, $p = 0.63$), HHMI investigators (K-S, $p = 0.44$), and NDPA finalists (K-S, $p = 0.62$). Figure 10 shows box plots for the number of citations per researcher at time of award. NDPA Pioneers had similar citation distributions as the matched R01 PIs (K-S, $p = 0.63$), HHMI investigators (K-S, $p = 0.44$), and NDPA finalists (K-S, $p = 0.62$). Figure 11

[23] Two of the 32 NDPA finalists from the FY 2004–2006 application processes were removed to compose our comparison group of 30 finalists. One 2004 finalist was excluded because she received an NDPA in 2005 and is evaluated as an NDPA awardee. A second investigator (Steve Benner) was a finalist in both 2004 and 2006. We considered him only once as a 2004 finalist because the information we had for his application was more complete for that year.

shows box plots of the researcher h-index at time of award. The NDPA Pioneers had similar h-index values as the matched R01 PIs (K-S, $p = 0.80$), HHMIs (K-S, $p = 0.62$), and NDPA finalists (K-S, $p = 0.19$).

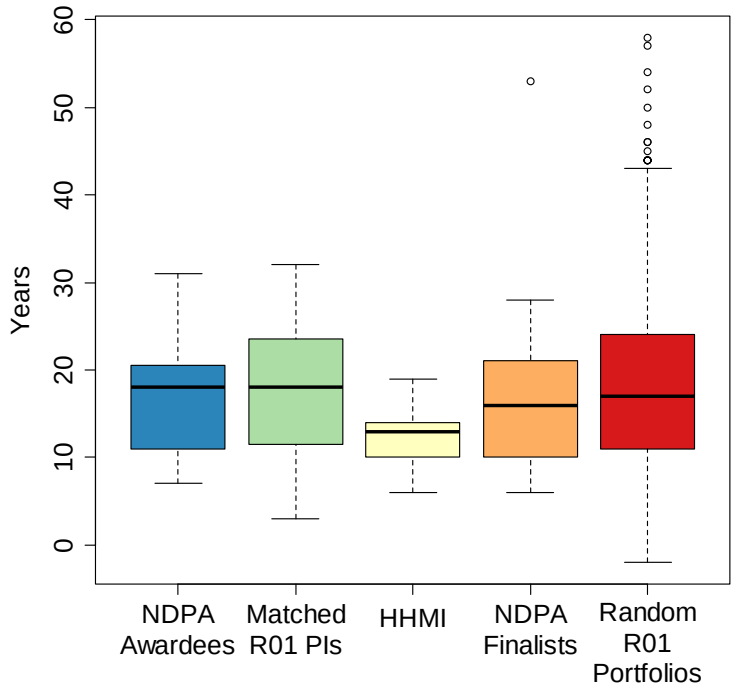



**Figure 7. Years since Degree at Time of Award**

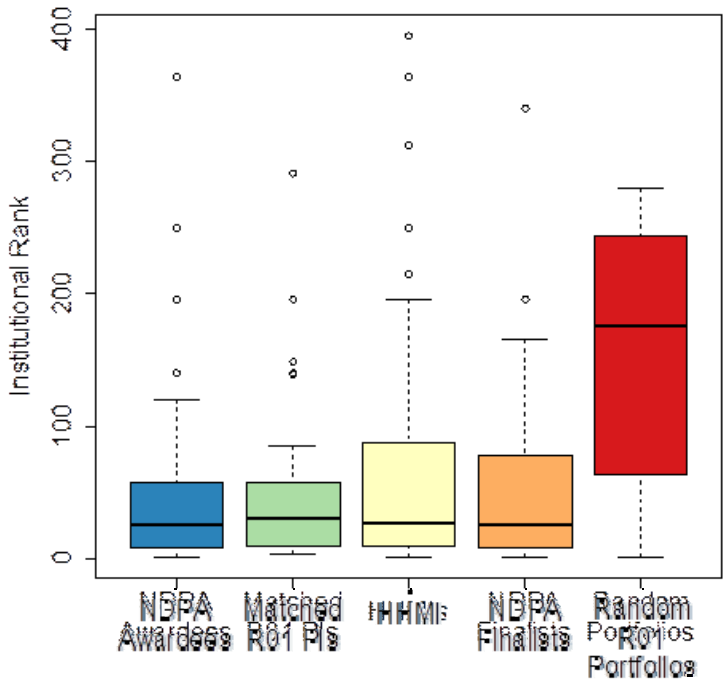



**Figure 8. SCImago Institutions Rankings at Time of Award**

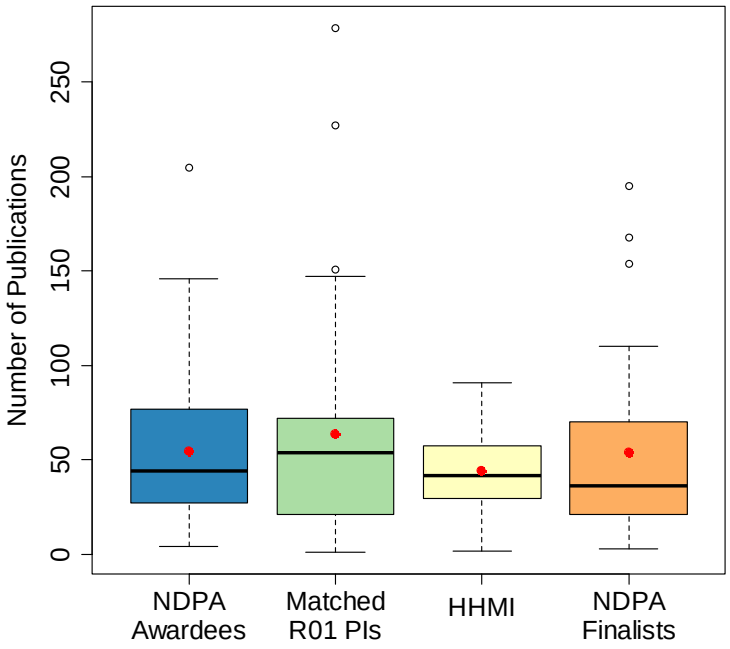
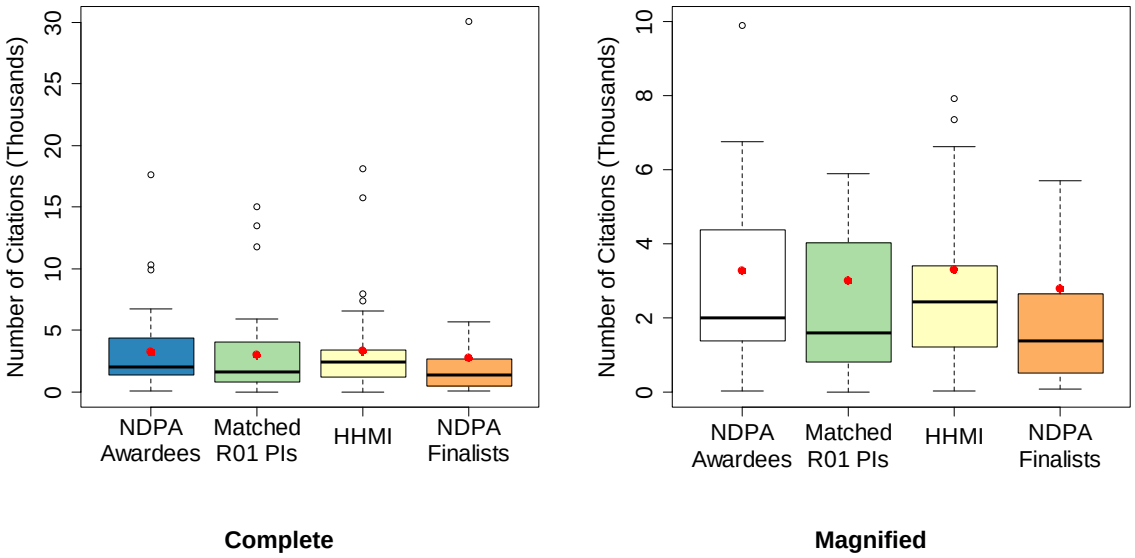



**Figure 9. Number of Publications at Time of Award**




**Complete** **Magnified**

**Figure 10. Number of Citations at Time of Award (Complete and Magnified)**

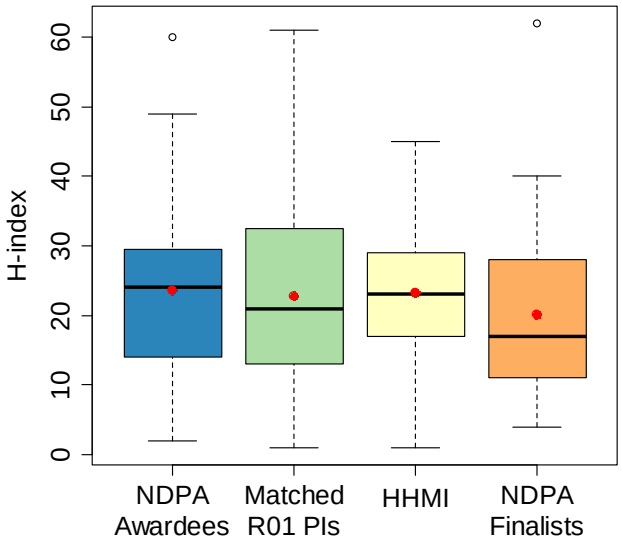



**Figure 11. H-index at Time of Award**

Figure 12 shows box plots of the integration score at time of award. Integration score is a measure of interdisciplinarity and is discussed in detail in Chapter 3. NDPA Pioneers are similar to the matched R01 PIs (K-S, $p = 0.64$) and the NDPA finalists (K-S, $p = 0.23$), but are more interdisciplinary than HHMI investigators (K-S, $p = 0.03$).

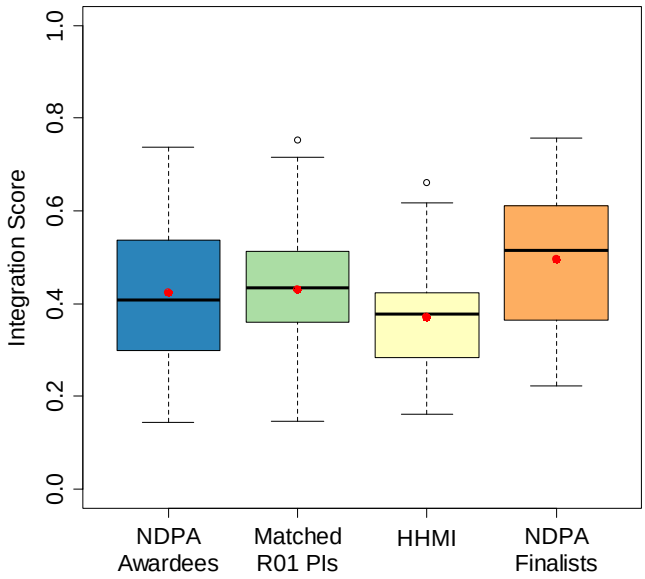



**Figure 12. Integration Score at Time of Award**

Figure 13 shows box plots of the number of unique co-authors working with each researcher prior to time of award. NDPA Pioneers are similar to the matched R01 PIs, HHMI investigators, and NDPA finalists (K-S, $p = 0.63, 0.82, 0.21$).

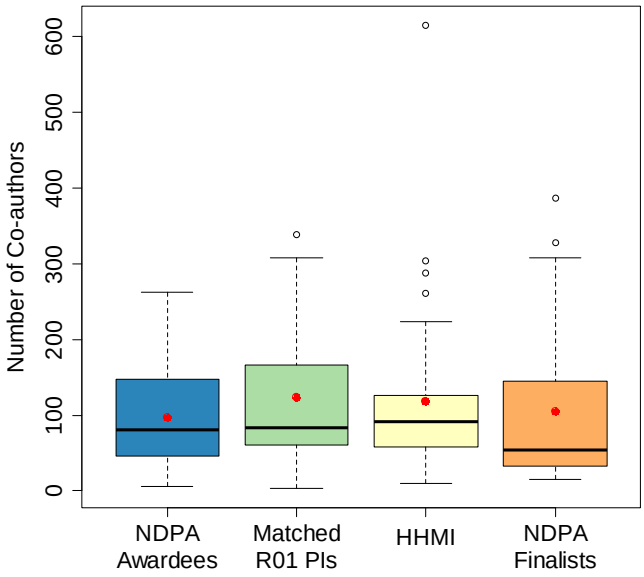



**Figure 13. Number of Unique Co-authors at Time of Award**

Figure 14 shows box plots of the number of unique institutions associated with co-authors working with each researcher prior to time of award. NDPA Pioneers are similar to the matched R01 PIs, HHMI investigators, and NDPA finalists (K-S, $p$ = 0.78, 0.11, 0.72).

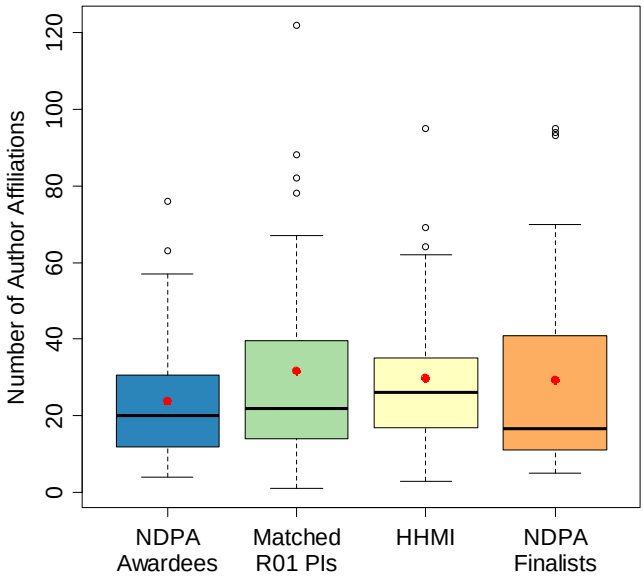



**Figure 14. Number of Unique Co-author Affiliations at Time of Award**

Table 3 summarizes all comparison groups, the study questions they specifically address, and their advantages and limitations. Table 4 summarizes the differences across the groups.

**Table 3. Summary of the Five Comparison Groups**

| ***Comparison Group*** | ***Description*** | **Study *Question*** | ***Advantages*** | ***Limitation*** |
|---|---|---|---|---|
| *Matched R01 grants*<br>*N = 35* | *R01 grants matched on PI characteristics within similar research areas* | *To what extent do the Pioneer award outcomes have more (or less) impact compared with traditional NIH grants given to similarly qualified researchers?* | *Controls for PI-related characteristics that may impact outcome* | *Does not control for award size* |
| *Matched R01 grantees*<br>*N = 35* | *Grantees of the matched R01 grants* | *To what extent do the Pioneer outcomes have more (or less) impact compared with similarly qualified researchers from traditional grant programs?* | *Controls for PI-related characteristics that may impact outcome* | *Does not control for award size* |
| *Random R01 Portfolios*<br>*N_av = 85* | *R01s with a portfolio direct cost comparable to that of the NDPA portfolio* | *To what extent* do *Pioneer award outcomes* have *more (or less) impact compared with portfolios of randomly selected R01* grants *the same size as the NDPA portfolio?* | *Controls for the portfolio award sizes that may impact outcome* | *Portfolios contain* different numbers of grants; do*es not control for PI characteristics* |
| *HHMI*<br>*N = 39* | *2005 Howard Hughes Medical Institute Investigators* | *To what extent* do *Pioneer award outcomes* have *more (or less) impact compared with a similarly high-prestige research program?* | *High-prestige program that funds high-risk high-reward research in a way that is similar to NDPA in many aspects; reputation for innovative investigators* | *Does not explicitly control for PI characteristics or award size* |
| *NDPA Finalists*<br>*N = 30* | *Individuals who were invited to interview, but were not awarded an NDPA from 2004–2006* | *To what extent are Pioneer award outcomes more (or less) impactful as compared with researchers who were almost as qualified as the Pioneers but did not get the Pioneer award?* | *Examines the outcomes of PIs who are "almost as exceptionally creative," and capable of producing high impact outcomes* | *Variable post-application funding amounts* |

**Table 4. Summary of the Differences across the Comparison Groups**

| *Comparison Group* | *Funding Level* | *Subject Area* | *PI Characteristics* | *High-Risk Research/ Flexibility* |
|---|---|---|---|---|
| *Matched R01 grant* | *Lower* | *Matched* | *Similar* | *No* |
| *Matched R01 grantee* | *Unknown* | *Matched* | *Similar* | *No* |
| *"Random" R01 Portfolios* | *Equal* | *Not matched* | *Different* | *No* |
| *HHMI* | *Higher* | *Not matched* | *Different* | *Yes* |
| *NDPA Finalists* | *Unknown* | *Not matched* | *Similar* | *N/A* |

## C. Collection of Data

Bibliometric data were collected in order to analyze the publication-based research outcomes of the matched R01 set, the FY 2005 HHMI investigators, the FY 2004–2006 NDPA finalists, and the FY 2004–2006 R01 portfolio set in comparison to the NDPAs and Pioneers. The primary sources of bibliometric data were the Thomson Reuters Web of Science (WOS) database and the internal NIH electronic Scientific Portfolio Assistant (eSPA) database.[24] Additional information was gathered from the SCImago Journal Rank (SJR), a measure of journal impact (SCImago Research Group 2001b). Appendix H provides the protocols for downloading and cleaning of the bibliometric data. Expert data were collected from topic-specific experts through a review of the five publications with the most impact ("top five") by each NDPA, matched R01 grants, and HHMI investigator.

### 1. Publication Data (Matched R01 PIs, HHMI Investigators, and NDPA Finalists)

We collected bibliometric data at three levels: researcher, grant, and top five publications. Researcher-level publications are defined as research articles where the PI of interest appears as an author.[25] Researcher-level data were collected from 1980 through the end of 2011 for the Pioneers, matched R01 PIs, HHMI investigators, and NDPA finalists in our comparison groups.[26] Grant-level publications are research articles that can be attributed to the grant number of interest through funding acknowledgments

[24] For more information on WOS, visit the Thomson Reuters product page at: http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/. For more information on eSPA, visit: http://thomsonreuters.com/content/press_room/science/468364.

[25] Research articles are publications produced from original research output. They exclude reviews, letters, notes, proceedings papers, etc.

[26] The WOS publication database extends only as far back as 1980.

in WOS, PubMed, or eSPA. Grant-level data were collected for NDPAs, matched R01 grants, and R01 portfolios. Top-five publications are the set of publications used in the expert review analysis. Top five data were collected for NDPAs, matched R01 grants, and HHMI investigators. The process for identifying the top five publications is described in greater detail in Section 2.C.4 where the expert review is discussed. Review papers were excluded from grant and researcher level analyses.

Top five publications were included in the grant-level set if they were research articles and published prior to 2012. Grant-level publications were included in the researcher-level set if the PI was an author on the publication. For each publication record, we collected the following information: publication title, authors, publication year, journal, citations accumulated each year from 1980 to 2011, document type, WOS category, and SCImago Journal Ranking.[27]

We used VantagePoint software to clean the WOS publication data[28] for use in analyses of the matched R01 grants and PIs, HHMI investigators, and NDPA finalists. The software was used to identify and remove publication records that were incorrectly attributed to the PIs and grants of interest, disambiguate author names and institutional affiliations, and match cited journals to WOS categories in an effort to understand interdisciplinarity.

Our data set contains 12,232 total publications of which 510 are top five, 813 are grant level, and 12,172 are researcher level.

### 2. Publication Data (R01 Portfolios)

Publication and citation data for each portfolio of grants (random R01 portfolios, NDPA, and the matched R01 groups) were collected via eSPA. Grant-attributed publications in eSPA were found via the SPIRES+ algorithm[29] and constrained to research publication using the "Research" flag in eSPA data export. Publication data coverage in eSPA is truncated at the end of 2010. For each publication record, we collected the following information: publication title, authors, journal, attributed NIH

---

[27] WOS categories compose a framework for understanding the topical coverage of a journal that is indexed by WOS. Several subject categories may be assigned to one journal. For a list of the 249 WOS categories, visit: http://images.webofknowledge.com/WOKRS56B5/help/WOS/hp_subject_category_terms_tasca.html

[28] For more information about VantagePoint, a text mining software for cleaning and analyzing search results from patent and literature databases, visit: http://www.thevantagepoint.com/products/vantagepoint.html.

[29] A combination of the NIH's Scientific Publication Information Retrieval and Evaluation System (SPIRES) and a Thompson Reuters-developed algorithm for matching MEDLINE articles to NIH projects. The algorithm is detailed here: https://espa.niaid.nih.gov/eSPA/Help/Documents/eSPAWG_PublicationMatching_20091210a.pptx.

grant numbers, and total citations.[30] Data for the R01 portfolio analyses did not undergo further cleaning after download. Publication data collection via eSPA yielded 266 grant attributed research publications for both the NDPA and matched R01 portfolios, of which 206 and 213, respectively, had citation count values in eSPA. Each of the 30 random R01 portfolios have more grant attributed research publications than the NDPA and matched R01 portfolios. Combined, the random R01 portfolios had a total of 14,352 unique research publications, of which 11,262 had citation count values in eSPA.

### 3. Publication Data (Text Mining)

Publication titles and abstracts were extracted from grant-attributed publications in the NDPA, matched R01 grants, random R01 portfolios, and HHMI investigator groups. The data sets were limited to those publications that have PubMed IDs and were in the SPIRES database between 2007 and the first March 2012. Publication data from the random R01 portfolios were further constrained by eSPA to publication dates of 2007 to 2010.

### 4. Expert Review Data

The expert review was conducted using the top five publications with the most impact. For the HHMI group, these publications were selected by each of the 39 HHMI investigators as part of their 2011 review. For NDPAs, we requested that the Pioneers select these publications, and 26 of them provided selections. For the matched R01 grants, we asked the program officers to make the selections, and 12 of them provided publications. The grant mechanism of the publications was blinded before the expert review, although the authors and journal titles were not.

For NDPAs and matched R01 grants for which the PIs and program officers provided fewer than five publications, we developed an algorithm to select publications from their remaining bodies of work. The algorithm chose two titles with the highest journal impact factors, two titles that were most highly cited, and one title with the highest citation rate, which was calculated by dividing total number of citations and number of years since publication. We compared this algorithm to other algorithms with different combinations of bibliometric impact by using the PI- and program-officer-chosen publications as a training set, but minimal differences were observed. We also compared the algorithms' retrieval to the PI- and program officer-selected data and sub-selected among algorithms that produced closest results. In addition, we preferred this algorithm on a theoretical basis; we reasoned that journal impact factor was more appropriate than total number of citations as a bibliometric proxy for impact in short

[30] According to eSPA help desk, eSPA was last updated in April 2011. It is unclear if the citation count are truncated at Dec 2010, or if they are from April 2011.

timeframes after publication. Where grants had fewer than five grant-attributed publications, we used as many as were available in the expert review. Overall, there were 518 papers, of which 23 were review papers and 4 were review book chapters. Of these 27 reviews, 11 were authored by Pioneers and 16 by the matched R01 PIs.

STPI designed the expert review protocol to collect expert ratings for both the impact and innovation of a researcher's publications. In addition, we included three questions related to the expert reviewer's personal and educational background. In total, we collected the following data:

- The impact of a researcher's individual papers on a five-point scale from extremely to not at all
- The impact of a researcher's set of papers, taken as a whole, on a five-point scale from extremely to not at all
- The innovativeness of the approaches of a researcher's individual papers on a five-point scale from extremely to not at all,
- The innovativeness of a researcher's set of papers, taken as a whole, on a five-point scale from extremely to not at all,
- The alignment between the research in each set of papers and the expert reviewer's own research area
- Year of birth
- Gender
- Educational degrees awarded

Refer to Appendix D for information on the characteristics of the reviewers, Appendix E for the protocol design, and Appendix F for data summaries of the expert reviewer's ratings of impact and innovation.

# 3. Methods

We used bibliometric techniques and expert review to measure the scientific impact and innovativeness of the approaches used by researchers, research groups, and grants.

## A. Bibliometric Methods

The application of bibliometrics in a comparative evaluation can provide a partial indication of impact (Cozzens 1996; Moed 2005; Martin and Irvine 1983).[31] In addition to communicating research findings, publications provide information on the author and co-authors, affiliations, journal, publication year, citations, cited references, and the journal's scientific area.

As Table 5 indicates, we conducted bibliometric analyses[32] on the comparison groups at the grant-level for NDPAs, matched R01 grants, and random R01 portfolios, and at the researcher-level for all comparison groups except the random R01 portfolios.

**Table 5. Bibliometric Analyses for Groups and Metrics Calculated**

| Group (award years) | Grant-Level | Researcher-Level |
|---|---|---|
| NDPA Pioneers (FYs 2004–2006) | X | X |
| Matched R01 PIs (FYs 2004–2006) | X | X |
| HHMI Investigators (FY 2005) | | X |
| NDPA Finalists (FYs 2004–2006) | | X |
| Random R01 Portfolios (FYs 2004–2006) | X | |

### 1. Background

A substantial body of literature demonstrates the use and limitations of bibliometrics for scientific evaluations. Bibliometrics are most often used to evaluate research

[31] Martin and Irvine (1983, 66) noted the use of publications and citations as partial indicators of the "level of scientific progress made by the individual or group." For example, researchers are motivated to publish "not only to present valuable results, but also for social, political, and career reasons" and other factors.

[32] Bibliometric analysis refers to the study of publication data including the authors, journals, and cited references, from scholarly literature to produce quantitative or statistical analysis. For further on bibliometric analysis, see van Raan (2003).

performance on basic research (Melkers 1993). The technique allows for a variety of metrics that focus on different portions of publication data and associated meta-data. For example, bibliometrics can be used to assess the scientific production of a country, an institution, a program, or an individual researcher. In 2010, the National Science Board used a country-specific metric to benchmark the U.S. national science capacity against other countries of interest (NSB 2010; Moed et al. 1995). The technique has also been used to measure the research performance of universities and research groups and associated institutions (Moed et al. 1985). Furthermore, the NIH and the NSF have long used publications counts and citations as evaluative indicators of a program's scientific activity (Narin 1976; Cozzens 1996).

## 2. Approach

STPI adapted Godin and Doré's (2004) definition of scientific impact to operationalize the bibliometric analysis. They assert that scientific impact can be measured in the following ways:

- Contributions to research through publications, including diffusion and appropriation of new knowledge, theories, methodologies, models, and facts
- Formation and development of specialties and disciplines
- Diversification of the type of research conducted (basic, applied, and strategic)
- Development of interdisciplinary, intersectoral, and international research

Many of these dimensions can be quantified through both simple and sophisticated analysis of information provided through publications and citations (National Academy of Sciences 1999, 2011; Cozzens 1996; van Raan 2003). After reviewing the scientometric and bibliometric analysis literature, we identified 11 traditional and emerging metrics within five broad categories—(1) productivity, (2) citation impact, (3) journal impact, (4) interdisciplinarity and diversity, and (5) collaborations and networks.

Table 6 lists the metrics STPI used, explains the rationale for using them, and indicates the data requirements for each.

**Table 6. Metrics used in Bibliometric Analysis by Category**

| Category | Metric | Rationale | Data Requirements |
|---|---|---|---|
| Productivity | 1. Number of publications | Indication of the production and diffusion of knowledge, theories, methodologies, models, and facts | Publication counts |
| | 2. Number of publications per grant funding amount | Indication of the productivity per funding dollars provided to the award | Publication counts<br>Total direct funding |
| Citation Impact | 3. Number of citations per awardee | Indication of actual appropriation of diffused knowledge in the scientific community | Publication citations |
| | 4. Number of citations per grant funding amount | Indication of the appropriation of knowledge per funding dollars provided to the award | Publication citations<br>Total direct funding |
| | 5. Number of citations per publication | Indication of the actual appropriation of diffused knowledge per individual publication | Publication citations<br>Publication counts |
| | 6. H-index | Indication of unusually high appropriation of knowledge relative to other scholarly work in a similar field | Publication counts<br>Publication citations |
| Journal Impact | 7. SCImago Journal Rank (SJR) | Indication of the publication's potential for diffusion; possibly accounting for the publication's visibility in the scientific community and citation counts | Journal titles<br>Journal ranking* |
| | 8. Proportion of publications by journal percentiles | Indication of the aggregate publications' potential for diffusion; possibly accounting for the publication's visibility in the scientific community and citation counts | Journal titles<br>Journal ranking* |
| Interdisciplinarity and Diversity | 9. Integration score | Indication of the integration of diverse knowledge or the diversification of the type of research being performed | Journal subject areas^ |
| | 10. Number of co-authors | Indication of the growth in number of co-authors in publications | Co-authors |
| Collaborations and Networks | 11. Number of co-author affiliations | Indication of the growth in number of affiliations associated with co-authors in publications | Institutional affiliations of co-authors |

* SCImago Research Group (2011b), retrieved February 5, 2012 from http://www.scimagojr.com/journalrank.php.

^ Subject areas are assigned by Thomson Reuters as Subject Categories. See Thomson Reuters, Journal Search, http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=D.

## 3. Analyses Conducted

Three of the five comparison groups—NDPA Pioneers, matched R01 PIs, and random R01 portfolios—were used in the grant-level analyses. All bibliometric indicators were calculated for the NDPA Pioneer grants and the matched R01 grants, while only three metrics (1, 3, and 5) relating to impact were calculated for the random R01 portfolios due to time constraints and data availability.

Four of the five comparison groups—NDPA Pioneers, matched R01 PIs, HHMI investigators, and NDPA finalists—were used in the researcher-level analyses. Nine of the 11 bibliometrics measures (1, 3, 5, 6, 7, 8, 9, 10, and 11) were analyzed. Metrics 2 and 4 (number of publications per grant funding amount and number of citations per grant funding amount) were calculated for the Pioneers and HHMI investigators, but not for the other two groups.

For researcher-level analysis, a distinction was made between the bodies of work published prior to the grant or grant application (i.e., 1980 to year of grant) and the bodies of work published post-award/application (i.e., year of grant to 2011). See Chapter 2 for details on the differences among these data. This distinction enabled us to create a baseline for the comparison groups across many of the metrics. The random R01 portfolios were not included in the research-level analysis due to time constraints.

STPI performed various statistical tests (where applicable) to ascertain the significance of differences between the NDPA Pioneers and the comparison groups across the metrics. Statistical tests performed included the Kolmogorov-Smirnov test to determine if the group distributions differ significantly, and the chi-square test to observe significant differences for categorical data.

### a. Productivity Metrics

STPI calculated two productivity metrics:

- Number of publications
- Number of publications per grant funding amount

Van Raan (2003) notes that “communication, i.e. the exchange of research results, is the driving force in science” and considers publications as direct research outputs that are critical to understanding the diffusion of knowledge. The number of publications over a researcher’s career indicates the level of activity or productivity of a researcher in a particular field. However, there are limitations when comparing publication counts across fields since some fields have a stronger tradition of publishing than others (Cozzens 1996). For the purposes of this study, comparisons of publication counts are less problematic because the NIH and comparison group researchers are all conducting

research within the biomedical and health-related domains. Additionally, publication counts depend on the length of the researcher's career and their respective institutional support. We have shown that four of the five comparison groups, for which the data were available, have similar publication counts prior to the award date (see Chapter 2).

### b. Appropriation of Knowledge

Four metrics represent the appropriation of knowledge:

- Number of citations
- Number of citations per grant funding amount
- Number of citations per publication
- H-index

Garfield (1955) finds that citations to a publication indicate an association between two publications that can be used to measure the publication's scientific impact. Moreover, some studies have concluded there is a positive relationship between citation counts and the outcomes of peer review (Oppenheim 1995 and 1997). While there are widely acknowledged issues with the use of citations for understanding scientific impact, citation counts have been accepted in the literature due to the relative ease of data collection and its representation of the direct research impact on the scientific community (Bornmann and Hans-Dieter 2008).[33] The total number of citations received may be influenced by the funding the researcher received as well as the number of publications that they produced. Thus, STPI also compared distributions of the number of citations normalized by the award funding amount and citations per publication in order to provide other perspectives of the research quality.

One limitation to analyzing citations is that they are accumulated over time, thus recent publications will have had less time to accumulate citations. To understand the citation distributions over time across the groups, STPI calculated the time to citation for grant-level NDPA and matched R01 publications. Other metrics, such as the h-index, are derived from citations and attempt to improve upon the limitations of simple citation counts for researcher-attributed analyses.

The h-index draws upon the idea that knowledge diffusion is dependent upon and integrates both publication and citation counts into one metric (Hirsch 2005). Variations on the h-index attempt to render impact more comparable across researchers by accounting for differences in the length of a researcher's career (Burrell 2007) and

[33] The unifying factor in these limitations is the lack of discrimination for the nature of the citation. An article receives one citation count regardless of whether the citation was based on a positive, negative, scientific, or non-scientific rationale. In addition, the citation could also have been the author's self-citation.

differences in citation patterns from diverse scientific fields (Iglesias and Pecharroman 2007). The h-index was further validated as a metric for research quality and impact when van Raan (2005) concluded that the h-index is positively correlated with the peer judgments for articles published by 147 chemistry research groups in the Netherlands.

In the bibliometrics literature, the h-index has been calculated at various degrees of aggregation: from a single researcher, to a laboratory, to a research field, to a set of journals, to a country (Hirsch 2005, Alonso et al. 2009, Jascó 2011). STPI calculated the h-index for the subset of grant-attributed publications as well as broad sets of publications (1980–2011) from a researcher's career.

### c. Journal Impact

STPI used two metrics using journal impact scores:

- SCImago Journal Rank (SJR)
- Proportion of publications by journal percentiles

We used the SJR indicator to measure the prestige of the journals in which the researchers from the comparison groups publish. The concept of a journal's impact is derived from citation analysis. It provides a proxy measure of the journal's effectiveness to communicate research results (Garfield 1979). The number of citations a journal receives per articles published provides an indication of the journal's impact on the scientific community (Garfield 1972). Moreover, there is a greater capacity for knowledge diffusion in journals with higher citations and impact.

Bibliometric researchers and companies have developed several journal impact indicators, including the Journal Impact Factor (JIF) developed by Thomson Reuters[34] and the SJR developed by the SCImago Research Group[35] (Garfield 2005; Moed 2010). The JIF is calculated by dividing the number of citations in the current year to all items published in a journal the previous 2 years with the number of publications in the journal during the previous 2 years. The SJR not only considers the citation counts but also the quality of the citation measured by the prestige of the journal from which the citation is coming. The importance of the journal is thereby measured by the importance of the citations they receive. SCImago researchers González-Pereira, Guerrero-Bote, and Moya-Anegón (n.d.) assert there is a strong correlation between the JIF and SJR as tested for journals in the biochemistry, genetics and molecular biology, and psychology fields. STPI chose the 2011 SJR to measure journal impact due to the public availability of the metric (SCImago Research Group 2011b).

---

[34] About the JIF: http://thomsonreuters.com/products_services/science/free/essays/impact_factor.

[35] About SJR: http://www.scimagojr.com/aboutus.php.

Two important considerations when comparing journal impact measures include (1) the time horizon used in the calculations and (2) the inclusion of reviews, editorials, and letters. Both the JIF and the SJR measure a journal's relative importance in a particular field in a specific time period and should be compared with impact factors for other journals calculated based on that same time period. Additionally, journals include reviews, editorials, and letters that often receive many more citations than other research or experimental publications (Seglen 1997). These publication types can easily distort the journal impact metric.

### d. Interdisciplinarity and Diversity

STPI used two interdisciplinarity measures to capture the diversity of the researcher's body of work:

- Integration score
- Subject area analysis

While interdisciplinarity has many meanings in scholarship, we define interdisciplinarity as the integration of traditional disciplines of knowledge into newly synthesized fields or niche areas of research within an existing field.[36] Scientists have developed emerging metrics, such as the integration score, to measure the level in which researchers have integrated knowledge from various disciplines into their published work. The integration score measures the knowledge within a body of research based on the journal's subject area from a publication's cited references; therefore, it is a backward-looking metric (Porter et al. 2007; van Raan 2002). STPI used a set of 221 journal subject areas assigned by Thomson Reuters to WOS indexed journals. While there are other measures of diversity, Rafols and Meyer (2010) noted that the integration score is the only measure that integrates the three aspects of diversity—variety, balance, and disparity—into one index. Variety represents the number of disciplines cited by the publication, balance signifies the distribution of citations among the disciplines, and disparity shows the similarity of dissimilarity of the subject areas to one another (Porter and Rafols 2009).

We also used the proportions of subject areas from the cited reference's journals to qualitatively represent the diversity of researchers' publications. We organized the 221 subject area assignments from Thomson Reuters into 18 macro-subject area clusters based on the factor analysis performed in Leydesdorff and Rafols (2009). We used the macro-subject areas in two analyses: (1) the frequency of cited references in each subject area normalized such that each cited reference and all publications had the same weight,

[36] For further clarification, refer to the discussion of "content integration" and the distinction between "interdisciplinary" and "integrative" in the social science in Klein (1990, 26–27).

and (2) the frequency of cited references in each subject area normalized such that each researcher had the same weight, regardless of number of publications. Both analyses represent a different perspective on the diversity of the research being funded by the program. The latter analysis enables us to observe whether a highly productive researcher exerts disproportionate influence on the differences in the distributions across or within any one of the 18 macro-subject areas.

The use of subject areas based on Thomson Reuters in this study presented a couple of difficulties: (1) the low accuracy of the subject area categories assigned to journals and (2) the aggregation of subject areas into macro-subject areas. Thomson Reuters assigns from one to six subject areas to WOS indexed journals based on a combination of factors, including the cited references in the journal publications. Rafols, Porter, and Leydesdorff (2010) measured the error associated with this assignment to be around 50%, meaning nearly half the indexed journals have at least one subject category assigned that is mismatched (Rafols and Leydesdorff 2009). However, most of the mismatched journals appear to fall in subject areas within the close vicinity of the macro-subject area categories (Rafols, Porter, and Leydesdorff 2010). In addition, the multidisciplinary sciences subject area is problematic since it includes too diverse an array of publications to categorize within any other subject area. This subject area is assigned to journals such as *Nature* and *Science*. In aggregating the 221 subject areas into 18 macro-subject areas, the multidisciplinary sciences subject area is grouped into the biomedical science macro-discipline. We encourage a judicious look be taken when comparing the biomedical science macro-discipline across the groups.

### e. Collaborations and Networks

STPI used two metrics to gain insight into the individual researcher collaboration networks among the groups:

- Number of co-authors
- Number of co-author institutional affiliations

De Solla Price (1970) explained that not only do publications provide information, but they also can reveal relationships among the people publishing the scholarly work. Given a researcher's body of work, the number of co-authors and institutional affiliations associated with that work reveals the magnitude of the researchers' co-author and institutional collaboration networks after application for or receipt of the award.

## B. Expert Review

### 1. Expert Panel Experimental Design and Analysis

The expert panel assessment was conducted on a total of 510 papers in 108 sets of "top five" papers ("packets") from the NDPAs, the matched R01 grants, and the HHMI investigators. The packets were divided into groups of between three and eight papers of similar research areas. An expert read at most four packets within a particular group, and each packet was read by three or four experts. Within the groups, the experts were assigned to papers using an alias-optimal design to make it as easy as possible to estimate separate expert and paper effects. Overall, 336 packet evaluations and 1,587 paper evaluations were included in the analysis. All 94 recruited experts completed the evaluation. One matched R01 PI was inadvertently used as an expert, but his ratings were excluded from our analysis to avoid any conflict of interest.

The analysis of the data was based on the multi-rater ordinal data methodology developed in Johnson and Albert (1999). We assumed that each paper and packet has an underlying impact (innovativeness) that can be measured. When an expert assesses the impact of a packet, we assumed that this "measurement" is made with some variability. We did not observe this measurement directly; instead, we observed a discretized version, much like the grade on an examination being discretized to a letter grade. Each expert discretized the impact (innovativeness) scale differently, and the "grade" cutoffs were estimated from the data. The underlying impact (innovativeness) was also estimated, as were the measurement variations for each expert. Since each packet was read by multiple experts, large estimated measurement variations indicated that the expert assessed packets differently from other experts. Similar assumptions were made for paper "measurements," with the additional assumption that paper impact (innovativeness) would be distributed around packet impact with some variation.

