Feasibility Study of an Outcome Evaluation of the National Institutes of Health's New Innovator Award Program

May 12, 2011

Prepared for the
National Institutes of Health

G. Stephane Philogene, Ph.D.
Project Officer
Office of Behavioral and Social Sciences Research
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director, NIH

Prepared by
Bhavya Lal – Task Leader
Mary Beth Hughes, Jamie Doyle, Alison Riggieri, Elizabeth Lee, Amy Marshall, & Stephanie Shipp

Institute for Defense Analyses
Science and Technology Policy Institute
1899 Pennsylvania Avenue NW, Suite 520
Washington DC 20006

Executive Summary

The National Institutes of Health's (NIH) New Innovator Award (NIA) program was created in Fiscal Year (FY) 2007 to support creative early-stage investigators who have not yet received a large research award such as an R01 grant, the typical NIH funding mechanism used for individual investigator research projects. The NIA was created to stimulate highly innovative research by funding high risk high reward (HRHR) research and to support promising new investigators by granting funds to conduct highly innovative research. Based on discussions with NIH staff and review of the literature, we use the terms "innovative research" and "HRHR research" interchangeably.

The NIA is the second program within the High-Risk Research Initiative operated by the NIH Office of the Director to support innovative biomedical and behavioral research that cuts across the 27 Institutes and Centers of the NIH. NIH asked the IDA Science and Technology Policy Institute (STPI) to analyze the feasibility and desirability of performing an outcome evaluation of the NIA program, and to propose a design for such an evaluation for the FY 2007–2008 NIA awards.

To determine whether an evaluation is warranted and feasible, STPI interviewed NIH staff who currently run the program, and staff who could explain its original intent. In consultation with NIH staff, STPI developed study questions, an evaluation design, and a logic model. A literature review was conducted to identify output and outcome indicators that could be used to answer those study questions. Next, the feasibility of using comparison groups with other Early Stage Investigators (ESIs) was assessed by scanning other programs at the NIH and beyond. In addition, we investigated methods to ensure valid comparisons, and collected pilot data.

Study Questions

Based on program leaders' priorities, the core study question relates to the program outcome goal of *stimulating highly innovative research*, above and beyond what a traditional funding mechanism for new investigators can do. The core study question and related subquestions are:

- Did the NIA program stimulate highly innovative (HRHR) research?
  - To what extent was the research conducted by NIA awardees more *innovative* and *high risk* (where high risk is defined as research that has an

inherent high degree of uncertainty) than research conducted by other early stage investigators (ESIs)?

  - To what extent did the *outputs* and *outcomes* of NIA-funded research lead to or were they likely to lead to advances in biomedical and behavioral research? How do these advances compare to those of a traditional NIH program that funds early stage investigators (ESIs)?

- In addition to the core study question, we also propose the evaluation explore spillover benefits, principally the impact of the NIA on the awardees' career, five years after receipt of the NIA, compared with other early stage investigators (ESIs). The secondary study question and related subquestions are:
- What were the program's spillover benefits, especially on the careers of NIA grantees?
  - What fraction of the awardees remains in biomedical-related fields as compared with other ESIs?
  - What is the nature of the research (whether continued HRHR research or other) and total funding received by NIA-funded researchers, as compared to other ESIs?
  - What fraction of NIA awardees are becoming leaders in their fields, as compared with other ESIs?
- STPI recommends *the award* as the main unit of analysis for the core study question, and *the individual researcher* (i.e., the awardee) as the main unit of analysis for the secondary study question.

NIA Outcome Evaluation Design

The design and methods proposed for the NIA Outcome Evaluation were informed not only by a literature review, and by parallel outcome evaluation of the FY 2004–2005 awardees of the NIH Director's Pioneer Award (NDPA). The NDPA outcome evaluation used a longitudinal design to examine the research activities, publications, collaborations, and other measures of impact and productivity of the awardees before and after the receipt of the NDPA award. The evaluation had no comparison group due to the small number of awardees, the diversity of the awardees, and the broad interpretation of the term "pioneering."

Unlike the NDPA, the number of NIA awardees is sufficiently large, and the scale of research more similar to traditional funds to warrant a quasi-experimental design. Several comparison groups were considered, and we ultimately determined that R0I Early Stage Investigators (R01 ESI) awardees formed the best comparison group given the core study question. There are 61 NIA awardees selected in FY2007–2008, which means that

approximately 61 R01 ESI awardees will be sampled to match the characteristics of the NIA awardees. This will be challenging and likely be done using a stratified sampling approach by choosing the criteria (characteristics) upon which to choose the sample. Diagnostics will need to be performed to determine whether stratified sampling improves the similarity of the comparison group or if another sampling approach will need to be taken.

Once the two groups are in place, the primary method for addressing the core study question is expert review of NIA-funded outputs and outcomes. Experts will be selected for their knowledge of the particular domain of research, general biomedical research expertise, and reputation in the community for being a radical thinker. For each of the NIA awardees and ESI R01comparison awardees, we propose that three experts review three publications (or other relevant outputs) chosen by the awardee. This approach was tested in the companion NDPA Outcome Evaluation, and while resource intensive, it worked well.

Other methods will supplement the expert review, including a survey of NIA and R01 ESI awardees, bibliometric analyses, and case studies of selected awardees. We propose two versions of the evaluation: a full scale evaluation for about $875,000 and limited version for about $590,000.

A pilot test of the methods showed that an outcome evaluation of the NIA program is feasible, yet is not without challenges. These challenges include defining “innovative research,” sampling R01 ESI awardees that will be included as the comparison group, and collecting data, especially for the comparison group. However, through its pilot analysis, STPI has found that each of the challenges appears to be surmountable, and an outcome evaluation is recommended.

Contents

List of Illustrations

1. Introduction

A. Purpose of the Feasibility Study

In Fiscal Year 2007, the National Institutes of Health (NIH) launched the NIH Director's New Innovator Award (NIA) as the second program under the High-Risk Research Initiative of the NIH's Roadmap for Medical Research. Since its inception, a series of four competitions have been held, and 167 awards have been made.[1] The purpose of the NIA, as stated in the FY 2007 Request for Applications (RFA), was to stimulate highly innovative research and support promising new investigators.[2]

The primary purpose of this study was to determine if an outcome evaluation of the New Innovator Award (NIA) program is *feasible*.[3] The specific objectives that were set forth to make that determination were to:

- Create a logic model to describe inputs, activities, outputs, and contextual factors of the NIA program;
- Identify the outcome domains, measures, and indicators and data sources that could be used to determine the impact of the NIA program;
- Recommend an appropriate design for the NIA outcome evaluation (including comparison group options); and
- Develop data collection tools and approaches for determining the impact of the NIA program.

B. Activities and Methods

In order to achieve the objectives set forth in the feasibility study, the IDA Science and Technology Policy Institute (STPI) undertook the following activities to assess whether an outcome evaluation of the New Innovator Award is feasible:

- **Conducted interviews with program staff and stakeholders**. Interviews included NIA program staff and NIGMS leadership, and select NIA awardees. In particular, two focus groups were held at the annual NIH Director's Pioneer

[1] There were 30 awards in FY 2007; 31 in FY 2008; 54 in FY 2009; and 52 in FY 2010.

[2] See http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-07-009.html.

[3] A companion process evaluation of the first three years of the NIA program being conducted by STPI is underway and will be available on the New Innovator Award website upon the evaluation's completion.

Award (NDPA) symposia in September of 2009 and September of 2010. Awardees from FY2007, 2008, and 2009 were included in these focus groups.

- **Refined the logic model**. A provisional logic model of the NIA program was developed as part of the companion Process Evaluation of the program, currently being conducted by STPI. As part of the feasibility study, STPI refined the logic model to better describe the inputs, activities, outputs, and outcomes of the NIA program as currently understood.
- **Performed a literature review** of "innovative research" and of programs designed to foster it, including evaluation designs for these programs. A review of the literature on how to foster innovative research, and evaluations of these programs was performed to examine possible approaches for an appropriate outcome evaluation design. Discussions were also held with experts who have performed evaluations of NIA-like programs, such as the Howard Hughes Medical Investigators, the Burroughs Wellcome Career Award in the Biomedical Sciences, as well as social scientists studying creative scientists in Nanotechnology and Human Genetics.
- Reviewed and analyzed existing data on NIA and possible comparison groups. Information sources for NIA and possible comparison groups included:
  - RFAs and other historical documentation
  - Applications and other scoring data
  - Annual progress reports
  - Publications
- **Developed an Outcome Evaluation Design.** Insights gathered through the activities and analyses described above were used to determine whether an outcome evaluation was feasible, and to develop recommendations for the design of an outcome evaluation. The design of the outcome evaluation includes:
  - Overall approach and timeline
  - Study questions
  - Recommended outcome indicators and data sources, and
  - Recommended analytic methods

The feasibility study was conducted between January 2009 and January 2011.

C. About this Report

The remainder of the report is organized as follows:

- Chapter 2 provides an overview of the literature review, highlighting indicators that have been used to measure "innovative research." (An overview of the literature that discusses indicators to measure the spillover effects to "support of promising new investigators" is in Appendix D.)
- Chapter 3 describes the NIA program and explains the logic model that describes the program and discusses their various components of this model.
- Chapter 4 covers the proposed design of the outcome evaluation, including the study questions, the unit of analysis, the type of evaluation, possible comparison groups, and the proposed approach and methods.
- Pilot data collection is the subject of Chapter 5, which provides the results of the pilot along with recommendations for data sources and collection techniques.
- Chapter 6 summarizes our recommendations.
- Supporting documentation is included in eight appendices.

2. Literature Review

A. Introduction

Formal evaluations of programs like the NIA that aim to support innovative investigators are limited. The purpose of the literature review was to look for methods and approaches that could inform the design of a potential outcome evaluation of the NIA program. The NIA Request for Application described the two goals of the program as stimulating highly innovative research and supporting promising new investigators. Thus, this literature review describes how innovative research (and related terms) has been defined and operationalized in the literature. Discussions with program leaders revealed that NIH equates the 'supporting promising new investigators' goal through the granting of funds to conduct highly innovative research, so the primary goal for the proposed outcome evaluation is to assess whether the NIA is stimulating highly innovative research.[4]

B. Defining "Innovative" Research

The goal of the NIA program is to "stimulate highly innovative research." Multiple terms, including "creative," "high-risk, high-reward," and "transformative" have been used interchangeably to describe innovative research in the literature and in documents of R&D funding programs.[5] For instance, the National Science Foundation uses "potentially transformative" as a criterion in all solicitations, while "creative" and "innovative" appear in both the NIA and the NIH Director's Pioneer Award program documents. The Office of the Director of National Intelligence's Intelligence Advanced Research Projects Activity (IARPA) describes all of their research as "high-risk, high-payoff." Although it is clear that there are many terms used, there has been little clarification in the literature on the relationship between the terms or if they are in fact used to describe the same concept. The following sections provide short overviews of these terms, from the literature and from program evaluations, with a brief discussion on the nuances that appear among them. Since there is not one accepted definition, the literature provides the foundation for defining 'innovative research' using multiple yet overlapping definitions. Each is defined in turn.

---

[4] See Appendix D for review of literature on Operationalizing "support of promising new investigators."

[5] Program leaders consider innovative research to be high risk, potentially high reward research (HRHR). We will use the terms innovative research and HRHR interchangeably.

Innovative: In the literature on innovation in science, "innovative"[6] has traditionally been defined as being related to, but distinct from, creativity (described below). Amabile et al. defines "innovation as the successful implementation of creative ideas within an organization." (Amabile et al. 1996) Researchers often see innovation as the *usage* or *diffusion* of creative ideas.

High-risk: The term "high-risk" research has not been well-defined, although it is commonly used in R&D funding program language. The HRHR Demonstration Oversight Group (DOG), comprising senior NIH officials, defined high risk high reward research as "research with an inherent high degree of uncertainty and the capability to produce a major impact on important problems in biomedical/behavioral research" (Austin 2008). The DOG elaborates further that the "HRHR definition has 2 independent components:

- First consideration must be "capability to produce high reward/impact (…more than solid, incremental science)
- If project is high reward/impact, then consider level of risk"

In a speech in 2004, Rita Colwell, former head of the National Science Foundation, put forth a typology of high risk projects in the context of research funding portfolios: a project could be risky because:

- The ideas underlying it are at odds with prevailing wisdom (conceptual risk)
- It requires the use of equipment or techniques that have not been proven or are extraordinarily difficult (technical risk)
- It is being undertaken by a scientist who has not demonstrated expertise in the area (experience risk); or
- It involves a unique combination of disciplines (multidisciplinary risk) (Colwell 2003)

It is noted that this is not a measure of the level of risk, but rather a categorization of types of risks.

Transformative: This term has been used by the National Science Foundation for use in official NSF documents. According to the NSF definition, "Transformative research involves ideas, discoveries, or tools that radically change our understanding of

---

[6] There are numerous definitions of innovation, although many apply to the transition from laboratory to market. See *Measuring Innovation and Intangibles: A Business Perspective*, by Alexandra Stone, Susan Rose, Bhavya Lal, and Stephanie Shipp, IDA Science and Technology Policy Institute, IDA Document D-3704, December 2008. NIA research focuses on earlier stage research that still must undergo several more research steps before being scaled up for commercialization. Hence the literature is leaner for describing innovation at these earlier stages.

an important existing scientific or engineering concept or educational practice or leads to the creation of a new paradigm or field of science, engineering, or education. Such research challenges current understanding or provides pathways to new frontiers."[7]

Creativity: Of the terms identified above, creativity is the most commonly used though no single authoritative definition or description of creativity exists. Simonton defines creativity as "the output of ideas that are both original and adaptive" (Simonton 1997). Alternatively, Ochse incorporates the idea of utility and originality into his definition of creativity and includes the production of an object or idea in his definition (Ochse 1990). Finally, Amabile et al. expand upon both of these definitions by asserting that creativity involves heuristic (encouraging discovery of solutions) rather than algorithmic tasks or thinking (Amabile et al. 1996).

In the context of scientific research, Heinze has developed a typology of creative research outcomes, which include (Heinze et al. 2007):

- Formulation of a novel idea (or set of ideas) that could instigate a new cognitive frame or advance theories to a new level of sophistication
- Discovery of new empirical phenomena that could stimulate the generation of new theories
- Development of a new methodology, enabling empirical testing of theoretical problems
- Invention of novel instruments that could instigate new search perspectives and research domains
- New integration of formerly disparate ideas into general theoretical laws enabling analyses of diverse phenomena within a common cognitive frame

Commonly agreed upon definitions of innovative research and its synonyms in the literature are lacking. Only a fraction of the literature has focused on innovation and creativity in science and engineering. There is evidence in the literature that terms such as innovative and creative are social constructs: an agreed upon, or implicit, idea or set of measures coming from within a community. Sternberg states that creativity is community-specific and acknowledges that the characteristics of creativity in the humanities may differ from the sciences (Sternberg 1990). Simonton asserts that input from the respective communities are necessary to develop the definition (Simonton 2003). Thus, as evaluators, one may have look to more specific, empirical measures that

[7] Definition of Transformative Research. Accessed from http://www.nsf.gov/about/transformative_research/definition.jsp.

have been used to define "innovative research." These measures are described in the next section.[8]

C. Empirical Measures of "Innovative" Research

There have been studies and evaluations in recent years that have provided some empirical measures of creativity and innovativeness (summarized in Table 1). While some of these measures are not based on underlying theory, they prove to be more functional than the definitions given in the literature for the purposes of this feasibility study.