In addition, several covariates were incorporated into the analysis to determine whether they are predictive of the expert-assessed impact (innovativeness). At the packet level, we considered grant mechanism, number of citations at time of award for the PI, number of publications at time of award for the PI, h-index at time of award for the PI, SIR at time of award, years since degree, receipt of early career award, and, for the matched R01 packets, total direct costs. At the paper level, we considered journal impact factor and total citations to date. Details of the analyses are discussed in Appendix I.

### 2. Analysis of Qualitative Data from the Expert Panel

STPI initiated a qualitative analysis of the open-ended responses collected from the experts to provide insight on how impact and innovation were assessed. The expert review protocol is reproduced in Appendix E. The goals of this exercise were to determine whether the experts used the given typologies when evaluating the research

and whether the experts considered impact and innovation to be related to one another, given that the two factors were found to be highly correlated through the quantitative analysis.

### a. Data

Data for the qualitative analysis came from open-ended responses from experts following the impact and innovation questions, which were asked in separate sections via the protocol. Each section of the protocol begins with a sentence on how each factor relates to the overall goals of the NDPA program. For example, the impact section begins with the following: "NDPA was created to promote highly innovative and potentially transformative approaches that have the potential to produce extremely high impact on a broad area of biomedical or behavior research." Reviewers were provided with a list of factors to consider when assessing impact and innovation within the context of this evaluation. These factors were based on the Heinze et al. (2007) and Colwell (2003) typologies of research outcomes and research risks.

For impact, "extremely high impact" refers to research that accomplishes one or more of the following:

- Radically changes present understand of an important existing scientific or engineering concept
- Leads to the creation of a new paradigm or field of science or engineering
- Challenges present understanding in the field(s)
- Provides pathways to new frontiers
- Challenges conventional wisdom
- Leads to unexpected insights that enable new techniques or methodologies
- Redefines the boundaries of science or engineering

For innovation, "extremely innovative" refers to approaches that can be characterized in one or more of the following ways:

- The ideas underlying the research are at odds with prevailing wisdom
- The research requires the use of equipment or techniques that have not been proven or are considered extraordinarily difficult
- The research involves a unique combination of disciplines

After assessing each paper and the packet as a whole, reviewers were asked to write responses to the following question in both sections: "What about these papers, individual or as a whole, made you choose your answers above?" Responses were

voluntary and had no character limits. Out of the 340 reviews that were collected, 334 and 309 provided answers for the impact and innovation questions, respectively.

### b. Method

Responses were coded inductively using content analysis assisted by NVivo, a software tool used for analyzing qualitative data. Content analysis is a systematic approach widely used to extract common themes from unstructured data (Lasswell and Leites 1968).

An initial coding framework, or codebook, was developed from the rubrics listed previously and from examining samples of responses. In contrast to quantitative data analysis, qualitative data analysis is an evolving process dependent on themes emerging from the text. Categories were added and refined after all responses were coded using the initial coding framework. The final codebook is available in Appendix G, which also provides representative examples of responses coded under each theme.

## C. Potential Limitations

The evaluation has limitations along three dimensions: measurement validity (which relates to accuracy of measurement), measurement reliability (which relates to repeatability of measurement), and internal validity (which relates to establishing causality).

With respect to measurement validity, the biggest concern is how well the abstract concepts of impact and innovativeness of research approaches were turned into empirically observable indicators. A key assumption was that publications capture the intended effects. It is possible that transformative research is more likely to be rejected by peer-reviewed journals or that, with our database limitations, we were unable to capture all of the publications attributed to the grant or researcher. As a specific illustration, we assumed that interdisciplinarity of publications is a close-enough proxy for the interdisciplinarity of research. It is likely that researchers are bringing interdisciplinarity into their research (for example, through formal or informal interactions with colleagues from other disciplines) without it reflecting in the subject area distribution of their publications. Publications, as the “currency” of science, however, were the most appropriate research outcome to assess impact and innovativeness in our evaluation.

Going beyond the limitation of using publications as a proxy for research, there are other limitations related to measurement validity. For example, we assumed that citations to research publications, researchers’ h-index, and journal reputations were good proxies of impact. The assumption is supported in the literature, but it is an assumption nevertheless that should be made explicit. Similarly, we assumed that the number of co-authors was a good proxy for collaborations, and the integration score was a good proxy

for interdisciplinarity. For the latter case especially, this may not be the case. The calculation of these scores is based on the diversity of subject categories assigned to journals by Thomson Reuter's Institute for Scientific Information (ISI) and the degree to which a subject category (SC) relates to a particular journal. This is somewhat problematic. First, the validity of the I-Score is dependent on the accuracy of the publication dataset used to describe a PI's publication history. The I-Score analysis used the publications from the WOS database, which indexes a limited selection of conference papers and proceedings, with coverage varying by research area. This point is particularly important in calculating the proposal I-Score from cited references, since many of the PI records contained conference proceedings that were not included in the WOS database and, thus, not matched to a SC. Second, the I-Score calculation is based on the diversity of SCs matched to the journals of a publication dataset; however, the accuracy of the matched SC(s) to classify a journal may vary (Boyack, Klavans, and Borner 2005). Boyack, Klavans, and Borner also note that for approximately 50% of the SCs, there is a high level of accuracy regarding the matched journal and SC. For the other 50%, there is less accuracy in the attribution.

With respect to measurement reliability, one potential limitation of the evaluation is that the impact and innovativeness assessments are based on experts' ratings of researcher publications. This is a subjective approach. Presumably a different set of evaluators might have given the research different ratings.

We addressed many of the issues related to measurement validity and reliability by using a multi-method approach that combined bibliometrics with expert reviews and text mining.

The three most important limitations that pose threats to internal validity of the evaluation relate to the appropriateness of the matched R01 set, the selection of the publications for analysis, and the timing of the evaluation.

With respect to the matched set, every attempt was made to ensure that the researchers in the comparison group were as similar as possible to the NDPA group. They were matched on areas of research, demographics, and other attributes of importance. An analysis showed the two groups to be highly comparable, including on bibliometric measures that were not used to select the comparison group. Nevertheless, there is still a chance that the matched set of R01s is not an apt comparison group, in which case any differences or lack of differences in outcome cannot be attributed to the receipt of the NDPA award.

Publications selected—both for the bibliometric analysis and expert assessment—can pose a threat to internal validity (i.e., our ability to assign causality) as well. While extreme effort was expended on accurate data disambiguation, download, and cleanup, an assumption was made that Thomson Reuter's WOS database and the NIH's eSPA,

SPIRE, and SPIRE+ databases accurately capture research publications. This may not have been the case.

Similarly, for the expert review, top five publications with the most impact were selected. While most of the Pioneers and HHMI investigators selected these publications themselves, for some of the Pioneers and the matched set of R01 PIs, these publications were selected either by the NIH program officers or STPI (using a carefully designed and validated algorithm to ensure comparability). There is a possibility, however, that the publications selected were not those with the most impact.

With respect to timing, the challenge is that transformative research is often not evident for years (even decades) after being conducted. This evaluation was done for research only 5 years after award of funds. There is a good chance, therefore, that at least for some of the research, the evidence base—whether the number of citations or expert judgment as to impact—is absent. Conversely, research considered to have high impact now, may, in fact, be disproved in a few years. And what we may be measuring is noise rather than signal.

We addressed these potential limitations by operationalizing concepts in multiple ways and using multiple methods of analyses.

# 4. Impact and Innovation: Grant-Level Findings

This chapter presents our findings with respect to the impact of research and the innovativeness of research approaches. We compare the NDPAs first to the matched R01 set of grants, and then to the random R01 portfolios. For the matched R01 grants, we discuss the first eight bibliometric indicators; the final four are discussed in Chapter 7. In addition, we present the results from analyses of expert assessments of the impact and innovativeness of the top five publications. For the random R01 portfolios, we discuss the findings from the analyses of three bibliometric indicators.

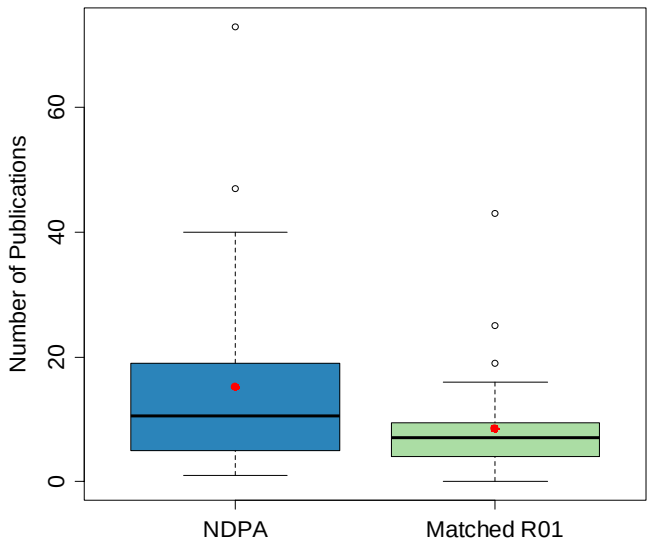
## A. Matched R01 Grants

We compared impact and innovation of the matched R01 grants to that of the NDPAs through bibliometrics and expert review. The results are presented graphically as box plots that show the relative distributions. We performed Kolmogorov-Smirnov (K-S) tests to determine whether the group distributions differ significantly.

### 1. Bibliometric Findings

#### a. Number of Publications

NDPAs produced more grant-attributed publications than the matched R01 grants (*K-S*, $p = 0.04$) (Figure 15).

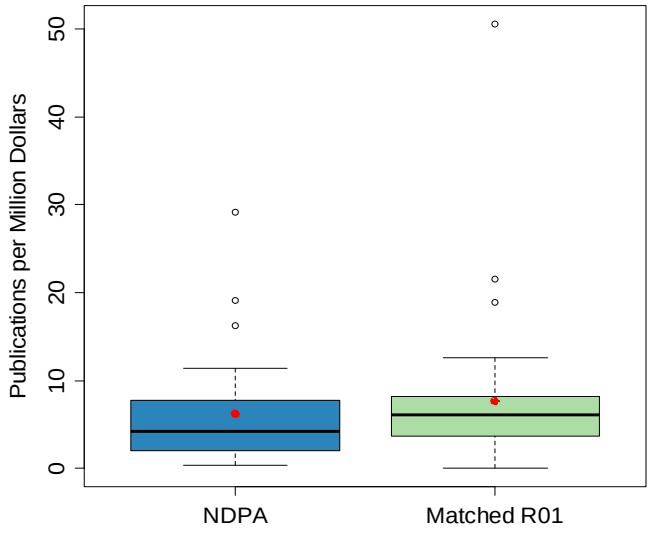



*Note:* One observation equals one grant.

**Figure 15. Grant-Level Publications**

### b. Number of Publications per Grant Funding Amount

When accounting for grant size (as measured by direct cost), there was no difference in the number of publications output by the two grant mechanisms (K-S, $p = 0.31$) (Figure 16).

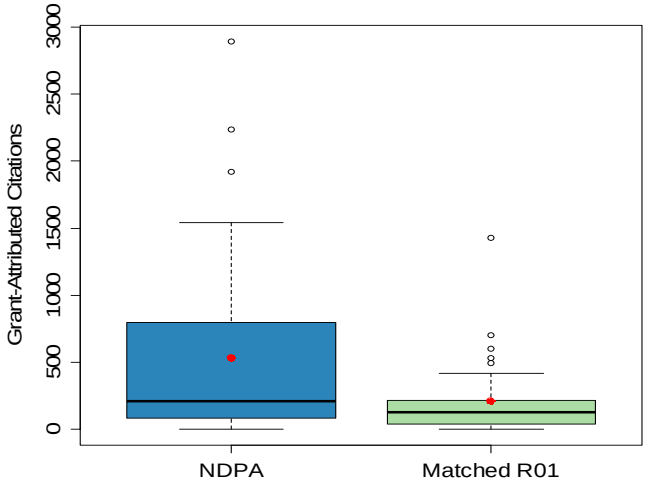



*Note:* One observation equals one grant.

**Figure 16. Grant-Level Publications per Million Dollars in Direct Costs**

### c. Number of Citations

There was no statistical difference between the Pioneers and the matched R01 PIs in citations to grant-attributed publications (K-S, $p = 0.06$) (Figure 17).

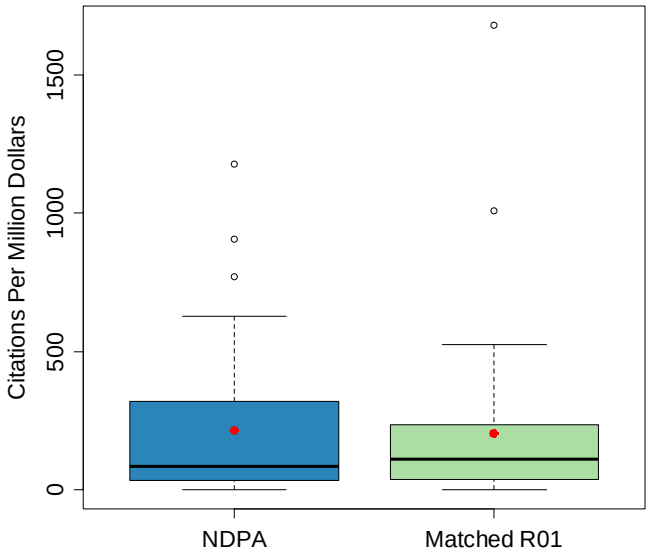



*Note:* One observation equals one grant.

**Figure 17. Grant-Level Citations**

### d. Number of Citations per Grant Funding Amount

When accounting for grant size, there continued to be no difference in the citations to grant-attributed publications (K-S, $p = 0.89$) (Figure 18).

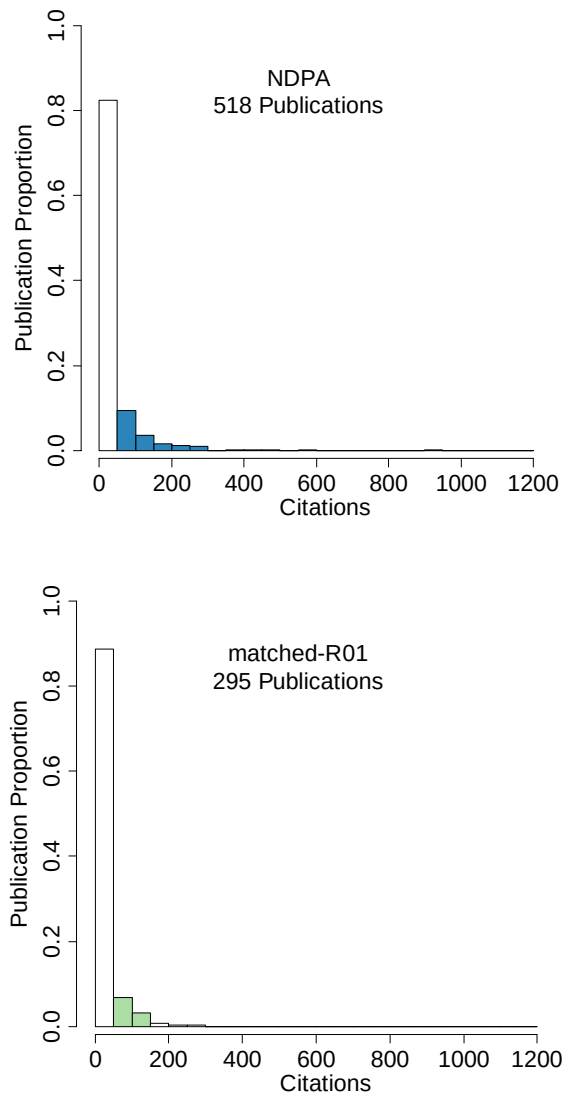



*Note:* One observation equals one grant.

**Figure 18. Grant-Level Citations per Million Dollars in Direct Costs**

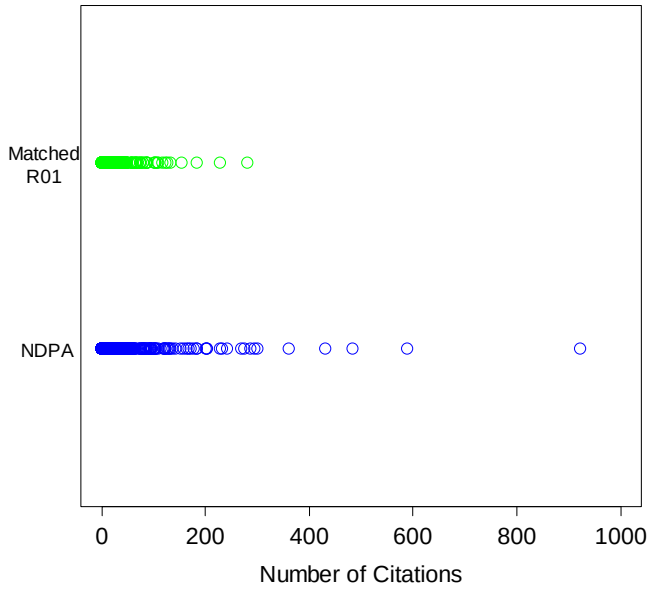
### e. Number of Citations per Publication

At the publication level, NDPAs and matched R01 grants had statistically similar citation distributions (K-S, $p = 0.34$) (Figure 19 and Figure 20). Nevertheless, it is notable that the NDPA distribution of citations has a longer positive tail. Several NDPA publications had 400 or more citations, while the matched R01 publications had no more than 281 citations. This might suggest that the most successful NDPAs had a greater impact than the most successful matched R01 grants.




*Note:* One observation equals one publication.

**Figure 19. Grant-Level Citations per Publication—Histogram**




*Note:* One observation equals one publication.

**Figure 20. Grant-Level Citations per Publication**

### f. Additional Analysis on Publications and Citations

Calculations show that NDPAs produced publications more slowly than the matched R01 grants (chi-sq, $p < 0.0001$) (Figure 21), but their publications, once released, were cited more rapidly (chi-sq, $p < 0.0001$) (Figure 22). These plots are a snapshot of publications and citations through the end of 2011. Figure 22, in particular, will continue to change as publications accrue citations.

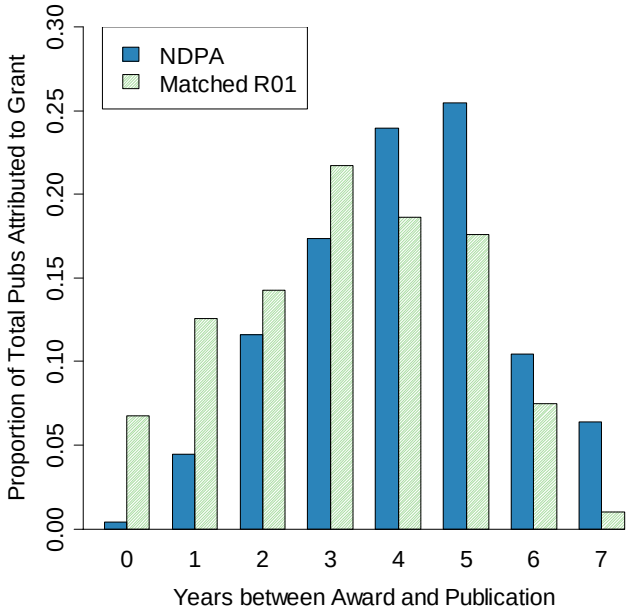



**Figure 21. Lag time between Publication and Award Year**

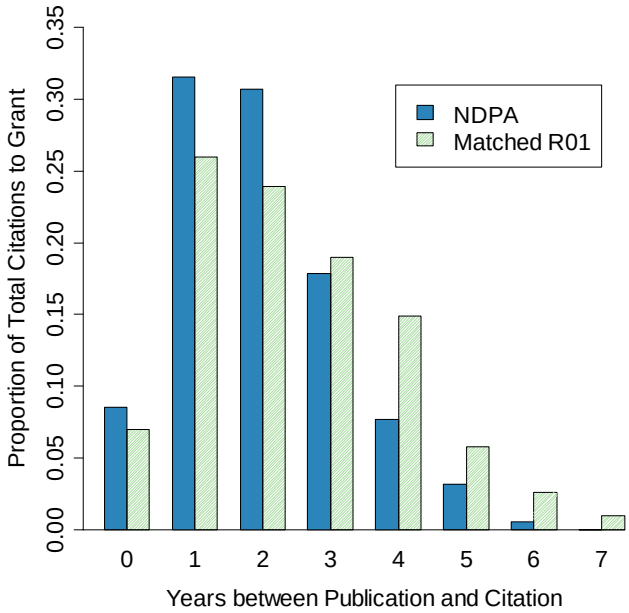
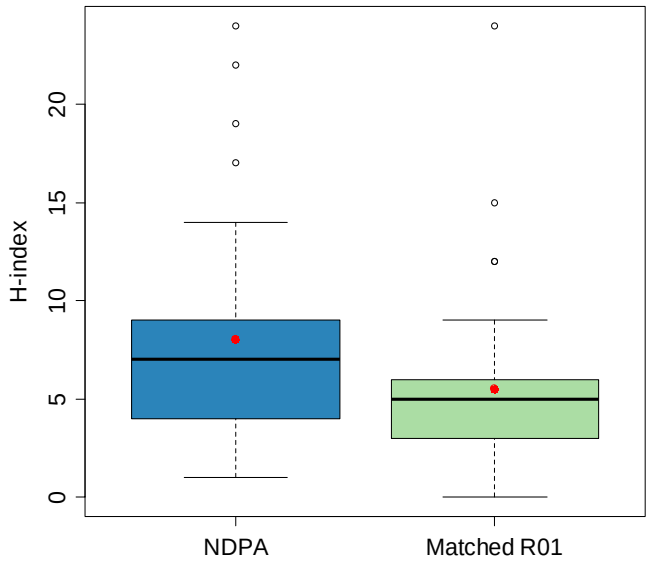



**Figure 22. Lag Time between Citation and Publication Year**

### g. H-index

NDPAs have larger h-index values than matched R01 grants (K-S, $p = 0.003$) (Figure 23).

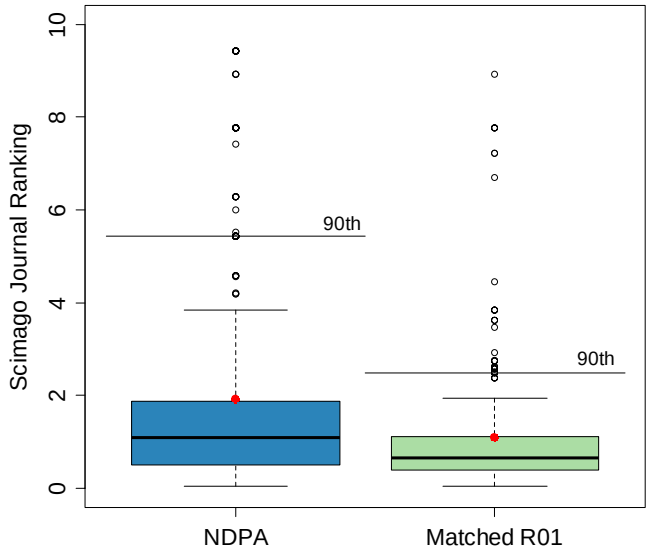



*Note:* One observation equals one grant.

**Figure 23. Grant-Level H-Index**

### h. SCImago Journal Rank

NDPA publications appeared in journals with higher journal impact factors than those of matched R01 publications, and there was a notable difference between the 90th quantiles in the distributions of the two groups (K-S, $p < 0.0001$) (Figure 24).




*Note:* One observation equals one publication.

**Figure 24. Grant-Level Journal Impact Factors**

### i. Proportion of Publications by Journal Percentiles

Approximately 25% of NDPA-attributed publications appeared in the top 0.5% of all SCImago-indexed journals, approximately 50% of NDPA publications appeared in the top 1% of all SCImago-indexed journals, and just over 30% of the matched R01 publications appear in the top 1% (chi-sq, $p < 0.0001$) (Figure 25).

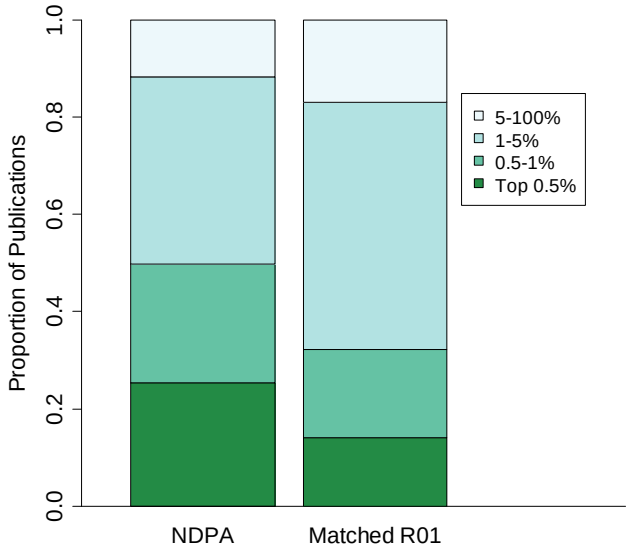
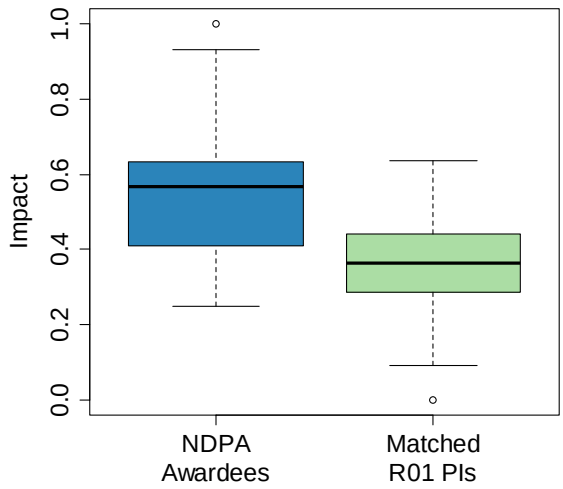



**Figure 25. Grant-Level Proportion of Publications by Journal Percentiles**

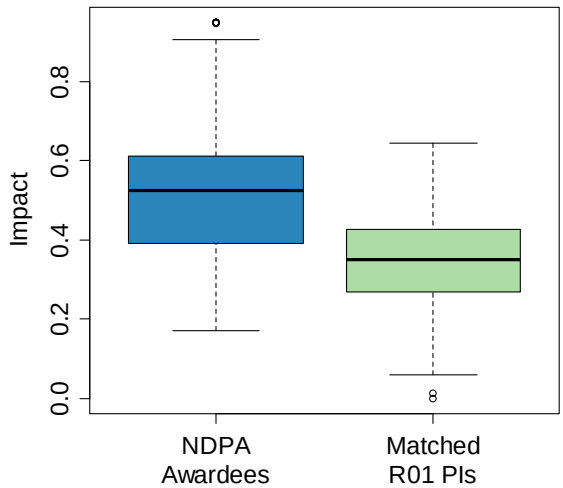
## 2. Expert Review Findings

The expert review assessed impact and innovation of both packets and papers. Figure 26 is a box plot of the estimated impact for NDPA and matched R01 packets. The Pioneers had a significantly higher impact (t-test, $p < 0.0001$, K-S $p = 0.0002$). Similar results are seen for papers (t-test, $p < 0.0001$, K-S, $p < 0.0001$) in Figure 27.




*Note:* One observation equals one grant.

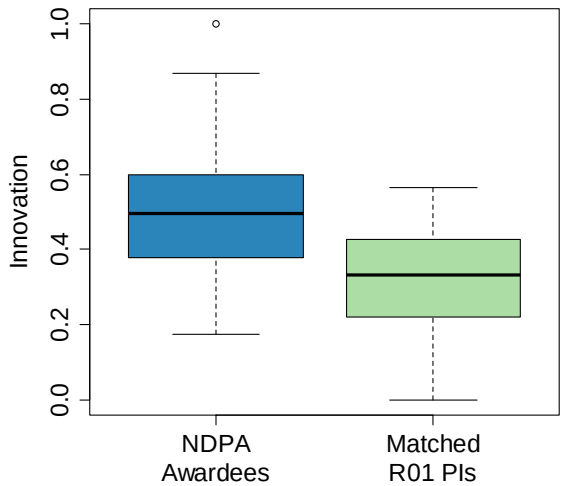
**Figure 26. Expert-Assessed Packet Impact**




*Note:* One observation equals one paper.

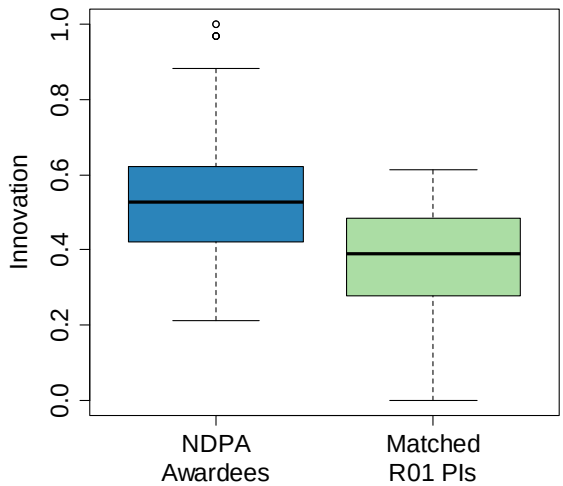
**Figure 27. Expert-Assessed Paper Impact**

Figure 28 and Figure 29 summarize the results for packet and paper innovation. Again, the experts assessed NDPA packets (t-test, $p < 0.0001$, K-S, $p = 0.0005$) and papers (t-test, $p < 0.0001$, K-S, $p < 0.0001$) to be significantly more innovative than the matched R01s.




*Note:* One observation equals one grant.

**Figure 28. Expert-Assessed Packet Innovation**




*Note:* One observation equals one paper.

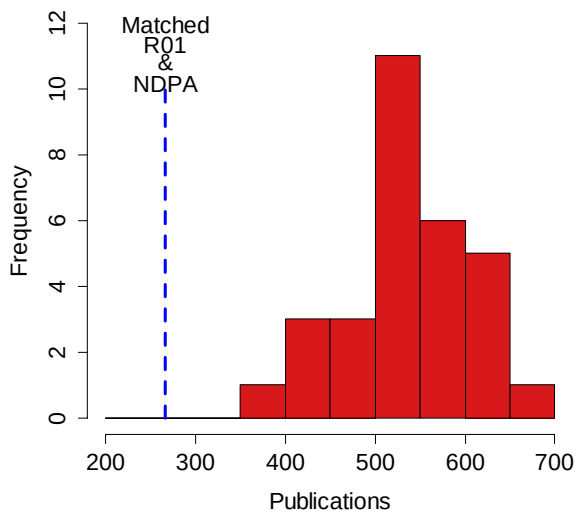
**Figure 29. Expert-Assessed Paper Innovation**

## B. Random R01 Portfolios

The impact of the random R01 portfolios was compared to that of the NDPA and matched R01 portfolios in a limited bibliometric analysis. Figure 5 and Figure 6 (presented previously in Chapter 3) are plots of the total direct costs of each randomly selected portfolio and the number of grants in each portfolio (the range is from 66 to 96 with a median of 86).

### 1. Number of Publications

Each random R01 portfolio consists of a greater number of grants than the NDPA portfolio (Figure 6) with similar direct costs (Figure 5). Figure 30 shows the number of publications in the 30 random R01 portfolios, compared to that of the NDPA portfolio and the matched R01 portfolio.

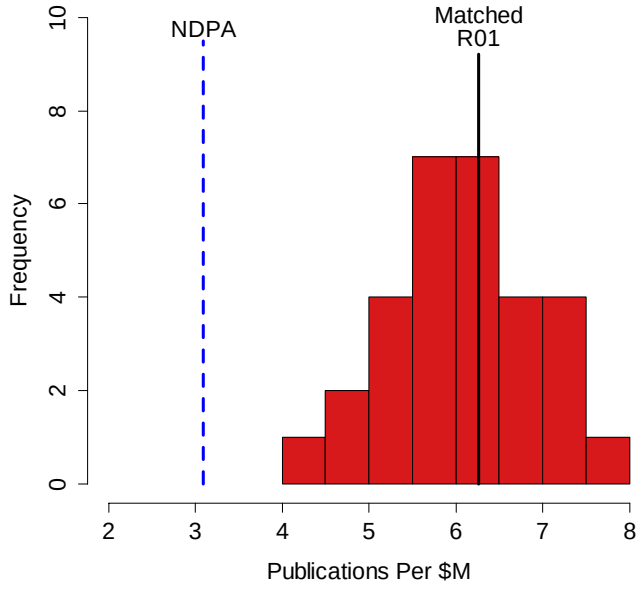



*Note:* One observation equals one portfolio.

**Figure 30. Random R01 Portfolio Grants and Publications**

### 2. Number of Publications by Portfolio Funding Amount

Figure 31 shows the number of publications for NDPA and the matched R01 portfolios when divided by the total direct costs of the portfolio. While the NDPA portfolio has fewer publications per dollar, the matched R01 portfolio is quite comparable to the random portfolios.

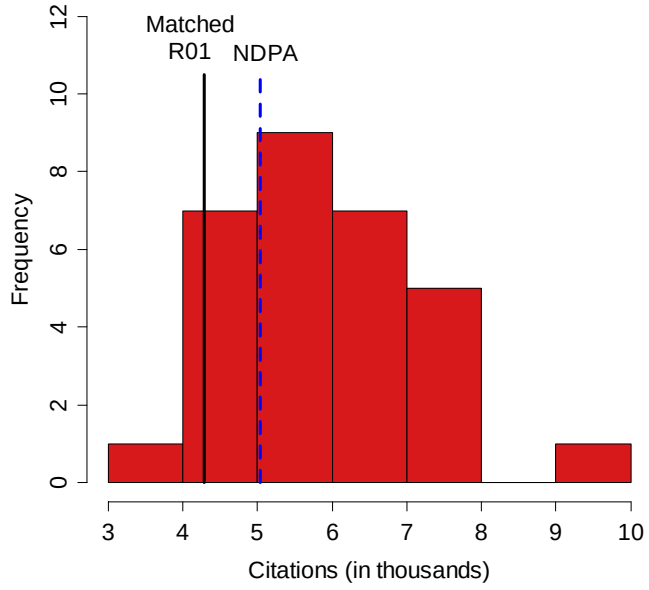



*Note:* One observation equals one portfolio.

**Figure 31. Number of Publications Divided by Total Direct Costs of Portfolio**

## 3. Number of Citations

Despite having fewer publications, the NDPA portfolio of grants received a similar number of citations as the matched R01 portfolio (Figure 32).

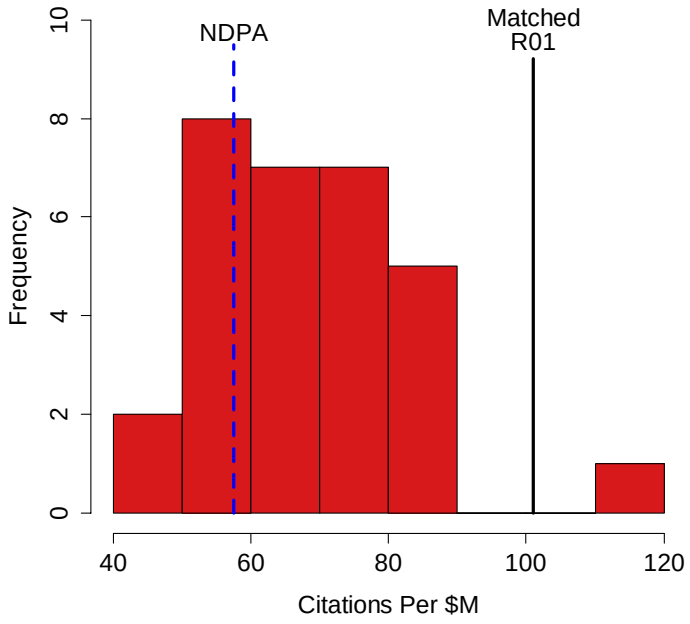



*Note:* One observation equals one portfolio.

**Figure 32. R01 Portfolio Citations**

## 4. Number of Citations per Portfolio Funding Amount

Figure 33 shows that when normalized by total direct costs of the portfolio, the NDPA and matched R01 portfolios still have similar numbers of citations as the random R01 portfolios.




*Note:* One observation equals one portfolio.

**Figure 33. Number of Citations Divided by Total Direct Costs of Portfolio**

## 5. Citations per Publication

The density estimations of the distributions of citations per unique publication displays a shift in the right side of the distributions for the NDPA and matched R01 portfolios when compared to the combined random R01 portfolio distribution (Figure 34).

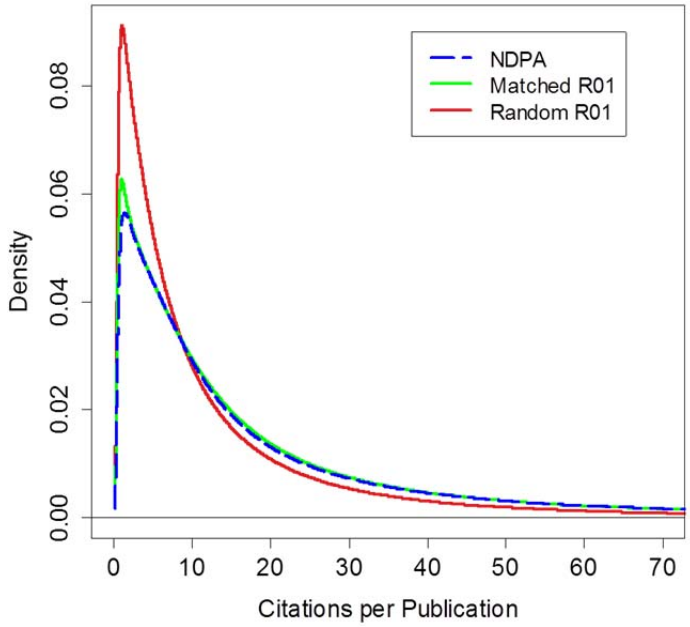
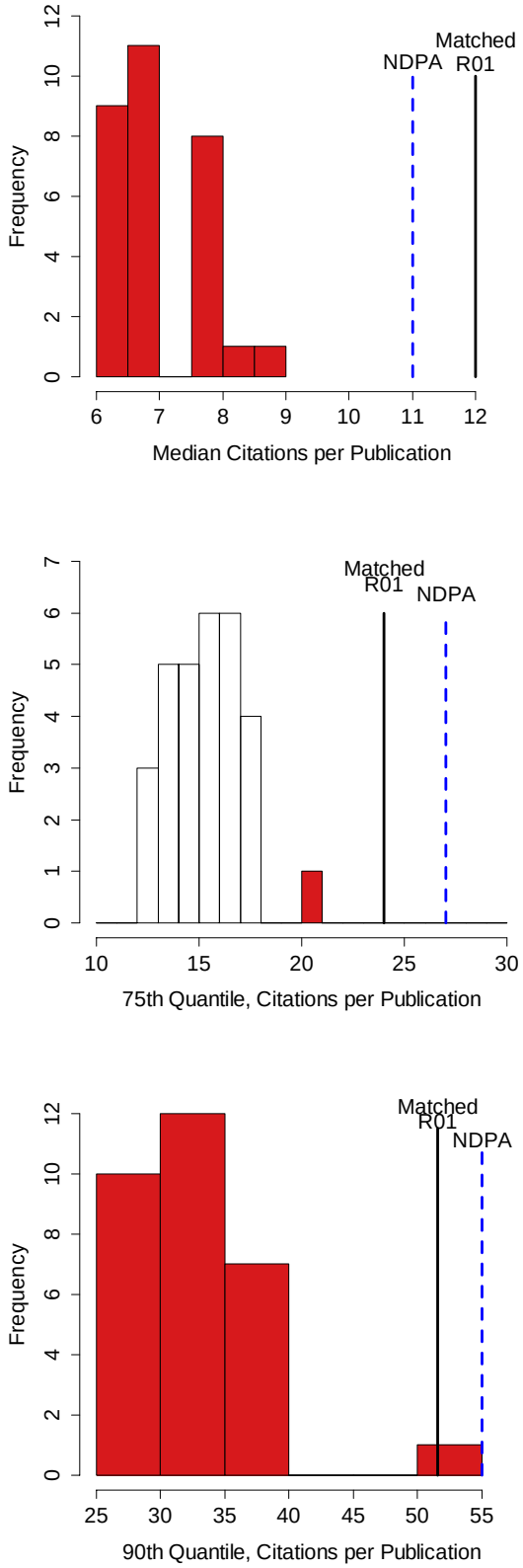



**Figure 34. Density Estimations of the Distributions of Citations per Unique Publication for the NDPA, Matched R01, and Random R01 Portfolios**

When considering the distribution of citations over publications, the NDPA portfolio had a higher median, 75th quantile, and 90th quantile than the R01 portfolios (Figure 35). The matched R01 portfolio, similarly, had a higher median, 75th quantile, and 90th quantile than the random R01 portfolios; the NDPA portfolio performed better than the matched R01s at the higher quantiles. Table 7 shows the spread in values (minimum/maximum) at each of the reported quantiles for the random R01 portfolio, the NDPA, and the matched R01 portfolios.