Table 1. Summary of Recent Evaluations of "Innovative" Research Programs or Scientists

Authors	Year	Program/Group Evaluated	Examples of Measures Used	Comparison Group Used
Pion and Cordray	2008	Burroughs Wellcome Career Award in the Biomedical Sciences (CABS)	Faculty in top-25 ranked institutions in NIH funds PI on NIH R01 or other NIH grant Age at first R01 Publications Total articles Articles in top-ranked journals Average citations per article	Non-awardees of the CABS program, matched using Propensity Score Analysis
Heinze and Bauer	2007	Nominated Creative Scientists in Nanotechnology	Overall productivity Citation rates Degree Centrality Integration Score	Peer scientists with the same publishing frequency
Azoulay, Zivin, and Manso	2009	Howard Hughes Medical Investigators	Publications in top percentiles of citations Nobel Prizes won Elected to National Academy of Sciences or Institute of Medicine Trained an early career award winner Novel keywords tagging publication	R01 MERIT Awardees, Early Career Award Winners, weighted using Propensity Score Analysis

[8] This supports the proposed approach to use expert review to assess the innovativeness of NIA-funded outcomes.

1. Publication-based indicators

***Productivity*:** Simonton argued that creative scientists are more productive than their peers, but that they also publish a higher number of ignored works. He posits that the likelihood of a researcher's peers finding his or her work creative is a probabilistic consequence of quantity. Under this model, the total number of publications can be used as an indicator for creativity (Simonton 2003). Heinze and Bauer tested this hypothesis and found that the total number of publications appeared to be a significant predictor of creativity. They claimed, however, that this operationalization is overly simplistic and suggested that other indicators, such as measures of impact, are preferred (Heinze and Bauer 2007).

Impact: In addition to productivity, researchers have found impact to be an indicator of creativity. Most commonly, impact is operationalized using citation rates (Heinze and Bauer 2007; Azoulay, Zivin, and Manso 2009). Evaluations of programs similar to NIA, such as HHMI, have used this as an indicator of creativity/innovativeness. An alternative to citation counts is the h-index, a bibliometric indicator that has risen in popularity since its introduction in 2005 (Hirsch 2005).[9] The h-index is a combined measure of the productivity and impact of a scientist. The h-index captures the career-long achievements of a researcher in the sense that it is insensitive to un-cited or lowly cited papers as well as to one or several highly cited papers. Several variations on the h-index have been used, some of which are discussed in Appendix A. Journal Impact Factor, a measure of the extent to which articles in a journal are cited, may be used to characterize the potential exposure of an article published in a specific journal.

Brokerage: Another indicator of innovativeness is brokerage. Brokerage is defined as a measure of an individual's connections to other scientists. The theory is that people with more connections to distinct social networks are considered brokers and hypothesized to have more innovative outputs since they are exposed to more diverse ideas (Burt 2004). Heinze and Bauer consider this theory by examining the association between the number of disparate authors and groups brokered by a researcher and his or her citation rate. They find that the connection of isolated researchers is a stronger predictor of creative work than number of co-authors alone (Heinze and Bauer 2007). This finding suggests that brokerage promotes creativity.

***Degree Centrality*:** Degree Centrality refers to the size of a researcher's network. This variable has been operationalized by Heinze and Bauer as the size of a scientist's co-authorship network in any three-year time period. They found that degree centrality does not predict a creative event, but does correlate with the number of author-level citations (Heinze and Bauer 2007).

---

[9] A scientist has an h-index h if h of his or her Np papers have at least h citations each and the other (Np – h) papers have fewer than ≤ h citations each.

Multidisciplinarity/Interdisciplinarity: As defined by Heinze et al, one dimension of research creativity is Multidisciplinarity (Heinze et al. 2007). While there is little clarification in the difference between multidisciplinarity and interdisciplinarity in the literature, in general, they both refer to the number of disparate bodies of specialized knowledge (Porter et al. 2007; Wagner et al.). In terms of publication data, interdisciplinarity may be analyzed via either (1) cited references of a publication set (the body of knowledge drawn from), (2) the publication set itself (body of knowledge), or (3) works citing the publication set in question (body of knowledge citing).

NDPA Outcome Evaluation Use of These Metrics. The NIH Director's Pioneer Award outcome evaluation conducted bibliometric analyses using the above metrics as proxies for innovativeness (Lal et al. 2011). These measures include productivity, creativity, impact, and collaboration. The analyses provided interesting information about the awardees, but conclusive results about the innovativeness of the research were not evident, perhaps because the evaluation is only 5 years out from the receipt of the award.

2. Non-publication-based indicators

Awards: Some evaluative research has used awards to supplement publication-based measures as indicators of innovativeness. For instance, Azoulay Zivin, and Manso studied elections to prestigious scientific societies, also known as metrics of scientific excellence, as an indicator of creativity (Azoulay, Zivin, and Manso 2009).[10] Although not often used in applied work, Simonton suggests a comparison of honors and awards listed on researchers' curricula vitae to measure creative impacts. Simonton identifies four categories of awards: (1) international recognitions, such as the Nobel prize, (2) national recognitions, such as election to the National Academy of the Sciences, (3) discipline-specific honors, such as the Distinguished Scientific Contribution Award of the American Psychological Association, and (4) society-level recognitions, such as 'fellow' status within a society (Simonton 2003).

Lab-level indicators: Another non-publication-based indicator of innovativeness is the number of students and fellows trained at a researcher's lab that go on to win Pew, Searle, Beckman, Packard, and Rita Allen scholarships (Azoulay, Zivin, and Manso 2009). This indicator is hypothesized to be an indicator of innovativeness, since more innovative labs can attract higher quality students. This indicator, however, has questionable validity and causation because the trainees' achievement may be caused by both the lab head's creativity and/or the students' own creativity. This caveat is acknowledged by Azoulay et al.

[10] Azoulay et al. acknowledge, however, that a lot of the members of these societies are previous HHMI members and therefore might skew results.

Patents: The last indicator of creativity used in applied evaluation research is patent activity. The two most common measures used are patent counts and patent citations. Previous studies have found that patents are granted by the most productive individuals in research(Stephan et al. 2007) and that patents are preceded by a flurry of scientific publications (Azoulay, Ding, and Stuart 2006). Patents, however, are typically only pursued in situations in which the research may have commercial value, and thus may not have the same interpretation across various fields.

D. Expert Review to Identify Innovative Research

In addition to the use of publication and non-publication indicators of innovative research identified in the literature, the use of consensual assessment techniques, or expert review, is another method described in this study.

Consensual assessment technique is a method that relies on the subjective judgments of appropriate observers, often experts, to determine whether a research product is creative (Amabile 1982). The Committee on Science, Engineering, and Public Policy, a joint committee across branches of the National Academies,[11] corroborated Amabile's conclusions in identifying expert review as the most effective means of evaluating federally-funded research programs in comparison to economic-impact studies, and bibliometric analyses on publication and patent data (National Academy of Sciences 1999). The committee discussed three forms of expert review which could be valuable for program evaluation: (1) quality review, which judges the quality of the scientific research, (2) relevance review, which judges the relevance of the research to the agency's mission, and (3) benchmarking review judges the international leadership status of the United States in the context of a program.[12]

Grant and Allen (1999) used a novel expert review approach to compare Wellcome Trust Showcase awards with a sample of standard project grants to assess the whether the grants were risky, novel, speculative, adventurous, and innovative on a 5-point scale. The evaluators selected ten research summaries; five from each group, from 40 summaries. These 10 summaries were then sent to the 48 members of the expert panels, which potentially could have yielded 12 reviews per project. In fact, each summary was reviewed by an average of 7.7 panel members. The authors stated this method, which they called a "masked randomized trial," eliminated much of the systematic error that might have occurred using standardized expert review, making the results more robust.

---

[11] The National Academies Committee on Science, Engineering, and Public Policy is a joint unit of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. See http://sites.nationalacademies.org/pga/cosepup/index.htm.

[12] The use of expert review supports Sternberg's statement that creativity is a community specific concept (Sternberg 1990).

In 2008, a Working Group on Peer Review of the Advisory Committee to the Director of NIH recommended that NIH shorten the length of the application and "engage more persons to review each application…optimally 4 or more."[13] The Working Group left the actual number of expert reviewers ambiguous, so Kaplan, Lacerta, and Kaplan (Kaplan D. 2008) conducted a statistical analysis to provide guidance on the optimal number of expert reviewers. They had 10 short proposals scored by an average of 48 reviewers. They then conducted a sensitivity analysis and found that "funding decisions will vary widely with the number of reviewers in considering proposals that are closely scored." They noted that the length of the application affects how many reviewers can be used for scoring; the shorter the application, the more reviewers that can be used. They conclude that the NIH peer review process should be designed to meet statistically significant criteria. [14]

The recently completed NIH Director's Pioneer Award (NDPA) Outcome Evaluation successfully used expert review to determine whether and how the awardees' research was pioneering (Lal et al. 2011). The goals of the expert review were to assess the impact of the work conducted under the Pioneer Award and to provide insight regarding the effects of the NDPA program. Awardees were asked to suggest 3 to 5 potential experts, with a mix of supporters and critics.[15] They were also asked to review a one-page summary of their research conducted under the NDPA, written by the STPI team, which could be sent to the expert reviewers. Finally, the awardees were asked to suggest three publications for the expert review. The expert reviewers were sent these materials as well as the NDPA Program Notice from the year of the award,[16] and a feedback form to record their assessment of the Pioneer project. Following submission of the completed feedback form, selected experts were invited to share additional feedback during a phone interview, and to clarify answers where necessary. This expert review process was used successfully to identify whether the awardees' research was pioneering and why. The drawbacks were the time it took to prepare the review materials, code the assessment forms, and to conduct follow-up phone meetings.

---

[13] NIH (2008). Enhancing peer review: A self-study by the NIH in partnership with the scientific community to strengthen peer review in changing times. Available: http://enhancing-peer-review.nih.gov/meetings/EnhancingPeerReviewACD2-21-08.pdf.

[14] A review by a STPI statistician did not agree with the Kaplan analysis, stating that the sample sizes are inappropriately small, highly selective, and susceptible to bias. There is a tradeoff of providing a more-in-depth proposal that fewer experts can review than a shorter proposal reviewed by many reviewers, perhaps not expert in the proposer's field.

[15] STPI did not use a suggested expert if their appeared to be a conflict of interest, such as being a co-author on NDPA related publications.

[16] The Program Notice was included to provide a reference for the experts when answering questions regarding the goals of the program.

E. Summary and Conclusions

In general, literature focusing on innovative or HRHR research is sparse. However, some recent studies and program evaluations have provided indicators and methods for measuring innovative outputs and outcomes. A summary table describing the indicators and methods of innovative research used in the literature is presented in Appendix A. In the literature review, the NDPA Outcome evaluation use of bibliometric indicators as one method to assess the innovativeness of the research is also discussed. The results, five years after receipt of the award, did not clearly show that the outputs produced by the awardee were more innovative than 5 years before the receipt of the award.

In addition to an overview of the use of expert review in evaluating programs, including high-risk, high-reward programs, the literature review discussed the successful use of expert review for the NDPA outcome evaluation and the use of expert review in other studies, as well as explore new approaches to expert review. The findings from this literature review informed the evaluation design.

3. Description of the NIA Program and Logic Model

A. Introduction

The NIA program is first described followed by the description of the logic model. The goal in constructing the NIA program logic model (Figure 1) was to accurately represent the inputs, activities, outputs, and outcomes of the NIA program as well as other factors that may affect its success. A preliminary logic model had been developed as part of the Process Evaluation of the NIA program STPI conducted. It was developed through an iterative process with NIA program staff and through reviews of administrative documents (e.g., the program RFAs).

In Fiscal Year 2007, the National Institutes of Health (NIH) launched the NIH Director's New Innovator Award (NIA) as the second program under the High-Risk Research Initiative of the NIH's Roadmap for Medical Research. Since its inception, a series of four competitions have been held, and 167 awards have been made. [17]

The purpose of the NIA, as stated in the FY 2007 Request for Applications (RFA), was to stimulate highly innovative research and support promising new investigators.[18]

Aside from the program goals, the NIA had a number of key features that defined the program:

- Eligibility criteria:
  - Applicants were required to meet the definition of "early stage investigator," which means having received their most recent doctoral degree, or completed their medical internship and residency, no earlier than 10 years from the release date of the RFA and no later than the receipt date for applications. Applicants could apply for a waiver of this requirement in the case of a lapse in the research period, for reasons including medical concerns, disability, family care responsibilities, extended periods of clinical training, natural disasters, and/or active duty military service.
  - Applicants were also required to meet the definition of "new investigator," which is defined as those investigators who have never applied successfully

[17] There were 30 awards in FY 2007; 31 in FY 2008; 54 in FY 2009; and 52 in FY 2010.

[18] See http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-07-009.html.

as a Principal Investigator on an R01 (or equivalent[19]) NIH grant or leader of a multi-project grant (e.g. P01).

- Activities:
  - Awards were for $300,000 per year in direct costs, an amount similar to the annual value of R01 grants, but are disbursed up front, instead of on an annual basis.
  - Awards were for five years, a period somewhat longer than the average R01 grant[20].
  - The use of funds by awardees was flexible, with no detailed budget submission required.
- Process:
  - Unlike other programs for Early Stage Investigators,[21] the NIA program was run centrally, out of the Office of the Director (OD), and administered by the National Institute of General Medical Sciences (NIGMS).
  - *Ad hoc* committees of extramural reviewers were used for evaluating applications, as opposed to study sections in the Center for Scientific Review. Extramural review was conducted independently with no face-to-face interaction among reviewers.
  - The application was relatively brief and allowed to be no more than 10 pages in length. Application materials consisted of a two-page biographical sketch of the applicant and an essay describing the proposed research that addresses each of three review criteria in detail:
    - The scientific problem to be addressed,
    - Innovativeness of the research proposed, and
    - Investigator qualifications.
  - Preliminary data were not required, but could be included.

Additional information on the NIA program is available at: http://commonfund.nih.gov/newinnovator/.

---

[19] Grants considered "R01 equivalent" include R23, R29, R37, or U01.

[20] NIH Research Grants are generally awarded for 1 to 5 budget periods, each normally 12 months in duration. R01s are generally awarded for 3 to 5 years and the average length of R01s has gradually increased over the years. See http://grants1.nih.gov/grants/funding/r01.htm.

[21] New and Early Stage Investigator Policies. See http://grants.nih.gov/grants/new_investigators/index.htm.

B. NIA Program Logic

There are several differences between the NIA and a typical funding program for Early Stage Investigators. On the most basic level, the inputs are different. An innovative project—one that is not bounded by specific aims—and flexible funding are the two most distinguishing factors.

The second notable difference, which is related to the inputs, is the ability of the NIA awardee to explore other areas of research. For other NIH awards, there is an expectation that the research will follow the path laid out to accomplish the specific aims. However, the NIA gives the researcher flexibility to experience failure and pivot his/her research agenda to explore other avenues. In theory this flexibility should encourage innovation.