*Note:* One observation equals one portfolio.

**Figure 35. Distribution of R01 Portfolio Citations per Publication (Median, 70th, and 90th Quantile)**

**Table 7. Spread in Values (Minimum/Maximum) at Each Reported Quantile of Comparison among Groups**

| | Median | | | 75th Quantile | | | 90th Quantile | | |
|---|---|---|---|---|---|---|---|---|---|
| | Min | | Max | Min | | Max | Min | | Max |
| NDPA | -- | 11 | -- | -- | 27 | -- | -- | 55 | -- |
| Matched R01 | -- | 12 | -- | -- | 24 | -- | -- | 52 | -- |
| Random R01 | 6 | 7 | 9 | 13 | 15 | 21 | 26 | 32 | 51 |

## 6. SCImago Journal Ranking

NDPA eSPA publications have higher journal impact factors than those of matched R01 and random R01 portfolio eSPA publications (K-S, $p = 0.02$, <0 .0001 respectively) (Figure 36). In addition, the 90th percentile of NDPA publications has a higher impact factor than the 90th percentiles of the matched R01 and random R01 portfolios.

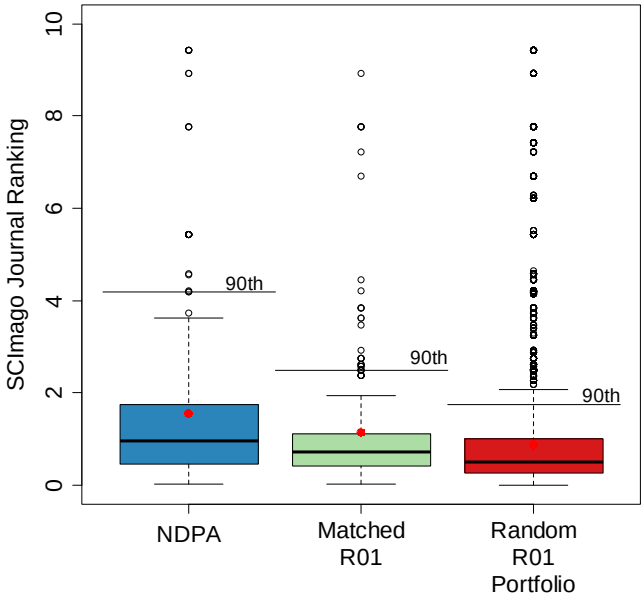



**Figure 36. Journal Impact Factors for NDPA, Matched R01, and Random R01 Portfolios**

## 7. Proportion of Publications by Journal Percentiles

Approximately 21% of NDPA eSPA publications appeared in the top 0.5% of all SCImago-indexed journals, and approximately 44% of NDPA publications appeared in the top 1% of all SCImago-indexed journals (Figure 37). The distribution for the random R01 portfolios differs from that of NDPA (chi-sq, $p < .0001$); only 7% and 21% of random R01 portfolio publications appeared in the top 0.5% and top 1% of journals,

respectively. NDPA and matched R01 publications had similar distributions, with 15% of matched R01 publications in the top 0.5% and 35% in the top 1% (chi-sq, $p = 0.14$). (Note that the comparison of NDPA and matched R01s differs from that presented previously in Figure 25 because of the different data set that was used for portfolio comparisons.)

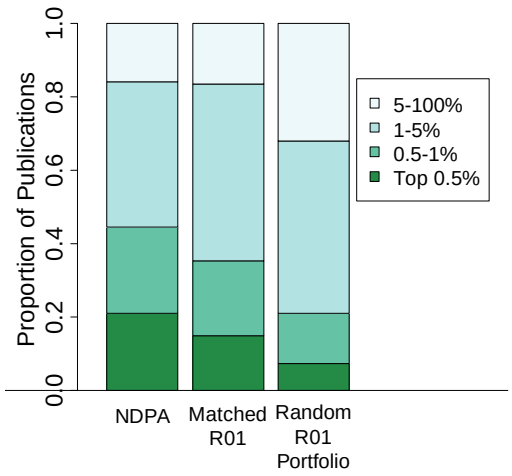



**Figure 37. Proportion of Publications by Journal Percentiles for NDPA, Matched R01, and Random R01 Portfolios**

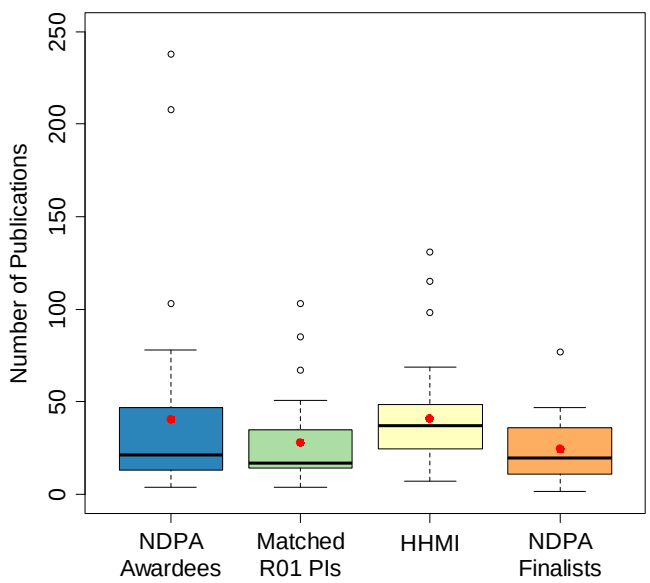
# 5. Impact and Innovation: Researcher-Level Findings

The impact of the matched R01, HHMI, and NDPA finalist researchers was compared to that of Pioneers through six bibliometrics. Additionally, the impact and innovation of HHMI researchers was compared to that of the NDPA and matched R01 grants through expert review.

## A. Bibliometric Findings

### 1. Number of Publications

In the years after award and application, Pioneers were similarly productive to the matched R01 PIs and NDPA finalists, and published fewer research articles than HHMI investigators (K-S, $p$ = 0.58, 0.37, 0.02 respectively) (Figure 38).

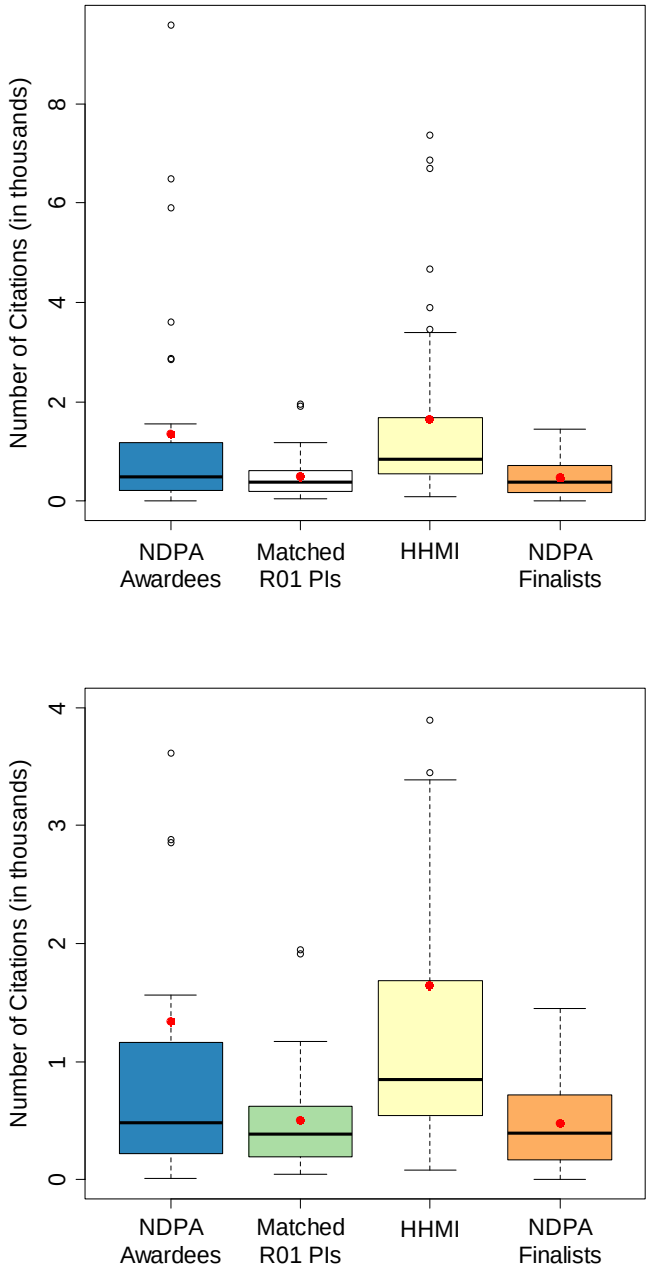



*Note:* One observation equals one PI.

**Figure 38. Researcher-Level Post-Award Publications**

## 2. Number of Citations

Citations to publications from the post-award period observed the same trend among the groups; NDPA awardee publications received similar numbers of citations as the matched R01 PI and NDPA finalist publications, and received fewer citations than HHMI publications (K-S, $p$ = 0.33, 0.16, 0.01 respectively) (Figure 39).

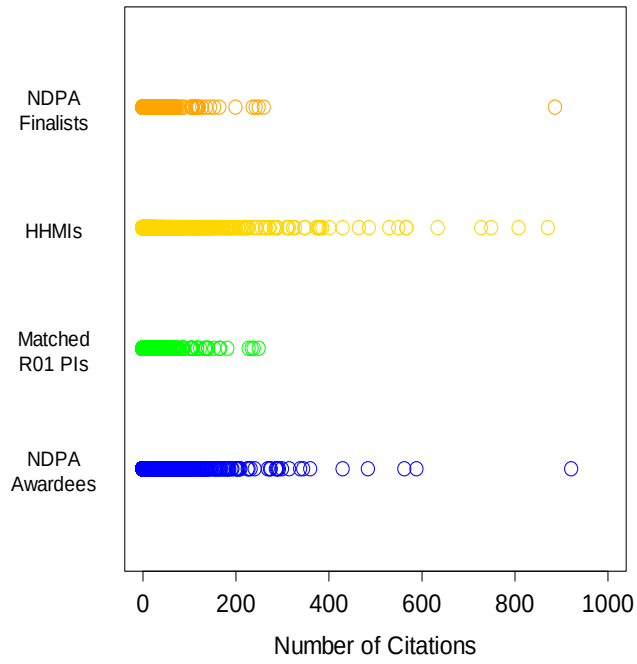



*Note:* One observation equals one PI.

**Figure 39. Researcher-Level Post-Award Citations**

### 3. Number of Citations per Publication

At the publication level, Pioneers had greater citation counts than matched R01 PIs and NDPA finalists, and similar citation counts as HHMI investigators in the post-award period (K-S, $p < 0.001$, $<0.001$, $= 0.12$ respectively) (Figure 40). The most notable similarity between the NDPA and HHMI distributions in comparison to those of the other groups is the distinctive tail of publications with large numbers of citations.

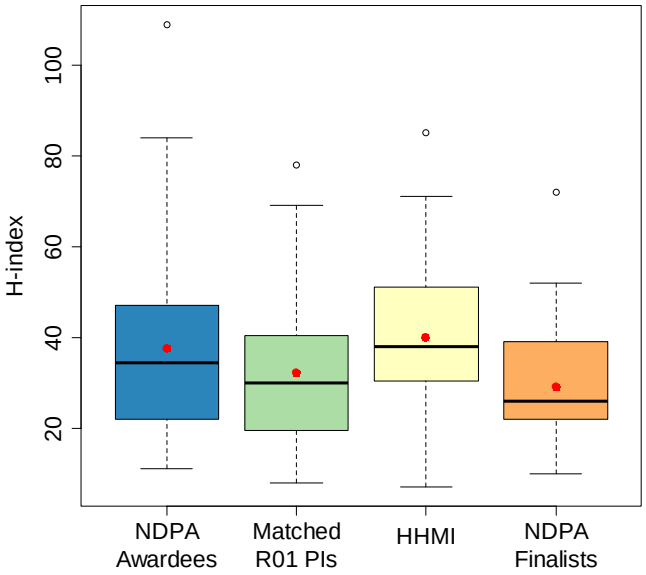



*Note:* One observation equals one publication.

**Figure 40. Researcher-Level Post-Award Citations per Publication**

### 4. H-index

Full career h-index, another proxy for a researcher's impact, showed no significant differences between Pioneers and matched R01 PIs, HHMI investigators, or NDPA finalists (K-S, $p = 0.42$, 0.72, 0.08 respectively) (Figure 41). H-index is cumulative over a researcher's career, so it is a less direct proxy of post-award research impact. Also, recall that there were no significant differences between the h-index values of the groups at time of award.

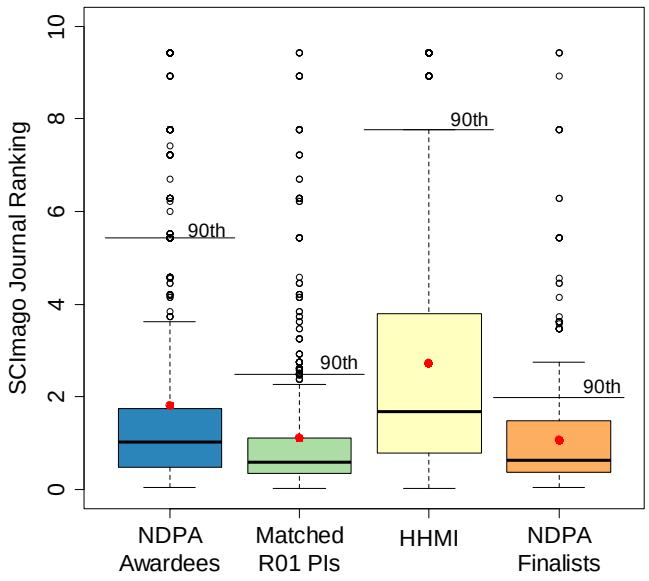



*Note:* One observation equals one PI.

**Figure 41. Researcher-Level H-Index to 2011**

## 5. SCImago Journal Ranking

Journal impact factor was another bibliometric proxy for research impact. NDPA researchers published in journals with higher impact factors than matched R01 PIs and NDPA finalists, and published in journals with lower impact factors than HHMI investigators in the post-award period (K-S, $p < 0.001$ for all comparisons) (Figure 42). The 90th quantiles for HHMI publications and NDPA publications are distinctly different than the 90th quantiles for the other two groups.




*Note:* One observation equals one publication.

**Figure 42. Researcher-Level Post-Award Journal Impact Factors**

## 6. Proportion of Publications by Journal Percentiles

Both the HHMI investigators and the NDPA Pioneers published more frequently than the matched R01 PIs and the NDPA finalists in the top 0.5% of SCImago-indexed journals in the post-award period. The HHMI investigators published in the highest ranking journals. A total of 41% of HHMI investigator publications were published in the top 0.5% of all SCImago-indexed journals. Pioneer publications in the top 0.5% accounted for 25% of the group's total publications, while the matched R01 PIs and NDPA finalists produced 12% and 11% in the top 0.5% of SCImago-indexed journals, respectively (Figure 43).

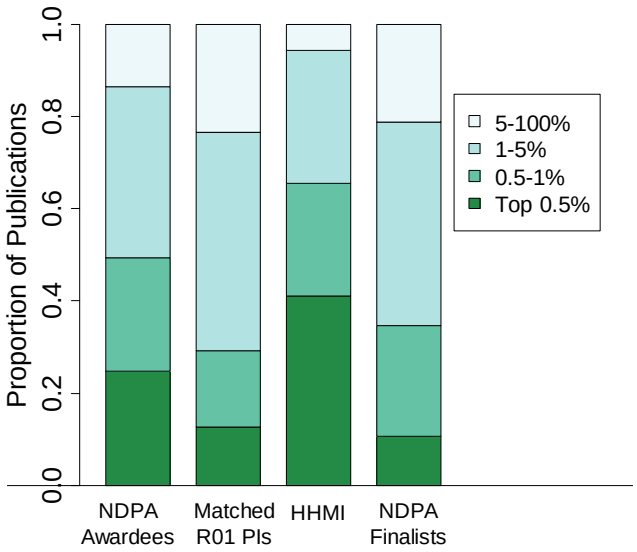



**Figure 43. Researcher-Level Post-Award Proportion of Publications by Journal Percentiles**

## 7. Additional Bibliometric Analyses: Outcomes per Selected Direct Costs

An additional set of analyses were performed on the Pioneer and HHMI investigator groups. These analyses compared the number of publications and citations per selected direct cost awards received by each PI. Due to the challenges of data collection and data availability for funding amounts, there are several limitations to this analysis. For each of the Pioneers and HHMI investigators, we collected information on all non-NDPA, NIH direct cost awards from the year of NDPA/HHMI receipt through the end of 2011, and used estimates for the following values based on information on the groups: NDPA direct cost amounts for the duration of the NDPA funding period, average annual half-salary[37] support for Pioneer awards, overall annual means for HHMI investigator operating and equipment budgets starting the year of NDPA/HHMI receipt through the end of 2011, and average annual salary support for HHMI investigator awards.[38]

Non-NDPA, NIH award amounts were included only if the Pioneer or HHMI investigator was the contact PI on the NIH application. We were not able to collect data for grants on which the PI of interest was a co-PI or otherwise affiliated with the grant.

NDPA direct cost amounts were uniformly set at $500,000 per year for the 5-year designated funding period, although NIH's QVR system reports that some investigators

[37] Fifty percent of the average annual salary support provided for NDPA awards.

[38] Non-NDPA NIH funding was collected from the NIH QVR system. Pioneer award funding and 50% salary amounts were estimated based on discussions with NDPA program staff. HHMI investigator average funding and salary amounts were estimated based on discussions with HHMI program staff.

received less than $500,000 in some years. The annual half-salary amount for Pioneer awards was estimated to be at $100,000. In consequence, each Pioneer was assigned $600,000 per year of NDPA funding.

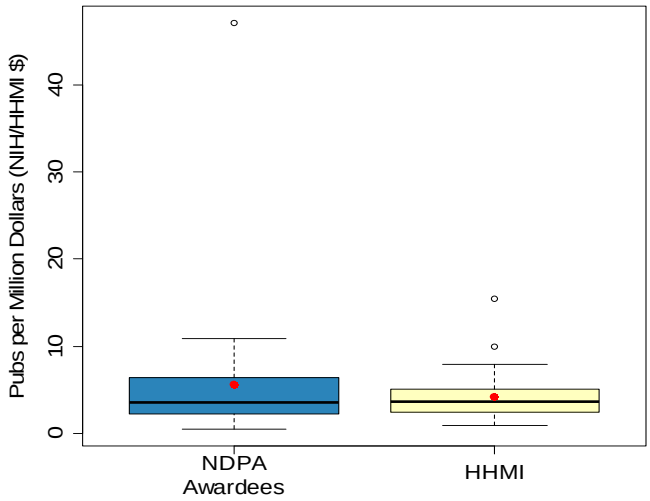
HHMI investigator direct cost awards were unknown for individual PIs. We were provided with the overall annual mean for the 5-year period from 2005 through 2010 for forty of the 2005 HHMI investigators. The average annual operating budget for forty 2005 HHMI investigators from 2005 to 2010 was $580,543. The average annual equipment budget for that same group was $179,169.[39] After adding the annual salary amount for HHMI investigators, approximately $200,000, each individual who received an HHMI investigator award (includes some Pioneers) was rounded to $960,000 per year of HHMI funding. It should be noted that all individuals in our study who received HHMI investigator awards were renewed for HHMI funding in 2010 and the $960,000 per year figure was extended to account for HHMI funding these individuals would have received during their renewal period in 2010 and 2011.

Due to data availability, NIH and HHMI investigator funding were the only sources included in this analysis, but it is known that these researchers have many other sources of funding (e.g., NSF awards, HHMI Early Career Scientist Awards, private foundation support). In addition, we cannot account for PI funding support from research collaborations or lab members.

### a. Number of Publications per Selected Direct Costs

In the years after award, Pioneers and HHMI investigators were similarly productive when accounting for NIH and HHMI investigator funding (K-S, $p = 0.47$) (Figure 44).

[39] HHMI program staff note that there are substantial variations in actual funding among the HHMI investigator awards.

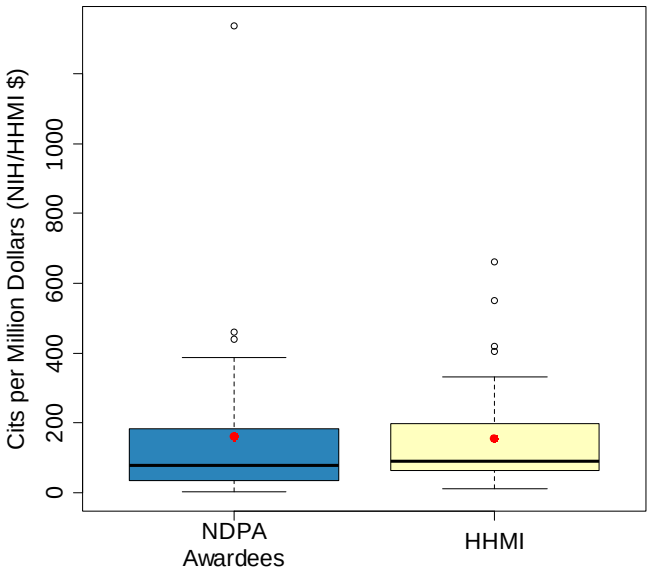



*Note:* One observation equals one PI.

**Figure 44. Pioneer and HHMI Investigator Publications per Selected Direct Costs**

### b. Number of Citations per Selected Direct Costs

Pioneers and HHMI investigators received similar numbers of citations to publications in the post-award period when accounting for NIH and HHMI investigator funding (K-S, $p = 0.32$) (Figure 45).




*Note:* One observation equals one PI.

**Figure 45. Pioneer and HHMI Investigator Citations per Selected Direct Costs**

## 8. Additional Analyses: Text Mining

In addition to bibliometrics and expert review, STPI conducted a text-mining analysis of publications that were attributed to grants across the comparison groups. The goal of this analysis was to determine how dense and similar the topical areas of the grant-attributed publications are. This emerging method of analysis might be able to predict areas of innovation. Areas of low density, for example, might indicate transformative research topics.

Two analyses were conducted on the Pioneers, the HHMI recipients, the matched R01 comparison group, and the random R01 portfolio group:

- A comparison of the differences (i.e., the divergence) between publications in each comparison group and the NIH-SPIRES background (2007–March 2012)
- A comparison of topics among the grant-attributed publications

For the first analysis, the pairwise Kullback-Leibler (KL) divergence between title and abstract corpa were provided by Edmund Talley's group. The divergences were calculated between each grant-attributed publication and the set of funded grant applications and publications in SPIRES (2007–March 2012). The 500 smallest pairwise KL divergences for each grant-attributed publication were aggregated by comparison group and filtered to include only pairwise publication comparisons (e.g., the 500 publications in SPIRES with the smallest divergence from each of the NDPA-attributed publications are grouped). The 10th quantile from each of these "top 500" groups were calculated and compared in order to see how each group differed from the SPIRES background.

is the analysis showed no statistical difference between the 10th quantiles of the KL pairwise divergence of the comparison group publications and SPIRES (Figure 46). This demonstrates, on the whole, that grant-attributed publications[40] have comparable numbers of similar papers (on a contextual basis) in SPIRES. Thus, any observed difference in citations among the comparison groups is due to something other than a difference in publication and citation habits or a paucity of research in an area.

[40] For HHMI, we used only "top five" papers.

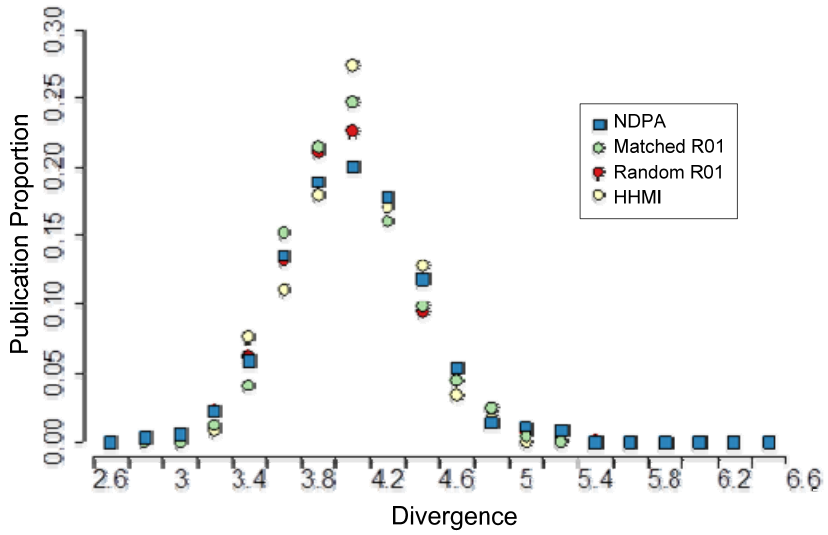
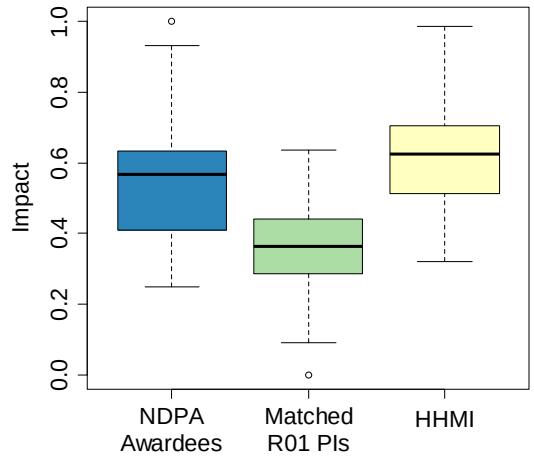



**Figure 46. Difference from SPIRES Publications**

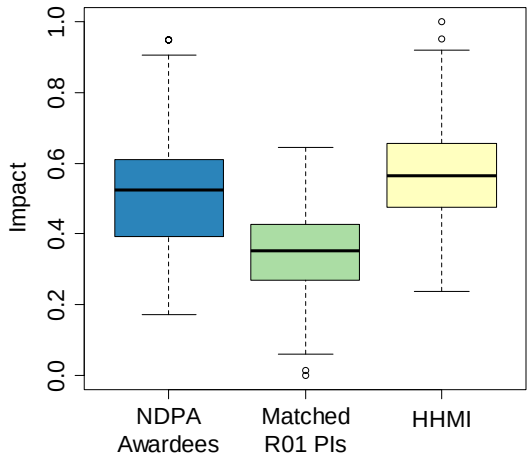
## B. Expert Review Findings

Figure 47 and Figure 48 display box plots of the expert-assessed impact of packets and papers. NDPA and HHMI packets have significantly more impact than the matched R01s (t-test, $p < 0.0001$, K-S, $p = 0.0002$), with no significant difference between NDPA and HHMI (t-test, $p = 0.08$, K-S, $p = 0.17$). However, for papers, HHMI grants have significantly more impact than the NDPAs (t-test, $p = 0.0008$, K-S, $p = 0.004$), which again have significantly more impact than matched R01s (t-test, $p < 0.0001$, K-S, $p < 0.0001$).




*Note:* One observation equals one grant.

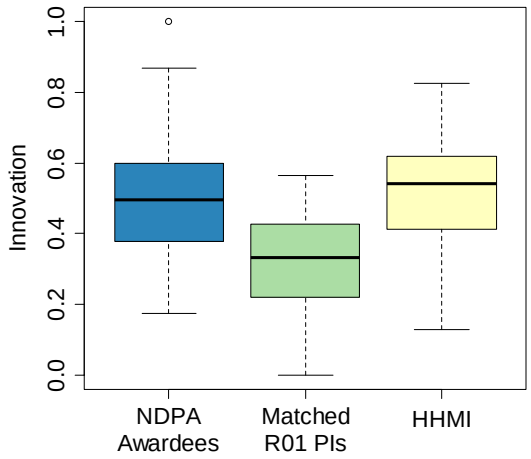
**Figure 47. Expert-Assessed Packet Impact**




*Note:* One observation equals one paper.

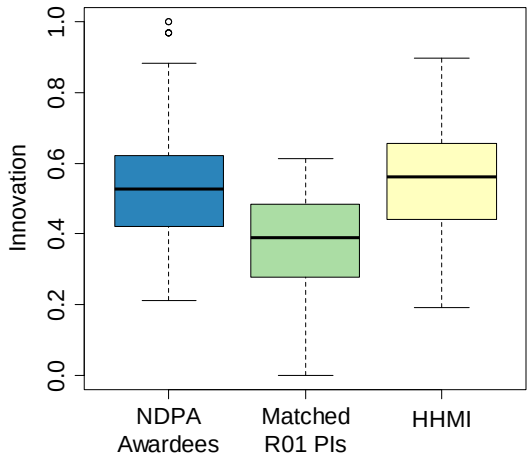
**Figure 48. Expert-Assessed Paper Impact**

Figure 49 and Figure 50 display box plots of the expert-assessed innovation of packets and papers. NDPA and HHMI packets are significantly more innovative than the matched R01s (t-test, $p < 0.0001$, K-S, $p = 0.0005$), with no significant difference between NDPA and HHMI (t-test, $p = 0.71$, K-S, $p = 0.88$). NDPA and HHMI papers are significantly more innovative than the matched R01s (t-test, $p < 0.0001$, K-S, $p < 0.0001$), with no significant difference between NDPA and HHMI (t-test, $p = 0.53$, K-S, $p = 0.14$).




*Note:* One observation equals one grant.

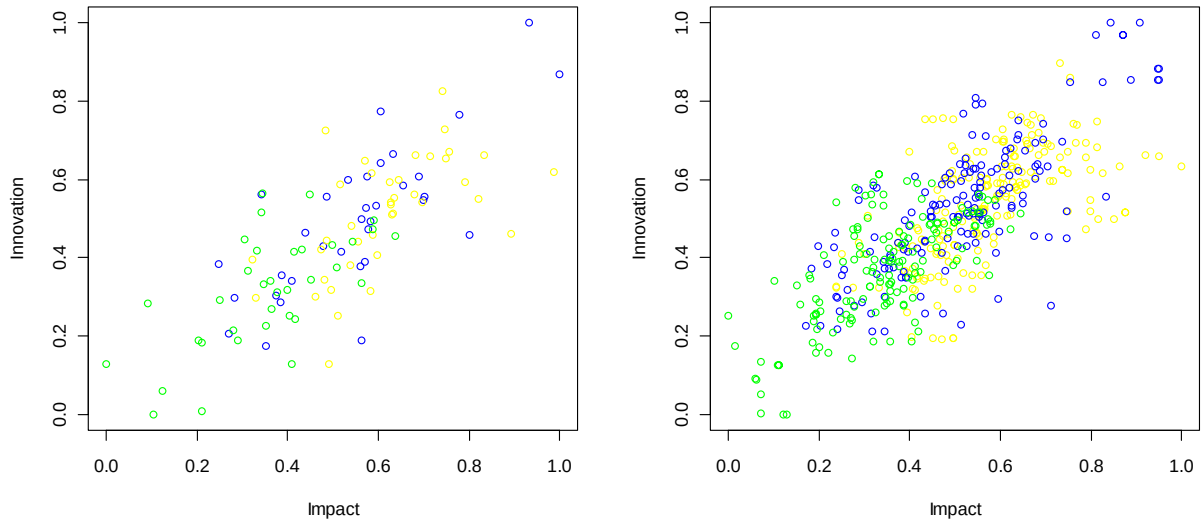
**Figure 49. Expert-Assessed Packet Innovation**




*Note:* One observation equals one paper.

**Figure 50. Expert-Assessed Paper Innovation**

Figure 51 shows a strong correlation between expert-assessed impact and innovation ($\rho = 0.76$ for packets and $\rho = 0.75$ for papers).




**Papers** **Packets**

**Figure 51. Correlation of Impact to Innovation for Papers and Packets**

Figure 52 and Figure 53 plot the estimated ranks packet impact and innovativeness. Note that the matched R01s are not among the top packets in the expert rankings. Figure 54 and Figure 55 plot the estimated ranks of paper impact and innovation.

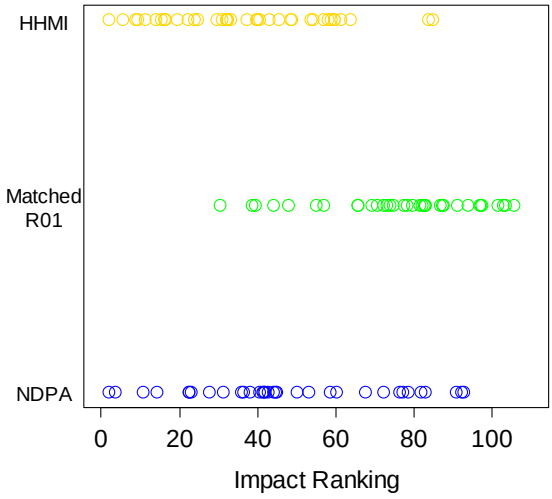



**Figure 52. Expert-Assessed Packet Impact Rankings**

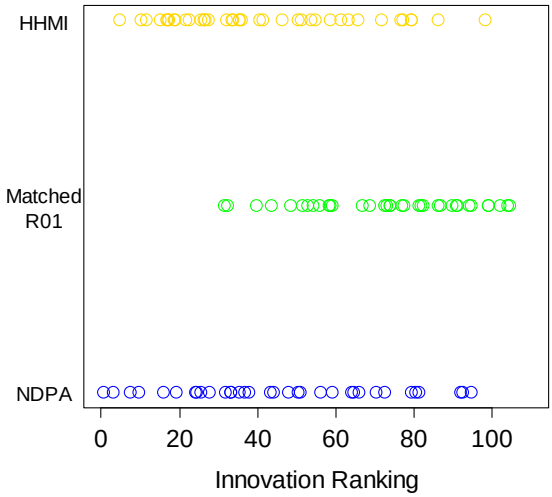



**Figure 53. Expert-Assessed Packet Innovation Rankings**

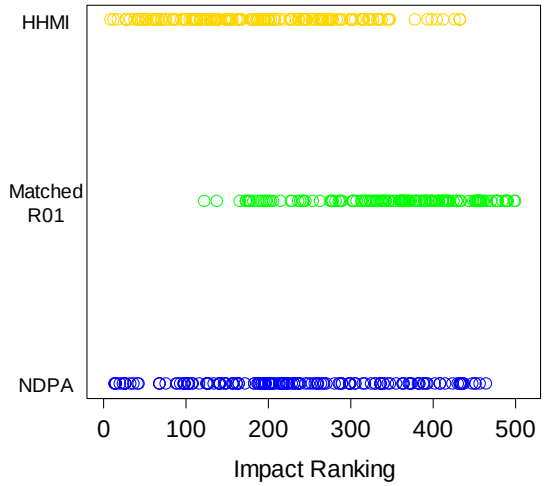



**Figure 54. Expert-Assessed Paper Impact Rankings**

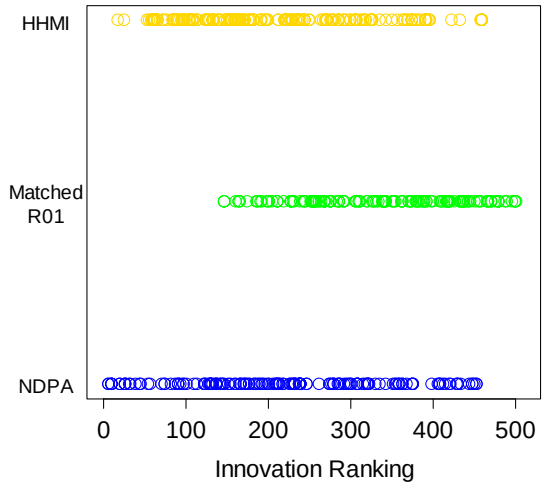



**Figure 55. Expert-Assessed Paper Innovation Rankings**

Additional discussion of the qualitative comments from the expert review can be found in Chapter 6.

## C. Regression Analyses

A regression model linking expert ratings, bibliometric indicators of performance, and information about individual awards was developed to determine if we can identify which variables are predictive of the underlying packet or paper impact as assessed by the experts. We had no expectation that the regression model would be a good surrogate for the expert ratings: it is unlikely that knowing the grant mechanism, the characteristics of the PI, and the journal that published the paper will fully characterize its level of impact

or innovation. Details of the model are discussed in Appendix I, and Table I-1 and Table I-2 summarize the coefficients of the fitted regression models.

The covariates considered were: granting mechanism (NDPA, R01, HHMI), number of pre-award citations for the PI, SCImago Institutions Ranking (SIR) for the PI at time of award, number of pre-award publications for the PI, PI h-index at time of award, years since degree, receipt of early career award, and, for the R01 awards, total direct costs.

For impact of the packet, the covariates that affected the rating (those that are significantly different from zero) are whether a PI received an HHMI or matched R01 award and the PI's pre-award lifetime number of citations. The impact of individual papers was affected positively by the impact factor of the journal in which the paper was published and the total number of citations to the paper. There was a negative impact when the paper was a review paper. Direct costs were included for matched R01 awards, and it is interesting to note that the size of the grant was not a statistically significant predictor of the impact rating of its five papers with the most impact.

With respect to ratings of the innovativeness of approaches, whether a PI got an NDPA or HHMI award was a factor in the rating, as was the PI's citations at award. Innovativeness of the approaches of the individual papers was (as with impact) affected positively by the impact factor of the journal in which it was published, and (again as with impact) there was a negative impact when the paper was a review paper. Again, it is interesting to point out that the size of the matched R01 grant was not a statistically significant predictor of the innovativeness rating of its five papers with the most impact.

# 6. Impact and Innovation: Qualitative Findings from Expert Assessment

As part of the expert assessment, reviewers were asked to provide comments about how they made their assessments of impact and innovation. Out of the 340 reviews our experts completed, 307 contained an open-ended discussion of the specific factors that influenced his or her assessment of impact, and 288 contained discussion of the specific factors influencing the assessment of innovativeness. This chapter discusses the findings from the analysis of those responses, which provide context for the quantitative analysis of the expert ratings.

## A. Assessing Impact

Experts were asked to rate the level of impact in the research on a five-point scale from extreme to no impact. They also responded to the following open-ended question:

> *What about these papers, individually or as a whole, made you choose your answers above?*

The following prompt was given within the survey:

> *NDPA was created to promote highly innovative and potentially transformative approaches that have the potential to produce extremely high impact on a broad area of biomedical or behavioral research. In this context, "extremely high impact" refers to research that accomplishes one or more of the following:*
>
> - *radically changes present understanding of an important existing scientific or engineering concept*
> - *leads to the creation of a new paradigm or field of science or engineering,*
> - *challenges present understanding in the field(s) involved,*
> - *provides pathways to new frontiers,*
> - *challenges conventional wisdom,*
> - *leads to unexpected insights that enable new techniques or methodologies, or*
> - *redefines the boundaries of science or engineering.*

Of the 340 responses, 307 contained a discussion of specific factors influencing the assessment of impact. Five themes, summarized in Table 8. Counts for Major Themes

Used by Experts to Discuss Impact of Packets, were commonly discussed. (Appendix G includes a detailed list of the themes identified and coded.) Table 8. Counts for Major Themes Used by Experts to Discuss Impact of Packets contains the number of responses that mentioned a particular theme broken out by the expert's quantitative ranking of the packet.

**Table 8. Counts for Major Themes Used by Experts to Discuss Impact of Packets**

| Theme | Extreme or Very High Impact (*n* = 168) | Moderate Impact (*n* = 107) | Slight or No Impact (*n* = 35) |
|---|---|---|---|
| Foundational research | 60 | 45 | 16 |
| Elucidation of pathways or mechanisms | 42 | 18 | 1 |
| Translational or clinical potential | 37 | 9 | 3 |
| Changing present understanding | 36 | 24 | 10 |
| Pathways to new frontiers | 28 | 11 | 0 |

"Foundational research" and "changing present understanding" were the general themes that experts used to express either positive or negative assessments of impact. They have large relative counts across all packet ratings.

"Elucidation of pathways or mechanisms" encompassed three primary ideas:

- Research that provided entirely new or important mechanistic insight
- Research that solved long-standing questions in a particular pathway or process and provided future opportunities for moving forward
- Research that definitely identified the role of a specific element in the pathway, which was previously unexplored or not well-understood

In research deemed to have extreme or very high impact, the foundational research theme appeared more often in the discussion of HHMI (28%) and NDPA (26%) packets than in assessments of the matched R01 packets (14%). Examples of experts' comments that fell under this category include:

> *"…Paper #5 was extremely impactful because it revealed a new connection between histone post-translational modifications and DNA replication, which is a largely unexplored but important aspect of eukaryotic biology."* (HHMI)

> *"The major papers in this group address a crucial issue in the development of the nervous system: how are specific connections formed between source and target structures during development. Two of the papers address this issue in the vertebrate retina, where beautiful lamination and sublamination provides an elegant example of specificity*

*of connections. One paper that I regard as extremely impactful shows how inputs segregate on to pyramidal neurons of layer 5 in the cerebral cortex—a major issue in development of the cerebral cortex. Another impactful paper shows the role of other semaphorins in formation of connections in the fruitfly drosophila. Together, they build an insightful picture of the role of a class of molecule, semaphorins, in the development of specific projection systems in the brain."* (HHMI)

*"Studies presented in these papers identified novel signaling components and signaling events in pathways mediating apoptosis, necroptosis and autophage. Two of these studies also identified inhibitors of these signaling pathways, which are potential lead compounds for anti-cancer drug. These findings made significant contributions to the understanding of the mechanisms of cell death and autophage, and advanced the field."* (NDPA)

*"Human TFG is a tumor suppressor or an oncogene in several human cancers but its mechanisms of action remain unclear. This paper demonstrates that the C. elegans homolog TFG-1 acts as both a novel apoptotic suppressor and an activator of cell growth." (R01)*

"Translation or clinical potential" encompassed four primary ideas:

- Research that identifies interventional or therapeutic targets
- Research that closes the gap between basic science and clinical delivery
- Utilization of animal models to demonstrate clinical potential
- Development of previously nonexistent culture systems

In research deemed to have extreme or very high impact, this theme appeared more often in the discussion of matched R01 (31%) and NDPA (26%) packets than in assessments of the HHMI packets (14%). Examples of experts' comments that fell under this category include:

*"The investigators have mapped several regulatory pathways that connect the binding of auto inducers at sites on the surface to functional changes within the cell. For public health a highly important result is the elucidation of the pathway that leads to expression of virulence factors when many Vibrio cholerae are present. They have also studied how the ability of the auto inducers to trigger quorum sensing relates to their chemical structure by synthesizing auto inducers with modified structures. This work has revealed ways in which the effect of the auto inducers could be blocked, suggesting routes to therapies. This work and studies like this are of the utmost significance for the mission of the NIH in the medium and longer term."* (HHMI)

*"The studies in aggregate are highly significant and made important advances in our understanding of the signaling pathways that operate in HSC [hematopoietic stem cells] and leukemia. These findings, although*

*[they] do not radically change our current understanding of the field, have led to significant scientific insights and the identification of pathways that could be manipulated for leukemia treatment."* (R01)

*"Development of a culture system for HPV is of very high impact, allowing therapeutic testing in culture, viral genetics, virus-host interactions, replication and in vitro models of oncogenesis. Very nice."* (R01)

*"This is an extremely impressive body of work. The high level of impact derives not only from the application of cutting edge techniques, several of which are novel and used for the first time, but also from the relevance to human disease. This work has high significance at the levels of basic and translational research and to clinical science. Adding to the impressiveness of this series of papers from the [Daley] lab is that they were all published over only a 3 year period. These papers truly advance understanding of stem cells and how they can be used in innovative ways in medicine."* (NDPA)

*"It has long been a goal to expand hematopoietic stem cells in culture, but the field has been marked by controversy and reports that were not reproducible. In most cases, the precious property of self-renewal has been quickly lost when the cells have been placed in culture. Many investigators have tried co-cultures with a variety of stromal cell types in hopes that one might provide a supportive stem cell niche. If particular endothelial cells work as described in the first of this group of papers, there will be tremendous impact and implications for clinical treatments. This is provided of course that the work is reproducible."* (HHMI)

*"These papers present a picture of the spread and genetics of the virus. They directly impact treatment."* (NDPA)

*"Collectively these papers helped identify key regulatory pathways involved in HSC function and or leukemic stem/progenitor cell progression. The primary impact of these papers is the identification of additional pathways that can be manipulated for therapy. Some of the work provided hard proof of pathway involvement that had been theorized before the work based on gene expression patterns in hematopoiesis/leukemogenesis. Each paper identified unrelated pathways and are therefore key advances in their own right. It is unusual for a single laboratory to make so many independent significant advances."* (HHMI)

"Pathways to new frontiers" was a theme presented to experts as a potential indicator of high impact research and was used when the experts' discussions specifically cited opportunities for future exploration. Specific ideas include:

- The specific methods chosen and research findings clearly define the path for moving forward

- New methods, techniques, or resources (clone libraries, genetic maps, etc.) were developed, especially for areas where there was a critical need
- The research pursued issues that were far ahead of mainstream, in an entirely new frontier
- The research findings have applications not just for a particular research topic, but with broad applicability
- Other researchers have followed suit, and are pursuing this line of investigation

In research deemed to have extreme or very high impact, the pathways to new frontiers theme was discussed in 19% of the NDPA responses, 17% of the matched R01 responses, and 14% of the HHMI responses. Examples of experts' comments that fell under this category include:

> *"This work forms a foundation for the use of viruses as genetically encoded nanoparticles. A nice array of applications is explored - fluorescent imaging, MR imaging, tumor targeting. Quite a few investigators in the field have subsequently jumped on this band wagon, and the exploitation of viruses as nanoparticles is really emerging as a significant thrust."* (R01)

> *"The inventive development and application of large-scale microfluidics and of high-throughput sequencing by the group of Stephen Quake have made these techniques among the most powerful discovery tools of molecular and cellular biology where the throughput is critically needed. The clever design and operation of his signature microfluidics have opened many research opportunities for his group and potentially others as well."* (NDPA)

> *"As a group, these papers are extremely impactful. The "Mapping and sequencing" paper provides pathways to new frontiers by providing the first high-resolution sequence map of human structural variation. The "A burst of segmental duplication" paper addresses a long standing question, "how humans evolved from the great apes." It leads to the creation of a new paradigm that a burst of genome duplication activity (rather than cytogenic rearrangements, single-base-pair changes, or retro-transposon activity) occurred during the period when humans diverged from the great apes. The "diversity of human copy number" paper provides pathways to new frontiers, by making ~ 1000 genes accessible to genetic studies of disease association (e.g., intellectual disability, autism, schizophrenia and epilepsy). The "duplication architecture" paper and the "punctuated cores of human genome evolution" paper, although slightly less impactful in their own right, set the stage for the subsequent extremely impactful papers."* (HHMI)

## B. Assessing Innovation

As part of the research assessment, experts were also asked to rate the level of innovation, specifically for the research approaches, on a five-point scale from extreme to no innovation. They also responded to the following open-ended question:

> *Regardless of its impact, how innovative are the approaches described in this set of papers, taken as a whole?*

The following prompt was given in the survey:

> *NDPA was created to promote highly innovative and potentially transformative approaches that have the potential to produce unusually high impacts on a broad area of biomedical or behavioral research. In this context, an approach may be considered extremely innovative if it accomplishes one or more of the following:*
>
> - *the ideas underlying the research are at odds with prevailing wisdom,*
> - *the research requires the use of equipment or techniques that have not been proven or are considered extraordinarily difficult, or*
> - *the research involves a unique combination of disciplines.*

288 of the 340 responses contained a discussion of specific factors influencing the assessment of innovativeness. Four themes, summarized in Table 9, were commonly discussed. Table 9 contains the number of responses that mentioned a particular theme broken out by the expert's quantitative ranking of the packet.

**Table 9. Counts for Major Themes Used by Experts to Discuss Innovation of Packets**

| Theme | Extreme or Very High Impact (*n* = 130) | Moderate Impact (*n* = 115) | Slight or No Impact (*n* = 43) |
|---|---|---|---|
| Research requires the use of equipment or techniques that have not been proven or considered difficult | 61 | 33 | 2 |
| Creative utilization or improvement of existing techniques | 41 | 23 | 6 |
| Standard methods | 19 | 60 | 28 |
| Changing present understanding | 24 | 27 | 6 |

"Changing present understanding" was a general theme that experts used to express either positive or negative assessments of innovation.

Use of "standard methods" was commonly cited when a paper or packet was not judged to be highly innovative. Examples of expert comments include:

> *"The use of Zebrafish genetics combined with electrophysiological and molecular approaches is not novel. They are well done in these papers, but do not demonstrate a novel or uniquely strong approach. The results have not significantly challenged prevailing wisdom."* (HHMI)

> *"While innovation is often in the eyes of the beholder, these are typical studies that are using standard methods (statistics for the epidemiological studies, clustering methods to trace evolutionary relatedness)."* (NDPA)

> *"While these studies were carefully undertaken, there is little in them in the way of innovation. There are very much examining well-studied areas using very conventional approaches."* (R01)

The "use of equipment or techniques that have not been proven or considered difficult" encompassed three primary ideas:

- Successfully completing technically sophisticated experiments, for which other investigators had previously been unsuccessful
- Research that completed laborious techniques, which was considered to be a "service" to other researchers in the field
- Research that was difficult due to inaccessibility of research materials

In "extremely" or "very" innovative work, difficult technique was mentioned in 55% of HHMI responses, 45% of matched R01s, and 38% of NDPA. Examples of the experts' comments that fell in this category include:

> *"Most of these achievements are not incremental advances achieved by applying safe, proven methods. Instead, they take rather radically new, unproven approaches."* (HHMI)

> *"Creativity was high and innovative, developing a robust culture system for HPV. Important for numerous researchers across the field."* (R01)

> *"This work is extremely novel and exciting. The techniques are very difficult, and prior to this group accomplishing their goals, most would have not considered most of the work in these papers to have been feasible. The work is highly interdisciplinary."* (NDPA)

> *"These manuscripts all make use of cutting-edge technologies that are very difficult. They all include a broad range of approaches to generate data that supports their conclusions."* (HHMI)

The "creative use of existing techniques" encompassed two primary ideas:

- Clever integration of complementary approaches
- Techniques applied in an innovative way to address the specific study question

In "extremely" or "very" innovative work, creative technique was mentioned in 29% of HHMI responses, 33% of matched R01s, and 35% of NDPA. Examples of the experts' comments that fell in this category include:

> *"It may not use any new techniques per se, but the demonstrated application is very powerful and constitutes a new method of fundamental biological research…."* (NDPA)

> *"I feel that the most innovative aspects of this research was the collaborative study on beta-catenin function. With that said, this study really does not represent a technical breakthrough, rather, the innovation aspect stems from clever integration of live cell imaging with biochemical and ultrastructural approaches to discover a completely new function for beta-catenin."* (R01)

> *"As noted in the previous section, each of these papers pinpoints a central biological question and employs excellent methodology to solve the question. The most obviously innovative contributions are the new telomere protection assay and the creation of the ts mutations in TRF2, but the use of live-cell monitoring of telomere movement in the 53BP1 paper was also stunning. Other techniques, such as FUCCI and the Shld1-regulated Pot1 were developed by others, but skillfully exploited."* (NDPA)

> *"These are clever applications of molecular genetics tools and approaches, including mutant fruitflies and mice, to address the role of several members of a class of molecule—the semaphorins—in brain development."* (HHMI)

> *"The techniques used here—flow cytometry, RNAi, gene profiling, for example—are not in themselves innovative. They have been applied to the study of mammalian stem cells and other developmental systems. What is unique and innovative about these studies is the collective application of these techniques to a model system, planaria, to which these techniques have not been systematically applied."* (HHMI)

## C. Summary of Qualitative Findings

When assessing impact, the experts used foundational research and changing present understanding as themes to distinguish work as having more or less impact. Changing present understanding was also the general theme used to distinguish innovative research.

However, additional specific themes emerged as important in describing high-impact work. The first specific theme was elucidation of pathways or mechanisms. This was more commonly cited among HHMI investigators and Pioneers than among the matched R01 PIs. The second theme was translational or clinical potential, which was more commonly cited among the Pioneers and matched R01 PIs than for HHMI

investigators. The third theme was pathway to new frontiers, which was cited similarly among the groups.

For innovation, the specific themes that emerged related to the difficulty and creativity of technique. HHMI investigators were cited for using difficult technique the most often, followed by Pioneers, followed by matched R01 PIs. Creative technique was mentioned in approximately similar proportions across the comparison groups.

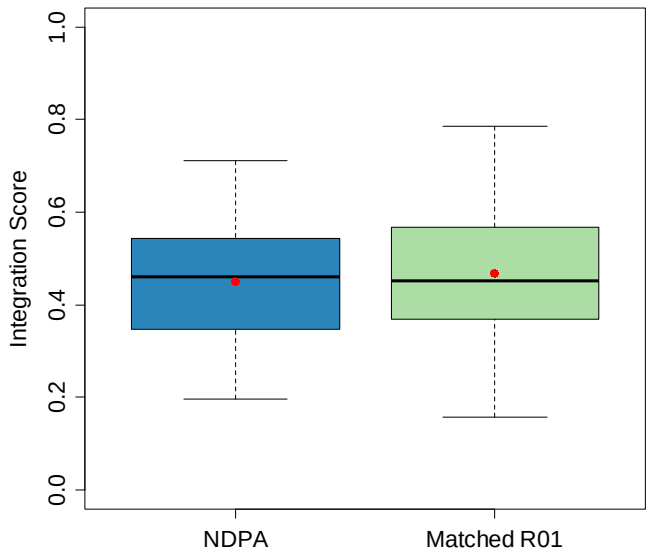
# 7. Interdisciplinarity and Collaborations Findings

The findings to the secondary study questions—around interdisciplinarity and co-authorship patterns—are presented for all comparison groups. This includes the assessment of four bibliometrics for the matched R01 set of grants and three for the HHMI investigators and NDPA finalists.

## A. Interdisciplinarity

### 1. Integration Score

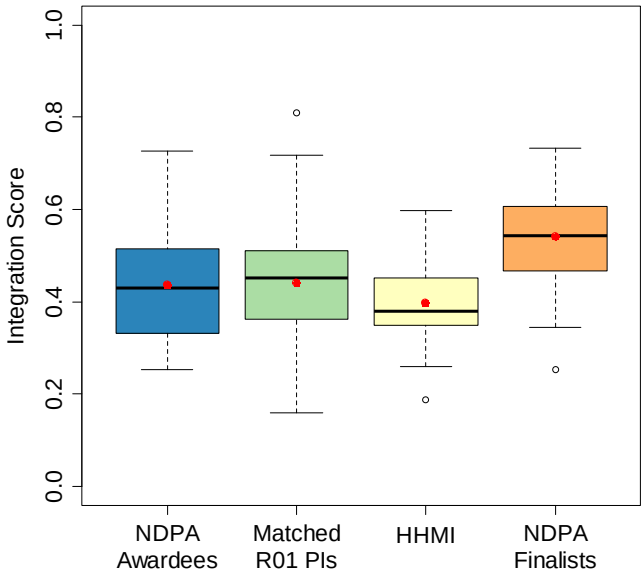
NDPAs and matched R01 grants have similar grant-level integration scores (K-S, $p = 0.82$) (Figure 56). We matched the R01 grants and NDPAs on the topical similarity of their applications, so the similarity in how these grant-attributed publications integrate knowledge suggests that the topics addressed by the grants may indeed be similar.




*Note:* One observation equals one PI.

**Figure 56. Grant-Level Integration Score**

In the period after the award, the researcher-level integration score (Figure 57) for Pioneers, matched R01 PIs, and HHMI investigators are similar (K-S, $p = 0.74$, 0.06). NDPA finalists, however, had higher integration scores than NDPA Pioneers during this time (K-S, $p = 0.01$).




*Note:* One observation equals one PI.

**Figure 57. Researcher-Level Post-Award Integration Score**

## 2. Additional Analyses: Proportion of Subject Area Categories

We conducted additional analyses to quantitatively represent the diversity of research publications by the NDPA and matched R01 grant groups, as reflected in the subject categories Thomas Reuters assigns to the publications the two groups cited in their published work. We organized the 221 subject area assignments from Thomson Reuters into 18 macro-subject area clusters, and used these clusters in two analyses: (1) the proportion of referenced articles in each subject area were normalized such that each NDPA and R01 publication has the same weight (Figure 58), and (2) the proportion of citations in each subject area were normalized such that each researcher has the same weight, regardless of their number of publications (Figure 59). Neither figure suggests strong differences in the subject areas cited as a result of NDPAs and matched R01 grants.

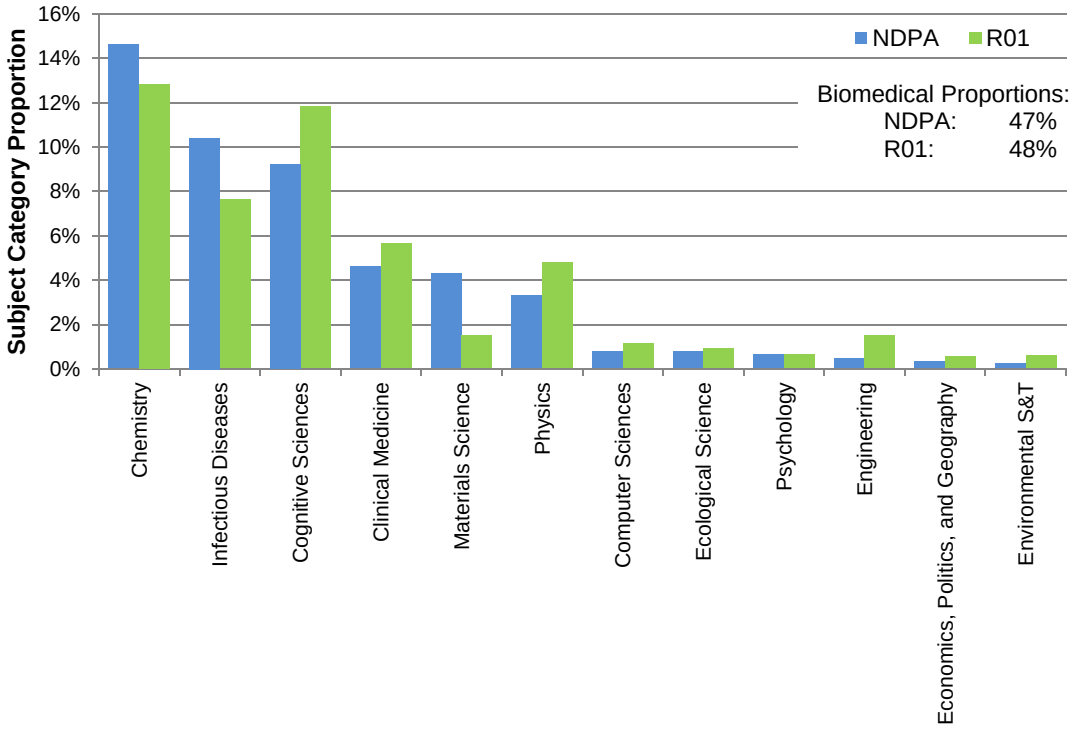



**Figure 58. Subject Categories, Aggregated by Publication**

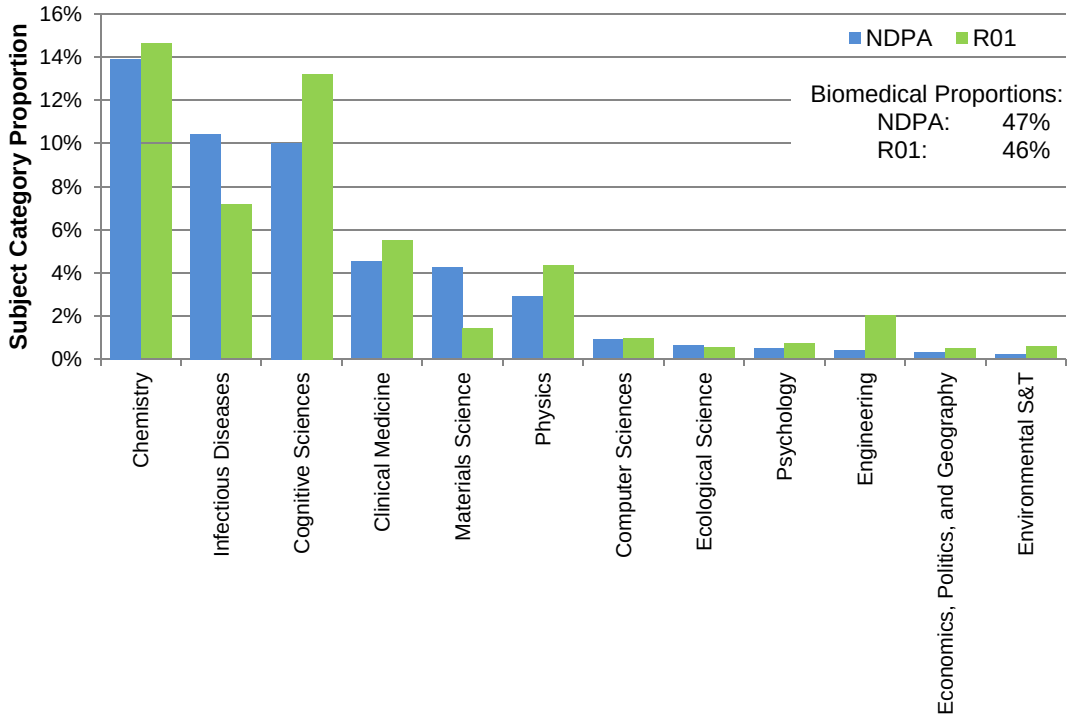


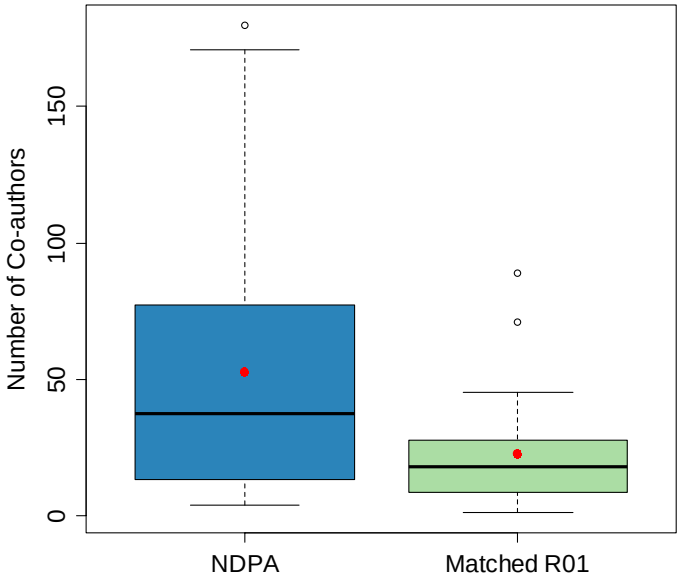

**Figure 59. Subject Categories, Aggregated by PI**

## B. Collaborations

### 1. Number of Co-authors

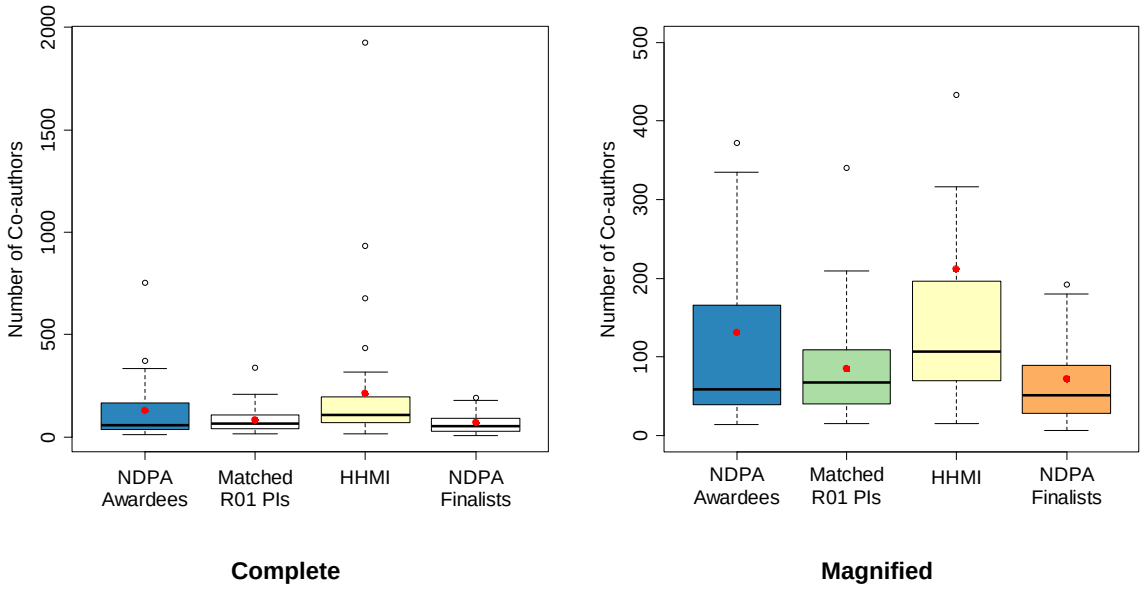
NDPA-attributed publications had greater numbers of authors than matched R01 grant-attributed publications (K-S, $p = 0.01$) (Figure 60).

NDPA Pioneers published with similar numbers of co-authors as matched R01 and NDPA finalist researchers in the post-award period (K-S, $p = 0.34$, 0.24) (Figure 61). HHMI investigators, however, published with a greater number of co-authors (K-S, $p = 0.007$). Since all of the groups published with a similar number of authors prior to the award period, it would appear that HHMI investigators formed a greater number of new collaborations than NDPA Pioneers after receiving the award.




*Note:* One observation equals one grant.

**Figure 60. Grant-Level Unique Co-authors**



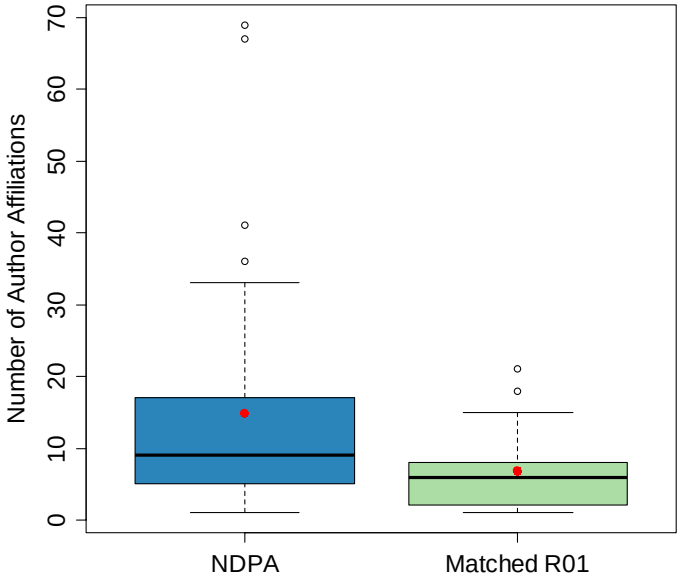

*Note:* One observation equals one PI.

**Figure 61. Researcher-Level Post-Award Unique Co-authors (Complete and Magnified)**

## 2. Number of Co-author Affiliations

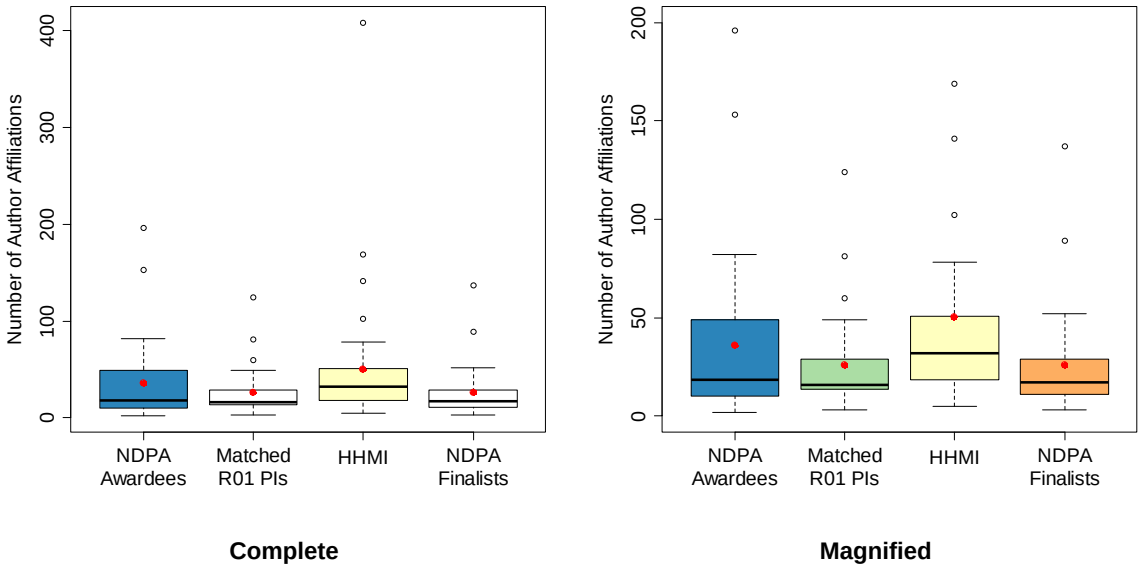
NDPA and matched R01 grant-attributed publications were published by researchers from similar numbers of institutional affiliations (K-S, $p = 0.07$) (Figure 62).

Unique author affiliations follow the same trend as co-authors; NDPA awardee publications have similar numbers of institutional affiliations as matched R01 and NDPA finalist researcher publications and fewer institutional affiliations than HHMI investigator publications in the post-award period (K-S, $p = 0.19$, 0.58, 0.04 respectively) (Figure 63).




*Note:* One observation equals one grant.

**Figure 62. Grant-Level Unique Author Affiliations**




*Note:* One observation equals one PI.

**Figure 63. Researcher-Level Post-Award Unique Author Affiliations (Complete and Magnified)**

# 8. Summary of Findings

This chapter summarizes the findings detailed in Chapters 4–7.

## A. Grant-Level Analyses of Impact and Innovativeness

Grant-level analyses were performed on three sets of grants—Pioneer grants from the first three cohorts (2004-2006), matched R01 grants in the same time period, and grant in 30 randomly selected portfolios of R01 grants (again from the same time period).

Bibliometric analyses indicate that the Pioneer grants resulted in a larger number of published papers (518) than the matched R01 grants (295). NDPA grants were larger than the matched R01 grants, and when considered per dollar, NDPA and matched R01 grants produced similar numbers of publications.

The 30 random R01 portfolios, each valued at $87.5 million, published more papers (median of 530) than the NDPA portfolio (266). When considered per dollar, the NDPA portfolio still had a slightly smaller number of publications than the random R01 portfolios.

Impact was assessed using three bibliometric (citations, journal impact and h-index) and two expert review-based ("packet" and paper level subjective assessment) indicators. Turning first to the number of citations per grant, NDPA awards received a similar number of citations as the matched R01 set of grants. The NDPA and matched R01 portfolios had a comparable number of citations (5,031 and 4,293, respectively) to the random R01 portfolios (median of 5,757), even though the random portfolios contained a greater number of publications.

On a per-publication basis, NDPA and the matched R01 grants have similar distributions. For example, 50% of NDPA and 56% of matched R01 grant-attributed publications received greater than or equal to 13 citations. It is important to note that there were several NDPA-attributed publications with over 400 citations and no matched R01-attributed publications with over 281 citations (none of the papers in the bibliometric analyses were review papers), indicating perhaps that a small number of Pioneers have been very successful.

Both NDPA and matched R01 grants, however, received more citations per publications than were attributed to the random portfolios. For example, 10% of NDPA and 7% of matched R01 publications, as opposed to only 3% of unique random R01 portfolio publications, had accumulated greater than 55 citations. While the Pioneers

tended to publish later in the grant period than the matched R01 PIs, their papers, once published, were cited more quickly.

Looking at the statistics on journal prestige, we found that the Pioneers' NDPA-attributed publications appeared in journals with higher impact factors than matched R01s. The Pioneers also had a larger fraction of their NDPA-funded publications in top journals than the other two groups—half of NDPA grant-attributed publications appeared in the top 1% of all science journals, compared to under one-third (32%) of the matched R01 publications appeared in the top 1%. Only 7% of the random R01 portfolio publications were in the top 0.5% of journals and 21% were in the top 1% of journals.

The h-index, an indicator that integrates both publication and citation counts into one metric, was greater for the NDPA-attributed publication portfolio than that of the matched R01s. For instance, 25% of NDPA grants and 15% of matched R01 grants had an h-index value greater than or equal to 9.

Expert assessment of impact (conducted on the five highest-impact publications of Pioneers and matched R01 publications) reflected the bibliometric analyses, finding NDPAs to have uniformly more impact than the matched R01 awards, at the level of both the packet and the individual papers.

Two methods were used to assess the innovativeness of research approaches in the grant-attributed publications. Expert assessment found NDPA "top five" publications to have research methods that were more innovative than the matched R01s, both at the "packet" and paper levels. Qualitative analysis of expert comments, summarized below, illustrates some systematic differences between the two groups. Text mining analysis did not find the NDPA research areas to be more "distant" from the NIH SPIRES background than the matched R01s.

## B. Researcher-Level Analyses of Impact and Innovativeness

Researcher-level analyses were performed on four sets of researchers—Pioneers and NDPA finalists from the first three cohorts, matched R01 grantees in the same time period, and HHMI investigators from 2005. The researcher level analysis was valuable in two regards. By including publications funded by a larger set of grants, going beyond just the Pioneer award, it provided a fuller picture of the accomplishments of Pioneers and the matched set of PIs. More importantly, it allowed for comparisons with HHMI investigators and NDPA finalists, which are groups that do not have specific grants outputs for comparison.

Comparisons with the researchers in the matched R01 set and finalist groups show the Pioneers to be similar to both groups with respect to the number of publications. Pioneers, however, had higher citations counts at both the PI and publication levels (for instance, 10% of Pioneer, 3% of matched R01 PI, and 3% of finalist post-award

publications had greater than or equal to 86 citations by the end of 2011). Pioneers also had higher journal impact factors, and a higher fraction of publications in journals with the most impact. In fact, twice as many (as a proportion of the total) Pioneer publications were in the top 0.5 % of all SCImago-indexed journals than was found for the other two groups. Pioneers were similar to both groups with respect to h-indices For instance, 50% of Pioneers, 43% of matched R01 PIs, and 27% of finalists had an h-index value greater than or equal to 34.5 at the end of 2011).

Comparisons with the HHMI investigators are more nuanced. HHMI investigators outperform Pioneers with respect to the number of publications. The HHMI investigators from 2005 produced a median of 37 papers, as compared with a median of 21.5 for the Pioneers.

When looking at the total numbers, HHMI investigators have a greater number of citations (25% of Pioneers and 38% of HHMI investigators had received greater than or equal to 1,157 citations to post-award publications by the end of 2011). However, when examined on a per *NIH and HHMI dollar* basis (see the detailed caveats in Chapter 5), the two groups produce a comparable number of citations. The two groups also have a similar number of citations *per publication* (50% of NDPA and 52% of HHMI post-award publications had greater than or equal to 15 citations by the end of 2011). One additional similarity between the NDPA and HHMI publications is the distinctively long tail of publications with large numbers of citations (10% of NDPA and 12% of HHMI post-award publications had greater than or equal to 86 citations by the end of 2011).

HHMI investigators publish in higher impact journals, and have twice the proportion of their publications in the top 0.5% of the journals as the Pioneers. Experts found the individual publications of the HHMI investigators to have more impact than those of the Pioneers, although there was no difference between the groups in the assessment of the body of work represented by the “top five” publications. There was no difference between the Pioneers and HHMI investigators with respect to career h-index (50% of Pioneers and 54% of HHMI investigators had h-index values greater than or equal to 34.5).

Expert assessments included assessment of HHMI Investigators as well (but not the NDPA finalists, because there was no single attributable “body of work” that could be assessed). With respect to innovativeness of research approaches, both in their numeric and qualitative assessment, experts found Pioneer publications to have research methods that were similarly innovative as the HHMI investigators, both at the packet and paper levels.

## C. Qualitative Findings from Expert Assessment

Out of the 340 reviews experts completed of the three sets of researchers (Pioneers, Matched R01 grantees and HHMI Investigators), 307 contained an open-ended discussion of the specific factors that influenced his or her assessment of impact, and 288 contained discussion of the specific factors influencing the assessment of innovativeness.

When assessing impact, the general themes used by the experts to distinguish work with more or less impact were how foundational the research was and how much it changed present understanding. The first specific scientific theme was elucidation of pathways or mechanisms. This was more commonly cited among HHMI investigators and Pioneers than among the matched R01s. The second theme was translational or clinical potential, which was more commonly cited among the NDPA and matched R01 PIs than for HHMI investigators. The third theme was pathways to new frontiers, which was cited similarly among the groups.

For innovativeness, the specific themes that emerged related to the difficulty and creativity of technique. HHMI investigators were cited for using difficult technique the most often, followed by NDPA, followed by matched R01. Creative technique was mentioned in approximately similar proportions across the comparison groups.

## D. Regression Analysis

A regression model linking expert ratings, bibliometric indicators of performance, and information about individual awards, was developed to determine which variables were predictive of the underlying packet or paper impact as assessed by the experts. For impact *of the packet*, there were two covariates that affected the rating: whether a PI received an HHMI or matched R01 award, and the PI's pre-award number of citations.

Impact of the *individual papers* was affected positively by the impact factor of the journal in which it was published, and the total number of citations to the paper, and there was a negative impact when the paper was a review article. Direct costs were included in the analysis for matched R01 awards, and it is interesting to note that the size of the grant was not a statistically significant predictor of the impact rating of the packet.

With respect to ratings of the *innovativeness of approaches*, whether a PI got an NDPA or HHMI was a factor in the rating, as was the PI's citations at award. Innovativeness of the approaches of the individual papers was (as with impact) affected positively by the impact factor of the journal in which it was published, and (again as with impact) there was a negative impact when the paper was a review paper. Again, it is notable that the size of the matched R01 grant was not a statistically significant predictor of the innovativeness rating of the packet.