Third, research funded by the NIA program is expected to lead to significant advances in biomedical and behavioral research.[22] The emphasis of the program on innovation and risk, in addition to the size of the award, should facilitate research progress.

Fourth, given the large and flexible funding amount and the innovativeness of the project, there is a belief that the NIA awardee will publish his/her findings in high-impact journals and apply for follow-on funding. The assumption is that the "safe science" that is usually funded through other NIH funding mechanisms would be published less frequently in high-impact journals as compared to research funded by NIA. In addition, since the NIA cannot be renewed, there is an expectation that the NIA project will eventually become more mainstream (less high risk as more is known about the research) and the research could be continued through support from traditional funding mechanisms. [23]

Because the early-stage NIA eligibility criterion is defined with respect to time from terminal degree rather than based on career stage, awardees are a diverse group with respect to career progression. As stated in the RFA, researchers must be less than 10 years out from their last terminal degree to be eligible for the program, and an analysis has shown that applicants and awardees show a broad distribution of years since terminal degree.[24] It is expected that the NIA funding will enhance the reputation of the NIA awardees and hence accelerate their careers, compared to ESIs.

---

[22] See http://commonfund.nih.gov/newinnovator/.

[23] These assumptions are based on interviews with NIH program leadership.

[24] New Innovator Award Process Evaluation. Draft. Will be available online at http://nihroadmap.nih.gov/newinnovator/. The average number of years since last degree is seven. The distributions are somewhat different between 2007 and 2008. In 2007, the awardees' time since the terminal (last) degree are distributed from 1 to 10 years, with13% being greater than 10 year (which is allowed if there is good reason why the awardee was out of the workforce for a period of time). In 2008, there are no awardees that are less than 3 years since terminal degree. Most (84%) of 2008 awardees are concentrated between 5 and 9 years since last degree.

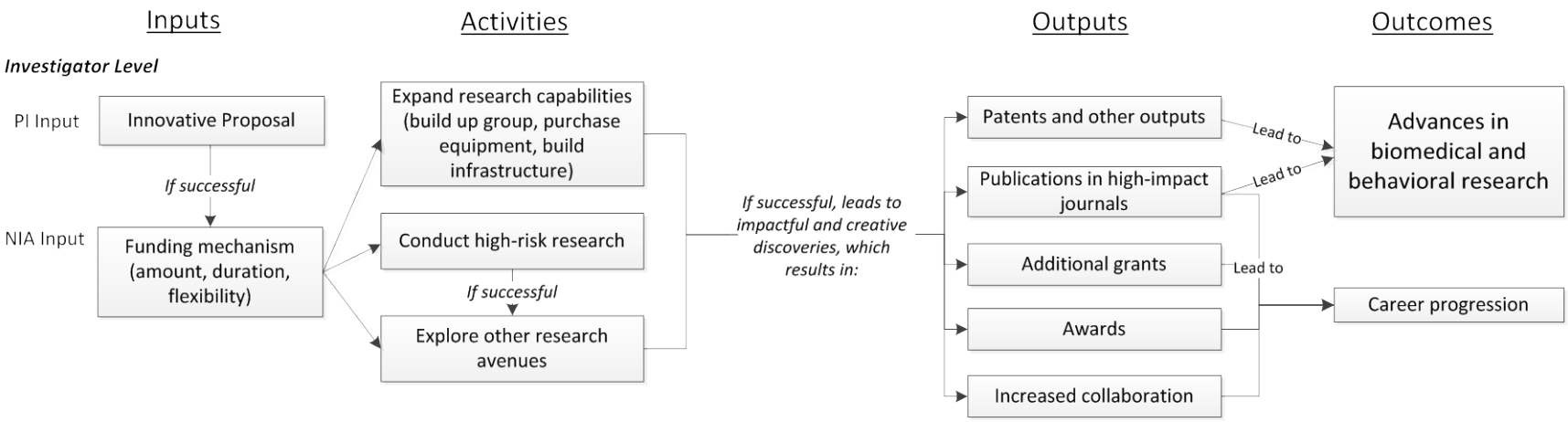

Figure 1. Investigator-Level Program Logic

4. Outcome Evaluation Design: Study Questions, Unit of Analysis, Comparison Groups, Data Collection Methods & Choice of Design

In this chapter, the possible types of evaluation designs are described, and proposed study questions, unit of analysis, and possible comparison groups are presented. Given this, we then discuss the methods to collect the data in an efficient and optimal way. We propose the use of multiple methods to capture the information needed for the primary data collection approach, which is the use of expert review to assess the innovativeness of the NIA awardees' research. In addition, the use of bibliometric analysis is proposed to also assess the innovativeness of the NIA awardees' research through the use of proxies for innovative research, such as productivity, creativity, collaboration (brokerage), and impact.

A. Study Questions

The primary rationale for an outcome evaluation is to test whether the New Innovator Award mechanism is necessary to facilitate early stage investigators' (ESIs) production of innovative research – or whether using the R01 would do just as well. Thus, based on the NIH program leaders' priorities, the study questions were framed around the core program **outcome goal of stimulating highly innovative** research, *beyond what a traditional funding mechanism can do*:

- Did the NIA program stimulate highly innovative (HRHR) research?
  - To what extent was the research conducted by NIA awardees more *innovative* and *high risk* (where high risk is defined as research that has an inherent high degree of uncertainty) than research conducted by other early stage investigators (ESIs)?
  - To what extent did the *outputs* and *outcomes* of NIA-funded research lead to or were they likely to lead to advances in biomedical and behavioral research? How do these advances compare to those of a traditional NIH program that funds early stage investigators (ESIs)?
- In addition to the core study questions, we also propose the evaluation explore spillover benefits, principally the impact of the NIA on the awardees' career, five years after receipt of the NIA, compared with other early stage investigators (ESIs).

- What were the program's spillover benefits, especially on the careers of NIA grantees?
- What fraction of the awardees remains in biomedical-related fields as compared with other ESIs?
- What is the nature of the research (whether continued HRHR research or other) and total funding received by NIA-funded researchers, as compared to other ESIs?
- What fraction of NIA awardees are becoming leaders in their fields, as compared with other ESIs?

Appendix B presents these study questions mapped to more specific subquestions and indicators, as well as identifies the potential data sources that would be used to answer them.

B. Type of Evaluation Design

1. Choices of Evaluation Design

There are three types of evaluation designs (Rossi 1993).

1. **Experimental designs** compare program outcomes in which the experimental and comparison groups are randomly selected. Experiments that use this approach are referred to as randomized or 'true' experiments. Fully experimental designs generally are not used in evaluating R&D funding programs because assignment of the awards is not at random (i.e. presumably the awardees differ from the non-awardees in some sense, which is why they were awarded the funds).
2. **Quasi-experimental designs** compare program outcomes to outcomes associated with another group of awards, activities, individuals, or institutions (typically either an external set of research awards/programs or the awardees themselves at a point in time prior to participation in the program). The primary differentiating design feature is that a comparison group for a quasi-experimental design is created retrospectively. For such a design, data collection efforts would center on the indicators and possible confounding variables described in the literature review. Identical data would be collected for a suitable comparison group. This comparison group would serve to provide a counterfactual to receiving an award and would allow the evaluation to develop inferential conclusions regarding the program's impact.
3. **Non-experimental designs:** Although, not ideal, there are methods available that can provide useful information to program evaluators. They include cross-

sectional analysis that provides a snapshot at one point in time or a longitudinal design, such as comparing the awardees at time of award and some number of years before and after receipt of the award (a pre-post design).[25]

C. Evaluation Design Selection

Based on the analysis and choice of study questions, unit of analysis, and data collection approaches, evaluation design options were considered for the primary program outcomes. Three criteria were used to assess each design family:

- *Feasibility of collecting required data.* How difficult and resource-intensive would it be to collect the data required to address the outcome using this design?
- *Potential payoff for funder in terms of utility and efficiency*. How useful would the results of an evaluation of this design be for the NIH?
- Ability to produce results that can be interpreted with confidence (internal validity). How likely is this design to lead to answers that can be reasonably trusted?

Program Outcome 1: Stimulating Highly Innovative (High Risk High Reward[26]) Research

A quasi-experimental study is feasible and appropriate for answering the study question for several reasons. First, selecting a comparison group retrospectively is feasible given (did the NIA program stimulate highly innovative (HRHR) research?) the size and diversity of the NIA awardee population and the potential comparison groups. Second, much of the data will come from NIH via IMPAC II/QVR databases and baseline characteristics can be assessed retrospectively by examining biosketches at the time the award application was evaluated. This feature will allow for a comparison of research outputs of NIA awardees and comparable R01 ESIs that followed the receipt of their respective grants.

Thus, a quasi-experimental design is proposed so that NIH can determine whether the NIA produces more innovative outcomes than other Early Stage Investigator (ESI) programs.

Program Outcome 2: Spillover Benefits – five years after receipt of the NIA

- Given that a quasi-experimental approach is proposed to evaluate the program outcomes, additional program impacts related to how the awardees can also be

[25] Some evaluators consider pre-post analyses as "quasi-experimental."

[26] As we noted in the literature review, we use "high risk high reward research" interchangeably with "innovative research."

evaluated to address the study question, *how did the NIA program support promising new investigators?*

Summary of Evaluation Design Recommendation

In sum, a quasi-experimental design is proposed for the NIA outcome evaluation. The design is feasible, yet resource intensive. The results, however, can inform NIH with some degree of certainty about whether the NIA is sufficiently different from traditional NIH funding mechanisms. This decision process, using the three criteria identified above is presented in Table 2.

Table 2: Criteria used to assess possible designs for NIA Outcome Evaluation

	Feasibility of collecting required data	Potential payoff for funder in terms of utility and efficiency.	Ability to produce results that can be interpreted with confidence
Experimental design (random selection of experimental and comparison groups)	Not possible, since the awardees are not randomly chosen	NA	NA
Quasi-experimental design (needs comparison group)	Feasible but resource intensive	Results will inform whether NIA is sufficiently different from traditional NIH funding mechanisms.	Strong – can claim with some confidence that program had unique characteristics
Non-experimental design			
Cross-sectional • No comparison group. • Data are for one point in time	Feasible, inexpensive	Snapshot of NIA awardees' research and career status, 5 years after receipt of NIA	Weak - Cannot claim that the program was needed to fund HRHR
Pre-post award analysis • No comparison group • Data are for two points in time	Feasible	Changes in NIA awardees' research and career status, 5 years after receipt of NIA,	In-between - Cannot claim with confidence that program had unique characteristics

NA = not applicable.

D. Unit of Analysis

The potential unit-of-analysis options for the study question are the individual researcher or the grant. To address the core study question, "Did the NIA program stimulate highly innovative research?," we proposed that the unit of analysis be the grant (award) itself. The award as the unit of analysis more accurately measures the impact of the program, as it isolates the funds that were provided by the New Innovator Award relative to other funds the researcher may have, although it may be difficult to precisely distinguish the attribution of the research to the NIA compared to other funding sources

received by the awardee. At a minimum, the publications that acknowledge the NIA are directly attributable to the NIA.[27]

To address the secondary study question, the proposed unit of analysis is the researcher. It is the only meaningful unit of analysis for evaluating the researcher's career, five years after receipt of the NIA.

E. Selection of Comparison Groups

1. Overview

A quasi-experimental design centers on the selection of a meaningful comparison group. While many programs at the NIH and beyond support early stage investigators, creating a meaningful comparison group remains a challenge. To find potential comparison groups, a search of both NIH and non-NIH programs for young investigator and creative young investigator programs was performed. A complete list, as well as key program characteristics, of programs that were considered can be found in Appendix C.

The list included:

- NIA applicants
- NIA finalists
- NDPA awardees[28]
- R01 awardees who are Early-Stage Investigators (R01-ESIs)
- Pathways to Independence awardees (K99/R00)
- NIEHS Outstanding New Environmental Scientists (ONES) and NIMH Biobehavioral Research Awards for Innovative New Scientists (BRAINS)
- NIDDK Type 1 Diabetes Pathfinder Award
- Mentored Research/Clinical Scientist Development Award (K01/K08)
- Independent Scientist Award (K02)
- HHMI Early Career Scientist Award
- PEW Scholars Program in the Biomedical Sciences
- Burroughs Wellcome Career Awards at the Scientific Interface

[27] Alternatively, one could argue that the NIA is a people-based program rather than a project-based program, so the analysis should be based on the person rather than on the project. For the reasons above, the proposed unit of analysis for the core study question is the award.

[28] Although the NDPA does not require that the investigator be an ESI, the program was considered in the initial group since the NIA is considered a "junior NDPA" program.

This list of comparison groups includes research funding programs, salary support programs, and hybrid research funding and salary support programs.

2. Comparison Criteria

Each of the possible comparison groups presents unique advantages and disadvantages, as presented in Appendix C. Some of the non-NIH comparison groups, such as the Howard Hughes Medical Institute (HHMI) Early Career Scientist Program and the Burroughs Wellcome Career Awards at the Scientific Interface**,** are similar in program characteristics, such as funding size and target population, but the use of these groups do not answer the question about whether NIA produces more innovative outcomes than do traditional R01 programs. For that reason, those programs were rejected.

If NIH programs were to be used as the comparison groups, data availability is less of a concern. Similar funding mechanisms for highly innovative early stage investigators, such as the NIDDK Type I Diabetes Pathfinder Award, pose problems because of their small sample size and their focus on specific subsets of biomedical research. This issue also applies to the ONES R01 program and the BRAINS R01 program. For that reason, these groups were rejected. The NDPA program also only grants a small number of awards each year, the amount of funds awarded are significantly higher than the NIA, and it does not solely fund early stage investigators. Thus it also was not considered further.

The awards that only support mentored career development, such as K08s and K01s, support earlier stage investigators than NIA and do not provide a similar level of research-related funds. For that reason, these groups were rejected for consideration. The Pathways to Independence (K99/R00) program was also rejected as it focuses on supporting additional career development and mentoring opportunities in a specific subset of biomedical research (muscular dystrophy).

Based on this assessment, a shortened list of best candidates for comparison groups was explored in greater detail: all NIA applicants, NIA finalists, and R01 ESIs.

These possible comparison groups were then further assessed on a variety of criteria to determine if the groups would be comparable, including:

- Whether the program/group is similar in *motivation* to the NIA program/awardees. Does the program have similar goals at the NIA in terms of stimulating highly innovative research and supporting promising new investigators?
- Whether the program/group is similar in *population* to the NIA program/awardees. Does the program fund researchers who are likely to have the same characteristics as the NIA awardees?

The possible comparison groups were also assessed on other dimensions to determine if it would be feasible to use them, include:

- Resources required for data collection
- Whether data would be available on the comparison program/group
- Whether members of the comparison group program would be likely to respond to a survey

The summary of this exploration is presented in Table 3.