## E. Analyses on Interdisciplinarity and Collaborations

Pioneers' grant-attributed and researcher-level publications were similarly interdisciplinary as all other groups other than the finalists (50% of Pioneers, 51% of matched R01 PIs, 33% of HHMI investigators, as compared to 83% of finalists, had post-award integration scores greater than 0.43). With respect to the subject areas integrated into the research, there appeared to be no strong differences in the areas cited by the Pioneer publications as compared with the matched R01s.

With respect to the number of co-authors on post-award publications, the Pioneers had a similar distribution to those of the matched R01 PIs (at the researcher level) and NDPA finalists, and fewer co-authors compared with HHMI investigators (50% of Pioneers, 57% of matched R01 PIs, and 43% of finalists, as compared to 85% of HHMI investigators, had greater than or equal to 60 unique co-authors). The Pioneers had a greater number of co-authors compared with the matched R01 PIs (grant level) (for instance, 50% of NDPA and 17% of matched R01 grant-attributed publications had greater than or equal to 38 unique co-authors).

With respect to institutional collaborations, the number of collaborating institutions for the Pioneers was similar to all groups except HHMI, which had a larger number of institutional collaborations (for instance, 50% of Pioneers, 43% of matched R01 PIs, and 47% of finalists, in contrast with 74% of HHMIs, had greater than or equal to 19 unique institutional collaborations).

## F. Conclusion

As the Sections above explain, a bibliometric and expert-review-based evaluation of the Pioneer program compared with other four programs of interest found several differences across the groups (note limitations of the analysis in Section 3C). We found, for example, that on a range of impact indicators, the Pioneers scored as well as or higher than the matched R01 PIs, both at the grant and researcher levels. On the same impact indicators, at a portfolio level too, the Pioneer portfolio scored as well as or higher than the random R01 portfolios. At the researcher level, the Pioneers scored similar to or higher than the NDPA finalists. Lastly, we found that the Pioneers had comparable performance on some impact indicators and lower on others as compared with HHMI investigators. In this section, we briefly consider the potential reasons behind some of the differences.

Addressing the first finding, that the Pioneers scored as well as or higher than the matched R01 PIs, we note that the Pioneers received more funding than the matched R01s (Figure 3) through a grant mechanism intended to provide more flexibility and to fund riskier ideas. The two groups were well matched on pre-award PI characteristics (Figure 7 through Figure 14). By construction, the NDPA and matched R01 grants worked in similar areas of science. **The differences between the matched R01 grants**

**and NDPA are likely not attributable to PI differences or differences in research area, but may be due to differences in funding or program characteristics (such as increased flexibility).**

The Pioneer portfolio also scored as well as or higher than the random R01 portfolios on the number of citations and journal impact factor. The portfolios were constructed to have similar direct cost amounts (Figure 5). The groups were otherwise randomly selected; they had different PI characteristics (Figure 7 and Figure 8), the grant mechanisms had different levels of funding and risk, and the areas of science were not controlled. **The differences between the random R01 portfolios and the NDPA portfolio are likely not attributable to funding, but may be due to differences in PI characteristics, research area, or program characteristics.**

The Pioneers did not outperform the HHMI investigators. While both programs provide flexibility and aim to fund riskier ideas, NDPA provided less funding than HHMI (how much less has been difficult to assess). In addition, there were differences in pre-award PI characteristics (Figure 7 through Figure 14), and the researchers may have been working in different areas of science. **The differences between HHMI and NDPA are likely not attributable to flexibility of research, or riskiness of ideas, but may be due to funding, differences in PIs, or differences in areas of science.**

The Pioneer also scored as well as or higher than the NDPA finalists. While the finalists were similar to the Pioneers with respect to PI characteristics (Figure 7 through Figure 14), we do not know enough about the funding profile of the finalists to comment on these similarities and differences.

As a final observation, it is interesting to note that the matched R01 portfolio performed better than the random R01 portfolios across bibliometric measures of impact. Matched R01s had more citations per publication than the random portfolios (Figure 34 and Figure 35), and the publications appeared in journals with higher impact factors (Figure 36 and Figure 37). While these groups shared a funding mechanism, the matched R01 portfolio received less funding, there are potential differences in the areas of science, and there were differences in PI characteristics—the matched R01s are elite researchers (Figure 7 and Figure 8). **The differences between matched R01 grants and the random R01 portfolios are likely not attributable to grant mechanism or funding, but may be due to PI characteristics or differences in research area.**

# Appendix A
# List of Illustrations

## List of Figures

Figure 1. Notional Manifestation of Two Types of Differences in Performance .....3
Figure 2. Distribution of Direct Costs for Type 1 R01 Grants, FY 2004–2006 .....9
Figure 3. Matched R01 Funding Compared to NDPA Funding .....10
Figure 4. Map of NDPA Research Areas within the NIH Landscape .....12
Figure 5. Distribution of Total Direct Cost for the 30 R01 Portfolios.....16
Figure 6. Distribution of Number of R01 Grants per Portfolio Compared with NDPAs ..16
Figure 7. Years since Degree at Time of Award .....20
Figure 8. SCImago Institutions Rankings at Time of Award .....20
Figure 9. Number of Publications at Time of Award .....21
Figure 10. Number of Citations at Time of Award.....21
Figure 11. H-index at Time of Award .....22
Figure 12. Integration Score at Time of Award.....22
Figure 13. Number of Unique Co-authors at Time of Award.....23
Figure 14. Number of Unique Co-author Affiliations at Time of Award.....23
Figure 15. Grant-Level Publications.....44
Figure 16. Grant-Level Publications per Million Dollars in Direct Costs.....44
Figure 17. Grant-Level Citations .....45
Figure 18. Grant-Level Citations per Million Dollars in Direct Costs .....45
Figure 19. Grant-Level Citations per Publication—Histogram.....46
Figure 20. Grant-Level Citations per Publication.....47
Figure 21. Lag time between Publication and Award Year.....48
Figure 22. Lag Time between Citation and Publication Year .....48
Figure 23. Grant-Level H-Index .....49
Figure 24. Grant-Level Journal Impact Factors.....50
Figure 25. Grant-Level Proportion of Publications by Journal Percentiles .....50
Figure 26. Expert-Assessed Packet Impact.....51
Figure 27. Expert-Assessed Paper Impact .....51
Figure 28. Expert-Assessed Packet Innovation .....52
Figure 29. Expert-Assessed Paper Innovation.....52
Figure 30. Random R01 Portfolio Grants and Publications .....53

Figure 31. Number of Publications Divided by Total Direct Costs of Portfolio .....54
Figure 32. R01 Portfolio Citations.....54
Figure 33. Number of Citations Divided by Total Direct Costs of Portfolio .....55
Figure 34. Density Estimations of the Distributions of Citations per Unique Publication for the NDPA, Matched R01, and Random R01 Portfolios .....56
Figure 35. Distribution of R01 Portfolio Citations per Publication.....57
Figure 36. Journal Impact Factors for NDPA, Matched R01, and Random R01 Portfolios.....58
Figure 37. Proportion of Publications by Journal Percentiles for NDPA, Matched R01, and Random R01 Portfolios.....59
Figure 38. Researcher-Level Post-Award Publications.....61
Figure 39. Researcher-Level Post-Award Citations .....62
Figure 40. Researcher-Level Post-Award Citations per Publication.....63
Figure 41. Researcher-Level H-Index to 2011 .....64
Figure 42. Researcher-Level Post-Award Journal Impact Factors.....65
Figure 43. Researcher-Level Post-Award Proportion of Publications by Journal Percentiles .....66
Figure 44. Pioneer and HHMI Investigator Publications per Selected Direct Costs.....68
Figure 45. Pioneer and HHMI Investigator Citations per Selected Direct Costs .....68
Figure 46. Difference from SPIRES Publications .....70
Figure 47. Expert-Assessed Packet Impact.....70
Figure 48. Expert-Assessed Paper Impact .....71
Figure 49. Expert-Assessed Packet Innovation .....71
Figure 50. Expert-Assessed Paper Innovation.....72
Figure 51. Correlation of Impact to Innovation for Papers and Packets.....72
Figure 52. Expert-Assessed Packet Impact Rankings.....73
Figure 53. Expert-Assessed Packet Innovation Rankings .....73
Figure 54. Expert-Assessed Paper Impact Rankings .....74
Figure 55. Expert-Assessed Paper Innovation Rankings.....74
Figure 56. Grant-Level Integration Score.....87
Figure 57. Researcher-Level Post-Award Integration Score .....88
Figure 58. Subject Categories, Aggregated by Publication .....89
Figure 59. Subject Categories, Aggregated by PI.....89
Figure 60. Grant-Level Unique Co-authors.....90
Figure 61. Researcher-Level Post-Award Unique Co-authors .....91
Figure 62. Grant-Level Unique Author Affiliations.....92
Figure 63. Researcher-Level Post-Award Unique Author Affiliations .....92

## List of Tables

Table 1. R01 Grants Selected for Comparison Group .....11
Table 2. Research and Clinical Degrees .....14
Table 3. Summary of the Five Comparison Groups .....24
Table 4. Summary of the Differences across the Comparison Groups.....25
Table 5. Bibliometric Analyses for Groups and Metrics Calculated.....29
Table 6. Metrics used in Bibliometric Analysis by Category.....31
Table 7. Spread in Values at Each Reported Quantile of Comparison among Groups .....58
Table 8. Counts for Major Themes Used by Experts to Discuss Impact of Packets .....78
Table 9. Counts for Major Themes Used by Experts to Discuss Innovation of Packets ...82

# Appendix B
# Potential Comparison Groups

We considered several samples and populations as potential comparison groups in our evaluation. Based on the availability of data and interest of the sponsor, we limited our pool to include primarily NIH funding mechanisms and groups. HHMI investigators were the only non-NIH mechanism that we considered, and they were included after much assistance from the HHMI. Table B-1 presents information on a select number of groups that were considered.

**Table B-1. Potential Comparison Groups**

| Potential Comparison Group | Description | Advantages | Disadvantages | Study Question | In Study? |
|---|---|---|---|---|---|
| Matched R01 Grantees (PI Characteristics) | R01 grantees matched on PI characteristics within similar research areas | Controls for PI-related characteristics that may affect outcome | Cannot control for award size | How do NDPA awardee outcomes compare to those of PIs with equally pioneering potential? | X |
| Random R01 Portfolios | R01 grantees with a portfolio direct-cost comparable to that of the NDPA portfolio | Controls for the portfolio award sizes that may affect outcome | Portfolios contain different numbers of grants; does not control for PI characteristics | How do NDPA awardee outcomes compare to those of R01 portfolios that have received similar amounts of funding? | X |

| Potential Comparison Group | Description | Advantages | Disadvantages | Study Question | In Study? |
|---|---|---|---|---|---|
| HHMI | 2005 Howard Hughes Medical Institute Investigators | High-prestige program that funds high-risk, high-reward research in a way that is similar to NDPA in many aspects; reputation for innovative investigators | Does not explicitly control for PI characteristics or award size | Does longer term, greater funding, and more flexibility result in more outcomes with greater impact than NDPA? | X |
| NDPA Finalists | Individuals who were invited to interview, but were not awarded an NDPA during FY 2004–2006 | Examines the outcomes of PIs who are "almost as exceptionally creative," and capable of producing high-impact outcomes | Variable post-application funding amounts | How do NDPA awardee outcomes compare to PIs of equally pioneering potential? | X |
| Random R01 Grantees | Randomly chosen Type 1 R01 grantees | Allows for comparisons to average PIs | Cannot attribute differences to population differences or funding mechanism | How do the Pioneers' outcomes compare to those of the average R01 grantee? | |
| MERIT (Method to Extend Research in Time) Awardees | PIs with superior research competence and productivity selected by program staff for long-term grant support; Less than 5% of NIH-funded investigators.* | Indirectly selects for PI productivity, PI character-istics, and possibly innovative-ness | MERIT awardees have a long track record with NIH funding on their existing projects and NDPA projects are more likely to be new ideas without preliminary data | How do the Pioneers' outcomes compare to the most productive NIH researchers? | |

| Potential Comparison Group | Description | Advantages | Disadvantages | Study Question | In Study? |
|---|---|---|---|---|---|
| R01 Grantees designated as high-performing by IC leadership | Selected for superior competence and outstanding productivity | May control for productivity and PI potential | Does not control for additional PI characteristics or award size | How do the Pioneers' outcomes compare to the "most reputable" R01 grantees? | |
| Matched R01 Grants (Size of Award) | R01 grants with similar direct costs | Controls for the impact of the size of award on outcomes | Cannot control for PI characteristics or similar research areas due to characteristics of the R01 population | How do the Pioneers' outcomes compare to those from comparable sized R01 grants? | |

* https://dcb.nci.nih.gov/Pages/MERITAwardees.aspx.

# Appendix C
# Researcher Profiles

This appendix provides basic information on each of the researchers across four groups—NDPA Pioneers, matched R01 PIs, HHMI investigators, and NDPA finalists. STPI collected the information from the NIH's internal database, QVR; the HHMI website and program leadership; and publicly available sources. Table C-1 contains the following fields: analysis group, researcher name, year of award or application receipt, grant number, institutional affiliation at time of award, year of doctoral degree attainment, doctoral degrees attained, and current lab website (as of June 2012).

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

This page has been redacted due to the personally identifiable information contained in this table.

# Appendix D
# Expert Reviewer Characteristics

STPI collected data to characterize the experts who participated in the expert review. These data were collected from researcher websites and self-reported by the expert reviewers in the expert review protocol. In addition, these data were supplemented by information gathered from the NIH's internal database, QVR. We present information on the following: age, gender, degree type, institutional affiliation, and job title.

## Age and Gender

The average age of the expert reviewer population is 52 years, and the minimum and maximum ages are 38 years and 72 years, respectively (Figure D-1). Of the 93 expert reviewers, 72 (77%) were male and 21 (23%) were female.

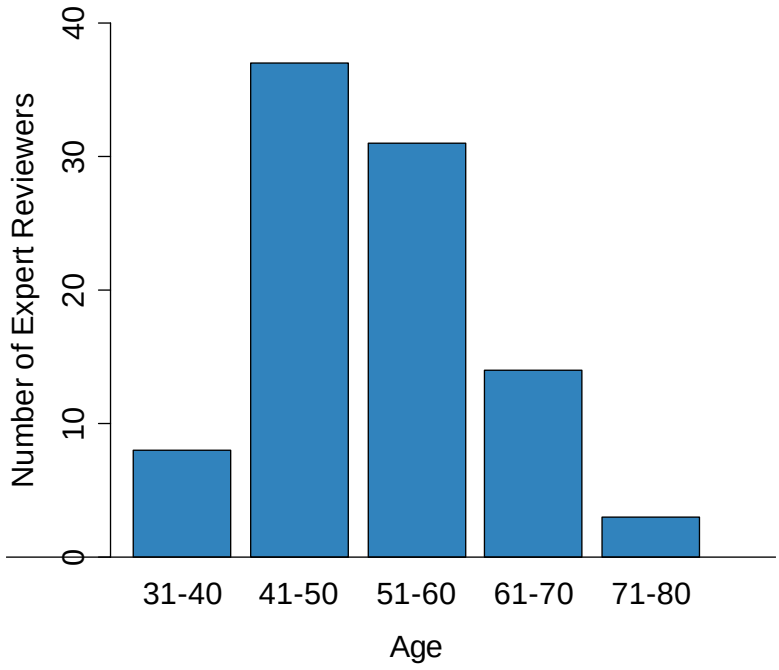



**Figure D-1. Age Distribution of the Expert Reviewers**

## Degree Type

STPI requested that the expert reviewers report the highest degrees they have attained. All the expert reviewers had received at least one type of doctoral degree. Out of 93 expert reviewers, 92 (99%) had received a research doctorate (PhD or equivalent). In addition, six (6%) had received a medical doctorate (MD or equivalent), and four (4%) had received a doctor of veterinary medicine (DVM or equivalent).

## Institutional Affiliations

Table D-1 lists the institutional affiliations of the 93 expert reviewers.

**Table D-1. Institutional Affiliations of the Expert Reviewers**

| Institution | Count | Institution | Count |
|---|---|---|---|
| Harvard University | 5 | La Jolla Institute for Allergy and Immunology | 1 |
| Emory University | 3 | Markey Cancer Center | 1 |
| The Scripps Research Institute | 3 | New York University | 1 |
| University of Pittsburgh | 3 | North Carolina State University Raleigh | 1 |
| University of Washington | 3 | Oklahoma Medical Research Foundation | 1 |
| Washington University | 3 | Princeton University | 1 |
| Carnegie-Mellon University | 2 | Rensselaer Polytechnic Institute | 1 |
| Johns Hopkins University | 2 | Rice University | 1 |
| Massachusetts Institute of Technology | 2 | Rutgers The State University of New Jersey New Brunswick | 1 |
| Northwestern University | 2 | Sanford-Burnham Medical Research Institute | 1 |
| Stanford University | 2 | Seattle Biomedical Research Institute | 1 |
| State University New York Stony Brook | 2 | St. Jude Children's Research Hospital | 1 |
| University of California Berkeley | 2 | Texas A&M University | 1 |
| University of California Davis | 2 | University California San Francisco-Gladstone Institute | 1 |
| University of California Irvine | 2 | University of Arizona | 1 |
| University of California San Francisco | 2 | University of California Los Angeles | 1 |
| University of Iowa | 2 | University of California San Diego | 1 |
| University of Pennsylvania | 2 | University of Chicago | 1 |
| University of Wisconsin Madison | 2 | University of Florida | 1 |
| Yale University | 2 | University of Maryland College Park | 1 |
| Baylor College of Medicine | 1 | University of Massachusetts Medical School Worcester | 1 |
| Beth Israel Deaconess Medical Center | 1 | University of Michigan Ann Arbor | 1 |

| Institution | Count | Institution | Count |
|---|---|---|---|
| Brandeis University | 1 | University of Minnesota Twin Cities | 1 |
| Brigham Young University | 1 | University of North Carolina Chapel Hill | 1 |
| Brown University | 1 | University of Oregon | 1 |
| California Institute of Technology | 1 | University of Texas at Austin | 1 |
| City University of New York | 1 | University of Texas MD Anderson Cancer Center | 1 |
| Columbia University | 1 | University of Texas Southwestern Medical Center Dallas | 1 |
| Cornell University | 1 | Vanderbilt University | 1 |
| Duke University | 1 | Virginia Polytechnic Institute and State University | 1 |
| George Mason University | 1 | Weill Cornell Medical College | 1 |
| Georgia Institute of Technology | 1 | Wellesley College | 1 |
| Indiana University School of Medicine | 1 | Yeshiva University Albert Einstein College of Medicine | 1 |

*Note:* One expert reviewer was counted twice because the individual held a joint appointment to Harvard and Beth Israel Deaconess Medical Center.

## Job Title

Job title information was collected for the 93 expert reviewers (Table D-2). A majority of the recruited experts are tenure-track faculty members at universities. Five of the 93 (5%) expert reviewers were not affiliated with universities, but held leadership positions at medical or research institutes. Seven of the 93 expert reviewers (8%) are chairs in their departments (Table D-3).

**Table D-2. Academic Rank or Position of the Expert Reviewers**

| Position or Title | Number of Reviewers |
|---|---|
| Distinguished Professor | 11 |
| Professor | 46 |
| Associate Professor | 21 |
| Assistant Professor | 7 |
| Adjunct Professor | 1 |
| Adjunct Associate Professor | 1 |
| Associate Research Professor | 1 |
| President and Chief Scientific Officer* | 1 |
| President* | 1 |
| Vice-President of Research* | 1 |
| Associate Member* | 2 |

* These positions were affiliated with medical or research institutes.

**Table D-3. Chair Positions for the Expert Reviewers**

| Chair Position | Number of Reviewers |
|---|---|
| Chair | 3 |
| Vice-Chair | 3 |
| Associate Chair | 1 |

# Appendix E
# Expert Review Protocol

The expert review protocol reproduced herein included an introduction and non-disclosure agreement, instructions, three sections of questions for each set of papers (Impact of Research, Innovativeness of the Approaches, and Research Experience), and three personal information questions.

**Introduction**

Dear Dr. {LastName},

The IDA Science and Technology Policy Institute (STPI) is conducting an expert review for the National Institutes of Health (NIH). The information collected in this review will be used to assess the impact and innovativeness of research funded by NIH's Director's Pioneer Award (NDPA), NIH R01 award, and the Howard Hughes Medical Institute (HHMI) Investigators program.

*Your Participation:*

We selected you for participation in this review based on your expertise in biomedical research. We have determined that you have no known conflicts of interest with any of the authors you have been assigned to evaluate. If you believe you have a conflict of interest, please contact Amy Richards, who may be reached at arichard@ida.org or (202) 419-3731.

We expect the review to take about eight hours. Please use only the papers provided to you to answer the questions, and do not do outside research.

*Due Date and Compensation:*

Please finish the expert review by {DueDate}. STPI will process your compensation within four weeks of receiving your full set of responses.

*Financial Information:*

You will be paid an honorarium of $750 by check for your participation following the completion of all questions in the review.

*Confidentiality:*

Unless required by law, only IDA Science and Technology Policy Institute (STPI) staff will have access to your responses to this online questionnaire. STPI staff are required to maintain confidentiality regarding your identity. Results of this review may be used for research, publications, or presentations; however, if your individual results are discussed, your identity will be protected by using a code number rather than your name or other identifying information. Review responses will be collected and downloaded from secure, web-based software, and stored on our secure systems, and all responses will be password-protected.

*Your Rights:*

Your participation is completely voluntary, and you are free to withdraw from the review at any time. However, the honorarium would be paid only upon completion of all questions in the review.

*Contact Information:*

Direct any questions about this expert review to Amy Richards at STPI, arichard@ida.org or (202) 419-3731.

*Non-Disclosure Agreement:*

All information that I receive from the IDA Science and Technology Policy Institute (STPI) shall be deemed proprietary information. I understand that until either (a) the information is made public or (b) STPI grants me specific written approval, I will, both during the review and after the review, treat the information as confidential, not use the information for any purposes other than to answer the questions below, and not disclose the information to a third party.

☐ ☐ I agree. (You will be directed to the INSTRUCTIONS page)

**[If selected "I agree," skip to Section 2. Instructions.]**

☐ ☐ I do not agree. (You will EXIT the website, and no longer participate in the study)

**[If selected "I do not agree," display the question.]**

If you marked I do not agree in error, please select PREVIOUS to change your selection to I agree and proceed with the NIH review.

To exit the review and no longer participate, select below and SUBMIT.

☐ ☐ I would not like to participate in the NIH expert review.

**[EXIT SURVEY]**

**Instructions**

You have been provided with {Total Packets} sets of papers, each set representing a grant or funding mechanism. Please review each set of papers, and then answer questions related to:

- the impact of each paper, and the set of papers, taken as a whole
- the innovativeness of the approaches of each paper, and the set of papers, taken as a whole
- the alignment between the research in each set of papers and your own research area.

At the end of the review, you will be asked three questions about yourself.

To navigate the instrument please use the PREVIOUS and NEXT buttons and not press the browser's back and forward buttons. If you have cookies enabled on your computer, your responses will be saved allowing you to pause, if needed, or continue on the same computer.

The questionnaire has been tested on Internet Explorer, Chrome and Firefox browsers. If you are using a different browser or mobile device or find navigating difficult, please let us know, and we will email you a copy of the questionnaire and acrobat files of the papers.

Please press the NEXT button when you are ready to proceed to the first set of papers.

**Expert Review Questions**

**Impact of Research**

NDPA was created to promote highly innovative and potentially transformative approaches that have the potential to produce extremely high impact on a broad area of biomedical or behavioral research. In this context, "extremely high impact" refers to research that accomplishes one or more of the following:

- radically changes present understanding of an important existing scientific or engineering concept
- leads to the creation of a new paradigm or field of science or engineering,
- challenges present understanding in the field(s) involved,
- provides pathways to new frontiers,
- challenges conventional wisdom,
- leads to unexpected insights that enable new techniques or methodologies, or

- redefines the boundaries of science or engineering.

*1. How would you categorize the impact of the research in each of the following papers?*

Choose one answer for each paper. Click on the paper's title to download and, if needed, print the paper.

> **Note:** You will not be able to proceed to the next page of questions unless you have responded to this question.

| | Extremely Impactful | Very Impactful | Moderately Impactful | Slightly Impactful | Not At All Impactful |
|---|---|---|---|---|---|
| [Title of Paper 1 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 2 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 3 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 4 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 5 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |

*2. How impactful is the research described in this set of papers, taken as a whole?*

> **Note:** You will not be able to proceed to the next page of questions unless you have responded to this question.

☐ Extremely Impactful

☐ Very Impactful

☐ Moderately Impactful

☐ Slightly Impactful

☐ Not At All Impactful

*3. What about these papers, individually or as a whole, made you choose your answers above?*

There are no character limits in the space provided and you may expand the text box to better view your response.

[open-ended text entry]

**Innovativeness of the Approaches**

NDPA was created to promote highly innovative and potentially transformative approaches that have the potential to produce unusually high impacts on a broad area of biomedical or behavioral research. In this context, an approach may be considered extremely innovative if it accomplishes one or more of the following:

- the ideas underlying the research are at odds with prevailing wisdom,
- the research requires the use of equipment or techniques that have not been proven or are considered extraordinarily difficult, or
- the research involves a unique combination of disciplines.

*4. Regardless of the impact, how would you categorize the innovativeness of the approaches taken in the research described in each of the following papers?*

Choose one answer for each paper. Click on the paper's title to download and, if needed, print the paper.

**Note:** You will not be able to proceed to the next page of questions unless you have responded to this question.

| | Extremely Innovative | Very Innovative | Moderately Innovative | Slightly Innovative | Not At All Innovative |
|---|---|---|---|---|---|
| [Title of Paper 1 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 2 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 3 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 4 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |
| [Title of Paper 5 hyperlinked to actual paper] | ☐ | ☐ | ☐ | ☐ | ☐ |

*5. Regardless of its impact, how innovative are the approaches described in this set of papers, taken as a whole?*

**Note:** You will not be able to proceed to the next page of questions unless you have responded to this question.

☐ Extremely Innovative

☐ Very Innovative

☐ Moderately Innovative

- ☐ Slightly Innovative
- ☐ Not At All Innovative

*6. What about these papers, individually or as a whole, made you choose your answers above?*

There are no character limits in the space provided and you may expand the text box to better view your response.

[open-ended text entry]

**Research Experience**

*7. How aligned is your own research expertise with that of the research presented in the papers as a whole?*

**Note:** You will not be able to proceed to the next page of questions unless you have responded to this question.

- ☐ Extremely Aligned
- ☐ Somewhat Aligned
- ☐ Not Very Aligned
- ☐ Not At All Aligned

You have completed a review of the first set of ${e://Field/Total%20Packets} sets of papers.

At this time, if you need to review your answers, please select PREVIOUS. Once you proceed, you will not be able to return to this section.

To continue, select NEXT.

**[Repeat questions in Section 3. Expert Review Questions and Section 4. Research Experience for each of {Total Packets}. There is a maximum of four Total Packets per expert.]**

**Personal Information**

This information will be used only to contextualize your responses in the review.

**[Questions 8–10 responses are optional.]**

*8. What is your year of birth? {YYYY}*

*9. What is your gender?*

- ○ Male
- ○ Female

*10. What are the highest degrees that you have obtained? Check all that apply.*

- ☐ PhD or equivalent
- ☐ MD or equivalent
- ☐ DDS or equivalent
- ☐ DVM or equivalent
- ☐ Other (please specify): [open-ended text entry]

You have now completed all responses to the NIH expert review. Thank you for your time and input.

Select the SUBMIT button to record your responses.

If you wish to revise your responses to the questionnaire, please contact Amy Richards (arichard@ida.org, 202 419-3731) to resend you the link.

**[EXIT SURVEY]**

# Appendix F
# Expert Review Data

The expert reviews were completed by all 94 experts that agreed to participate in the expert review. The paper- and packet-level ratings for each group were distributed evenly between the impact and innovativeness questions (1,587 paper-level ratings and 336 packet-level ratings for each of the two sections).

Of the 3,174 paper-level ratings, the NDPA Pioneers, matched R01 PIs, and HHMI investigators received 1,024, 978, and 1,172 ratings, respectively.

Of the 672 packet-level ratings, the NDPA Pioneers, matched R01 PIs, and HHMI investigators received 216, 220, and 236 ratings, respectively.

Figure F-1 and Figure F-2 present the paper-level rating distributions, and Figure F-3 and Figure F-4 show the packet-level rating distributions for each of the three groups. The data are represented without the statistical corrections that STPI used to test for significant differences among the groups.

Figure F-5, Figure F-6, and Figure F-7 show the distributions of rating proportions for each of the researchers in the three groups for impact paper ratings, and Figure F-8, Figure F-9, and Figure F-10 show these distributions for innovation paper ratings. Figure F-11, Figure F-12, and Figure F-13 show rating proportions for each of the researchers in the three groups for impact packet ratings, and Figure F-14, Figure F-15, and Figure F-16 show these proportions for innovation packet ratings.