Table 3. Assessment of Possible Comparison Groups

Comparability and Feasibility	All NIA Applicants	NIA Finalists	ESI R01 Awardees
Ability to address the study question – value of NIA in the NIH portfolio	(-)	(-)	(+)
Proposal is similar in motivation	(+) Focus on innovative (HRHR research)	(+) Focus on innovative (HRHR research)	(-) Less focus on innovative(HRHR) research
Similar in Population	(-) Likely difference in quality of researcher/proposal (-) Did not receive funds	(+) Little difference in quality of researcher/proposal (-) Did not receive funds	(+) Similar career stage (+) Received roughly same amount of funds
Cost Effectiveness of Data collection	(-) Large population, requiring sampling	(+) Similar Population	(-) Large population, requiring sampling
Data availability	(+) Substantial data available from NIH database	(+) Substantial data available from NIH database	(+) Substantial data available from NIH database
Likelihood of responding to a survey	(-) Potential low response rate for outcome data requests	(-) Potential low response rate for outcome data requests	(-) Potential low response rate for outcome data requests

Based on this second more considered exploration, the NIA applicants were eliminated for consideration. The reasons are the likely difference in quality of the proposal, reflected by the fact that they did not receive NIA funding, and the potential for low response rates for outcome data requests. This left the NIA finalists and the R01 ESIs.

Our hypotheses concerning each of the two comparison groups are as follows:

- NIA awardees should produce more highly innovative research and have indicators of better career progression, five years after receipt of the NIA, compared to the NIA finalists, because of the presence of the NIA funds.
- NIA awardees should produce more highly innovative research than R01 ESIs and have the freedom to pursue new research directions because of the flexibility in the use of the funding.

Ultimately, the evaluation is designed to help NIH decide if the NIA program added value to the NIH portfolio of funding programs above and beyond traditional mechanisms. In other words, if a traditional program can support innovative (HRHR) research, then the value of the NIA program becomes more suspect. As a result, it makes most sense to compare the innovativeness of NIA outputs and outcomes with those of outcomes from more traditional mechanisms. Therefore, STPI recommends that Early Stage Investigator R01 awardees (R01 ESIs) be used as the primary comparison group in the proposed evaluation.

F. Evaluation Approach and Methods

1. Overall Approach

The overall approach for the evaluation is as follows:

- Selection of R01 ESIs for comparison, using sampling or other techniques
- Collection of output indicators for NIA awardees and R01 ESI comparison group from NIH databases and CVs
- Collection of other output indicators for NIA awardees and R01 ESI comparison group via a survey
- Analysis of quantitative output indicators using bibliometric and descriptive statistical methods
- Recruitment of expert reviewers and implementation of expert review process
- Case studies of a subset of NIA awardees and R01 ESI comparison group via phone interviews

Below each of these steps is briefly described.

2. Selection of R01 ESI Comparison Group Awardees

The choice of the proposed source population (R01 ESIs) is described in the previous section. The R01 ESI population is much larger (e.g. 456 awardees versus 54 NIA awardees in 2009) and likely differs in characteristics that impact their research outcomes. Therefore, a systematic way to select the R01 ESI comparison group based on

these researcher characteristics such as years since degree, degree type, post-doctoral training, and title (as an indicator of tenure status) will reduce these differences and remove observed confounding effects. The most likely way to do this will be to sample from the RO1 ESI population to form a constructed comparison group so that they more closely resemble the NIA group.

Two alternatives for constructing a comparison group are appropriate for this population: stratified sampling and propensity score matching. Both of these methods will be tested, and the sampling technique which provides a comparison group most like the NIA awardees will be used for the evaluation. Stratified sampling uses categories or "strata" to bin the comparison group for the purposes of sampling. Thus, all individuals in the comparison group are first stratified by one or more of the baseline traits; and then awardees are randomly selected within each stratum. Diagnostics will need to be performed to determine whether stratified sampling improves the similarity of the comparison group or if another sampling approach will need to be taken.

Alternatively, propensity score matching may be used. This method finds matches that are similar in their likelihood of being selected as awardees. This involves constructing a logistic regression model where the outcome is receiving an NIA award using individual characteristics that are thought to impact both receiving the award and the research outcomes. For each person, the predicted probability of receiving the NIA is calculated and these predicted probabilities are the propensity scores. These propensity scores are used to create the comparison group.

The goal of these exercises is for the treatment and the control group to be statistically similar in how their observed characteristics are distributed. (Rosenbaum, 1983) Stratified sampling, although effective, works better with large populations. Since each observation can only be assigned to one bin, some exact matches may not be found for each awardee. Propensity score analysis has been used in several recent evaluations, including those of HHMI investigators (Azoulay, 2009) and the Burroughs-Wellcome Career Award (Pion, 2008). Because a composite is used to construct the comparison group, more covariates can be used. One these sampling methods are performed, a review of these constructed comparison group's distribution of the covariates will inform the final selection of the comparison group used in the evaluation.

3. Collection of Output Data via NIH Databases and CV Analysis

Certain indicators that are proposed to answer the evaluation study questions can be found in existing sources, primarily the NIH IMPACII database and through researchers' CVs. After the R01 ESI comparison group is selected, these indicators would be collected and stored for analysis. The pilot data study (in the next chapter) describes examples of the indicators that are likely be able to be found via this method.

4. Collection of Additional Data via Survey

Some of the data are not likely to be found in existing data sources. For example, questions around whether the research project pursued was the same as that proposed, or questions around the nature of the risk and creative outcomes (e.g. using the Colwell and Heinze Typologies, respectively) cannot be addressed through CV analysis or the NIH database. For this reason, a short, web-based survey is proposed. See Appendix G for the proposed survey questions.

5. Analysis of Quantitative Output Indicators

Once the output indicators are collected, analysis will be performed to assess the outputs of the NIA awardees relative to the R01 ESI comparison group. Many of the output data collected will have to be further analyzed to make the comparisons (e.g. calculations of citation indices, etc.).

6. Selection of Expert Panel and their Review and Analysis

An expert panel is proposed to review the innovative / high risk, high reward nature of the research undertaken by each of the 61 NIA awardees and 61 R01 ESI awardees in the comparison group. Based on our experience with evaluating outcomes for the NDPA (Lal et al. 2011), we propose using at least three independent assessments.[29] Thus, there would be a total of 366 reviews of the 61 NIA and 61 R01 ESI awards. If each expert reviews 10 awardee packages, 37 to 40 experts will be needed for the expert review.

To prepare for the expert review, the PI survey of outcomes will verify their NIA (or R01 ESI) attributed publications, and request them to list other award outputs such as materials or software, and indicate what percent of the research is attributable to the NIA (or R01 ESI) program. They will also be asked to select the three publications (or other outputs) that best represents their NIA (or R01 ESI) work for review by expert reviewers, as well as to list the names of at least 3 experts (2 supporters and 1 critic) of their research.[30] We will also independently seek experts to review the research.

Within their research areas, each expert panel member will be assigned 10 award packages to review, and given specific instructions on how to rate them. Each expert will receive a random combination of NIA and R01 ESI packages; for example, one expert might receive 10 NIA and no R01 ESI packages, while another might receive 5 NIA and 5 R01 ESI packages. Each package will include a description of the NIA program, the

---

[29] This feasibility study proposes the use of the traditional 3-expert review, based on success of using this approach for the NDPA (Pioneer) outcome evaluation. See literature review for discussion of expert review.

[30] For the NDPA Outcome Evaluation, STPI did not use a suggested expert if their appeared to be a conflict of interest, such as being a co-author on NDPA related publications.

expert reviewer protocol, and the 3 publications with author names and other identifying information (such as references) deleted to blind them. From the expert reviewers' perspective, each of the packages, whether an NIA or R01 ESI package, will be blinded and presented as an awardee of an NIH program.

Experts will be asked to review 3 publications chosen by each awardee and to assess the risk, innovativeness and potential/actual outcomes from the award. The protocol is included as Appendix F. The names, affiliations and demographic characteristics of the NIA/R01 ESI awardees will be blinded to the expert panel members to ensure anonymity.[31]

Part of the protocol will ask experts to describe risk and outcomes for high risk research. [32] We propose that the experts choose which categories fit the project being reviewed, and also to provide an indication of the strength of that assignment. For example, a sliding scale from "highly unlikely" to "highly likely," could be used to assess the likelihood of an outcome occurring, such as the likelihood that the formulation of a novel idea (or set of ideas) could lead to a new cognitive frame or advance theories to a new level of sophistication.

7. Case Study Analysis

In addition to the methods above, a case study approach could be utilized for a sample of the NIA awardees and the comparison group to capture broader impacts such as career progression and impact of advances, and to explain variations observed in the expert and descriptive analyses. If, for instance, NIA awardees had fewer publications than the comparison group, none of the metrics described previously would allow an evaluator to understand why. This case study approach will allow evaluators to identify any factors that may contribute to outcomes but are not easily quantified, such as a change in career focus or priorities, and institutional level variables such as access to additional funds.

The case study approach will utilize semi-structured interviews, based on the detailed study questions. Given the resource intensity of case study, case studies of only about 15 NIA awardees and 15 comparable R01 ESIs could be conducted.

---

[31] Although it is difficult to fully blind publications, we will ask the reviewers to not attempt to identify the awardee via a search of PubMed, Google Scholar, or other bibliometric database.

[32] See Section 2B Literature Review, Defining "Innovative Research" for a description of the risk and outcome typologies.

5. Pilot Data Collection

STPI conducted a pilot of data collection in order to determine the feasibility of the proposed outcome evaluation design. There were two major components of the pilot data collection—one to collect data that would inform the selection of a comparison group, and the second to determine which data sources would be needed to collect output and outcome variables.

A. Pilot Data Collection to Select Final Comparison Groups

The data needed to develop comparison groups are primarily characteristics of the NIA awardees and possible comparison groups that may affect the outcomes observed—or "input data." The difficulty in data collection at this stage is that there are potentially a large number of researchers to be examined, prohibiting resource-intensive collection methods.

STPI piloted "input data" collection on NIA awardees, finalists, and FY 2009 R01 grant recipients who held the designation of an early stage investigator. In order to quickly collect a vast amount of already existing information, STPI developed a visual basic script to automate the downloading of NIH biosketches from QVR, which they wrote in Python to extract the information from the NIH biosketches in PDF format and convert it to fields in a Microsoft Excel spreadsheet. The fields that could be collected into an analyzable format via this method included:

- eRA commons name
- Degree years
- Post-doctoral training
- Title

Based on these data, STPI analyzed the FY 2009 NIA awardee and the FY 2009 R01 ESI samples on mean years since degree, post-doctoral training, degree type, and title, using Fisher's exact test. The results of this comparison are shown in Table 4. The analysis shows that the two groups are not similar enough to use the R01 ESIs as a comparison group *as they currently exist*.[33] The R01 ESIs are farther since PhD and more diverse in their backgrounds than are NIA awardees. The next step would be to further

[33] The statistical results for mean years since last degree, type of degree, and title (e.g. associate professor) indicate that the two groups are not similar.

clean the data, eliminate outliers amongst the ESI R01 awardees, re-run the statistical tests for similarity, and then sample from the group if necessary.

Table 4. Comparison between FY2009 NIA Awardees and ESI-R01 Awardees

	R01 ESI 2009		NIA 2009	
	N	**%**	**N**	**%**
Mean Years Since Last Degree/Training	8.5		7.0	
Standard Deviation	3.8		2.4	
Missing	4		5	
% Postdoc	465	95.5	49	90.7
Degree				
MD	66	13.6	2	3.9
MD/PHD	77	15.8	5	9.8
PHD	343	70.6	44	86.3
Missing	1		3	
Title				
Professor	3	0.6	0	0.0
Associate Professor	43	8.8	1	1.9
Assistant Professor	359	73.7	42	77.8
Instructor/Non-tenure Track Professor/Research Fellow	67	13.8	8	14.8
Other	14	2.9	3	5.6
Post-Doctoral Fellow	1	0.2	0	0.0
Difference in Means Test (R01 ESI 2009 vs. NIA 2009)				
Years Since Last Degree/Training	0.01			
Total	487		54	
Fisher's Exact p-values for the Following Tests (R01 ESI 2009 vs. NIA 2009)				
Postdoc	0.17			
Degree	0.05			
Title	<0.001			

Although we propose using the R01 ESIs as a comparison group, as part of this study, we initially considered the NIA applicants and finalists and thus present the comparisons here. Table 5 compares the finalists to the awardees across the first three years of the program. No statistically significant differences are seen among the groups on the variables of gender, degree type, years since degree, average overall score on their NIA applications, and research area. Therefore, an advantage of using the NIA finalists

could be that they are comparable to NIA awardees on these particular traits at baseline. However, as noted above the comparison group selected for an outcome evaluation of the NIA program is the R01 ESI awardees.

Table 5. Comparison between FY2007-2009 NIA Awardees and NIA Finalists

		2007		2008		2009	
		Finalists	**Awardees**	**Finalists**	**Awardees**	**Finalists**	**Awardees**
Gender (Number)	Female	14	12	9	11	13	19
	Male	26	18	38	20	32	35
Degree Title (Number)	MD	3	2	4	1	6	2
	MD/PHD	5	7	11	7	5	9
	OTH	0	0	2	0	1	0
	PHD	32	21	30	23	33	43
Seniority (Years)	Min	2	1	2	4	3	3
	Mean	7.03	7.5	7.09	7.23	7.75	6.87
	Median	6	8	7	7	7	7
	Max	12	16	13	12	15	14
	Stdev	2.74	3.43	2.82	2	2.81	2.35
Average Overall Score	Min	4	4	3.67	4	3.33	3.33
	Mean	4.7	4.74	4.26	4.56	4.17	4.3
	Median	4.67	5	4.33	4.67	4.33	4.33
	Max	5	5	4.67	5	5	5
	Stdev	0.25	0.31	0.29	0.29	0.35	0.38
Research Area (Number)	Behavioral and Social Sciences	8	4	5	2	6	5
	Chemical Biology	0	2	1	5		6
	Clinical and Translational Research	7	5	8	2	12	8
	Epidemiology	1	0	3		1	2
	Immunology	N/A	N/A	6	2	3	5
	Instrumentation and Engineering	4	3	3	2	3	5
	Molecular and Cellular Biology	10	5	7	8	11	3
	Neuroscience	N/A	N/A	8	4	4	9
	Pathogenesis	5	9	N/A	N/A	N/A	N/A
	Physiology and Integrative Systems	2	1	4	4	3	3
	Quantitative and Computational Biology	3	1	2	2	2	8

B. Pilot Data Collection for Output and Outcome Variables

Output and outcome data are needed from a variety of different sources. STPI selected five NIA awardees and five R01 ESIs at random and explored the feasibility of data collection with respect to outputs. Details on the data collection are provided in Appendix E.

This data collection effort revealed several key points that will inform the evaluation. First, although NIH biosketches have a standardized format which aids data collection, they may not be the most updated source since they are only submitted during the NIH application phase. Therefore, if the applicant has not resubmitted a grant application to the NIH relatively recently, the information contained in the biosketches may not reflect the most recent information. Therefore, biosketches should only be used to collect information for pre-award data—the "input data" described above.

Secondly, although CVs will provide evaluators with much of the output and outcome data, they were difficult to obtain without direct contact with the researcher. Therefore, substantial effort to obtain recent CVs will be required.[34]

Thus, multiple sources are needed to collect the required information. Below are our recommendations for data sources including some discussion about data quality.

- **Publications**: Web of Science, PubMed, NIH SPIRES, and CVs seem to be the most efficient and complete method of gathering this data. Although it is noted that none of the sources alone gives a complete set of publications, combining the four will provide the most complete set available.
- **Patents**: Discrepancies between CV and the United States Patent and Trademark Office (USPTO) exist. Since CV and NIH iEdison patent information is self-reported and therefore has the potential to be inaccurate, USPTO is deemed more reliable than CVs. Although patents and patents pending can be collected for the evaluation, the process is estimated to be time-intensive. In addition, since the number of authors who patent is expected to be low given the evaluation time frame, the results of the analysis may not provide insights into the impacts of the NIA program. Based on the data in the pilot study, only half of the researchers appeared to have patenting activity and the majority of these had less than five patents total.
- **Employment**: Pre-award employment data could be collected using biosketches, but post-award employment information could be collected manually from CVs. The pilot study indicates that employment information is readily available from several sources. Listings in faculty directories may also supplement CVs.

[34] CV information could be collected in the survey.

- **Grants**: There is not one reliable source for total research funding. Because grant information on CVs are inconsistent, and because it is not an official measure, the use of QVR is recommended for measuring NIH funds. Non-NIH funds are harder to assess without supplementary data collection via a survey Based on the results of the pilot study, CVs can be utilized to count the number of extramural funding projects. If additional specificity, such as funding amount, is required, the utility of CVs as a primary data source declines. QVR does however, provide, data that would be important for future potential analyses such as year of first R01. Research funding data collection from CVs had the limitations of manual data collection and time-sensitive information to that of the employment section.
- **Awards**: Researchers do not report awards on CVs in a systematic manner. Instead of using the CV as a starting point, evaluators could identify major awards in the biomedical sciences and check to see if any recipients won these awards. In addition, a survey could be performed to collect the data directly from the grant recipients. Finally, biosketches could be used as a data source for the pre grant award information.

Summary: No adequate single secondary data source was identified for these output and outcome indicators. *A survey is therefore recommended to supplement data collection.*

C. Potential Challenges for Data Collection

The pilot data collection revealed several challenges for the outcome evaluation data collection.

1. Capturing All Publications

Since it has been demonstrated that both Web of Science and CVs have incomplete publication data for individual researchers, as demonstrated in Appendix E, it is possible that even a combination of these sources will not have a complete list of publications.

2. Handling Common Names

Common names pose a difficult and time-intensive challenge for collecting publication data. For example, one of the subjects' Web of Science query resulted in 912 records that covered over 100 subject categories. There are several methods to cope with common names when searching for specific authors in Web of Science. Web of Science provides an enhanced filter for authors called "distinct author sets."[35] This feature identifies "sets of papers likely written by the same person" using "citation data." While

[35] Web of Science, http://images.isiknowledge.com/WOKRS410B4/help/WOS/h_summda.html.

not precise, it displays top source titles, publication years, and top subject areas in order to help users identify the correct individuals. For our common name search, the "distinct author sets" tool returned 76 potential authors, many of which overlapped by years and institutions**.**

Another common name filtering tool, the text-mining software VantagePoint, can import Web of Science query data. Publications that share common author names can be filtered by any field exportable from Web of Science, such as author affiliation, publication year, source, and subject category. If there is existing background knowledge of the individual being researched, scanning titles for familiar topics and looking at co-collaborator names may also enhance the accuracy of the filtering process. In our pilot study, STPI chose the most appropriate distinct author sets and imported the data into VantagePoint to identify which articles belonged to the researcher in our study. The data cleaning process includes some human error. Web of Science on its own returned 20 of 24 (83%) CV publications. When the data were cleaned using the aforementioned techniques in VantagePoint, there was an accuracy of 90%; 18 of the 20 CV publications found in Web of Science were identified.

Common names are also an issue with patents. The names of the inventors are listed, but initials instead of full names may be used on first and middle names. Institutional affiliations are not at all included; city and state locations are used instead. When dealing with common names, ample foreknowledge of the individual's research and institutional locations will be necessary.

3. Manual Data Collection

Employment data must be collected manually from CVs. QVR allows one to export PI institution and position title, but data cleaning may still need to be performed. Pre-award position titles may be gleaned automatically from application biosketches using the Python program described above. In addition, collection of patent data from both CVs and the USPTO databases has the potential to be time-consuming because it must be done manually. USPTO provides no data export capabilities. Furthermore, patents must be examined one by one to ensure that the correct inventor has been identified and to count the number of patent citations.

4. Time-Sensitive Information

To perform the most accurate comparisons between the employment information before the award and after the award, data across researchers must refer to the same time. While CVs included the employment years, positions currently held often indicated that they were held "to present." If CVs are used for the data source of this information, they should ideally be collected from researchers of both groups at the same time.

5. Findings on Award Information and Potential Limitations

Neither CVs nor biosketches are a completely reliable source of awards and honors information, since considerable variation was observed in terms of the type of awards reported. It is likely that CVs will include more awards information than biosketches, however, since those documents tend to be longer.

6. Identification of Awards That Could Serve As Indicators

Scientific awards and honors will vary greatly among different researchers. There may be difficulty in determining which of the awards could serve as indicators of creativity and career status, five years after receipt of the NIA, particularly since awards may be given by different stakeholder groups (i.e. popular media, scientific societies, universities).

7. Measurement Issues

The variation in awards included was significant. Since the researchers are early-career investigators, most of investigators examined do not have well-known or prestigious awards. Of the pilot group, the majority of awards fell into three categories: fellowships or other grants that have "award" in the title (i.e. HHMI early career award, American Cancer Society post-doctoral fellowship, Walter H. Coulter Foundation Early Career Translational Research Award in biomedical engineering), membership in societies (i.e. Alpha Omega Alpha Honor Medical Society, Phi Lambda Upsilon National Chemistry Honor Society), or awards given by their universities (i.e. Young Mentor of the Year Award from Harvard Medical School, Excellence in Teaching Award from Stanford Medical School).

D. Summary

This pilot data collection revealed that although a multiple data sources will provide a significant amount of data for the evaluation, some data cannot be collected using these sources. Therefore, the use of a survey is recommended to fill in data gaps and answer the research questions that will not have adequate data from these data sources. These research questions include program level research questions, the impact of the award on the awardees' research groups, and the multiple alternative indicators such as mentorship activities and the development of new courses. (See Appendix G.)

6. Summary and Recommendations

A. Is an NIA Outcome Evaluation Feasible?

This feasibility study included multiple components, starting with the study question of primary interest to NIH: Did the NIA program stimulate highly innovative (HRHR) research? This question breaks down into two specific sub-questions:

- To what extent was the research conducted by NIA awardees more *innovative* and *high risk* (where high risk is defined as research that has an inherent high degree of uncertainty) than research conducted by other early stage investigators (ESIs)?
- To what extent did the *outputs* and *outcomes* of NIA-funded research lead to or were they likely to lead to advances in biomedical and behavioral research? How do these advances compare to those of a traditional NIH program that funds early stage investigators (ESIs)?

In addition to the core study question, we also propose the evaluation explore spillover benefits, principally the impact of the NIA on the awardees' career, five years after receipt of the NIA, compared with other early stage investigators (ESIs). In this context, the secondary study question is: What were the program's spillover benefits, especially on the careers of NIA grantees? The sub-questions include:

- What fraction of the awardees remains in biomedical-related fields as compared with other ESIs?
- What is the nature of the research (whether continued HRHR research or other) and total funding received by NIA-funded researchers, as compared to other ESIs?
- What fraction of NIA awardees are becoming leaders in their fields, as compared with other ESIs?

In order to address the study questions, first, we explored the literature to operationalize the term high risk high reward (HRHR) research. NIH defines it as "[a] highly innovative [approach] that [has] the potential to produce an unusually high impact on a broad area of biomedical or behavioral research." Such research may also have a higher probability of failure, or be "high-risk." In some ways therefore, high risk is a surrogate for high reward. We propose that the NIA outcome evaluation examine both the nature and intensity of risk in the awardees' research portfolios, along four dimensions: Conceptual Risk, Technical Risk: Experience Risk, and Multidisciplinary Risk.

We also propose that the "high reward" nature of NIA research be examined more thoroughly. The term implies that research must be more rewarding than solid, incremental science. With no precedent of measurement in place, we put forth a typology developed for "creative" science, with a higher burden of proof, and again, assess the nature and intensity, along five dimensions: New Idea, New Phenomenon, New Methodology, New Technology, and New Framework. Both frameworks are discussed in depth in this report.

The literature also points to some bibliometric measures. In our review, three particular metrics emerged: productivity of NIA researchers, networks before and after receipt of award, and impact of research. However, based on lessons learned from the recently concluded outcome evaluation of the NDPA program, we propose that there be less emphasis on bibliometric approaches.

In order to ensure that the evaluation remains focused, we propose that the unit of the analysis for the core study question be the NIA *grant*, and for the secondary study question the individual or *awardee*.

Many comparison groups were considered, and given the intent of the evaluation – to explore the "value added" of the NIA program versus others -- the proposed comparison group is a sample of R01 ESI awardees. Using sampling approaches discussed and tested in this feasibility study, we propose to select about the same number of R01 ESI awardees (about 60) and compare with NIA awardees on a range of dimensions identified by the study questions above.

Third, we propose that the core data collection be via expert review, supplemented by other methods. The rationale is based both on the literature – which claims that HRHR is a social construct, and can be judged only by the community – and our assessment of objectivity. Only experts, albeit a special kind of expert who can rise beyond the mainstream, can independently assess if the research funded was high risk and high reward. We therefore propose that three *blinded* publications (or other relevant outputs) of the 61 NIA awardees and about the same number of R01 ESI awardees be reviewed for their ability to be HRHR. Experts will be selected for their knowledge of the particular domain of research being assessed for its HRHR content, general biomedical research expertise, and reputation in the community for being a radical thinker. The approach is cost intensive, and will required the consent of about 40 experts.

The feasibility study also examined the availability of data from internal NIH databases, including CVs and biosketches of the awardees. We concluded that the use of these resources and a web-based survey would be needed to collect sufficient data to address the study questions. Case studies of selected NIA and R01 ESIs as well as some bibliometric analysis around citations and network development will also add context to

the evaluation, if sufficient time and funds are available (see time and cost estimates below).

Based on our research and preliminary data collection and analysis, STPI found that an outcome evaluation of the NIA program is feasible. The outcome evaluation will not be without challenges. NIA awardees are a diverse group, representing many fields generally related to biomedical and behavioral research. Thus standard outcomes may not be uniformly compared across the group. Nevertheless, STPI believes the proposed outcome evaluation design could provide information that would be useful for NIH to assess the value of the NIA program.

B. Timing of Execution of Evaluation

The outcome evaluation design as proposed above would measure short-term outcomes of the NIA award, in order to provide useful feedback to the NIH. However, if transformative scientific advances were funded through the program, it is possible that they would not be observed through these short-term outcomes. Thus, a later-stage evaluation may be reasonable.

The first cohort of the NIA awardees (FY2007) will not end their award term until the fall of 2011. For the evaluation of the short-term outcomes of the program, the earliest time to start the evaluation would be 2012. Waiting a few additional years would assure that most publications arising from the award are published. In any case, whether the study starts in 2012 or later, the timing to execute the study remains about the same.

- 0-10 months:
  - Develop and submit package for clearance from the Office of Management and Budget[36] for survey instrument and case study interview protocol. Collect, clean, and analyze "input data" on the FY07-09 NIA awardees and the proposed comparison group, R01 ESIs awardees. If sampling is required, develop sampling plan and perform until satisfactory comparison group is constructed.
- 11-14 months: field survey and prepare for expert panel
- 1-18 months: collect bibliometric and funding data.
- 13-18 Conduct expert panel reviews and code results
- 10-18 months: Conduct case studies of randomly selected NIA awardees and comparison group.
- 19-24 months: synthesize results and write evaluation report.

[36] It takes 9 to 12 months to obtain OMB clearance after the package is submitted.

Because of the cost of this study,[37] various options can be considered:

- Option 1 would be to conduct the study based on the full number of 2007-2008 NIA awards and R01 ESIs (61 each) and 10-15 case studies from each group. This approach but would give greater confidence in any findings of difference between the NIA program and the comparison ESI R01 program (estimated cost: ~$875K). The components of the study would include expert review, data collection from NIH databases, Web of Science, and Survey of Outcomes, bibliometric analysis, and case study. The full population of 2007-2008 NIA and R01 ESI awardees would be included in the analysis.
- Option 2 would involve conducting the study based on a smaller sample of NIA R01 ESI awardees – up to 20 awardees from each program and 5-7 case studies from each group. Limiting the study might run the risk that it might not be possible to generalize the 'innovativeness' findings from the expert panel to the full program (estimated cost: ~$590,000). The components of the study would include expert review, data collection from NIH databases and Web of Science, bibliometric analysis, and case study. A survey would not be undertaken.

C. Conclusion

STPI found that an outcome evaluation of the NIA program is feasible. The goal of the study is to understand the extent to which the research conducted by NIA awardees is more *high risk* (where high risk is defined as research that has an inherent high degree of uncertainty) and *high reward* than research conducted by other early stage investigators (ESIs). To address this, we propose a quasi-experimental design using awardees from the R01 ESI program as a comparison group. Given NIH's interest in outcomes of the NIA program as compared with traditional programs, we propose that the primary evaluation method be expert review, supplemented by bibliometric analysis, administrative data, survey, and case studies.

STPI believes the evaluation proposed in this document will likely be of interest to those outside of the NIA program as it will add to the understanding of the effect of different funding mechanisms for biomedical and behavioral research, including the R01 mechanism, the foundation of NIH funding.

[37] See Appendix H.