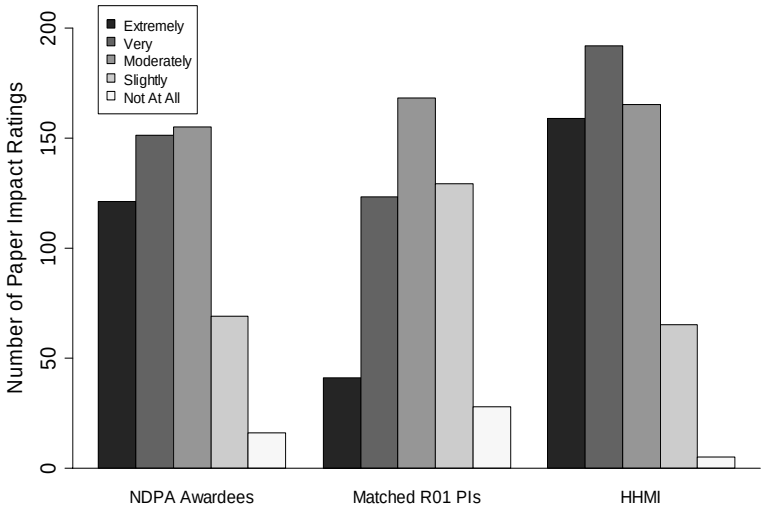



**Figure F-1. Impact Distributions for Paper-Level Ratings**

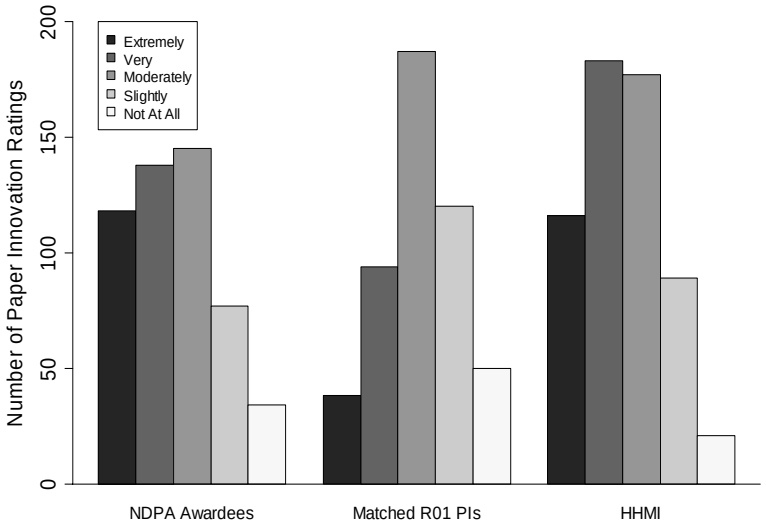



**Figure F-2. Innovation Distributions for Paper-Level Ratings**

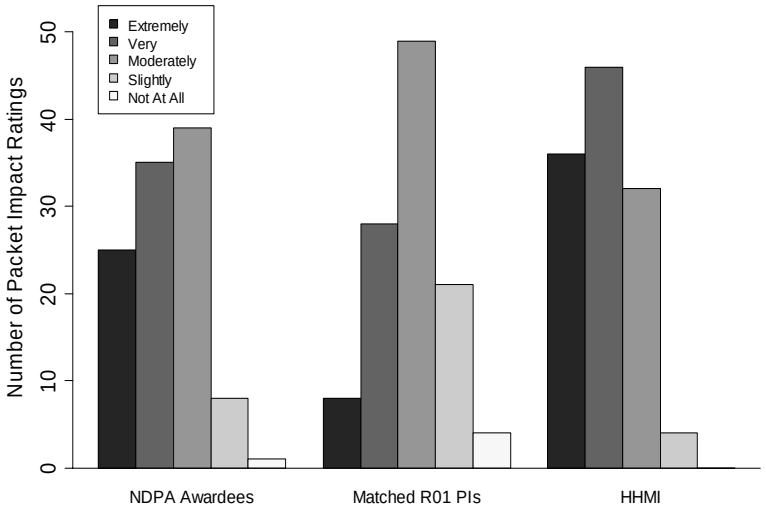



**Figure F-3. Impact Distributions for Packet-Level Ratings**

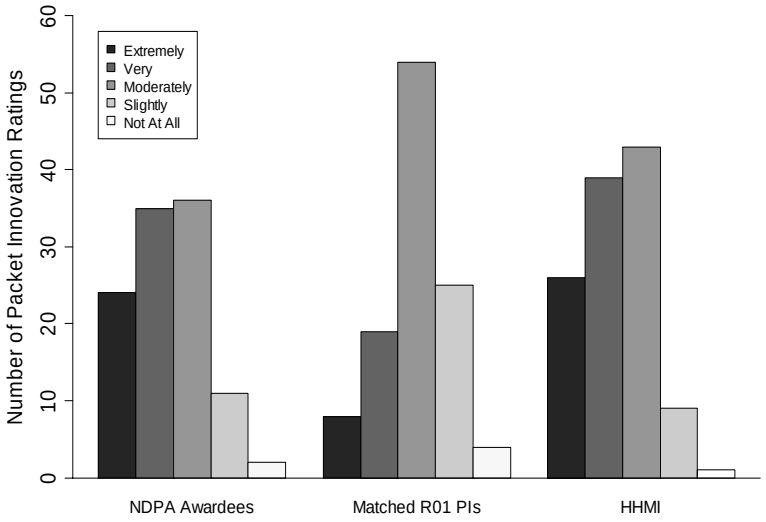



**Figure F-4. Innovation Distributions for Packet-Level Ratings**

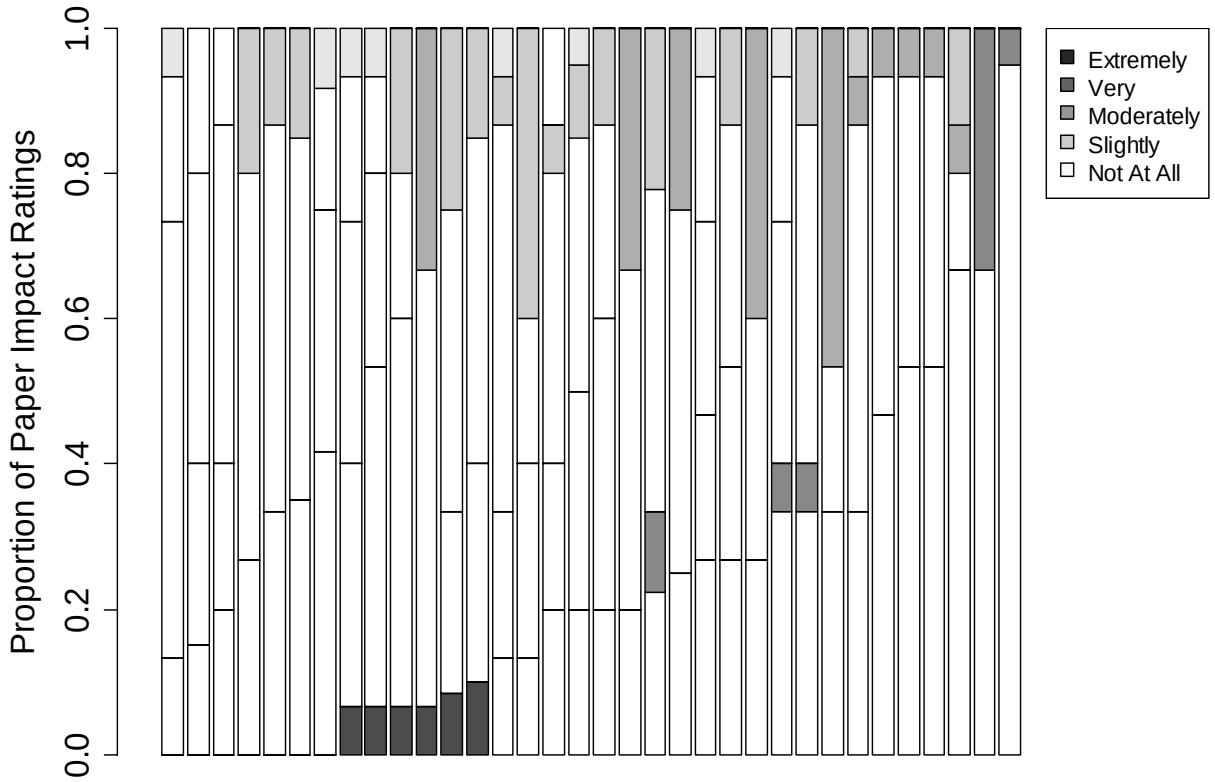



**Figure F-5. Impact Distributions by Researcher for Paper-Level Rating Proportions for NDPA Pioneers**

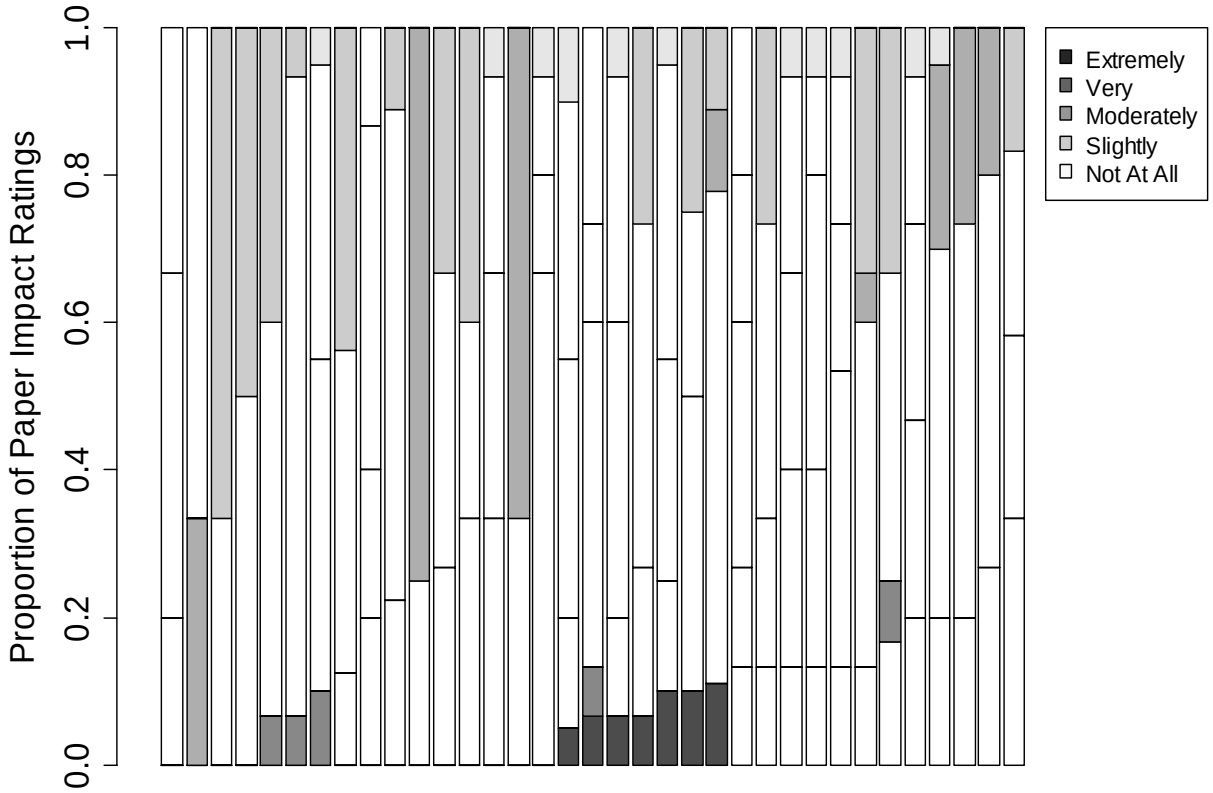



**Figure F-6. Impact Distributions by Researcher for Paper-Level Rating Proportions for Matched R01 PIs**

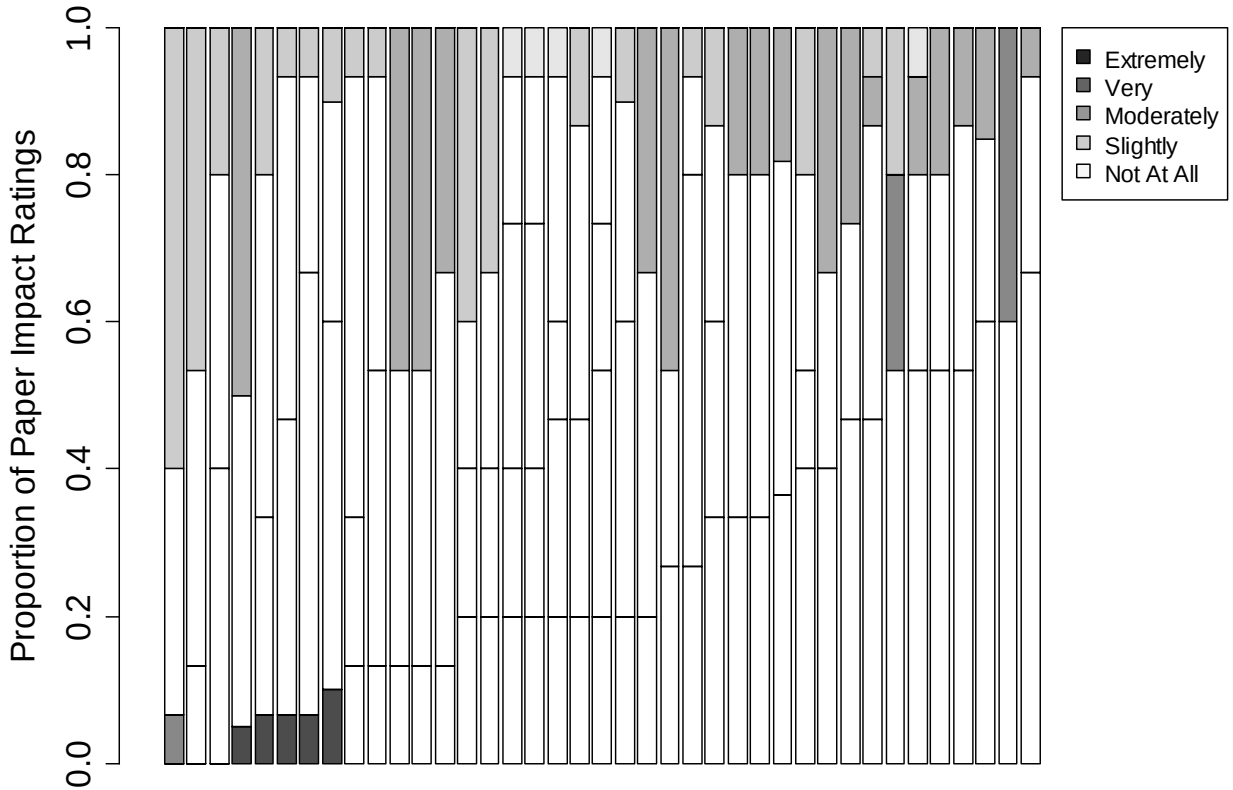



**Figure F-7. Impact Distributions by Researcher for Paper-Level Rating Proportions for HHMI Investigators**

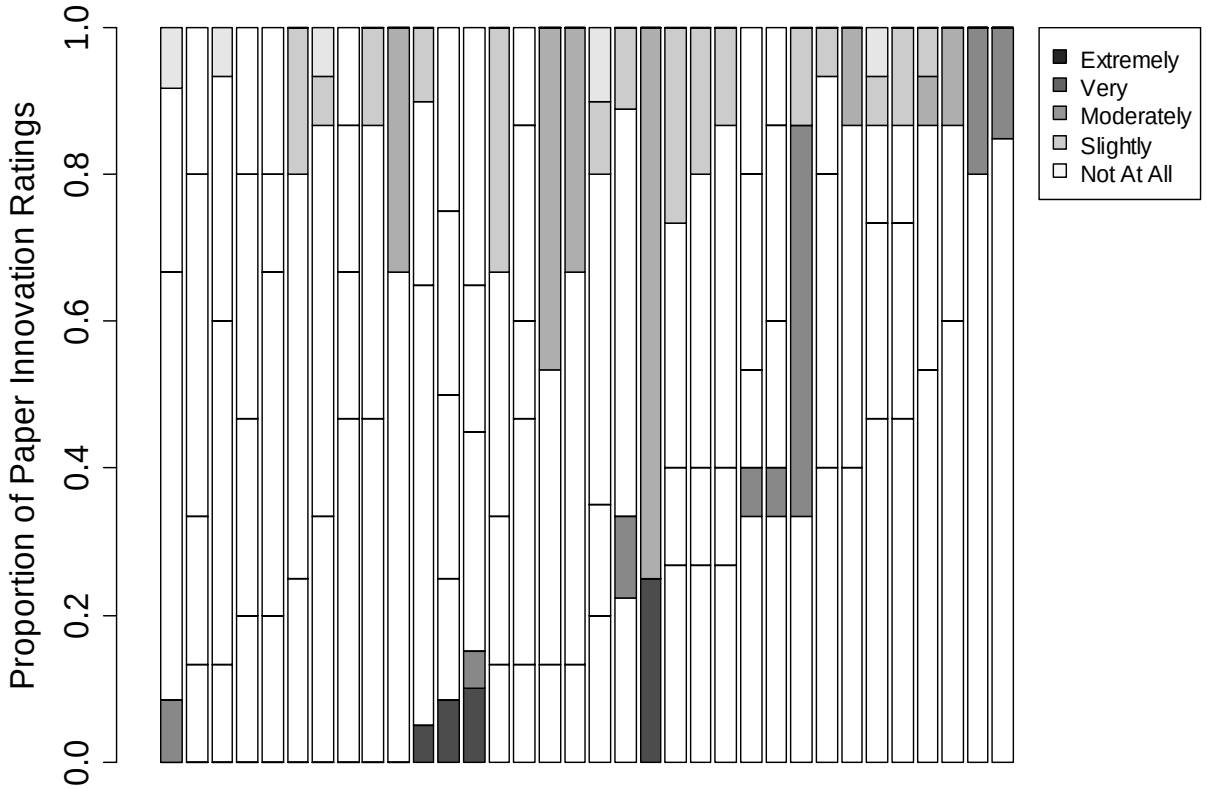



**Figure F-8. Innovation Distributions by Researcher for Paper-Level Rating Proportions for NDPA Pioneers**

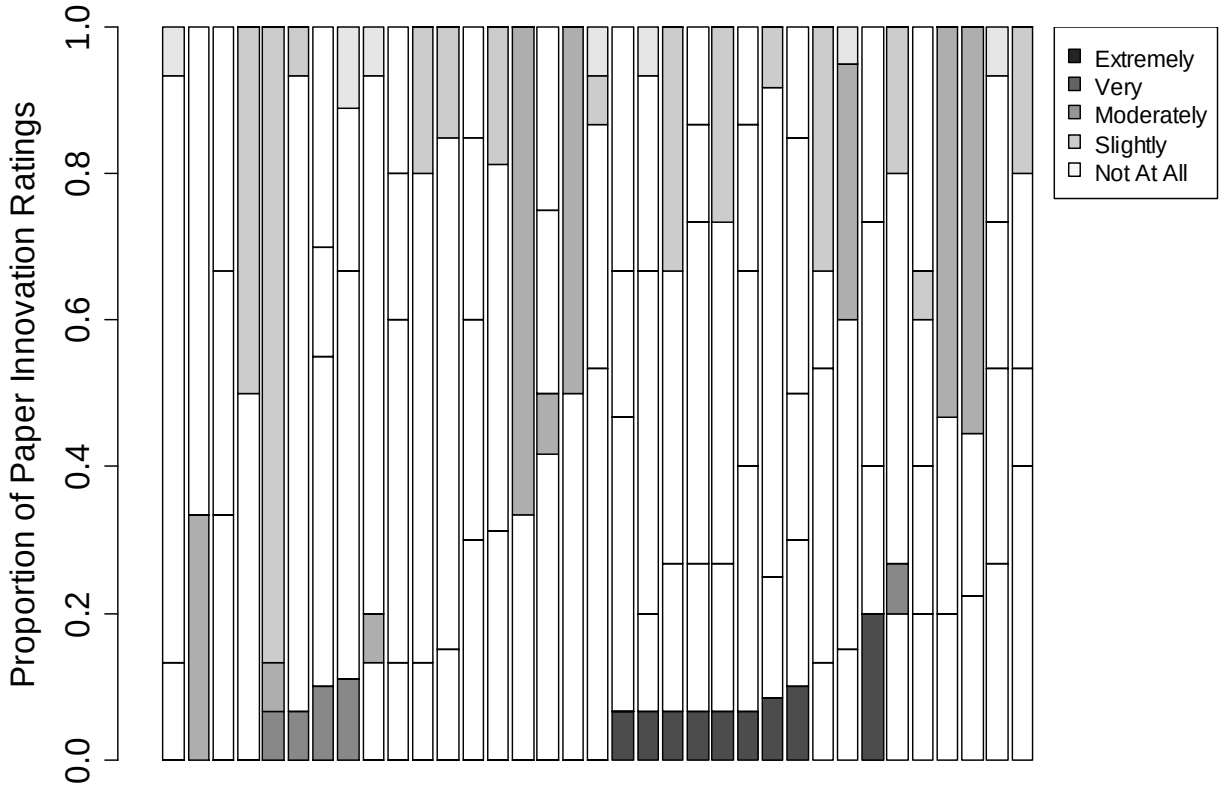



**Figure F-9. Innovation Distributions by Researcher for Paper-Level Rating Proportions for Matched R01 PIs**

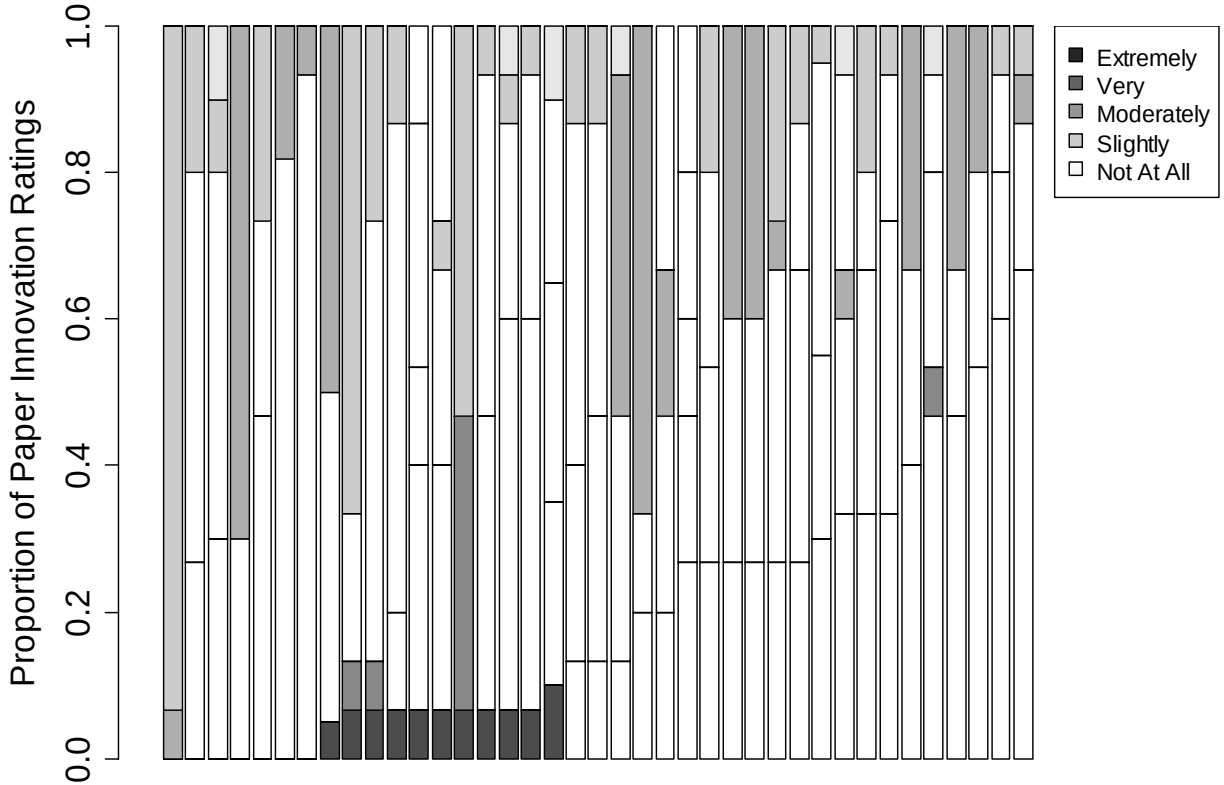



**Figure F-10. Innovation Distributions by Researcher for Paper-Level Rating Proportions for HHMI Investigators**

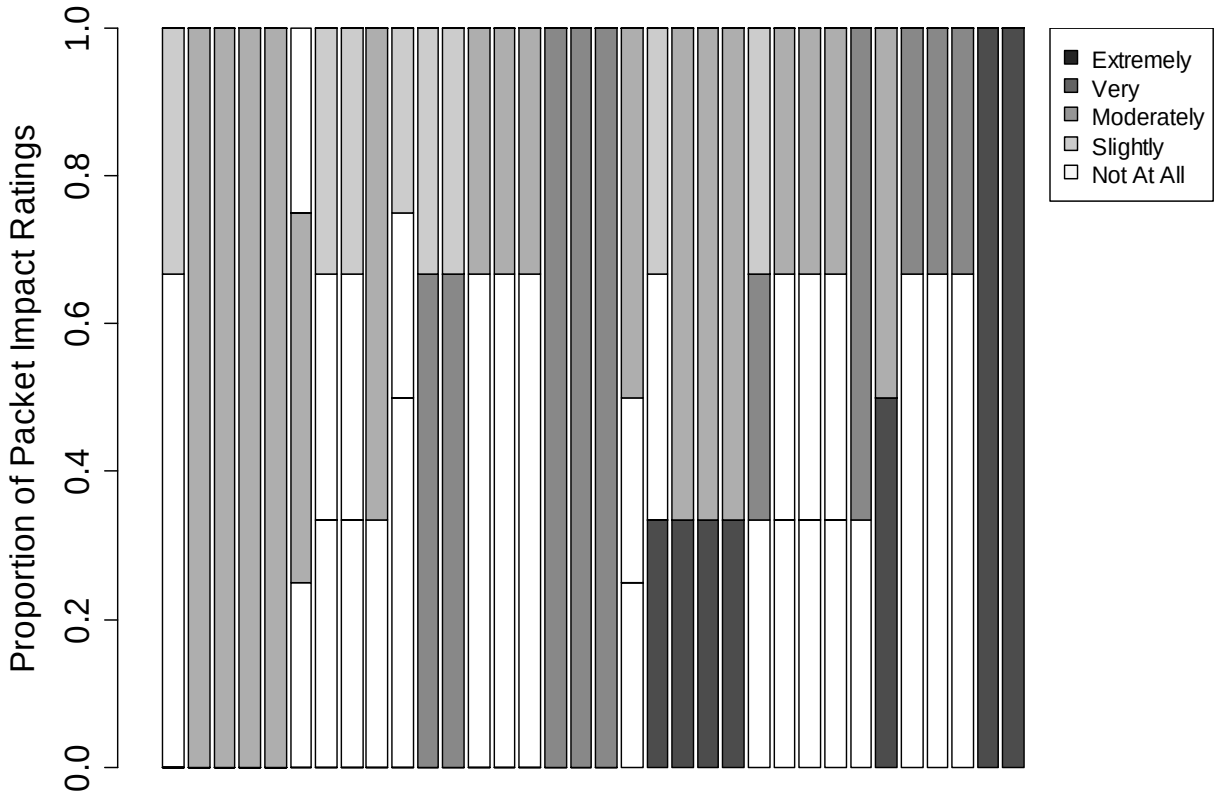



**Figure F-11. Impact Distributions by Researcher for Packet-Level Rating Proportions by NDPA Pioneers**

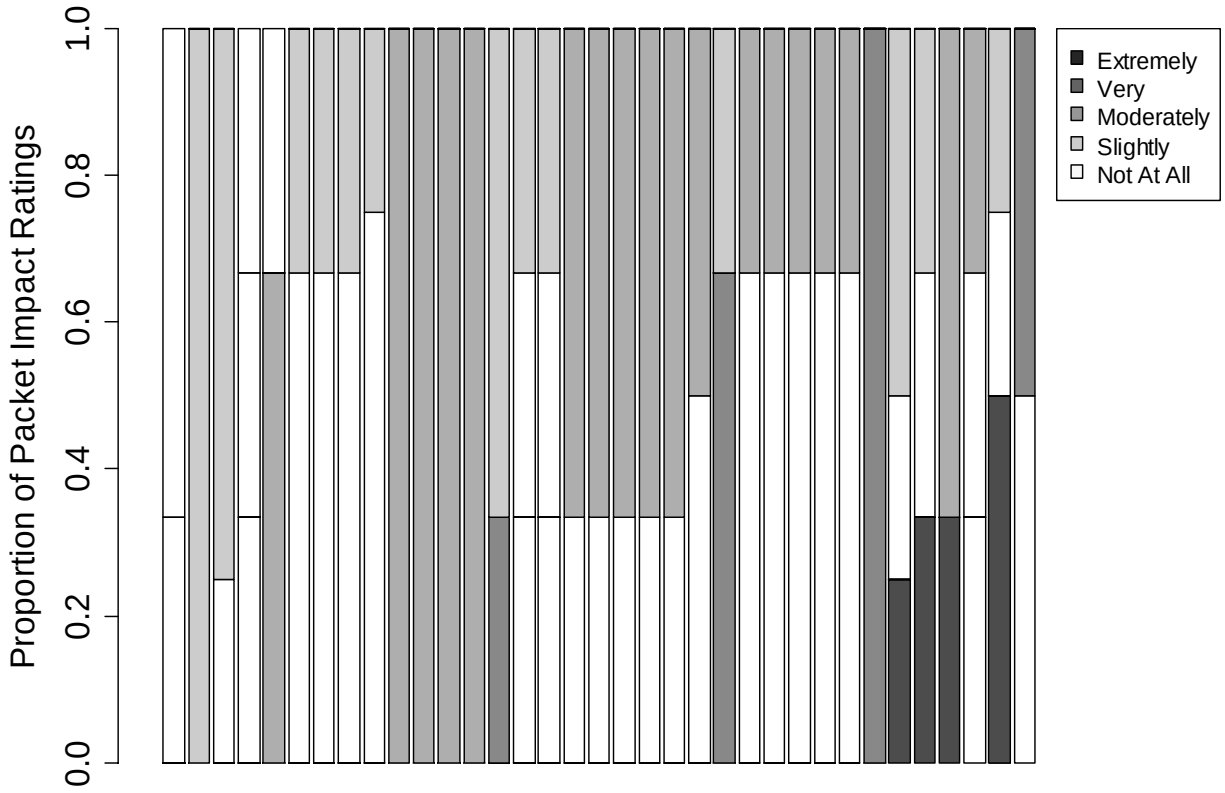



**Figure F-12. Impact Distributions by Researcher for Packet-Level Rating Proportions by Matched R01 PIs**

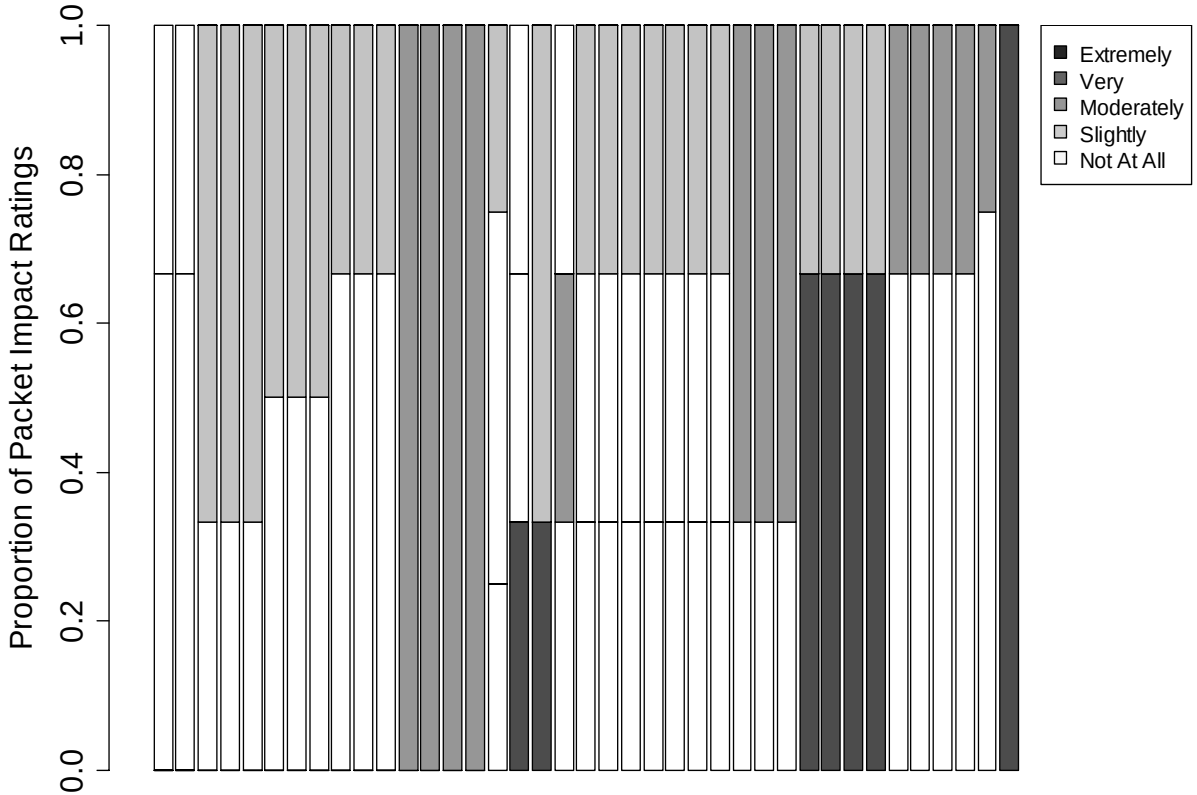



**Figure F-13. Impact Distributions by Researcher for Packet-Level Rating Proportions by HHMI Investigators**

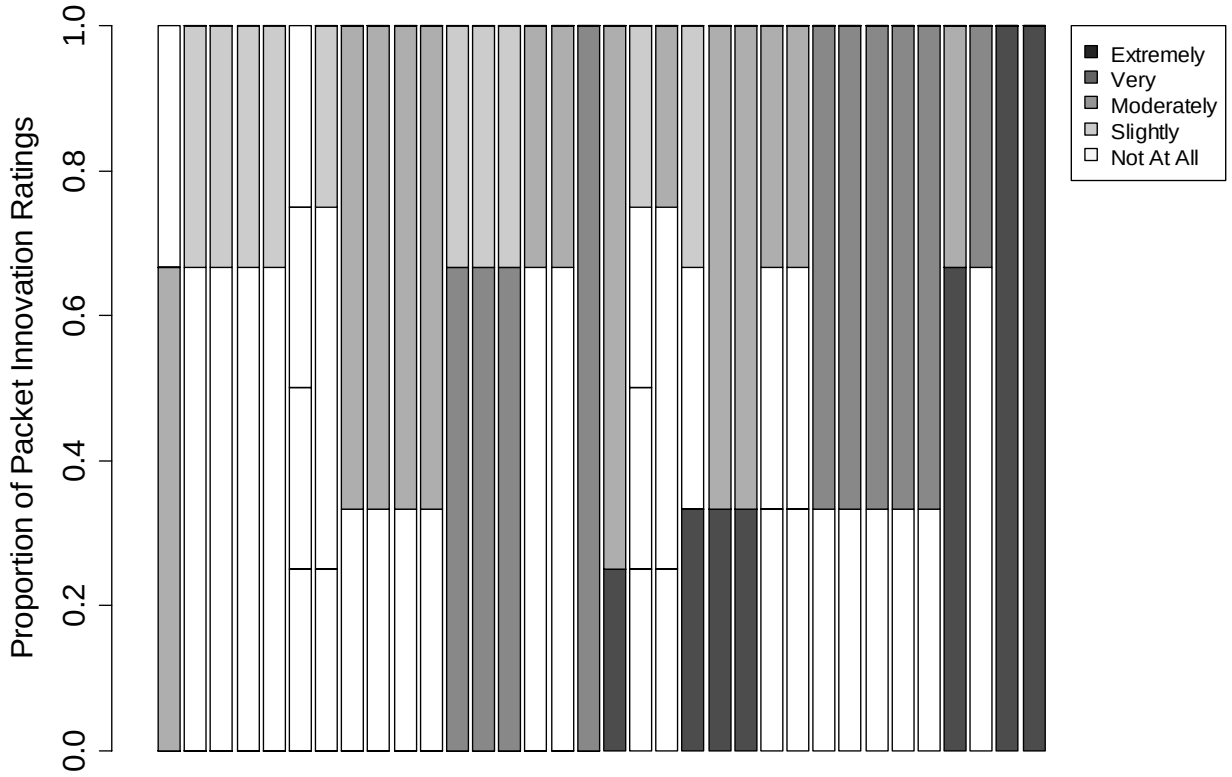



**Figure F-14. Innovation Distributions by Researcher for Packet-Level Rating Proportions by NDPA Pioneers**

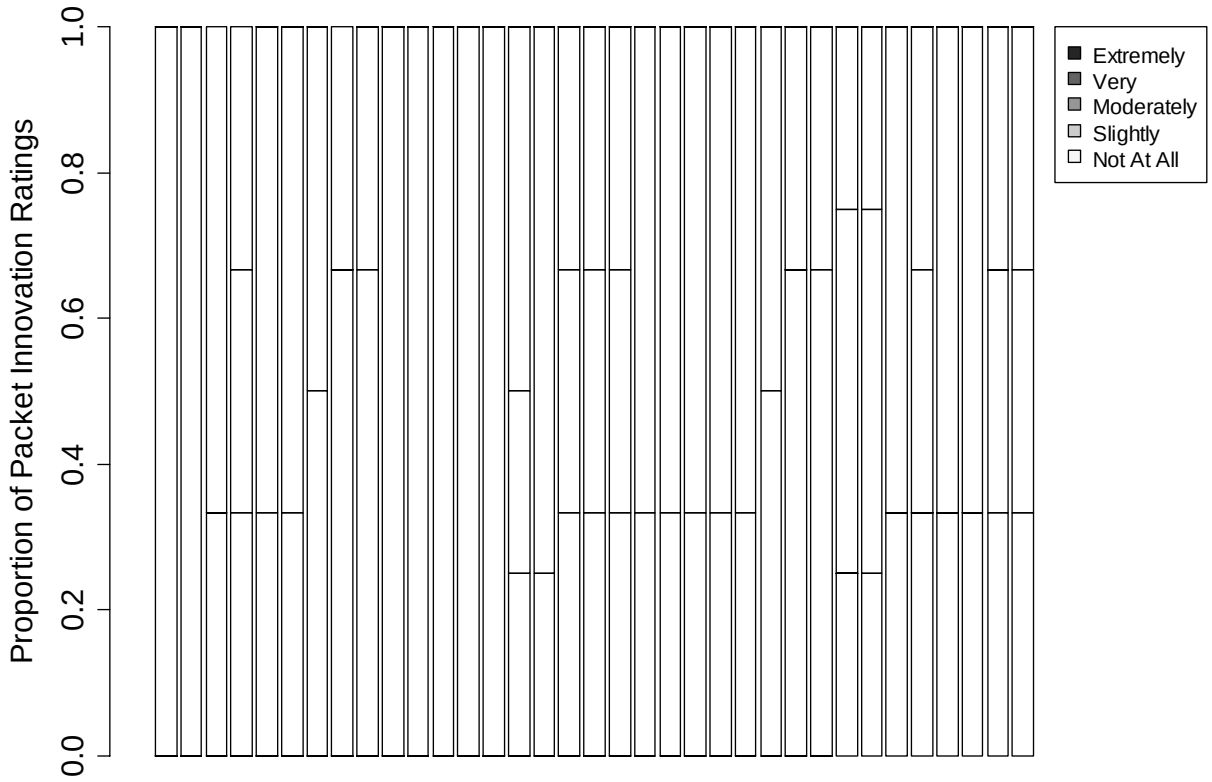



**Figure F-15. Innovation Distributions by Researcher for Packet-Level Rating Proportions for Matched R01 PIs**

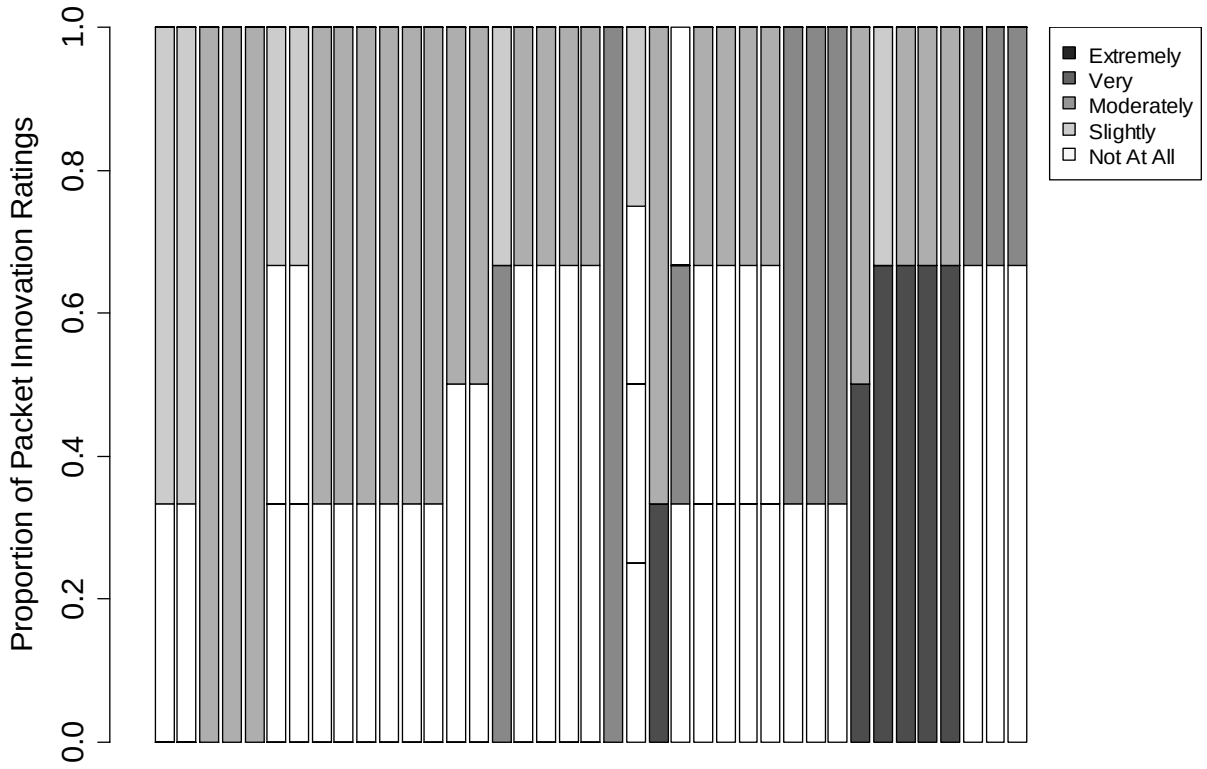



**Figure F-16. Innovation Distributions by Researcher for Packet-Level Rating Proportions for HHMI Investigators**

# Appendix G
# Qualitative Analysis Codebook

To better understand why experts characterized papers as having impact or being innovative, qualitative coding was performed on the expert reviewer open comments. Codes and subcodes were developed to find patterns across expert comments. For responses related to the impact of the research, statements were categorized under one of the following broad codes: evaluation strategy, impact, or impact decision (Table G-1). For responses related to the innovativeness of the research, statements were categorized under one of the following broad codes: evaluation strategy, innovation, or innovation decision (Table G-2). Subcodes under each of the broad codes further characterized the nature of the statement. Representative examples from experts are provided for each of the impact and innovation subcodes (Table G-3 and Table G-4).

**Table G-1. Code Descriptions for Impact**

| Code/Subcode | Description |
|---|---|
| **EVALUATION STRATEGY** | Reviewers' approach to providing rationale for impact/innovation ratings. |
| Entire packet | Rationale centered on the packet as a whole. |
| Multiple papers | Rationale included discussion of innovation/impact of more than one paper as a group. |
| Single paper | Rationale included discussion of innovation/impact for a single paper. |
| Unclear | Unclear whether rationale was based on a single paper, multiple papers, or the entire packet. |
| **IMPACT** | Discussion of research impact. |
| Impactful | Research was considered impactful. |
| Moderate or incremental | Research was considered to have moderate impact or is considered incremental science. |
| No response | No explicit discussion of impact. |
| Not impactful | Research was considered not impactful. |
| **IMPACT DECISION** | Discussion of how impact assessments were made. |
| **Scientific Discovery** | Discussion of research impact focused on the scientific discovery. |
| *Elucidation of pathways or mechanism of action* | Elucidation of pathways or understanding mechanism of action were the primary reasons without further descriptions that would allow for further coding into the rubric categories. |

| Code/Subcode | Description |
|---|---|
| *Foundational research* | Decision was based on the scale of impact on a specific field or across fields, including whether other researchers have built on the research, with no other explanation lending to placement inside one of the rubric categories. |
| *Reproduced/Reproducible* | Decision was based in whole or in part on whether the findings have been reproduced or whether they are reproducible. |
| *Translational or clinical potential* | The research was/was not translational or had/did not have the potential to directly impact patient health or outcomes without further descriptions that would allow for further coding into rubric categories. |
| *Scale-crossing* | The research crossed/could not cross scales of organism complexity (cells, tissues, organs, organ systems, and organisms) without further descriptions that would allow for further coding into rubric categories. |
| *Rubric* | Discussion of impact of the scientific discovery was based on the rubric given for the impact questions. |
| Challenges present understanding | The research challenged or did not challenge present understanding. |
| Changes present understanding | The research changed or did not change present understanding. |
| Insights that enable new techniques | The research provided insights that enabled new techniques or technologies. |
| New paradigms | The research provided pathways to new paradigms. |
| Pathways to new frontiers | The research provided pathways to new frontiers. |
| Redefines the boundaries of science and engineering | The research redefined the boundaries of science and engineering. |
| **Scientific Method** | Discussion of impact focused on the scientific method or approach. |
| *Standard or established methods* | Reviewer mentions that standard or established methods were used. |
| *Rubric* | Discussion of impact of the scientific method was based on the rubric given for the Innovation questions. |
| Underlying research is at odds with prevailing wisdom | [Not used] |
| Research involves a combination of disciplines | The method or approach used was drawn from multiple disciplines. |
| Research requires the use of equipment or techniques that have not been proven or considered difficult | The method or approach used equipment or techniques that have not been proven or are considered difficult. |
| **Scientific Team Composition** | Composition of the team, including the reputation of the PI, was mentioned as part of the justification. |
| **Other** | Other factors influencing assessment of research. |
| *Citations* | Citations were discussed as part of the rationale. |

| Code/Subcode | Description |
|---|---|
| *Claims outweigh data* | Decision was based on the conclusions not being supported by the results or approach used. |
| *Journal prestige* | Journal prestige was discussed as part of the rationale. |
| *Not published yet or recently published* | Impact was too soon to tell because the manuscript was recently published or not yet published. |
| *Review* | The publication being a review paper, as opposed to an original scientific publication, was mentioned as part of the rationale. |

**Table G-2. Code Descriptions for Innovation**

| Code/Subcode | Description |
|---|---|
| **EVALUATION STRATEGY** | Reviewers' approach to providing rationale for impact/innovation ratings. |
| Entire packet | Rationale centered on the packet as a whole. |
| Multiple papers | Rationale included discussion of innovation/impact of more than one paper as a group. |
| Single paper | Rationale included discussion of innovation/impact for a single paper. |
| Unclear | Unclear whether rationale was based on a single paper, multiple papers, or the entire packet. |
| **INNOVATION** | Discussion of research innovation. |
| Innovative | Research was considered innovative. |
| Moderate or incremental | Research was considered to be moderately innovative. |
| No response | No explicit discussion of innovation. |
| Not innovation | Research was considered not innovative. |
| **INNOVATION DECISION** | Discussion of how innovation assessments were made |
| **Methodology** | Discussion of Innovation focused on the scientific method or approach. |
| *Standard or established methods* | Reviewer mentions that standard or established methods were used. |
| *Rubric* | Discussion of innovation of the scientific method was based on the rubric given for the innovation questions. |
| Underlying research is at odds with prevailing wisdom | The underlying research used was at odds with prevailing wisdom. |
| Research involves a combination of disciplines | The method or approach used was drawn from multiple disciplines. |
| Research requires the use of equipment or techniques that have not been proven or considered difficult | The method or approach used equipment or techniques that have not been proven or are considered difficult. |
| *Creative utilization or improvement of existing techniques* | Discussion of innovation was based on the creative use or improvement of existing techniques. |
| *Other* | Discussion of innovation included other aspects of the methodology which were not consistent with the rubric, or with the other methodology categories. |
| **Discovery** | Discussion of research innovation focused on the scientific discovery. |
| *Elucidation of pathways or mechanism of action* | Elucidation of pathways or understanding mechanism of action were the primary reasons without further descriptions that would allow for further coding into the rubric categories. |

| Code/Subcode | Description |
|---|---|
| *Foundational research* | Decision was based on the scale of impact on a specific field or across fields, including whether other researchers have built on the research, with no other explanation lending to placement inside one of the rubric categories. |
| *Reproduced/Reproducible* | Decision was based in whole or in part on whether the findings have been reproduced or whether they are reproducible. |
| *Translational or clinical potential* | The research was/was not translational or had/did not have the potential to directly impact patient health or outcomes without further descriptions that would allow for further coding into rubric categories. |
| *Scale-crossing* | The research crossed/could not cross scales of organism complexity (cells, tissues, organs, organ systems, and organisms) without further descriptions that would allow for further coding into rubric categories. |
| *Rubric* | Discussion of innovation of the discovery was based on the rubric given for the Impact questions. |
| Challenges present understanding | The research challenged or did not challenge present understanding. |
| Changes present understanding | The research changed or did not change present understanding. |
| Insights that enable new techniques | The research provided insights that enabled new techniques or technologies. |
| New paradigms | The research provided pathways to new paradigms. |
| Pathways to new frontiers | The research provided pathways to new frontiers. |
| Redefines the boundaries of science and engineering | The research redefined the boundaries of science and engineering. |
| **Creative Hypotheses or Ideas** | Decision was based on the creativity in the concepts or hypotheses underlying the research. |
| **Team Composition** | Composition of the team, including the reputation of the PI, was mentioned as part of the justification. |
| **Synonymous with Impact** | The reviewer referred to his or her discussion for the impact question for justification of his or her innovation rating, or specifically cited that innovation was perceived to be synonymous with impact. |
| **Other** | Other factors influencing assessment of research innovation. |
| *Review* | The publication being a review paper, as opposed to an original scientific publication, was mentioned as part of the rationale. |