Appendix A. Indicators of Innovative Research (A-1) and Researcher's Career (A-2)

Table A-1. Example Indicators of Innovative Research (not all were referenced in the literature review)

Outcome	Metric	Data Needed	Description	Selected Reference
Creativity	Consensual assessment	Expert Review	This method uses a panel of judges to rate the creativity of an artistic product by using their own subjective definition rather than any given objective criteria. (Amabile 1982)	Amabile 1982; Heinze, Shapira et al. 2007;Grant & Allen 1999, National Academies 1999)
Productivity	Overall productivity	Number of author publications	This measure provides a coarse measure of the potential impact of a researcher's work, according to the theory that when a researcher has published more, there is a greater likelihood that his or her work will be considered high-impact by peers.	Heinze and Bauer 2007; Simonton 2003
Impact	Citation rates	Citation count for each publication, Year of publication	This is a measure of research impact. Creative and impactful publications should be cited more frequently, which can be normalized for the age of the publication.	Heinze and Bauer 2007; Azoulay 2009
	Fractional counting of citations	Citation count for each publication, Number of cited references of each citing publication	This method normalizes for differential citation patterns between research fields.	(Leydesdorff 2010)
	H-index	Number of publications, Citation count for each publication	The H-index is a combined measure of the productivity and impact of a scientist. It is insensitive to never or seldom cited papers as well as to one or several highly cited papers.	Hirsch 2005; (Cronin and Meho 2006); Azoulay 2009
	H-rate	Number of publications, Citation count for each publication, year of publications	This measure builds on the H-index and addresses variations in publication frequency and citation rates over the length of a researcher's career. It is based on the average annual increase in an	(Burrell 2007)

Outcome	Metric	Data Needed	Description	Selected Reference
			author's H-index .	
	G-index	Number of publications, Citation count for each publication	This measure builds on the H-index by giving more weight to highly-cited articles in an attempt to capture the difference in citation patterns of the most highly cited papers	(Egghe 2006)
	AR-index (AWCR)	Number of publications, Citation count for each publication, year of publication	This measure normalizes the citation rate of each article in an author's publication set for the age of the article.	(Jin 2007)
	Journal impact factor	Journal Impact Factor (calculated by other parties)	This is a proxy for journal quality and characterizes the potential exposure and impact of an article published in a specific journal. A journal is considered "high-impact" if it is cited often by other highly cited journals.	(Saha 2003); (Seglen 1997); (Bergstrom 2010)
Network Brokerage	Network brokerage	Authors of each publication in data set	This is a measure of an individual's connections. People with more connections to distinct social networks are considered brokers and hypothesized to have more creative outputs since they are exposed to more diverse ideas.	Burt 2004; Heinze and Bauer 2007
Interdisciplinarity	Degree centrality	Number of authors for each publication in data set	This measures the size of a researcher's network. While degree centrality does not predict a creative event, but it does correlate with impact (via the number of author-level citations).	Heinze and Bauer 2007
	Integration score	List of References in Publications, Subject Categories of each publication	This measures the diversity and distribution of knowledge used to create the research by using cited references. Multiple applied works have used integration scores as a sign of creativity. It ranges from zero (low diversity of cited references) to one (high diversity of cited references)	Porter, Cohen et al. 2007; Heinze and Bauer 2007
	Specialization score	List of publications, Subject Categories for each publication	This is an indicator of the multidisciplinarity of a publication set. This score is based on the span and distribution of subject categories of a researcher's publications.	Porter, Cohen, et al. 2007

Table A-2. Example Indicators of Career Status (not all were referenced in the literature review)

Outcome	Metric	Data Needed	Description	Selected Reference
Researcher Recognition	Awards/honors	List of awards and honors received by PI	Honors and awards listed on researchers' curricula vitae as a proxy of the impact of creative research, as well as of career independence.	Simonton 2003, Westat 2007
	Awards/honors to lab trainees	List of awards and honors received by students from a given lab	Students and fellows trained at the lab who go on to win honors/awards recognizing creative researchers such as Pew, Searle, Beckman, Packard, and Rita Allen scholarships.	Azoulay 2009
	Patents	Patents	Patents are filed in the case when an invention has been made that may be of use to protect.	Stephan, Gurmu et al. 2007; Azoulay, Ding et al. 2006, Westat 2007
	Professional Development	Service on Review Committees	Service on Review Committees can be a measure of the involvement in the research community	Westat 2007
Researcher Career Path	Employment Status	Tenure	The status of the researcher as to whether they have been awarded tenure or not, and the time to tenure is a measure of researcher independence	(NRC 2006), Westat 2007
		Prestige of Faculty Position	If the researcher is employed at a top-tier university	(NRC 2006)
	Funding Status	Time to Obtain first R01 grant	The obtaining of an NIH R01 grant is seen as a stage in researcher independence.	(NRC 2006), Westat 2007

Appendix B. Detailed Study Questions

Table B-1. Detailed Study Questions: Innovative Research

Key Study Question	Study Question	Detailed Study Question	Expert Review	Publication Databases	Progress Reports	Survey/Interview	Case Study	NIH Staff	NIH QVR	USPTO
1. Did the NIA program stimulate highly innovative (HRHR) research?	To what extent was the research conducted by NIA awardees more high risk than research conducted by other early stage investigators (ESIs)? (where high risk is defined as research which has an inherent high degree of uncertainty)	Was the research viewed by the community as being innovative and high risk?	X	X		X	X			
		Did the research lead to a new area of science, either through the formulation of new ideas or the synthesis of existing ideas?	X	X	X	X	X			
		Were the research outputs high-impact?	X	X	X	X	X	X		
		Did the research result in the discovery of a new empirical phenomenon?	X	X	X	X	X	X		

Key Study Question	Study Question	Detailed Study Question	Expert Review	Publication Databases	Progress Reports	Survey/Interview	Case Study	NIH Staff	NIH QVR	USPTO
		Did the research result in the development of a new methodology that enables the empirical testing of theories?	X	X	X	X	X	X		
		Did the research result in the invention of novel instruments that opened up new research possibilities?	X	X	X	X	X	X		
	To what extent did the outputs and outcomes of NIA-funded research lead to or were likely to lead to advances in biomedical and behavioral research? How do these advances compare to those of a traditional NIH program that funds ESIs?	Were follow-on grants received to continue the research?			X	X	X		X	
		Did the publications from the research appear in biomedical and behavioral research fields or outside these areas?			X	X				

Key Study Question	Study Question	Detailed Study Question	Expert Review	Publication Databases	Progress Reports	Survey/Interview	Case Study	NIH Staff	NIH QVR	USPTO
		Did the research result in patents? Were the patents in the biomedical and behavioral research fields or outside these areas?			X	X				
		Were other research outputs produced? Were they in the biomedical and behavioral research fields or outside these areas?			X				X	
•What were the program's spillover benefits, especially on the careers of NIA grantees?	What fraction of the awardees remains in biomedical-related fields as compared with other ESIs?	Did they stay in same disciplinary area?		X	X	X	X		X	
		How integrated are they in their fields or in the scientific community?		X		X	X			
		Have they developed new courses? Are they in new areas or disciplines?		X	X	X	X			

Key Study Question	Study Question	Detailed Study Question	Expert Review	Publication Databases	Progress Reports	Survey/Interview	Case Study	NIH Staff	NIH QVR	USPTO
	What is the nature of the research (whether continued HRHR or other) and total funding received by NIA-funded researchers, as compared to other ESIs?	Did they apply for and receive research funding? What are the areas of funding? Did they change?			X	X	X	X		
	What fraction of NIA awardees are becoming leaders in their fields, as compared with other ESIs?	How are they perceived by the scientific community?		X	X	X	X			
	What is the career progression of the NIA awardees compared with other ESIs?	How quickly do they publish as lead faculty member (i.e. without former advisors)?		X	X	X	X			
		How quickly do they obtain tenure?			X	X	X			
		Do they move or receive offers to move to other institutions?			X	X	X	X		

Appendix C. Comparison of NIA Program with Other Funding Programs that Could Provide Potential Comparison Groups

Table C-1. Detailed Comparison of Other Young Investigator Programs

Program	Key Program Characteristics	Similarities to NIA Awardees	Differences from NIA Awardees	Research Question Appropriateness
NIA finalists (DP2)	N/A	N/A	Possible difference in quality of researcher/ proposal	Innovativeness and Impact Career progression
NIA Applicants (DP2)	N/A	N/A	Likely difference in quality of researcher/proposal	Innovativeness and Impact Career progression
NDPA (DP1)	Not a Young Investigator Program Yearly funding Large award aimed at innovative researchers Similar time requirement (51% of research effort) Preliminary data not required	Focus on creativity Cuts across disciplines Similar flexible funding No preliminary data requirements	Different career stage Small number of awardees Different funding stream (yearly instead of up front)	Innovativeness and Impact
R01-ESIs	No innovative research goals Lower average funding amount Yearly disbursement of funds Possible extension of funds Identical career stage requirements	Similar career stage requirement Cuts across disciplines	Review process varies by IC and RFA High-risk research not emphasized Similar award size Different funding stream (yearly instead of up front) Larger group of awardees	Innovativeness and Impact Career progression

Program	Key Program Characteristics	Similarities to NIA Awardees	Differences from NIA Awardees	Research Question Appropriateness
Pathways to Independence (K99/R00)	Includes mentorship requirement Different research experience requirement Must have less than 5 years post doc training) Cannot have tenure track faculty position # of years of grant funding is smaller Yearly amount is significantly smaller		Shorter research duration Smaller total research funding High-risk research not emphasized Different career stage	Career progression
ONES/BRAINS (R01)	Small number of grantees Specific discipline Different previous research restrictions (no requirement in terms of past awards) 50% time requirement Access to additional funds for career development Money not upfront Is in the R01 system Advisory committee Large award aimed at innovative researchers ESI requirement	Funding size similar Data availability Similar career stage Focus on innovative research	Small number of awardees Limited research focus Larger time requirement Disbursement of funds	Innovativeness and Impact Career progression
Type 1 Diabetes Pathfinder Award (DP2)	Disbursement of funds done annually Small number of grantees Restricted to diabetes research DP2 funding mechanism Large award aimed at innovative researchers Preliminary data not required	Focus on innovative (Is the funding flexible?) Similar funding mechanism	Small number of awardees Limited research focus	Innovativeness and Impact Career progression

Program	Key Program Characteristics	Similarities to NIA Awardees	Differences from NIA Awardees	Research Question Appropriateness
The Mentored Research Scientist Development Award (K01)	Meant for researchers who propose projects in new areas from past research Awardees must have a "full time" appointment at the academic institution that is the applicant institution Salary support only Aimed at new investigators	Large sample size Similar career stage	Variation of awardee characteristics based on IC level criteria Not a research grant No career stage restriction	Career progression
Independent Scientist Award (KO2)	Salary support only Aimed at newly independent investigators The candidate must have a doctoral degree and independent, peer-reviewed research support at the time the award is made	Data availability Similar career stage	Variation of awardee characteristics based on IC level criteria Not a research grant	Career progression
Mentored Clinical Scientist Research Career Development Award (K08)	Meant for clinical degree holders Salary support only Aimed at new investigators	Data availability Similar career stage	Variation of awardee characteristics based on IC level criteria Not a research grant Different population	Career progression
HHMI Early Career Scientist Program	Tenure track position for 2-6 years Award focused on person not award Large award aimed at innovative researchers Focus on innovative research	Similar sample size Focus on innovative research Similar career stage Similar award size Ability to change research directions	Data availability Different population (Social and Behavioral scientists not included)	Innovativeness and Impact Career progression
PEW Scholars Program in the Biomedical Sciences	Size of award is smaller Small number of awardees Focused on outstanding young investigators in the biomedical sciences For new tenure track professors	Similar career stage	Data availability Difference population (Social and Behavioral scientists not included)	Innovativeness and Impact Career progression

Program	Key Program Characteristics	Similarities to NIA Awardees	Differences from NIA Awardees	Research Question Appropriateness
Burroughs Wellcome Career Awards at the Scientific Interface	Small research award Small number of awardees Aimed at new investigators in technical disciplines who want to work in the biomedical sciences Awardees cannot have a tenure track position but have a committed institution	Focus on innovative research	Data availability Small number of awardees Different award size Slightly difference career stage	Innovativeness and Impact Career progression

Appendix D. Operationalizing "Support of Promising New Investigators"

The second study question for the proposed New Innovator Award (NIA) outcome evaluation is how did the NIA "support promising new investigators." There are several ways in which this goal could be operationalized. The program could be supporting new investigators merely by giving them the funds to do the research, for example, and the outcomes looked for would be based on the research the awardees accomplish. Based on discussions with NIA program staff, the interpretation STPI recommends comparing NIA and R01 Early Stage Investigators(ESI) awardees by assessing the fraction that remain in biomedical-related fields five years after the receipt of the NIA, the nature of their subsequent funding, whether they are becoming leaders in the fields, and their career progression. Some of the indicators that have been used to measure career status are:

Employment and tenure*:* Although a few researchers have included alternative career paths in their evaluations (Mavis and Katz 2003), the majority of the evaluations have focused on promotion into a tenure track position at an institute of higher education. Employment has been measured in various ways, including: time to becoming an assistant professor (Pion and Cordray 2008; Williard and O'Neil 1998; Gaughan 2009; NRC 2006; Pion and Hammond 2005) and time to promotion to full professorship (NRC 2006). Indicators of the quality of the institution qualified and added to these employment measures; they included: the rank of the institution in terms of NIH funds (Pion and Cordray 2008; NRC 2006; Pion and Hammond 2005) and the National Research Council program rating (Mavis and Katz 2003).

Research funding: This category refers to the ability of the awardee to obtain external research funding to continue their research. This variable has been operationalized as: time to first NIH grant (Gaughan 2009; Pion and Hammond 2005), time to first NIH grant as a PI or first R01 (Gaughan 2009; Pion and Cordray 2008; NRC 2006; Mavis and Katz 2003), and time between attaining professorship and receiving their first R01 grant (NRC 2006). More generally, this has been operationalized as the age at receipt of first federal funding (Williard and O'Neil 1998; Mavis and Katz 2003).

Research outcomes: This measure refers to the scientific productivity of the awardee as an indicator of research independence. The measure refers to the quality, quantity, and subject areas of publications. In terms of quality and quantity, measures have included: total number of articles (NRC 2006), number of articles published in top-ranked journals, average citations per article (Pion and Cordray 2008; NRC 2006; Mavis

and Katz 2003), citations per individual (NRC 2006; Mavis and Katz 2003), and average impact factor of the author (Pion and Hammond 2005). Evaluators often use the 224 ISI Web of Science subject categories to categorize the journals in which awardees have published (Pion and Hammond 2005).

Other Indicators: Additional indicators include career roles; research and teaching loads; service on review committees; membership in research-oriented professional associations; grants applied for and received; NIH priority scores; salary level; and the reception of their scientific work by the research community (serving as journal editors and writing book chapters) (Westat 2007).

Appendix E. Details on Pilot Data Collection

Descriptions of Data Sources

Curriculum Vitae of Awardees and Comparison Researchers

The CVs of awardees and comparison researchers can serve as a rich data source for information on career independence, and for selected research outputs. CVs unfortunately are not standardized and can vary considerably in the information they contain. Many researchers post their CVs on public websites, and they can be collected in this fashion. For those researchers who do not have publically-available CVs, email queries can be sent to request electronic copies of their most recent CVs. Information that can typically be gathered from CVs includes department, highest degree and year of receipt, field of highest degree, publication data and awards.

Application Biosketches of Awardees and NIH Comparison Researchers

All applicants to NIH grants are required to submit a biosketch as part of their application. This short document resembles abridged curriculum vitae. It includes selected publication data, demographic data, employment, and education data. Biosketches may be obtained for all applicants of NIH grants.

Progress Reports of Awardees and Comparison Researchers

The progress reports of NIA awardees contain information on research activities, and research outputs such as publications and patents. The data contained in progress reports are unstructured, leading to challenges in automating extraction of key information.

Bibliometric Databases

The following bibliometric databases can be used in conjunction to provide a relatively complete picture of the publication and patent output of NIA awardees and comparison researchers:

- Web of Science
- Scopus
- PubMed
- Journal Impact Factor Databases
- United States Patent and Trademark Office (USPTO) databases

Web of Science is an online academic citation indexing tool provided by Thomson Reuters. It provides access to over 10,000 of the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings and is used by most researchers working in bibliometrics.[38] SciVerse Scopus is an alternative abstract and citation database whose services are comparable to those of Web of Science. It contains nearly 18,000 titles from 5,000 publishers worldwide and over three million conference papers.[39] Since Scopus has a smaller database than Web of Science, Web of Science is considered the superior alternative. Preliminary queries of the NIA awardees suggested that PubMed had fewer publications than Web of Science and no unique articles to add to what could be found in Web of Science. Consequently, further pursuit of PubMed as a source was deemed unnecessary.

Searches will likely have to be done on individual PI names, as searching only on funding acknowledgements is likely to miss many publications, as was demonstrated in the pilot data collection discussed below. Information that can be gathered from these sources includes number of publications, number of patents, journals of publication, and some citation indices (although others will have to be calculated elsewhere).

To help rate impact of publications, journal impact factor databases exist. Journal impact factor information can be obtained from the free online database eigenfactor.org. This information helps to identify the potential impact of subjects' research outputs, since it is an indicator of potential citation rates of a publication. Journal impact factors may alternatively be obtained from *Journal Citation Reports*, a service provided by Thomson Reuters

To identify subjects' patenting level, the United States Patent and Trademark Office (USPTO) has two free online databases, the Patent Full-Text and Image Database (PatFT) and the Patent Application Full-Text and Image Database (AppFT) to search for issued patents and published patent applications. [40] Patent applications that have not been approved for publication by their inventors or have not been pending for over 18 months are not searchable at all. Several alternative free patent databases exist online. Exploratory evaluation of Patent Lens and Free Patents Online revealed that the USPTO databases provide the most refined results when searching by inventor name.

---

[38] See http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/.

[39] See http://www.info.sciverse.com/scopus/about/.

[40] A published patent application is a pending patent application that has been filed over 18 months ago and approved by the inventor for public release. There is a benefit to the inventor for allowing their applications to be published; if their patent application matures into a patent (i.e., is approved), they may collect royalties for the period of time between the date of application publication and the date the patent is issued.

QVR/IMPAC II

The internal NIH data system accessed through Query/View/Report (QVR) contains information on applicants to NIH grant programs. QVR pulls can be done to ascertain grant application history and NIH funding information for NIA awardees and potential comparison groups. In addition, employment data is found in QVR as well as all past and current NIH funding. The internal NIH database also contains information on race/ethnicity and gender.[41]

Survey of Awardees and Comparison Researchers Outputs and Outcomes

A survey or interviews with NIA awardees and comparison researchers would function to gather information on the research activities, and to supplement missing data from the above sources. Interview questions should be structured around the study questions, and some of them may need to be qualitatively coded.

Interview of NIH Staff

Interviews with NIH program officers and NIH leadership will likely be semi-structured, and answers to the questions will be used to inform the evaluation team about changes within NIH.

Expert review

A panel of three experts per awardee will review 3 publications that best showcase the awardees work. There 54 NIA awardees that corresponds to an equal number of ESI R01 awardees who serve as the comparison groups.

Pilot Data Collection Results

Publication Data

Pilot Process

Publication data was compiled for each of the ten researchers in the pilot group using Web of Science. To evaluate the completeness of Web of Science (WoS) queries, we identified how many publications listed on CVs were found in our Web of Science data pulls by researcher. This process enabled us to approximate the percentage of the researchers' publications that database would capture. CVs were not considered as a source of publication information; two of ten in the pilot group indicated that only selected publications were listed on their CVs, and the other eight researchers made no

[41] Some but not all data are available in the publically-available RePORTER . To conduct the NIA Outcome Evaluation will require access to the internal data through the QVR/IMPAC II system.

indication whether their list of publications was complete. Progress reports and application biosketches were not seriously considered for publication data collection because it was evident that publication data in these sources was not complete.

Results of Pilot Study

The mean and median percentage of CV publications found in Web of Science (WoS) was 85% (See Table E-1.) It should be noted that it is assumed that publication data is found to be incomplete, because of 20% of the pilot had "Selected Publication" as the title in their CV. WoS is shown to be an incomplete source, since there is considerable variation in how many of the publications listed on a CV are found in WoS.

Table E-1. Results of Publication Data Collection

Group	Unique Identifier	Indication of "Selected" or "All" Publications on CV	Number of publications listed on CV	Percentage of CV publications found in WoS
NIA	1	unspecified	14	71%
NIA	2	unspecified	12	75%
NIA	3	selected	22	95%
NIA	4	selected	24	83%
NIA	5	unspecified	23	78%
R01	6	unspecified	7	86%
R01	7	unspecified	19	89%
R01	8	unspecified	17	76%
R01	9	unspecified	23	100%
R01	10	unspecified	64	97%

Source: CVs, Web of Science.

Patent-Based Bibliometric Indicators

Pilot Process

Search queries within the USPTO database were performed by inventor name, or more specifically, researcher name. Each patent was examined individually in order to determine if the patent was written by the correct researcher. Patent number, date filed, and date issued were easily found on this page. By clicking on a "cited by" references link, the number of times and where the patent was cited could be found. When researchers provided patent information on their CVs, this information was cross-checked against the patent information found via USPTO search queries.

Results of Pilot Study

Five of ten awardees had at least one patent or published patent application registered in the Patent Full-Text and Image Database (PatFT) and the Patent Application Full-Text and Image Database (AppFT) . Three of the five individuals with patents or patents pending noted patent information on their CVs. A comparison of the number of patents issued and pending from both sources is shown in Table E-2. The number of patent citations for a researcher's collective patents is also displayed.

Table E-2. Results of Patent Data Collection

Group	Unique Identifier	Number of Patents and Patents Pending Listed on CV	Number of Patents and Patents Pending on USPTO	Number of Patent Citations, for patents on USPTO
NIA	1	189	217	5
NIA	2	0	0	N/A
NIA	3	0	1	1
NIA	4	0	0	N/A
NIA	5	0	0	N/A
R01	6	0	1	0
R01	7	3	1	0
R01	8	0	0	N/A
R01	9	0	0	N/A
R01	10	3	0	N/A

Source: PatFT and AppFT on USPTO website, CVs.

Employment

Pilot Process

STPI had access to employment data through a number of independent sources. This data could be found in application biosketches, CVs, and QVR. Position title, institution, and years of employment were the fields of interest. If researchers had multiple appointments at the time of receiving the award, the time of the start of the evaluation, or both, all of the appointments were collected.

Results of Pilot Study

All sources considered had some form of employment information for 100% of the appropriate pilot group members. Position title and institution had to be collected manually from CVs and application biosketches. Using the Python program, these fields could be gathered automatically from application biosketches. Some pre-award employment information from QVR could be downloaded automatically into Excel, but

this information tended to be incomplete. For instance, position titles were not consistently found in QVR. Table E-3 displays the information available from each source.

Table E-3. Results of Employment Data Collection

Source	Pre-Award Job Title	Pre-Award Institution	Years of Pre-Award Employment	Post-Award Job Title	Post-Award Institution	Years of Post-Award Employment
Application biosketches (NIA only)	100% (5)	100% (5)	100% (5)	0%	0%	0%
CVs	100% (10)	100% (10)	100% (5)	100% (10)	100% (10)	100% (5)
QVR	30% (3)	100% (10)	0%*	0%*	0%*	0%*

*Values in table indicate that this information field was available for the noted source. Years of employment was not available from QVR. Post-award information was not consistent in QVR because it could only be gathered if researchers had applied to another NIH grant after having received the award. Furthermore, the information for different researchers would come from different times. Source: NIA application biosketches, CVs, QVR.

Education

Pilot Process

STPI accessed education information data through application biosketches, CVs, and QVR. The comparison group's data was available through their CVs and QVR. The fields of interest were: doctoral degree type, institution from which degree was received, and year of degree receipt. Multiple doctoral degrees were noted for researchers where appropriate.

Results of Pilot Study

All three sources had some form of education information for 100% of the appropriate pilot group members. Five of five NIA awardees had all relevant education fields on their application biosketches. Ten of ten pilot researchers had all relevant education fields on their CVs. Education data availability was less complete in QVR. Information had to be collected manually from application biosketches and CVs while information was collected automatically from QVR. Information found in multiple sources was the same. Table E-4 displays the results of the pilot.

Table E-4. Results of Education Data Collection

Source	Type of Doctoral Degree	Institution Where Degree Was Attained	Year When Degree Was Attained
Application biosketches (NIA awardees only)	100% (5)	100% (5)	100% (5)
CVs	100% (10)	100% (10)	100% (10)
QVR	100% (10)	0%*	90% (9)

Source: NIA application biosketches, CVs, QVR.

Note: Values in table indicate that this information field was available for the noted source. Institution where degree was obtained was not a downloadable field in QVR.

Research Funding and Grants

Pilot Process

STPI examined CVs for the availability of research funding information. Fields of interest included: project title, grant number, funding organization, funding period, funding amount, and PI's role in the grant application. QVR was examined to determine the ease of accessibility of NIH funded and unfunded applications.

Results of Pilot Study

Nine out of ten CVs included some form of research funding information, but the types of information included varied greatly. Of the nine researchers who had a funding section, nine included current research support, seven included completed research support, five included pending support, and one included unfunded grant applications. Project titles, funding organizations, and the funding periods were almost universally included. The results of the CV data collection are in Table E-5.

Ten of ten researchers had downloadable NIH funding information from QVR. Information available on QVR includes: project number, total award cost, project start date, project end date, early-stage investigator and new investigator eligibility at time of grant application, fiscal year of first R01 receipt.

Table E-5. Funding Data Availability on CVs

Group	Unique Identifier	Inclusion on CV	Project Title	Organi-zation	Funding Period	Funding Amount	Grant Number	Researcher Role
NIA	1	—	—	—	—	—	—	—
NIA	2	X	X	X	X	—	—	X
NIA	3	X	X	X	X	—	X	X
NIA	4	X	X	X	X	X	X	—
NIA	5	X	—	X	—	—	—	—
R01	6	X	X	X	X	X	—	—
R01	7	X	X	X	X	—	X	X
R01	8	X	X	X	X	X	—	X
R01	9	X	X	X	X	X	X	X
R01	10	X	X	X	X	X	X	X

Source: CVs.

Awards

Pilot Process

STPI examined the availability of honors and awards information in application biosketches and CVs.

Results of Pilot Study

Nine of ten researchers included awards and honors on their CVs while four of five researchers included honors sections in their biosketches. While awards information would have to be collected manually from CVs, the STPI-developed Python tool could be used to export this data from biosketches into Excel. Table E-6 provides the results of the pilot exploration. Most awards reported are fellowships, other grants that are called awards (i.e., HHMI early career award, Hood Foundation early career award), induction into societies (i.e., Alpha Omega Alpha- honor medical society), and awards given out by their universities. Some researchers included awards from high school and undergraduate, making the summation of all reported awards not an accurate measure. Because of the inconsistencies in terms of which awards to include, meaningful data collection from these sources for this indicator is not suggested.

Table E-6. Results of Award Information Collection

Group	Unique Identifier	Awards Section on CV?	Awards Section in Biosketch?
NIA	1	Yes	No
NIA	2	Yes	Yes
NIA	3	Yes	Yes
NIA	4	Yes	Yes
NIA	5	Yes	Yes
R01	6	Yes	N/A
R01	7	Yes	N/A
R01	8	No	N/A
R01	9	Yes	N/A
R01	10	Yes	N/A

Source: NIA application biosketches, CVs.

Alternative Indicators

Pilot Process

STPI examined the data availability of alternative indicators on CVs and application biosketches. Alternative indicators examined were: membership on NIH study section, membership on other grant review committee, service as journal reviewer, service on journal editorial board, book chapter authorship, lectures given as an invited speaker, member on thesis committee, chair of thesis committee, number of graduate students trained, number of post-doctoral students trained, number of courses taught.

Results of Pilot Study

Table E-7 shows the availability of these indicators from the data sources. It should be noted that lack of information for one indicator could mean that these researchers did not engage in those activities. Researchers provided varying amounts of information on these indicators.

Table E-7. Data Availability of Alternative Indicators

Source	NIH Study Section Membership	Other Grant Review Committee	Journal/Manuscript Reviewer	Journal Editorial Board	Book Chapter Author	Invited Talks	Committee Member or Chair	Grad Students and Post-docs Trained	Course Taught
CVs	3	2	4	1	4	6	3	4	6
Application Biosketch	0	0	0	0	0	0	0	0	0

Source: NIA application biosketches, CVs.

Note: Numbers indicate number of pilot group researchers that provided this information.

Appendix F. Expert Reviewer Protocol

Purpose of Expert Review:
The goal of the expert review is to characterize the research undertaken with National Institute of Health (NIH) funds on the extent of its "innovativeness." Your evaluation, along with those of others, will be used to assess the outcomes of programs funded by NIH.

Definition of Innovative/High Risk, High Reward Research
The terms "innovative" and "high risk high reward research" are used to describe research that has the capability to produce a major impact on important problems in biomedical/behavioral research and that an inherent high degree of uncertainty (Austin 2008). [1]

Instructions:
Please read the three publications selected by the awardee to represent the research they accomplished with the support of NIH funds. Please keep in mind the definition of innovative/high risk, high reward research when reviewing the publications.
Please complete the enclosed Expert Reviewer Protocol for each awardee. If you prefer to complete this online, go to *(provide link)*
After submission, we *may* call you for a short (~30 minute) phone call with you to discuss your responses.

Non-Disclosure Agreement:
All non-public information that you receive as part of the NIH review shall be deemed proprietary information. It is understood that until either (a) the information is made public through publication in a journal or (b) NIH grants the expert with specific written approval, the expert will, both during the review and thereafter: treat the information as confidential; not use the information except to answer the questions below; and not disclose the information to a third party without prior written approval from NIH.
Expert's Initials:

[1] The report referenced in the Austin presentation is "National Institutes of Health Report to Congress on Certain Demonstrations Projects: Bridging the Sciences and High-Risk High Reward Research, September 2009.

I. Characterizing the risk of the research

a. Please indicate which of the following statements (if any) are true of the research described in the 3 publications (or other materials provided) by the researcher[2]:*(Choose all that apply.)*

☐ One or more of the fundamental ideas underlying the research were at odds with prevailing wisdom at the time (Conceptual Risk)

☐ The research required use of equipment or techniques that had not been proven or were extraordinarily difficult (Technical Risk)

☐ The research required knowledge of fields beyond the researcher's previously demonstrated area of expertise (Experience Risk)

☐ The research involved a unique and unprecedented combination of perspectives, disciplines, or approaches (Multidisciplinary Risk)

☐ None of these statements is true of the research

b. How would you rate research idea on (TYPE of RISK)? *(This question will appear for each type of risk checked.)*

(Choose one.)
☐ Not risky
☐ Of medium risk
☐ Very risky
☐ I do not know

c. Why did you choose the statement(s) above? *(Enter answer in paragraph form.)*

II. Characterizing the outcomes of the research

a. Please indicate which of the following potential and/or realized outcomes apply to the research:[3] *(Choose all that apply.)*

☐ The research resulted in the formulation of a new ideas or the advancement of theoretical concepts (New Idea)

☐ The research resulted in the discovery of new empirical phenomena (New Phenomena)

---

[2] Dr. Rita R. Colwell, Keynote Address to the International Life Sciences Summit of Georgetown University, Washington, D.C., October 20, 2003. Accessed from http://www.nsf.gov/news/speeches/colwell/rc031020lifesci_summit.htm.

[3] T. Heinze et al., "Identifying creative research accomplishments: Methodology and results for nanotechnology and human genetics." *Scientometrics*, Vol. 70, No. 1 (2007) 125–152.