**Table G-3. Code Examples for Impact**

| Code/Subcode | Description |
|---|---|
| **EVALUATION STRATEGY** | |
| Entire packet | "As a whole, the work is important because it provides mechanistic insights into the pathology of viral infections and suggests strategies for prevention." |
| Multiple papers | "I was less enthusiastic about the last two papers because, while solid, I just did not find the insights derived from the studies to be that profound." |
| Single paper | "The paper concerning malaria was a comparative genetic analysis of strains of Plasmodium that are species specific, and the conclusion was that the human-specific form had evolved from the chimpanzee-specific strain. Not very surprising." |
| Unclear | "Ok, but not outstanding." [Only response] |
| **IMPACT** | |
| Impactful | "The optogenetics technology is revolutionizing neuroscience. Some of the applications are more interesting than others, some papers are over-hyped, but it is a huge advance, perhaps the most important in the last two decades of neuroscience." |
| Moderate or incremental | "As with the previous set of papers these finding seem to modify incrementally our understand[ing] of these systems. I found none of them revolutionary." |
| No response | [Not applicable] |
| Not impactful | "Paper 3 is a very detailed analysis of the mechanism of terpene formation that I don't expect to have a big impact outside this limited field." |
| **IMPACT DECISION** | |
| **Scientific Discovery** | |
| *Elucidation of pathways or mechanism of action* | "In totality this work is highly impactful in understanding mechanisms regulating stem cell renewal and differentiation." |
| *Foundational research* | "The paper chosen as extremely impactful demonstrates an important function of myosin II in driving keratocyte cell motility through activation of actin disassembly. This paper will be highly cited and will serve as the basis for a substantial body of work in cell motility." |
| *Reproduced/Reproducible* | "There are aspects of these papers that are interesting, particularly the claimed finding of reprogramming pancreatic cells without transcription factors. However, this has not been reported by anybody else and raises questions about reproducibility. …I am really curious if that reprogramming protocol is reproducible. Overall, I would not say these reports have had a high impact. If that were the case, the claimed reprogramming protocol would have been in every local newspaper." |

| Code/Subcode | Description |
|---|---|
| *Translational or clinical potential* | "The gold-silver nanocage may be a novelty in nanofabrication, but it has no impact on drug delivery. There is no way such structures can be translated for drug delivery. .. The lack of a medical perspective significant dampens the impact of the work." |
| *Scale-crossing* | "The other two papers do not address questions that are as fundamental as the three articles cited above. While the quality of the work is high, I did not find these stories as compelling. The article by Cheng et al. (2009) seemed the least impactful because it focuses on very specific molecules and is not likely to have the major impact on the field or across species, that the other papers are likely to have." |
| *Rubric* | |
| Challenges present understanding | "I would not score this work as Extremely Impactful as it does not challenge wisdom, create a new paradigm or give unexpected insight. This work is excellent, timely, very impactful and rigorous and identifies upstream mechanisms of what is anticipated to be a critical new pathway for understanding neurodegenerative mechanisms." |
| Changes present understanding | "Each of these papers is really stellar, and changes the way we think about regulation of transcription and translation in cells." |
| Insights that enable new techniques | "The first paper demonstrates the use of plant viruses to image and differentiate arterial and venous capillaries. It introduces what appears to be a novel platform for imaging." |
| New paradigms | "The impact of this work is particularly high because essentially either nothing (or highly controversial information) was known on the effect of molecular crowding and nonspecific protein-protein interactions on protein stability, before these studies by Pielak. Impact and innovation are highly connected in these pioneering paradigm-shifting studies." |
| Pathways to new frontiers | "While each individual paper does not achieve a novel result (that had not already been achieved by other chemical backbones), when taken as a whole the body of work did explore a new chemical frontier and led to some unexpected insight, particularly into the physical basis for the stability of designed structures." |
| Redefines the boundaries of science and engineering | "However, I give it less impact than the whole body of work in the previous set, because it will not break down boundaries. It will develop new techniques, but the techniques will be accepted by the field." |

| Code/Subcode | Description |
|---|---|
| **Scientific Method** | |
| *Standard or established methods* | "The PLOS genetics paper using this same approach to uncover the underlying mutation in flincher and toppler strains in Lass1 is of significance with the impact on lipofuscin accumulation, but it does not reach the high impact of some of the others, most likely due to this same approach of using mouse mutants leading to ataxia and then finding underlying mutations." |
| *Rubric* | |
| Underlying research is at odds with prevailing wisdom | [Not used] |
| Research involves a combination of disciplines | "excellent mix of basic science molecular approaches, genome biology, physiology of bacteria as they associate with host, cause disease, etc. The breadth of knowledge and approaches is unusual for many labs [that] typically focus on one thing and do it over and over. This PI is cutting edge in many fields." |
| Research requires the use of equipment or techniques that have not been proven or considered difficult | "This is an extremely impressive body of work. The high level of impact derives not only from the application of cutting edge techniques, several of which are novel and used for the first time, but also from the relevance to human disease." |
| **Scientific Team Composition** | "Prof. [name removed] has applied molecular dynamics simulations more broadly than anyone else. He is prolific and his papers are characterized by a rigor and attention to detail." |
| **Other** | |
| *Citations* | "The five papers together have only been cited a total of 57 times, which isn't bad, but by no means highly impactful. " |
| *Claims outweigh data* | "Most of this work is highly speculative. The claims far outweigh the data." |
| *Journal prestige* | "For starters the authors address important aspects of stem cell development. Clearly the authors are at the cutting edge in their area. This is reflected by the type of journals they publish in and by the methods used to accomplish their goals. There is no doubt that this science has an impact in the organism system they work in." |
| *Not published yet or recently published* | "The Seiler et al. manuscript was just published in Nature Immunology, so its scientific impact has not yet been evaluated by the scientific community." |
| *Review* | "The review paper is well done, but cannot be held to have the impact of a primary science paper." |

**Table G-4. Code Examples for Innovation**

| Code/Subcode | Description |
|---|---|
| **EVALUATION STRATEGY** | |
| Entire packet | "My score for the set is modestly higher than for each taken alone — because together they do apply state-of-the-art (albeit not novel) methods in a new way to a new preparation. As stated above, this establishes a solid foundation for future and hopefully more innovative work." |
| Multiple papers | "Papers #1, #2, and #3 were moderately innovative both in the starting hypothesis and in the methods that were used. These studies were logical extensions of the knowledge in the field using mostly standard tools for molecular biology. All three of these papers involved collaborations with groups that had appropriate expertise for specialized methods. Paper #4 was highly innovative because it was the first demonstration of a new technology, including development of new data analysis tools. Paper #5 was very innovative because it combined cell sorting and Next Generation DNA sequencing to get at the question of which parts of the genome are dysregulated in a histone methyltransferase mutant." |
| Single paper | "The 'human movement...' paper develops a new model." |
| Unclear | "The questions poised are cutting edge and the approaches taken are relatively novel." [Only response] |
| **INNOVATION** | |
| Innovative | "These studies illustrate boldness in challenging conventional wisdom and a careful approach, using whatever means available to address issues of key importance. This includes the innovative creation and implementation of novel techniques and approaches, such as the use of better EM imaging techniques and genetic methods to create targeted knockouts in C.elegans." |
| Moderate or incremental | "Paper 1 combines structure analysis with information on the spatial and developmental expression of the enzymes and the functional effects of their knockout: moderate innovation. Paper 2 is a straightforward crystallographic analysis that I would consider slightly innovative. Paper 3 seemed like a pretty straightforward crystal structure analysis with some computational modeling that I rank slightly innovative. I rank paper 4 as moderately innovative for combining evolutionary mutation information with a chromatography based analysis of the enzymatic activity of the enzymes generated. I rank paper 5 as moderately innovative for combining the chemical metabolite analysis, crystal structure data and in vivo assays. I rank the group as a whole moderately innovative. |
| No response | [Not applicable] |

| Code/Subcode | Description |
|---|---|
| Not innovation | "Techniques here are pretty standard in the field...they are applied to a new circuit/system, but the approach is not dramatically innovative." |
| **INNOVATION DECISION** | |
| **Methodology** | |
| *Standard or established methods* | "These studies used well-established methods and a traditional approach. The hypotheses were largely derived from previous evidence. The authors performed a series of direct experimental tests of the mechanisms that might underlie specific behaviors. The behavioral connection could not be validated since all experiments were performed in vitro. The data are definitely novel and interesting for the cerebellar physiologists. The approach is conventional." |
| *Rubric* | |
| Underlying research is at odds with prevailing wisdom | "These studies illustrate boldness in challenging conventional wisdom and a careful approach, using whatever means available to address issues of key importance. This includes the innovative creation and implementation of novel techniques and approaches, such as the use of better EM imaging techniques and genetic methods to create targeted knockouts in C.elegans.." |
| Research involves a combination of disciplines | "These studies are all highly innovative as the investigators were able to combine theoretical and computational statistical physics, experimental immunology, and clinical data to address outstanding questions that are fundamental to the field. The approaches are unconventional and require both deep understanding of the biological questions and effective implementation of statistical and computational methods." |
| Research requires the use of equipment or techniques that have not been proven or considered difficult | "The approach is technically innovative, in the sense that authors did not only implement the real time MRI technology but also modified it for speech analysis. I cannot judge how much the authors contributed to the development of this new technology since I did not follow the history of how real time MRI developed and evolved, but nevertheless, those studies do present a significant technical advance." |

| Code/Subcode | Description |
|---|---|
| *Creative utilization or improvement of existing techniques* | "While many of the approaches are not necessarily innovative in and of themselves, the way they were used is highly novel. For instance, the clarity of fax sorting NG2+ SVZ versus cortical cells to evaluate EGFP expression and activity on migration was a very innovative approach at the time. The methods to evaluate direct interactions between NPCs and NSCs in the Notch/EGFR Nature 2010 paper is also a highly novel way to address the question posed. Finally in the chordin 2010 Nat Neurosci paper, the approach to use a demyelinating insult to address differences in migration and differentiation and site from which migration occurs was a novel use of system." |
| *Other* | "The approaches chosen are able to answer the questions the authors asked. All the approaches are creative and not limited to the organism in question." |
| **Discovery** | |
| *Elucidation of pathways or mechanism of action* | "Novel signaling components and protein modifications that mediate cell death and autophagy were identified in these papers. However, these studies did not rely on novel techniques, and the results are within the framework of existing pathways." |
| *Foundational research* | [Not used] |
| *Reproduced/Reproducible* | "#1 -Logic — very innovative — combines many approaches and bioinformatics; unique combination of methods is innovative. #2, 3,4 — Moderately innovative. Interesting concepts and results, but these are not out of the norm of typical research. #5, extremely innovative. This is a complex in vivo screen, but led to a hit from a subset of a compound library that could then be explored further. I would have bet against this screen working, but they found interesting and reproducible results with it." |
| *Translational or clinical potential* | "For the most part these papers use state-of-the-art mouse genetics to help define important roles for endothelial cells in stem cell niche formation. The most innovative paper dealt with the generation of multi potent stem cells from the germ line. It identified new markers that can potentially help create a whole new approach for fertility treatments. The other papers were innovative not for their technical approaches, which could be done in many labs, but for the unique insights and advances in defining the vascular niche." |
| *Scale-crossing* | "This paper is intellectually innovative, as its reach is so broad. While the findings are not at odds with prevailing wisdom, they push the envelope much further by charting brand new territory in terms of the functionality of the Shh pathway and the importance of the proteoglycans. A tremendous, lovely story." |

| Code/Subcode | Description |
|---|---|
| *Rubric* | "The PLOS genetics paper using this same approach to uncover the underlying mutation in flincher and toppler strains in Lass1 is of significance with the impact on lipofuscin accumulation, but it does not reach the high impact of some of the others, most likely due to this same approach of using mouse mutants leading to ataxia and then finding underlying mutations." |
| Challenges present understanding | "The work illustrates the power of combination of disciplines and the application of especially sophisticated techniques. The results challenge, or even overturn, some prevailing ideas about DNA structure, and offer new approaches that may actually be useful." |
| Changes present understanding | "There is only one paper in this group. The novelty of this work is 1) the discovery of cell death genes can also regulate cell size and 2) therefore the cell death process and cell size program can be coupled. These connections would not been predicated based on prior knowledge." |
| Insights that enable new techniques | "The most innovative contribution of these papers is the use of fluorinated side chains to better enable the visualization of proteins in intact cells. Although much of the spectacular resolution associated with traditional biomolecular NMR is lost in this approach, it does enable the detection of protein NMR signal in vivo, even for larger globular proteins, a feat that has largely defied methods based on traditional labeling schemes. Further innovation is provided by considering the specific character of proteins as the primary crowding agent within cells, and noting that they differ from traditional agents used to mimic cellular crowding in vitro." |
| New paradigms | "Paper 1 I consider very innovative in the sense that it conceived of a whole new field of regulation. Paper 2 — Extends the analysis to metabolites which was cutting edge technology at the time, so it gets slight innovation for that. Paper 3 — Extends the analysis to a functional endpoint. Beyond the first two papers, I would rate it high for impact but probably not for innovation. Paper 4 — Monitoring metabolites was pretty new technology so some innovation points should be awarded. Paper 5 — The findings are potentially of high impact but the approach is not particularly innovative. Overall it is a moderately innovative package, with the metabolism monitoring representing the most innovative of the technologies." |

| Code/Subcode | Description |
|---|---|
| Pathways to new frontiers | "In this work 'the whole is greater than the sum of its parts.' Heald's use of cytoplasmic extracts to reconstitute complex cellular functions was not new; however, it was essential for generating a data set robust enough to be used for mathematical modeling. It was the skill with which the studies leveraged these two complimentary approaches that was the most compelling innovative aspect for me. Especially interesting were the predictions of how "putative factors" that affect microtubule plus or minus end dynamics will affect spindle length. These predictions prompt quite a few interesting future experiments. It is trendy for investigators to attempt to combine basic cell biological research with mathematical modeling but the results are not typically as provocative as these are. " |
| Redefines the boundaries of science and engineering | "These papers exactly fall into the category of 'development of new frontiers' that 'redefine the boundaries of science and engineering.' This has been achieved through a combination of novel genetic engineering and transformative application of the technology to study awake behaving animals while selectively manipulating neuronal activity by light. Importantly, this novel technology is transferrable to other research groups, making it a true community resource with a huge impact." |
| Creative Hypotheses or Ideas | "The main innovations in these paper are conceptual. The experimental techniques applied are relatively standard, but the kinds of analysis and the posing of the questions allows the authors to re-frame long-standing questions about the visual system and to connect the biological features of retinal circuitry with computational problems faced by the visual system. Because these papers did not directly deal with treatment of a human disease, it is difficult to compare their long-term impact to that of papers that directly address the causes and effects of disease on human health. Nonetheless, I believe that these papers have [been] and will continue to be very important for our understanding of early sensory processing in a number of systems." |
| Team Composition | "This is a tough choice. These papers build off of many years of mouse ES cell biology to help suggest possible approaches for successful hES cultures. Human ES are much more technically challenging to work with and making these technical advances are very important to and necessary for the field. The challenges are often empiric and difficult to accomplish, but not massive departures from what has worked in other systems. Dr Keller is a leader in these efforts and helped develop much of the mouse ES biology that the human work is often patterned after." |

| Code/Subcode | Description |
|---|---|
| Synonymous with Impact | “Here again, innovation clearly parallels impact. The technique appears to be completely novel in the context of biomolecules — i.e. if you include the Science paper — but of necessity I am guessing. I assume that standardization is required for most of the systems studied — however for routine measurements, this technique appears to have great promise.” |
| Other | “About all sets of papers — Innovation (in approach) must be evaluated with careful validations. Please see my detailed comments on individual papers regarding the impact of the research. Only the 4th and 5th papers were done with reasonable validation. Individual paper — #1 Science Review. This is review paper and innovation cannot be evaluated. #2 Nature Neuroscience — This paper may be innovative, but I don’t think the approach was validated. #3 Science vol 324 — Same as #2 paper. #4 Science vol 332 — Several unique approaches, including the use of circadian arrhythmic mutant flies and an enriched environment (fly mall) are very innovative. #5 Nature — Standard methodology was used, but careful analysis revealed an innovative discovery which lead to an innovative idea for local sleep.” |
| Review | “The fact that it is a review article made it difficult to examine the innovativeness of the approaches.” |

# Appendix H
# Protocols for Downloading and Cleaning Publication Data for Bibliometric Analysis

## Background

STPI obtained and cleaned publication data at the researcher and grant levels. Researcher-level publications are defined as research articles where the PI of interest appears as an author. Researcher-level data were collected for NDPA Pioneers, matched R01 PIs, HHMI investigators, and NDPA finalists. Grant-level publications are defined as research articles that can be attributed to the grant number of interest through funding acknowledgments in Web of Science (WOS), PubMed, or NIH electronic Scientific Portfolio Assistant (eSPA). Grant-level data were collected for NDPAs, matched R01s, and R01 portfolios.

The protocols described in this appendix detail the data download and cleaning processes at the researcher level. At the researcher level, we used the WOS database to download publication data and disambiguate researchers. Then we used VantagePoint (VP), commercially available software, to clean and prepare the publication data for statistical analysis using R.

At the grant level, we downloaded publication data that corresponded to grant numbers as opposed to authors, and we altered the data-cleaning methods to render them appropriate for the grant-level data. We collected the grant-attributed titles using a combination of the eSPA database and WOS. For the matched R01 PIs, we performed queries by grant number, and for NDPA Pioneers, we queried the text "DP1," the NDPA activity code, and "Pioneer" to get a comprehensive list of titles. The complete publication data for these titles were then downloaded from WOS and cleaned in the manner described in the data-cleaning protocol.

In addition to cleaning the researcher-level and grant-level data with VP, we continued to fix errors that were observed throughout the bibliometric analysis process to develop the cleanest dataset possible. The errors that were observed arose from a combination of errors from the WOS database, the eSPA database, and human judgment. All analyses presented in the report reflect the most accurate dataset.

The downloading protocol and data-cleaning protocol follow.

## Start

To begin, you will need to access the Thomson Reuters WOS publication database.

1. Create a Researcher ID to keep track of and save queries.
2. Copy the most up-to-date fuzzy files, import filters, macros, and thesauri (VantagePoint file types) to the appropriate Program Files folder on your computer

You will also be provided with a folder of NIH application biographical sketches (biosketches) or curriculum vitae (CV) for the researchers in the comparison group that is assigned to you. Additional materials for data cleaning include: a spreadsheet of the PI's full name, award year, and STPI-assigned "STPICode" for each researcher; a spreadsheet that matches WOS indexed journals to WOS Subject Categories (WOS SC-Journals spreadsheet); and a summary spreadsheet at the PI data rate in which to record metrics of interest. Time estimates are provided as a rough guide for the amount of time that should be spent on each process.

## Data Downloading Protocol (Researcher Level)

### Download the Publications [TOTAL TIME: 10-20 Minutes]

3. Perform an "Author Finder" search using last name and first initial.
   - For example, Abbott L is the search query for Larry F. Abbott.
   - For people with common last names, refer to the biographical sketch/CV and the format of the name from the publications listed; use judgment in whether a middle initial should also be included in this search query. Including the middle name will help you sort through the distinct author sets more efficiently and determine which author is the correct one.
4. Select the appropriate "Distinct Author Set" by referring to the provided biographical sketch/CV (Figure H-1). Judgment should be used here in selecting the correct author set. Some things that might help you identify the correct distinct author set are:
   a. Look at the middle initial in a person's name.
   b. Match field of study with related "source titles for the author" in WOS (although sometimes this is not accurate because people don't perform research in the field in which they got their degree).
   c. Look at the year in which they received their PhD to ensure that the "publication years" in WOS roughly make sense (we assume that people will start publishing around the time they get their PhDs or a little before).

d. Match the “last known institution” in WOS to the institution of employment listed on the biosketch/CV (the biosketches and CVs are from 2004-2006 so this may or may not be useful if the individual has moved institutions in the last few years).

e. You may choose one or more author sets, based on your judgment of whether the publication years and subject areas listed are compatible with the information provided on the author.

**18 Author Sets for Pellegrino J***

Page 1 of 2 Go

Select one or more sets and click "View Records".

View Records | Clear All | How do I provide feedback?

| Select | Set | Author Names | # of Records | Source Titles (top 5 by record count) | Publication Years | Subject Areas (top 5 by record count) |
|---|---|---|---|---|---|---|
| | 1. | Pellegrino, JW<br>PELLEGRINO, J | 77 | - BULLETIN OF THE PSYCHONOMIC SOCIETY (11)<br>- MEMORY & COGNITION (7)<br>- JOURNAL OF EXPERIMENTAL PSYCHOLOGY-HUMAN LEARNING AND MEMORY (6)<br>- INTELLIGENCE (4)<br>- JOURNAL OF EDUCATIONAL PSYCHOLOGY (4) | 1974 - 2007 | - Social Sciences (76)<br>- Life Sciences & Biomedicine (10) |
| | 2. | Pellegrino, J<br>PELLEGRINO, JJ | 40 | - JOURNAL OF MEMBRANE SCIENCE (15)<br>- DESALINATION (5)<br>- ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY (4)<br>- SEPARATION SCIENCE AND TECHNOLOGY (4)<br>- INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH (3) | 1988 - 2008 | - Physical Sciences (34)<br>- Multidisciplinary Science & Technology (32)<br>- Life Sciences & Biomedicine (3) |
| | 3. | Pellegrino, J<br>Pellegrino, JJ | 40 | - TRANSFUSION (20)<br>- BLOOD (7)<br>- VOX SANGUINIS (6)<br>- JOURNAL OF CLINICAL LABORATORY ANALYSIS (4)<br>- BRAZILIAN JOURNAL OF MEDICAL AND BIOLOGICAL RESEARCH (1) | 1996 - 2006 | - Life Sciences & Biomedicine (40) |

**Figure H-1. Distinct Author Sets Resulting from the Author Finder Search in Web of Science**

5. Click on “View Records.”
6. Do a quick comparison of the publication titles listed in the biosketch/CV and the titles that have been returned in the search query in order to make sure that you have the correct person (keeping in mind that the CV/biosketch is from 2004-2006). If you do not find any of the publications listed in the biosketch/CV, there may be an error—go back and look at the distinct author sets again.
7. On the left, see the “Refine Results” box. Under the “Document Types” arrow, refine by the “article” document type.
8. Also under the “Refine Results” box, filter for publications up to and including 2011.
9. Return to the refined list of query hits. At the bottom of the page, find the box that says “Output Records” (Figure H-2).
10. Select the radio button with the option “Records __ to __”. In the blank boxes, input “1” and “*the number of results from the search query*” in order to download data for all of the publications you’ve found.

*Note: You may download a maximum of 500 records from WOS at any one time, so it may be necessary to download the records in parts.*

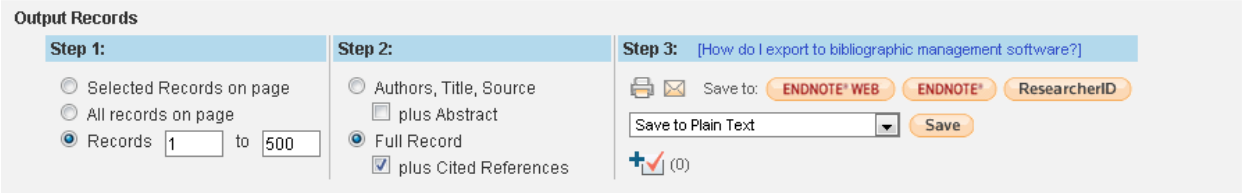



**Figure H-2. Output Records in Web of Science**

11. Select the radio button "Full Record" and check the box "plus Cited References".
12. Select the option "Save to Plain Text" in the drop down menu. Click the "Save" button.
13. Rename the downloaded file as the query you used in the "Author Finder" search (e.g., Abbott L.txt; e.g., Abbott L2.txt; etc.).
14. In the browser, return to the results page and click "Create Citation Report."
15. At the bottom of the resulting Citation Report page, select the radio button with the option "Records __ to __". In the blank boxes, input "1" and "*the number of results from the search query*" in order to download data for all of the publications you've found (Figure H-3).
16. Select the option "Save as Excel file," and click the "Save" button. Rename the resulting file with the STPICode. The citation report data will form the foundation for the publication data in spreadsheet form.

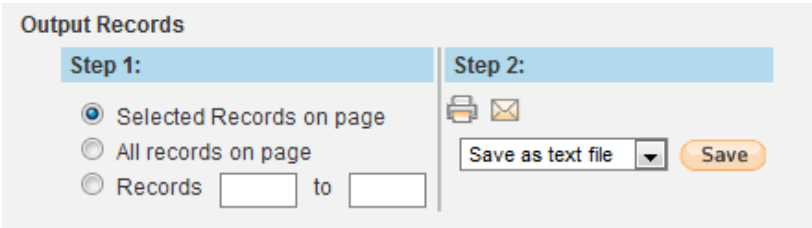



**Figure H-3. Output Citation Report in Web of Science**

## Data Cleaning Protocol

### Import Text Files into VantagePoint

Import each Text File individually into VP, saving the filename as the STPICode for each researcher (e.g., "Awardee1.vpt").

17. Select the files that need to be imported using the “Raw Data File (Use Import Filter)” (under File or when the window pops up when you first open VP).
18. The Import Filter that you will use is: “WOK-WOS (Field-Tagged text),” a filter custom-made for our purposes.

*Note: if the file does not initially show up—select “Select New Filter Directory” button in the bottom of the window and you can navigate to wherever you’ve saved the Import Filter file on your computer.*

19. Select all primary fields and add the secondary fields you want to import: select using Ctrl+
    a. Cited Journal
    b. Cited Authors
    c. Cited References
    d. Funding Award Numbers
    e. Funding Organization
20. Select Finish and let the file load.

**FIELD 1: Authors [TOTAL TIME: 10-15 Minutes]**

**For **Authors**—make sure that your final cleaned Authors field is labeled **“Authors FINAL”****

*Note: you can rename the final cleaned field or create a copy of the field by right-clicking on the field name in the Summary window—copy rather than rename the field if you have undergone a major cleanup process (used the most recent thesaurus) to create the field.*

21. Under Fields>List Cleanup—highlight the field “Authors” and use the General.fuz file in your Fuzzy folder (Figure H-4) (**do not** check “Verify Matches w/another Field”); you can specify the New Field to create and label this as the thesaurus or fuzzy function you are using (e.g., “Authors RunGeneral.fuz”).
22. In Cleanup Confirm window—make sure any groupings make sense and given your time allotted for cleaning, try to focus on those authors that have more rather than less Number of Records.
    f. Selecting All Items and Right-clicking to Sort>All Items>By Name will make it easier to see similar names.
    g. You can drag any names to a similar name to group names—please follow the standardization instructions below.

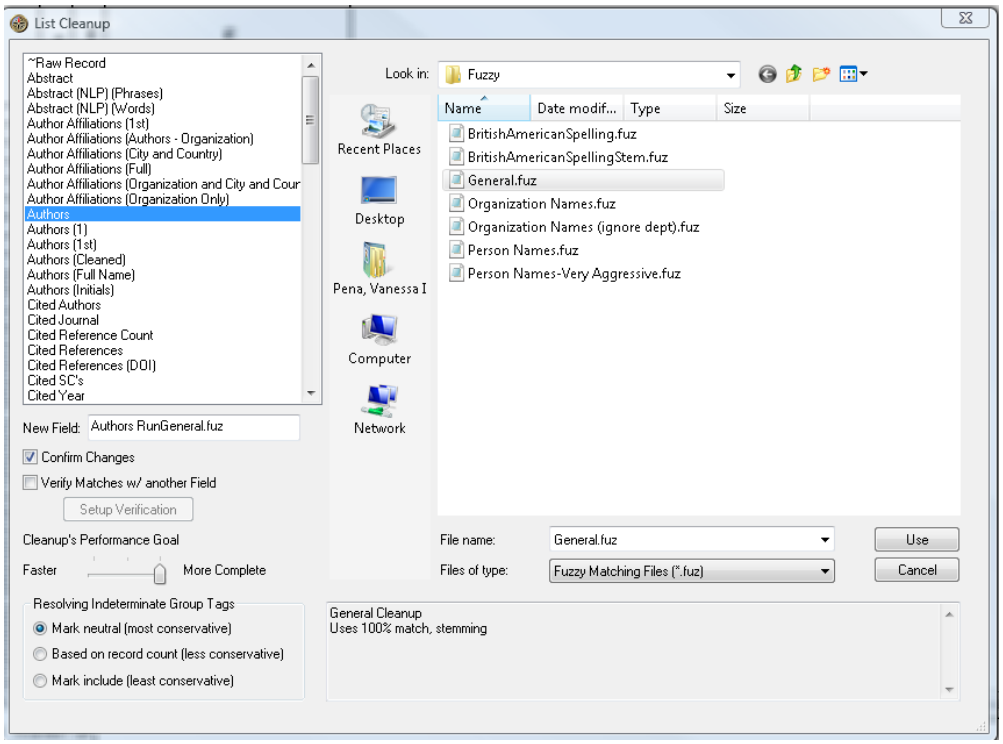



**Figure H-4. List Cleanup Window in VantagePoint**

*Note: When grouping names you think are similar, and if initials are missing and the last name is sufficiently "uncommon" (e.g., Turrigiano, G; Turrigiano G G, and Turrigiano, GG), then always group the one without the middle initial and without spaces into the name with the middle initial (e.g., new group name: Turrigiano, GG).*

*Note: When grouping similar names and one part of the name is spelled out and one isn't (e.g., Swinehart, Christian D and Swinehart, CD); then always group the spelled out name into the initial name (e.g., new group name: Swinehart, CD).*

*Note: For less common names (e.g., Zhang, Z and Zhang ZY) do not try grouping these until you have checked their affiliations in "(Optional) Step 24." It is best to stay conservative with your groupings.*

23. Once you are done grouping names, "Save as Thesaurus"—this will access your Thesaurus folder and save the File name as "NDPA-YourGroup-Authors" (e.g., NDPA- Pioneers -Authors.the).

*Note: You can now use this file in the future when cleaning authors for other PIs.*

24. (Optional) Making further changes — If there were similar names for which you would like to check their affiliations or look at all the publication record fields to determine if they are indeed the same individual, then you can view the new author list ("Authors RunGeneral.fuz") you have created and highlight the names you would like to investigate.

h. There is a default detail window to the left of VP that shows the Title of the publications for any highlighted cells in your list.

*Note: a general check may be to compare the publication titles of the individuals and check for similar keywords or context of the publications for those with similar names.*

i. You can add additional detail windows to the right of the interface by right-clicking and selecting "Create Detail Window"; you can choose a field to look at to compare affiliations, publication years, and subject categories of the names you've highlighted ( Figure H-5).

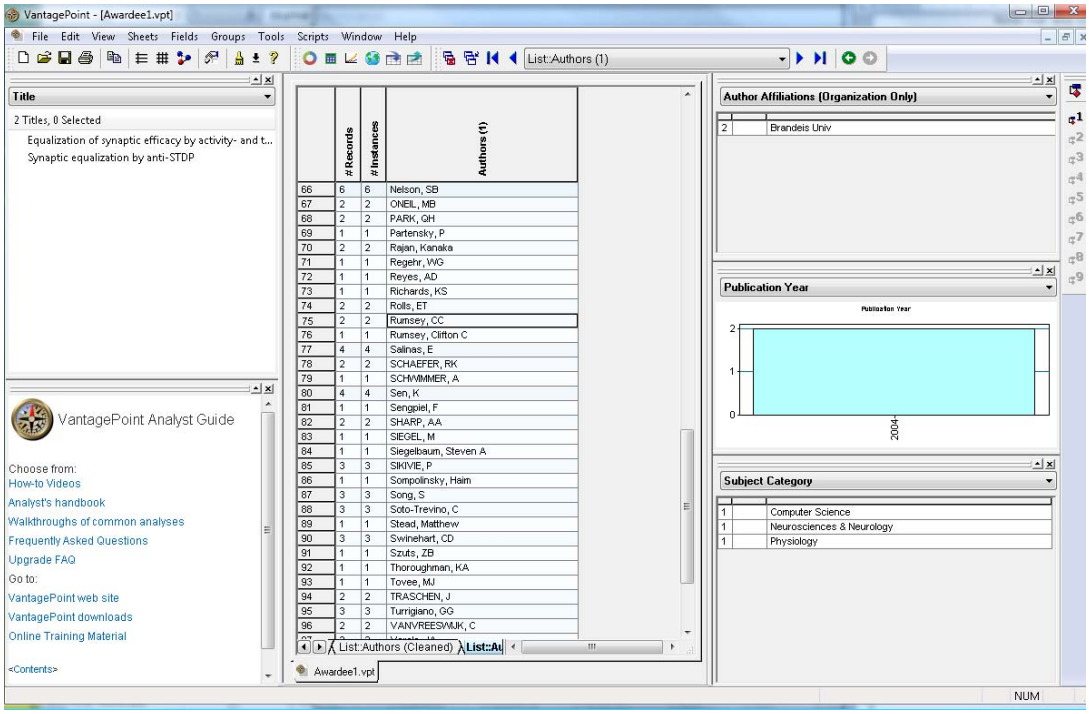



**Figure H-5. Highlighted Similar Names and Detail Windows in VantagePoint**

*Note: a general check on the Affiliations (Organization Only), Publication Year, and Subject Category should provide more than enough information to tell whether the similar names are really the same individual.*

j. If you would like to make additional changes to the list, you can edit and run the new Authors thesaurus ("NDPA- Pioneers -Authors.the")

k. To edit the Authors thesaurus—select Tools>Thesaurus editor…; open the Authors thesaurus, select the newest Authors field ("Author RunGeneral.fuz").

l. You will be adding a "Top Level Item" by right clicking on the second window that lists the contents of the thesaurus—you will then type in the

new label for the name (e.g., I have added Rumsey below by typing out the New Item name “Rumsey, CC” and highlighting the two Rumsey names in the Authors list (left) and dropping it into the “Rumsey, CC” top level item [ Figure H-6]). Save.

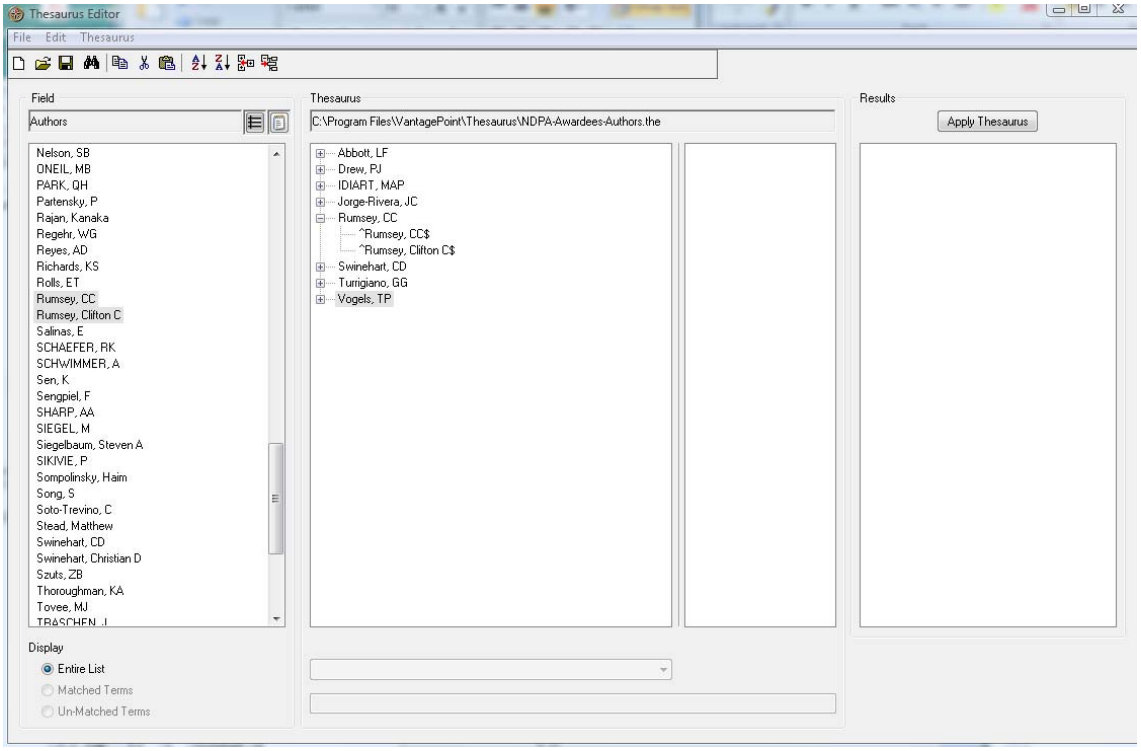



**Figure H-6. Thesaurus Editor Editing an Authors List and Thesaurus in VantagePoint**

*Note: The thesaurus editor automatically inserts any notations necessary when you drag and drop and there’s no need to further edit the names inserted in the group.*

*Note: You can check whether the thesaurus now recognizes Rumsey, Clifton C by selecting “Apply Thesaurus” and displaying Matched Terms (whether now this name is in the matched term list) or Un-matched Terms (whether the name is no longer in this list).*

m. Using the new edited thesaurus on your latest cleaned Authors list, you will specify a name for the new field.

25. If you are done with the Author cleanup process, then specify this field as “**Authors FINAL”**.

**FIELD 2: Author Affiliations [TOTAL TIME: 10-15 Minutes]**

**For Author Affiliations (Organization Only)—make sure that your final cleaned Affiliations field is labeled “Author Affiliations (Organization Only) FINAL”**

26. Review the list of Author Affiliations (to get a sense of similar affiliations).
27. Under Fields>List Cleanup…—highlight the field "Author Affiliations (Organization Only)" and use the "NDPA-Organization-Names.fuz" file in your Fuzzy folder (**do not** check "Verify Matches w/another Field"); you can specify the New Field to create and label this as the thesaurus or fuzzy function you are using (e.g., "Author Affiliations (Organization Only) RunNDPAOrg.fuz").
28. In Cleanup Confirm window—add or remove groupings if necessary as indicated in Field1: Authors—Step 2.
29. Once you are done grouping names, "Save as Thesaurus"—this will access your Thesaurus folder and save the File name as "NDPA-*YourGroup*-Affiliations" (e.g., NDPA- Pioneers -Affiliations.the).

*Note: You can now use this file in the future when cleaning affiliations for other PIs.*

*Note: Similarly, you can edit the new Affiliations thesaurus if you find after reviewing that your new Affiliations list requires further cleaning—Field1: Authors-(Optional) Step 4.*

30. If you are done with the Author Affiliation cleanup process, then specify this field as **"Author Affiliations (Organization Only) FINAL".**

**Grouping by Publication Years [TOTAL TIME: 20-30 Minutes]**

31. Refer to the provided data where the PI's award year is noted.
32. In the Summary window, double-click on the Publication Year field and right-click to sort by year.
    n. Highlight all the years previous to and including the year in which the PI received the award and right-click to "Add Selection to Group…". Label the New Group as "Before".
    o. Highlight all the years after the award year of the PI and right-click to "Add Selection to Group…", and label the New Group as "After".
33. Add Detailed Windows (right-click) to show the fields: (1) Authors FINAL and (2) Author Affiliations (Organizations Only) FINAL ( Figure H-7).
34. Count the unique authors and affiliations.
    p. Highlight the "Before" column created in your Publication Year list and count the number of unique authors and affiliations are associated with that time period. Record these numbers.

q. Highlight the “After” column created in your Publication Year list and count the number of unique authors and affiliations are associated with that time period. Record these numbers.

*Note: Save the excel file as their STPI_Code (e.g., “Awardee1VPData.xlsx”) for our records.*

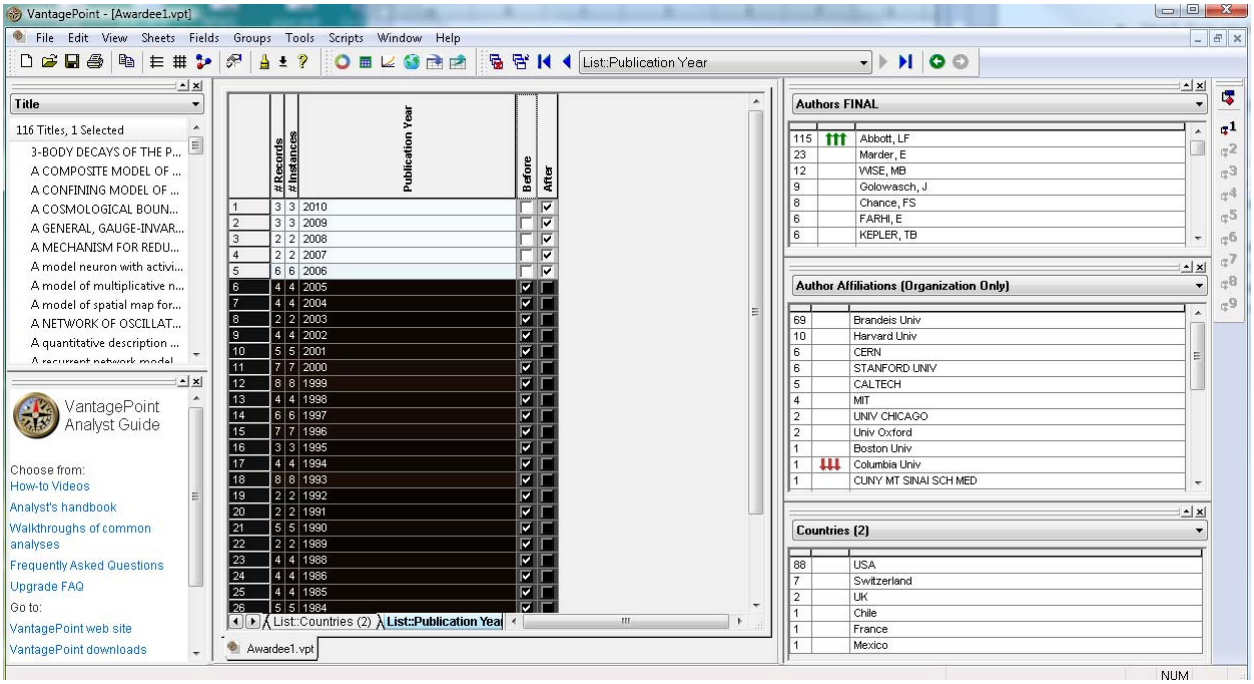



**Figure H-7. Detailed Windows and Before and After Groups in Publication Year Field in VantagePoint**

## FIELD 4: Cited Journals and Interdisciplinarity Calculations [TOTAL TIME: 20-30 Minutes]

**You will be creating new fields: “Cited SC FINAL” and creating two subsets of the data in new VantagePoint files “Before” and “After”.**

35. Open the provided spreadsheet with WOS Journals matched to WOS Subject Categories (WOS SC-Journals spreadsheet).

36. Clean the Cited Journals Field - Under Tools>Thesaurus editor…; open the most up-to-date Journal-Subject Category (J-SC) thesaurus, and select the field “Cited Journals”.

    r. Apply the thesaurus and check the Un-Matched Terms.

    s. Use the WOS SC-Journals spreadsheet to lookup the individual journals and check to see (1) if the journal is indexed in WOS; (2) if it is, then to edit the

thesaurus by dragging the journal to the subject categories that are linked to the journal. Save.

*Note: There may be multiple subject categories for the journal, so make sure to drop the journal into each one of the subject categories in the second window of the thesaurus editor.*

*Note: The WOS SC-Journals spreadsheet may have variations of the journal name so make sure to search variations of the journal or use wildcard notation when searching for the journal name.*

37. Once satisfied with the editing, use the J-SC thesaurus on the Cited Journal Field. IMPORTANT: check "Allow Multiple Matches". Specify the new field as "Cited SC RunJ-SC.the" Confirm that the thesaurus has not left out journals with high frequency (number of publications). If you need to make further changes—use the thesaurus editor (Step 2). If you are satisfied with the changes, then rename the field as "**Cited SC FINAL"**.

*Note: As a general rule of thumb, journals that are Un-Matched Terms in the J-SC thesaurus and with at least 5 publications should be searched in the WOS SC-Journals spreadsheet.*

38. Subset the file/data to calculate the Integration Score for "Before" and "After" publications:

    t. Open the Publication Year list and highlight the "Before" labeled grouping (this will highlight all cells for years grouped as "Before.")

    u. Go to File-Create Sub-dataset to create a new VantagePoint file with only the highlighted cells.

        1) Make sure in the Create Sub-dataset window that Selection is selected. Select Ok. ( Figure H-8)

        2) Save the file as the "STPICode_Before" (e.g., Awardee1_Before)

    v. Create a new VantagePoint file for the "After" labeled grouping in the Publication Year list in a similar manner and save as "STPICode_After" (e.g., Awardee1_After).

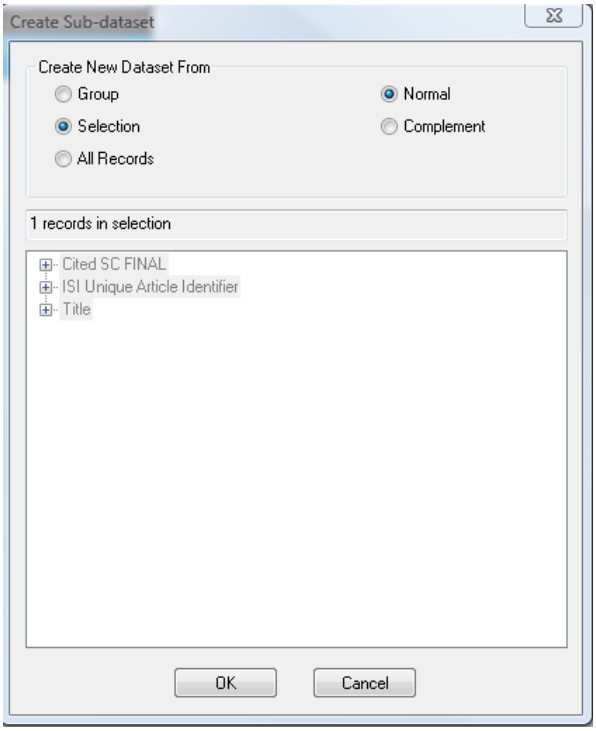



**Figure H-8. Create Sub-dataset Window in VantagePoint**

39. For each of the new "Before" and "After" VantagePoint files, calculate the Integration Score using the Calculate Integration v3.vpm Script (Scripts>Run Script…). When calculating the Integration Score, you will be prompted to enter:

    w. "Author": enter the "File Name" field.

    x. "Subject Category": enter the "Cited SC FINAL" field.

    y. "Cited Subject Category": also enter the "Cited SC FINAL" field.

40. The Integration Scores will show up in an excel file that is opened by the script; record this data for the "before" and "after" periods.

## Data Compilation

To create the final publication data spreadsheet, we compiled variables from the citation report data, VantagePoint field export, SCImago Journal Rank (SJR), PubMed, and the NIH's Query/View/Report database. For each publication record, we collected the following information for our analyses: publication title, authors, publication year, journal, citations accumulated each year from 1980 to 2011, document type, WOS category, and SJR.

We merged the researcher-level, grant-level, and “top five” publication records into a single file using a rough title match and removed duplicate records. Additional data-cleaning steps were performed. We checked that publication records were appropriately marked as researcher level and grant level; for instance, the author field for researcher-level publications needed to have the author’s last name, and the grant-level publications needed to be published after the year of award but before 2012, and be a research article (as opposed to a review, editorial material, etc.). We matched the correct SJRs to journal names using a combination of automated methods and by hand. Where systematic errors in the data were found for individuals or grants, we removed all publications associated with that individual or grant and downloaded and cleaned the publication data again. Once the data were corrected in the spreadsheet, we edited the corresponding researcher-level and grant-level VantagePoint files so that the data were consistent across all interfaces.

# Appendix I
# Statistical Details

## Basic Model

The model is discussed in terms of impact, but the same structure was used for both impact and innovativeness. The basic structure of the model is adapted from Johnson and Albert (1999).

Assume that each packet has an underlying impact that can be measured. Let $\mu_i$ denote the impact of packet i. We assume that the packet impacts are independent. To establish a scale for impact, we assume that $\mu_i \sim \text{Normal}(0, 1)$.

We assume that the expert has some "measurement error" ($\sigma_k^2$) when evaluating the impact of packet i. Let $t_{ik}$ denote the perceived impact of packet i by expert k. We assume that $t_{ik} \sim \text{Normal}(\mu_i, \sigma_k^2)$. We do not observe either $\mu_i$ or $t_{ik}$. Instead, we assume that the expert discretizes his individual impact scale, and we observe which "bin" contains the assessed impact. For each expert, let $\gamma_{kl}$ denote the bin cutoffs for expert k. We take the leftmost cutoff (l = 6) as negative infinity and the rightmost cutoff (l = 1) as positive infinity. The other bin cutoffs are assumed to be ordered. This notation is displayed in Figure I-1. The expert would assess the packet as a 2.

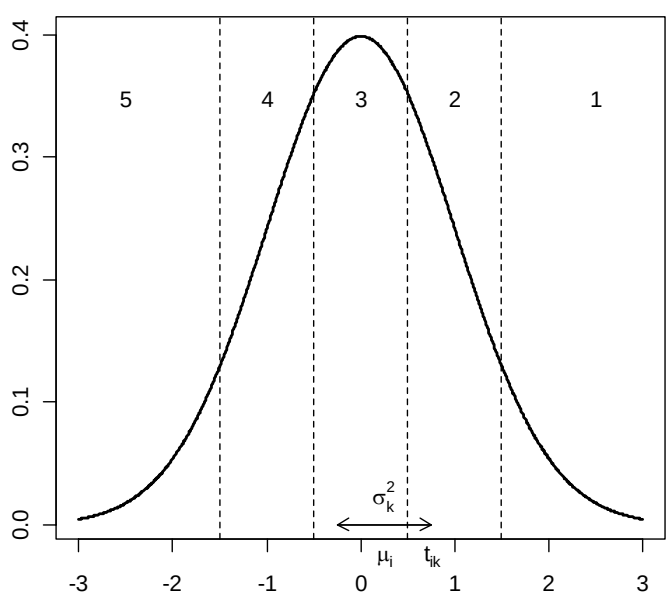



**Figure I-1. Bayesian Ordinal Data Model**

We use a similar structure for papers. Let $z_{ij}$ denote the impact of paper j from packet i. We assume that paper rankings are distributed around the packet ranking, so $z_{ij} \sim \text{Normal}(\mu_i, \tau_i^2)$. Let $\rho_{ijk}$ denote the assessed impact of paper j from packet i by expert k. We assume that the expert k has some "measurement error" when evaluating the impact of paper j from packet i, so $\rho_{ijk} \sim \text{Normal}(z_{ij}, \omega_k^2)$.

The rating for each packet i by expert k has likelihood $\Phi(\gamma_k, \mu_i, \sigma_k^2) - \Phi(\gamma_{k-1}, \mu_i, \sigma_k^2)$, where $\Phi$ denotes the standard normal cumulative distribution function. The rating of each paper j from packet i by expert k has likelihood $\Phi(\gamma_k, z_{ij}, \omega_k^2) - \Phi(\gamma_{k-1}, z_{ij}, \omega_k^2)$.

We fit the model using a Bayesian approach, which requires prior distributions to be specified for each parameter. We require that the $\gamma$ be ordered, but otherwise use a non-informative (flat) prior distribution. We use Inverse Gamma (10,3) distributions for the variance parameters $\sigma_k^2$, $\omega_k^2$, and $\tau_i^2$.

The model is fit using standard Markov chain Monte Carlo techniques. The figures in Chapters 2 and 5 are normalized so that the estimated impact and innovation ranged between 0 and 1; the figures in this appendix are on the original Normal (0, 1) scale.

## Regression Model

When we add covariates to the model, we are interested in seeing if we can identify which variables are predictive of the underlying packet or paper impact as assessed by the experts. The response variable ("true" packet or paper impact), however, is unobserved. We estimate the impact from the expert-rating data, and we must account for the variability inherent in the estimate when we fit our regression model. In addition, our observations are not independent: we have assumed that the packet ratings are informative about the paper ratings and vice versa. Consequently, the estimates are somewhat different from the ordinary least squares estimates.

We modify the model above by assuming that the expected value for packet impact ($\mu_i$) is $X_i\beta$, where $X_i$ is a vector of observed packet or investigator characteristics. In particular, we consider granting mechanism (NDPA, R01, HHMI), number of pre-award citations for the PI, SCImago Institutions Ranking (SIR) for the PI at time of award, number of pre-award publications for the PI, PI h-index at time of award, years since degree, receipt of early career award, and, for the R01 awards, total direct costs. A description of these variables is found in Chapter 2, Section A.

We also assume that the expected value for paper impact ($z_{ij}$) is $\mu_i + W_i\delta$, where $W_i$ is a vector of observed paper characteristics. In particular, we consider journal impact factor, whether the paper is a review paper, and total citations through 2011. These variables are also described in Chapter 2, Section A.

To complete our Bayesian specification, we choose prior distributions for our parameters $\beta$ and $\delta$. To keep the correct marginal distribution for the $\mu_i$, we assume that

$$\beta \mid \mu, \nu \sim \text{Normal}((X^TX)^{-1}X^T\mu, \nu^2(X^TX)^{-1})$$

and

$$\nu \sim \text{InverseGamma}(0.5(\text{number of grants} - \text{number of covariates}), 0.5*(\mu - X\beta)^T(\mu - X\beta)).$$

We choose a flat, non-informative prior distribution for $\delta$.

## Results

We have no expectation that the regression models will be a good surrogate for the expert ratings: it is unlikely that simply knowing the grant mechanism, characteristics of the PI, and the journal where the paper appears will fully characterize its impact or innovativeness. Table I-1 summarizes the estimated regression coefficients for impact. A normalized variable has had its mean subtracted and is divided by the standard deviation so that the covariate has mean 0 and variance 1.

**Table I-1. Regression Results for Impact**

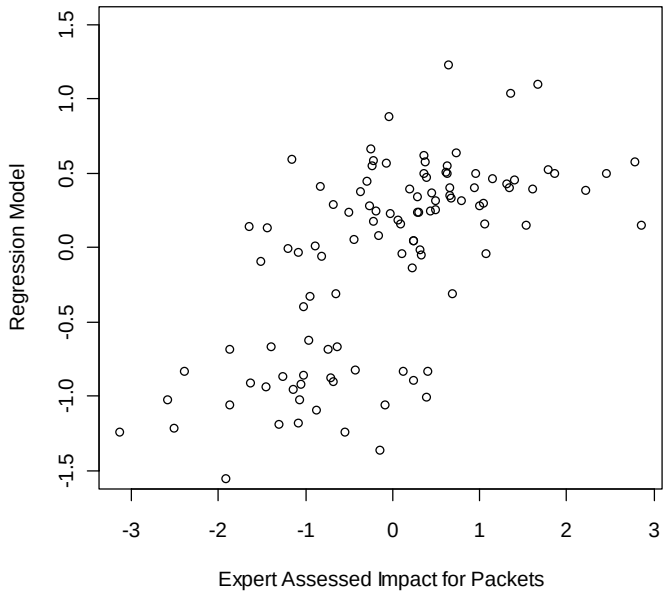
| Level | Metric | Estimate | 90% Credible Interval | Probability $\beta < 0$ |
|---|---|---|---|---|
| **PI-Level** | NDPA | 0.69 | (–0.21, 1.6) | 0.10 |
| | Matched R01 | –1.3 | (–2.6, 0.0041) | 0.95 |
| | HHMI | 1.0 | (0.17, 1.9) | 0.024 |
| | Pre-Award citations (normalized) | 0.39 | (0.039, 0.74) | 0.034 |
| | SCImago Institutions Ranking | –0.0012 | (–0.0035, 0.0010) | 0.82 |
| | Pre-award publications | –0.000094 | (–0.0092, 0.0091) | 0.51 |
| | H-index at award | –0.020 | (–0.061, 0.022) | 0.78 |
| | Years since degree | –0.0016 | (–0.046, 0.043) | 0.53 |
| | Early career award | 0.20 | (–0.26, 0.67) | 0.24 |
| | Direct Costs ($M) | 0.88 | (–0.14, 1.92) | 0.078 |
| **Paper-Level** | Journal impact factor (normalized) | 0.44 | (0.34, 0.55) | 0.00 |
| | Review paper | –0.88 | (–0.21, –0.53) | 1.0 |
| | Total citations (normalized) | 0.16 | (0.056, 0.27) | 0.0068 |

The linear model is:

predicted packet impact = 0.69*(Is NDPA?) – 1.3*(Is Matched R01?) + 1.0*(Is HHMI?)
+ 0.39*(normalized pre-award citations) – 0.0012*(SCImago Institutions Ranking)
– 0.000094*(pre-award publications) –0.020*(h-index at award)
– 0.0016*(years since degree) + 0.20*(early career award?)
+ 0.88(direct costs in $M)

predicted paper impact = predicted packet impact + 0.44*(normalized journal impact factor)
– 0.88*(review paper?) + 0.16*(normalized total citations through 2011)

The correlation between the expert-estimated "true" impact for packets and the regression model is 0.64. The correlation between the expert-estimated "true" impact for papers and the regression model is 0.66. See Figure I-2.



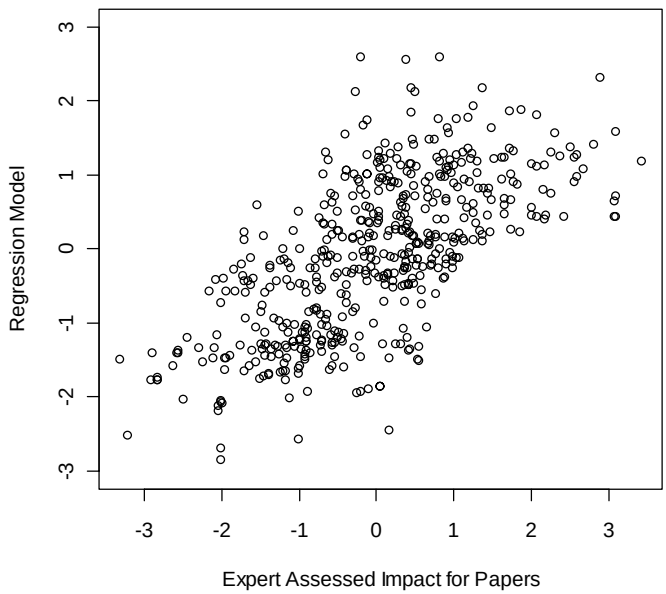




**Figure I-2. Regression Models for Impact**

Table I-2 summarizes the estimated regression coefficients for innovation. The linear model is:

predicted packet impact = 0.95*(Is NDPA?) – 0.19*(Is Matched R01?) + 0.98*(Is HHMI?)
+ 0.42*(normalized pre-award citations) – 0.0024*(SCImago Institutions Ranking) – 0.00095*(pre-award publications) –0.018*(h-index at award)
– 0.0099*(years since degree) + 0.089*(early career award?)
+ 0.11(direct costs in $M)

predicted paper impact = predicted packet impact + 0.34*(normalized journal impact factor) – 1.6*(review paper?) + 0.018*(normalized total citations through 2011)

The correlation between the expert-estimated "true" innovation for packets and the regression model is 0.56. The correlation between the expert-estimated "true" innovation for papers and the regression model is 0.55. See Figure I-3.

**Table I-2. Regression Results for Innovation**

| | | Estimate | 90% Credible Interval | Probability β < 0 |
|---|---|---|---|---|
| **PI-Level** | **NDPA** | 0.95 | (0.020, 1.9) | 0.047 |
| | Matched R01 | –0.19 | (–1.6, 1.2) | 0.59 |
| | HHMI | 0.98 | (0.085, 1.9) | 0.036 |
| | Pre-Award citations (normalized) | 0.42 | (0.041, 0.79) | 0.035 |
| | SCImago Institutions Rank | –0.0024 | (–0.0048, 0.000080) | 0.94 |
| | Pre-award publications | –0.00095 | (–0.010, 0.0086) | 0.57 |
| | H-index at award | –0.018 | (–0.063,.026) | 0.75 |
| | Years since degree | –0.0099 | (–0.057, 0.038) | 0.63 |
| | Early career award | 0.089 | (–0.39, 0.57) | 0.38 |
| | Direct Costs ($M) | 0.11 | (–0.97, 1.2) | 0.44 |
| **Paper-Level** | Journal impact factor (normalized) | 0.34 | (0.25, 0.44) | 0 |
| | Review paper | –1.6 | (–2.0, –1.2) | 1.0 |
| | Total citations (normalized) | –0.018 | (–0.11,0.08) | 0.61 |

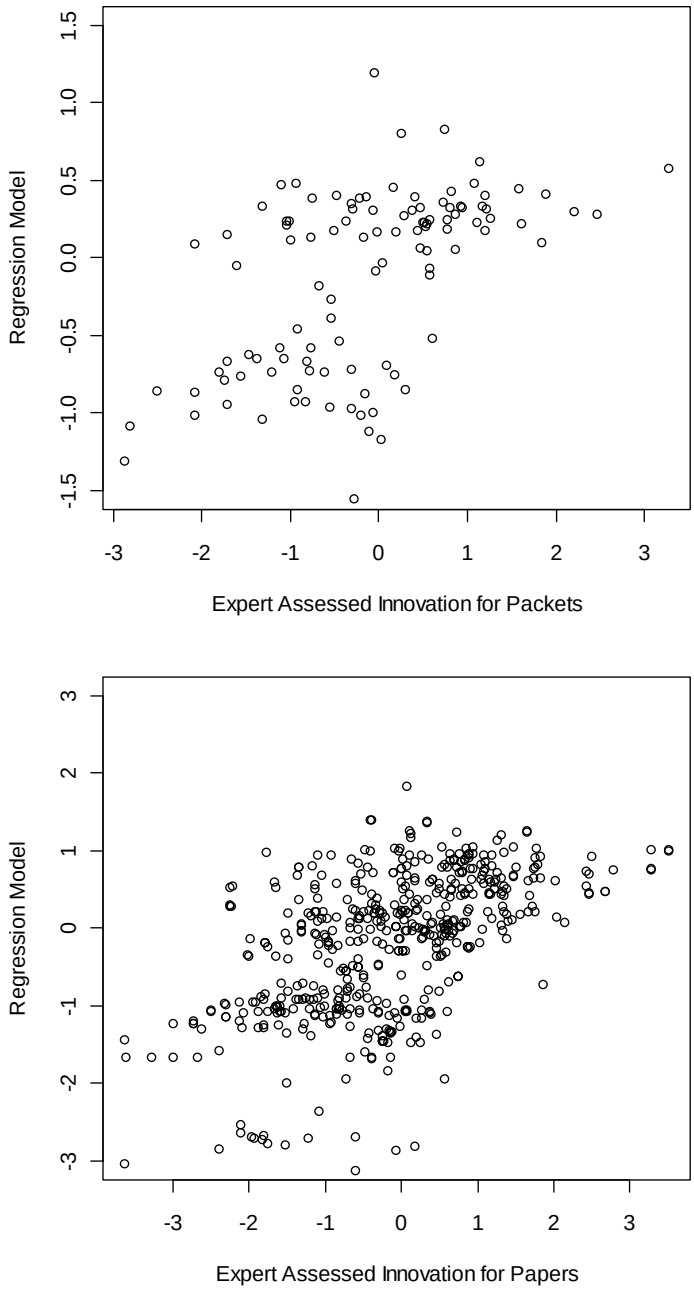



**Figure I-3. Regression Models for Innovation**

We conducted a preliminary analysis of an additional covariate: the number of grants cited on each publication. We had data on 200 of the 500 "top five" papers from NDPA and the matched R01s. The analysis suggested that having additional grants cited on the paper increases both impact and innovation.

# References


Alonso, S., F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera. 2009. "h-Index: A Review Focused in Its Variants, Computation and Standardization for Different Scientific Fields." *Journal of Informetrics* 3: 273–289. doi:10.1016/j.joi.2009.04.001.

Azoulay, P., J .S. G. Zivin, and G. Manso. 2011. "Incentives and Creativity: Evidence from the Academic Life Sciences." *RAND Journal of Economics* 42: 527–554.

Bornmann, L. and D. Hans-Dieter. 2008. "What Do Citation Counts Measure? A Review of Studies on Citing Behavior." *Journal of Documentation* 64(2008), 45–80. doi: 10.1108/00220410810844150.

Boyack, K. W, R. Klavans, and K. Börner. 2005. "Mapping the Backbone of Science." *Scientometrics* 64(3): 351–374.

Brenner, S. 1998. "The Impact of Society on Science." *Science* 282 (5393): 1411–12. doi: 10.1126/science.282.5393.141.

Burrell, Q. L. 2007. "Hirsch index or Hirsch rate? Some Thoughts Arising from Liang's Data." *Scientometrics* 73(1): 19–28.

Cameron, B. 2005. "Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications." Librarian and Staff Publications. Paper 3. http://digitalcommons.ryerson.ca/library_pubs/3.

Colwell, R. R. 2003. "Keynote Address to the International Life Sciences Summit of Georgetown University." Given at the International Life Sciences Summit of Georgetown University in Washington, D.C., on October 20, http://www.nsf.gov/news/speeches/colwell/rc031020lifesci_summit.htm.

Cozzens, S. 1996. "Assessment of Fundamental Science Programs in the Context of the Government Performance and Results Act (GPRA)." Washington, D.C.: RAND.

de Solla Price, D. J. 1970. "Citation Measures of Hard Science, Soft Science, Technology, and Nonscience." In *Communication among Scientists and Engineers*, edited by C. E. Nelson and D. K. Pollock, 3–22. Lexington, MA: D.C. Heath and Company.

Foo, J. Y. A. 2011. "Impact of Excessive Journal Self-Citations: A Case Study on the Folia Phoniatrica Et Logopaedica Journal." *Science and Engineering Ethics* 17 (1): 65–73.

Garfield, E. 1955. "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas." *Science* 122 (1955), 108–111. Reprinted in *Essays of an Information Scientist* 6 (1983), 468–471.

Garfield, E. 1972. “Citation Analysis as a Tool in Journal Evaluation.” *Science* 178: 471–479.

Garfield, E. 1979. *Citation Indexing. Its Theory and Application in Science, Technology and Humanities.* New York: Wiley.

Garfield, E. 2005. “The Agony and the Ecstasy-The History and Meaning of the Journal Impact Factor.” Presented at the International Congress on Peer Review and Biomedical Publication, Chicago, http://garfield.library.upenn.edu/papers/jifchicago2005.pdf

Godin, B., and C. Doré. 2004. “Measuring the Impacts of Science: Beyond the Economic Dimension.” Montreal: INRS.

González-Pereira, B., V. P. Guerrero-Bote, and F. Moya-Anegón. n. d. “The SJR Indicator: A New Indicator of Journals’ Scientific Prestige.” http://arxiv.org/ftp/arxiv/papers/0912/0912.4141.pdf.

Guo, S., and M. W. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications.* Sage Publications, Thousand Oaks, CA.

Heinze, T., P. Shapira, J. Senker, and S. Kuhlmann. 2007. “Identifying Creative Research Accomplishments: Methodology and Results for Nanotechnology and Human Genetics.” *Scientometrics* 70(1): 125–152.

Hirsch, J. 2005. “An Index to Quantify an Individual’s Scientific Research Output.” *Proceedings of the National Academies of Science* 102(46): 16569–16572.

Iglesias, J. E., and C. Pecharroman. 2007. “Scaling the h-index for Different Scientific ISI Fields.” *Scientometrics* 73 (3): 303–320.

Jascó, P. 2011. “The H-Index, H-Core Citation Rate and the Bibliometric Profile of the Scopus Database.” *Online Information Review* 35 (3): 492–501.

Johnson, V., and J. Albert. 1999. *Ordinal Data Modeling*. New York: Springer-Verlag.

Klein, J. T. 1990. *Interdisciplinarity: History, Theory, and Practice*. Detroit, MI: Wayne State University Press.

Lal, B., M. E. Hughes, S. Shipp, E. C. Lee, A. M. Richards, and A. Zhu. 2011. “Outcome Evaluation of the National Institutes of Health (NIH) Director’s Pioneer Award (NDPA), FY 2004–2005: Final Report.” IDA Science and Technology Policy Institute. http://commonfund.nih.gov/pdf/Pioneer_Award_Outcome%20Evaluation_FY2004-2005.pdf.

Lal, B., R. Chaturvedi, A. Zhu, M. B. Hughes, S. Shipp, C. Kang, A. Marshall, and E. Yglesias. 2010. “Process Evaluation of the National Institutes of Health Director’s Pioneer Award Program: FY 2004–2008,” IDA Document D-4014, January. http://commonfund.nih.gov/pdf/PioneerAwardProcessEvaluation_2004-2008.pdf.

Lasswell H., and N. Leites. 1968. [1949.] *Language of Politics: Studies in Quantitative Semantics*. Cambridge, MA: MIT Press. 2nd ed.

Leydesdorff, L., and I. Rafols. 2009. "A Global Map of Science Based on the ISI Subject Categories." *Journal of the American Society for Information Science and Technology* 60, no. 2 (2009): 14.

Martin, B. R. and J. Irvine. 1983. "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy." *Research Policy* 12 (1983): 61–90.

Melkers, J. 1993. "Bibliometrics as a Tool for Analysis of R&D Impacts." In *Evaluating R&D Impacts: Methods and Practice*, edited by B. Bozeman and J. Melkers, 43–62. Kluwer Academic Publishers.

Mervis, J. 2004. "Male Sweep of New Award Raises Questions of Bias." *Science* 306 (5696): 595. doi: 1126/science.306.5696.595a.

Moed, H. F. 2005. *Citation Analysis in Research Evaluation.* Vol. 9, *Information Science and Knowledge Management*. Springer.

Moed, H. F. 2010. "Measuring Contextual Citation Impact of Scientific Journals." *Journal of Informatics* 4, 265–277.

Moed, H. F., R. E. de Bruin, Th. N. van Leeuwen. 1995. "New Bibliometric Tools for the Assessment of National Research Performance: Database Description, Overview of Indicators and First Applications." *Scientometrics* 33: 381–422.

Moed, H. F., W. J. M. Burger, J. G. Frankfort, and A. F. J. van Raan. 1985. "The Use of Bibliometric Data for the Measurement of University Research Performance." *Research Policy* 14: 131–149.

Narin, F. 1976. *Evaluative Bibliometrics. The Use of Publications and Citation Analysis in the Evaluation of Scientific Activity.* Computer Horizons, Inc., Cherry Hill, New Jersey: Computer Horizons, Inc.

National Academy of Sciences (NAS). 1999. "Evaluating Federal Research Programs. Research and the Government Performance and Results Act." Washington, D.C.

National Academy of Sciences (NAS). 2011. "Measuring the Impacts of Federal Investments in Research: A Workshop Summary." Washington, D.C.

National Institutes of Health, 2004. "Proceedings of the 88th Meeting of the Advisory Committee to the Director," National Institutes of Health, May 6, http://www.nih.gov/about/director/acd/minutes/050604acd.htm.

National Institutes of Health, 2007. "The QVR System—An Example of CIT-eRA Cooperation." *Interface* 237 (Spring). Accessed on November 10, 2011. http://datacenter.cit.nih.gov/interface/interface237/QVR.html.

National Science Foundation. 2001. "A Descriptive Analysis of the Presidential Faculty Fellows Program: Contributions to Science and Engineering through Leadership in Research and Teaching." NSF Report 01118 (July). http://www.nsf.gov/pubs/2001/nsf01118/nsf01118.pdf.

National Science Foundation. 2010. "Doctorate Recipients from U.S. Universities: 2009." NSF 11-306. http://www.nsf.gov/statistics/nsf11306/.

National Science Board (NSB). 2010. *Science and Engineering Indicators 2010*. Arlington, VA: National Science Foundation (NSB 10-01).

Oppenheim, C. 1995. "The Correlation between Citation Counts and the 1992 Research Assessment Exercise Ratings for British Library and Information Science University Departments." *Journal of Documentation* 51: 18–27.

Oppenheim, C. 1997. "The Correlation between Citation Counts and the 1992 Research Assessment Exercise Ratings for British Research in Genetics, Anatomy and Archaeology." *Journal of Documentation* 53: 477–487.

Porter, A. L., A. S. Cohen, D. Roessner, and M. Perreault. 2007. "Measuring Researcher Interdisciplinarity." *Scientometrics* 72(1): 117–147.

Porter, A. L., and I. Rafols. 2009. "Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time." *Scientometrics* 81 (3): 719–745.

Rafols, I., and M. Meyer. 2010. "Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience." *Scientometrics* 82 (2): 263–287.

Rafols, I, A., L. Porter, and L. Leydesdorff. 2010. "Science Overlay Maps: A New Tool for Research Policy and Library Management." *Journal of the American Society for Information Science and Technology* 61 (9): 1871–1887.

Rafols, I., and L. Leydesdorrf. 2009. "Content-Based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects." *Journal of the American Society for Information Science and Technology* 60, no. 9 (2009): 1823–1835.

Rosenbaum, P. 2005. "An Exact Distribution-Free Test Comparing Two Multivariate Distributions Based on Adjacency." *Journal of the Royal Statistical Society*, Series B 67 (4): 515–530.

SCImago Research Group. 2011a. "SCImago Institutions Rankings: SIR World Report 2011, Northern America Supplement." http://www.scimagoir.com/pdf/sir_2011_world_report_northern_america.pdf.

SCImago Research Group. 2011b. "SCImago Journal & Country Rank." http://www.scimagojr.com/journalrank.php.

Seglen, P. O. 1997. "Why the Impact Factor of Journals should Not Be Used for Evaluating Research." *British Medical Journal* 314: 498–502.

Stuart, E. A. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* 25 (1): 1–21.

Talley, E. M., D. Newman, D. Mimno, B. W. Herr II, H. M. Wallach, G. A. P. C. Burns, A. G. Miriam Leenders, and A. McCallum. 2011. "Database of NIH Grants Using Machine-Learned Categories and Graphical Clustering." *Nature Methods* 8(6): 443–444.

Thomson Reuters. 1994. "The Thomson Reuters Impact Factor." http://thomsonreuters.com/products_services/science/free/essays/impact_factor.

van Raan, A. F. J. 2002. "Assessment of the Scientific Basis of Interdisciplinary, Applied Research: Application of Bibliometric Methods in Nutrition and Food Research." *Research Policy* 31: 611–632.

van Raan, A. F. J. 2003. "The Use of Bibliometric Analysis in Research Performance Assessment and Monitoring of Interdisciplinary Scientific Developments." *Technology Assessment-Theory and Practice* 1: 20–29.

van Raan, A. F. J. 2005. "Comparison of the Hirsch-Index with Standard Bibliometric Indicators and with Peer Judgment for 147 Chemistry Research Groups." http://arxiv.org/ftp/physics/papers/0511/0511206.pdf.

# Abbreviations

| | |
|---|---|
| CAREER | Faculty Early Career Development |
| eSPA | electronic Scientific Portfolio Assistant |
| FY | fiscal year |
| HHMI | Howard Hughes Medical Institute |
| ICs | Institutes and Centers (of the NIH) |
| ISI | Institute for Scientific Information |
| JIF | Journal Impact Factor |
| M | million |
| NCI | National Cancer Institute |
| NCRR | National Center for Research Resources |
| NDPA | NIH Director's Pioneer Award |
| NEI | National Eye Institute |
| NHLBI | National Heart, Lung, and Blood Institute |
| NIA | National Institute on Aging |
| NIAID | National Institute of Allergy and Infectious Diseases |
| NIAMS | National Institute of Arthritis and Musculoskeletal and Skin Diseases |
| NIBIB | National Institute of Biomedical Imaging and Bioengineering |
| NIDA | National Institute of Drug Abuse |
| NIDCD | National Institute of Deafness and Other Communication Disorders |
| NIDCR | National Institute Institutes of Dental and Craniofacial Research |
| NIDDK | National Institute of Diabetes and Digestive and Kidney Diseases |
| NIEHS | National Institute of Environmental Health Sciences |
| NIGMS | National Institute of General Medical Sciences |
| NIH | National Institutes of Health |
| NIMH | National Institute of Mental Health |
| NINDS | National Institute of Neurological Disorders and Stroke |
| NSF | National Science Foundation |
| NYI | NSF Young Investigator Program |
| PECASE | Presidential Early Career Award for Scientists and Engineers |
| PFF | Presidential Faculty Fellows Program |
| PI | Principal Investigator |
| PYI | Presidential Young Investigator award |
| QVR | Query/View/Response |
| SC | subject category |

| | |
|---|---|
| SIR | SCImago Institutions Ranking |
| SJR | SCImago Journal Rank |
| SPIRES | Scientific Publication Information Retrieval and Evaluation System |
| STPI | Science and Technology Policy Institute |
| WOS | Web of Science |

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

**1. REPORT DATE** *(DD-MM-YYYY)*
XX-08-2012

**2. REPORT TYPE**
Final

**3. DATES COVERED** *(From - To)*
Aug 2011 – Aug 2012

**4. TITLE AND SUBTITLE**
An Outcome Evaluation of the National Institutes of Health (NIH) Dirctor's Pioneer Award (NDPA) Program, FY 2004–2006

**5a. CONTRACT NUMBER**
OIA-0408601

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Lal, Bhavya
Wilson, Alyson
Jonas, Seth
Lee, Elizabeth
Richards, Amy
Peña, Vanessa

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**
OSTP-20-0004.87 (NH-20-6440)

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Science and Technology Policy Institute
1899 Pennsylvania Avenue, NW, Suite 520
Washington, DC 20006-3602

**8. PERFORMING ORGANIZATION REPORT NUMBER**
IDA Paper P-4899

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Office of the NIH Director
Division of Program Coordination, Planning, and Strategic Initiatives
Bldg. 1, Room 201
Bethesda, MD 20892

**10. SPONSOR/MONITOR'S ACRONYM(S)**
NIH

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
This Draft Final Publication has not been approved for distribution and release. Reproduction or use of this material is not authorized without prior permission from the Director of the IDA Science and Technology Policy Institute.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
This evaluation addressed two primary study questions: To what extent does the research supported by the Pioneer program produce unusually high impact, and to what extent are the research approaches used by the Pioneers highly innovative? Using complementary methods (bibliometric analysis and expert review), STPI compared outcomes for five groups: a matched set of NIH Research Project Grant Program (R01) grants, principal investigators on the same R01 grants, random R01 portfolios that received comparable amounts of funding to the NDPA portfolio, a group of Howard Hughes Medical Institute (HHMI) investigators, and a group of NDPA finalists. STPI found that the Pioneers performed as well as or better than the matched R01 PIs and NDPA finalists. We also found that the Pioneer portfolio performed as well as or better than the random R01 portfolios on bibliometric measures of impact. Lastly, we found that HHMI investigators scored as well as or better than the Pioneers on all metrics of interest.

**15. SUBJECT TERMS**
transformative research, outcome evaluation, high risk high reward research, comparison groups, observational study, quasi-experimental design evaluation

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Wilder, Elizabeth |
| Unclassified | Unclassified | Unclassified | Same as Report | | 19b. TELEPHONE NUMBER *(Include area code)* (301) 402-7617 |