☐ The research resulted in the development of a new methodology, enabling empirical testing of theoretical problems (New Methodology)

☐ The research resulted in the invention of novel instruments that open up new research possibilities (New Technology)

☐ The research resulted in new synthesis of existing disparate ideas (New Framework)

☐ None of these statements is true of the research.

b. How would you rate the research on creating a (TYPE of OUTCOME)? *(This question will appear for each type of outcome checked.)*

(Choose one.)

☐ Strongly agree
☐ Moderately agree
☐ Moderately disagree
☐ Strongly disagree

III. Degree of innovation

a. To what extent do you agree that the accomplished research was innovative?

(Choose one.)

☐ Strongly Agree

☐ Moderately Agree

☐ Moderately Disagree (skip to III.c)

☐ Strongly Disagree (skip to III.c)

b. How would you characterize the innovativeness? *(Check all that apply.)*

☐ Research introduced novel theoretical ideas

☐ Research created novel (original) systems that led to broader insights

☐ Research used cutting-edge experimental approaches resulting in major advances

☐ Research combined fundamental principles, modeling, and testable experiments in new ways

☐ Research translated scientific principals into laboratory practice

☐ Research pursued an antagonistic (contrary to the norm) approach with success

☐ Research will revolutionize field

☐ Research introduced new tools that are radically different than current tools

☐ Research cut across multiple disciplines and produced new ideas and technologies

☐ Other (Please describe) *(Enter answer in paragraph form.)*

c. If the research was not innovative, please select the reason why below.

☐ Another research group came up with similar technology

☐ Research was routine with modest and inconsequential results

☐ Research is solid, but impact on field is/will be average

☐ Other *(Enter answer in paragraph form.)*

Please provide examples of others who are doing innovative research in this area.
Other *(Enter answer in paragraph form.)*

Appendix G. NIH (New Innovator Award or ESI R01) Survey of Outcomes

Note: The survey will be administered to the 61 NIA awardees and 61 ESI R01 awardees.

Welcome to the NIH *(New Innovator Award or ESI R01)* Survey of Outcomes. Please provide responses to the following questions to the best of your ability. You may choose not to answer specific questions and it will not affect your ability to submit the survey.

Use the navigation bar at the bottom of each survey page to navigate the survey. Do not use your browser's navigation buttons. If you wish to review your responses at any point, click the "Review" button in the survey navigation bar.

If you would like to save the survey and come back to complete it another time, enable cookies in your browser, click the "Save" button at the bottom of any page, and use the original link you received to return to the survey. The survey should take approximately 15 minutes to complete.

To review the NIA Request for Applications, criteria, or processes, visit:
http://commonfund.nih.gov/newinnovator/
OR
To review the ESI R01 Request for Applications, criteria, or processes, visit:
http://grants.nih.gov/grants/new_investigators/resources.htm/ (and click on R01)

Note that participation in this survey is entirely voluntary. Your decision to participate will have no effect on your current or future NIH funding status. Respondent confidentiality will be protected to the extent provided by law, and only aggregate information concerning overall impressions will be reported to the NIH.

If you have questions or concerns regarding completing this survey, please contact us at NIAoutcomes@xxx.org.

OMB# 0925-xxx Exp: xx/20xx

Public reporting burden for this collection of information is estimated to average 15 minutes per respondent, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information.

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0534). Do not return the completed form to this address.

To begin the survey, click the "Next" button below.

Top 3 Research Outputs from *(NIA/ESI R01)*-funded research

Please provide your three most important outputs (often publications) from your *(NIA/ESI R01)* that will be used by expert reviewers to assess the outputs and outcomes from your *(NIA/ESI R01)* research.

1.
2.
3.

Potential Reviewers for Expert Review

Please provide the names of at least 3 possible reviewers (a mix of supporters and critics of your*(NIA/ESI R01)* research.

1.
2.
3.
4.
5.

Research Outputs from *(NIA/ESI R01)*-funded research

Other than the ones listed above, has the *(NIA/ESI R01)*-funded research resulted in other publication(s)?

(Choose one)

- ☐ No
- ☐ Manuscript(s) in preparation
- ☐ Yes

☐ **Publications, conference papers, and presentations**

Report only the major publication(s) resulting from the work under this award. There is no restriction on the number. However, agencies are interested in only those publications that most reflect the work under this award in the following categories:

☐ **Journal publications.** List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Include any peer-reviewed publication in the periodically published proceedings of a scientific society, a

conference, or the like. A publication in the proceedings of a one-time conference, not part of a series, should be reported under "Books or other non-periodical, one-time publications."
For each publication listed, identify author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no). Please list all your publications that you can attribute to (NIA/ESI R01) funding and the percent attribution.

☐ **Books or other non-periodical, one-time publications.** Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like.
Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (book, thesis or dissertation, other); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).
List all your books, etc. that you can attribute to (NIA/ESI R01) funding and the percent attribution.

☐ O**ther publications, conference papers and presentations**. Identify any other publications, conference papers and/or presentations not reported above.
Specify the status of the publication as noted above.
List all your other publications, etc. that you can attribute to (NIA/ESI R01) funding and the percent attribution.

Has your (NIA/ESI R01)–funded research resulted in other outputs? *(Choose one.)*
☐ Yes
☐ No

What products resulted from the project during the reporting period? *(Choose all that apply.)*
☐ Website(s) or other Internet site(s);
☐ Technologies or techniques;
☐ Inventions, patent applications, and/or licenses; and
☐ Other products, such as data or databases, physical collections, audio or video products, software or NetWare, models, educational aids or curricula, instruments, or equipment

If there is nothing to report under any of the following items, enter "Nothing to Report."

Website(s) or other Internet site(s)

List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Please provide URL for the website or other internet sites that you can attribute to (NIA/ESI R01) funding and the percent attribution.

Technologies or techniques

Identify technologies or techniques that have resulted from the research activities. Describe the technologies or techniques and how they are being shared.

List all your technologies or techniques that you can attribute to NIA funding and the percent attribution.

Inventions, patent applications, and/or licenses

Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

*List all your **inventions, patent applications, and/or licenses** that you can attribute to (NIA/ESI R01) funding and the percent attribution.*

Other products

Identify any other significant products that were developed under this project. Describe the product and how it is being shared. Examples of other products are:

☐ Databases;
☐ Physical collections;
☐ Audio or video products;
☐ Software or NetWare;
☐ Models;
☐ Educational aids or curricula;
☐ Instruments or equipment;
☐ Data & Research Material (e.g., cell lines, DNA probes, animal models); and
☐ Other.

*List all your **other products** that you can attribute to NIA funding and the percent attribution.*

Research Risk and Types of Outputs/Outcomes

Characterizing the risk of the research

How would you rate your NIA/ESI R01) research on risk? *(Choose one.)*

- ☐ Not risky
- ☐ Of medium risk
- ☐ Very risky
- ☐ I do not know
- ☐ I cannot recall

Which of the following statements (if any) are true of your (NIA/ESI R01) research outputs and outcomes?[45] *(Choose all that apply.)*

- ☐ One or more of the fundamental ideas underlying the research were at odds with prevailing wisdom at the time (Conceptual Risk)
- ☐ The research required use of equipment or techniques that had not been proven or were extraordinarily difficult (Technical Risk)
- ☐ The research required knowledge of fields beyond my area of expertise (Experience Risk)
- ☐ The research involved a unique and unprecedented combination of perspectives, disciplines, or approaches (Multidisciplinary Risk)
- ☐ None of these statements is true of the research

How would you rate research on (TYPE of RISK)? *(This question will appear for each type of risk checked. Choose one.)*

- ☐ Strongly agree
- ☐ Moderately agree
- ☐ Moderately disagree
- ☐ Strongly disagree

Why did you choose the statement(s) above? *(Enter answer in paragraph form.)*

[45] Dr. Rita R. Colwell, Keynote Address to the International Life Sciences Summit of Georgetown University, Washington, D.C., October 20, 2003. Accessed from http://www.nsf.gov/news/speeches/colwell/rc031020lifesci_summit.htm.

Characterizing the outcomes of the research

Please indicate which of the following potential and/or realized outcomes apply to your (NIA/ESI R01) research:[46] *(Choose all that apply.)*

☐ The research resulted in the formulation of a new ideas or the advancement of theoretical concepts (New Idea)
☐ The research resulted in the discovery of new empirical phenomena (New Phenomena)
☐ The research resulted in the development of a new methodology, enabling empirical testing of theoretical problems (New Methodology)
☐ The research resulted in the invention of novel instruments that open up new research possibilities (New Technology)
☐ The research resulted in new synthesis of existing disparate ideas (New Framework)
☐ None of these statements is true of the research.

How would you rate your (NIA/ESI R01) research on (TYPE of OUTCOME)? *(This question will appear for each type of outcome checked. Choose one.)*

☐ Strongly agree
☐ Moderately agree
☐ Moderately disagree
☐ Strongly disagree

Why did you choose the statement(s) above? *(Enter answer in paragraph form.)*

Comparing (NIA/ESI R01)-funded research with earlier research

Would you have chosen to seek NIH funding for your (NIA/ESI R01) research if the (NIA/ESI R01) program did not exist? *(Choose one.)*

☐ Yes
☐ No
☐ I do not know
☐ I cannot recall

Was the (NIA/ESI R01) research a significant departure from your previous research focus? *(Choose one)*

☐ Yes

---

[46] T. Heinze et al., "Identifying creative research accomplishments: Methodology and results for nanotechnology and human genetics." *Scientometrics*, Vol. 70, No. 1 (2007) 125–152.

☐ No

In what way(s) did your (NIA/ESI R01) project differ from your previous research focus? *(Enter answer in paragraph form.)*

In what way was your (NIA/ESI R01) research different from what is typically funded by the NIH? *(Choose all that apply.)*

☐ It had very little or no preliminary data
☐ It did not fall into the research interest of a single NIH Institute/Center
☐ It lacked an appropriate NIH study section
☐ As an investigator I had limited or no prior history in the proposed field
☐ I do not know
☐ My (NIA/ESI R01) idea was not different from what is typically funded by the NIH.
☐ Other

In your opinion, what is the likelihood that your (NIA/ESI R01) research would have been supported by any other funding sources (including from NIH awards and beyond)?

From the NIH *(Choose one.)*
☐ Very unlikely
☐ Somewhat unlikely
☐ Somewhat likely
☐ Very likely

From Other Sources *(Choose one.)*
☐ Very unlikely
☐ Somewhat unlikely
☐ Somewhat likely
☐ Very likely

Why do you think your (NIA/ESI R01) research could or could not have been supported by other NIH funding sources? *(Enter answer in paragraph form.)*

Why do you think your (NIA/ESI R01) research could or could not have been supported by other non-NIH funding sources? *(Enter answer in paragraph form.)*

To what extent was the idea proposed in your (NIA/ESI R01) application different than what you produced? *(Enter answer in paragraph form.)*

Since your (NIA/ESI R01) application, have you sought follow-on funding to your (NIA/ESI R01) project? *(Choose one.)*

- ☐ Yes
- ☐ No

What was the funding source(s)? *(Choose all that apply.)*

- ☐ NIH
- ☐ Other Government Agencies
- ☐ Private Foundation
- ☐ Institutional Resources
- ☐ Other

For each funding source checked,

Did you receive funding from this source?

- ☐ Yes
- ☐ No

If Yes, what program did you receive funding? (please specify)

What was the amount of the funding? $_____

Career progression since receipt of (NIA/ESI R01) award

Indicate which of the following important developments have taken place since your (NIA/ESI R01) application.

(Choose all that apply)

- ☐ I received an award(s)
- ☐ I was promoted
- ☐ I have received tenure
- ☐ I am applying for tenure
- ☐ I filed a patent application or have been granted a patent
- ☐ I received additional funding
- ☐ I expanded my research group
- ☐ I formed new partnerships/collaborations
- ☐ I changed my research focus
- ☐ I have expanded my research focus
- ☐ I changed institutions
- ☐ My work has been featured in the popular press and/or media

☐ I published an original, peer-reviewed article(s)
☐ Other

Which award(s) have you received since your (NIA/ESI R01) application? *(Enter answer in paragraph form.)*

Describe any other impacts of the (NIA/ESI R01) program not captured elsewhere in this survey. *(Enter answer in paragraph form.)*

Share any other thoughts related to the (NIA/ESI R01) program or to this survey. *(Enter answer in paragraph form.)*

You have reached the end of the survey. To view a summary of your responses before submitting, click the "Review" button below. Click "Finish" to submit your responses.

Exit Page

Thank you for completing the NIH Survey of Outcomes. We appreciate your help in the assessment of this important NIH program.

Please note that your responses will be kept strictly confidential. Only aggregate data from this assessment will be reported to NIH, and your participation will have no effect on your current or future NIH funding status.

Questions regarding this survey or the (NIA/ESI R01) program evaluation can be directed to:

NIHoutcomes@xxx.org

Please close this browser window to exit the survey and to ensure accurate recording of your responses.

For more information on the NIA program, please visit:
http://commonfund.nih.gov/newinnovator/
OR
To review the ESI R01 Request for Applications, criteria, or processes, please visit:
http://grants.nih.gov/grants/new_investigators/resources.htm/ (and click on R01)

Appendix H. Estimated Timeline and Cost for New Innovator Award (NIA) Outcome Evaluation

Conducting an outcome evaluation for the New Innovator Award program will involve multiple steps as outlined in Table H-1. A description of each step follows.

Table H-1. Timeline for Proposed NIA Outcome Evaluation, by number of months for entire NIA awardee population and comparison group (61+61) for 2007-2008

NIA Outcome Evaluation Timeline (Number of Months)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Option 1: Include all NIA awardees (61) & 61 R01 ESIs																									Total cost
Review & finalize study design (months 1-2)	■	■																							$ 27,500
Develop Protocols, OMB Clearance (Months 1-10)	■	■	■	■	■	■	■	■	■	■															$ 52,000
Select comparison group candidates (R01 ESIs)	■	■	■	■	■																				$ 26,000
Conduct survey (Months 11-14)											■	■	■	■											$ 87,000
Collect other data (Months 1-9)	■	■	■	■	■	■	■	■	■																$ 47,000
Pilot expert review (months 11-12)	■	■	■	■	■	■	■	■	■																$ 13,000
Expert Review for 122 awardees (Months 13-18)													■	■	■	■	■	■							$ 170,000
Bibliometric Analysis (122 awardees) (Months 1-18)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■							$ 55,800
Case Studies (30 case studies) (Months 10-18)										■	■	■	■	■	■	■	■	■							$ 201,000
Write report (Months 19-24)																			■	■	■	■	■	■	$ 150,000
Monthly meetings and udpates	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	$ 45,600
																									$ 874,900

Option 2: select sample of awardees																									
NIA Outcome Evaluation Timeline (Number of Months)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total cost
Review & finalize study design (months 1-2)	■	■																							$ 27,500
~~Develop Protocols, OMB Clearance (Months 1-10~~)																									$ -
~~Conduct survey (Months 11-14~~)																									$ -
Select comparison group candidates (R01 ESIs)	■	■	■	■	■	■	■	■	■	■															$ 26,000
Collect other data (Months 1-9)	■	■	■	■	■	■	■	■	■																$ 80,000
Collect 3 publications from WoS for each awardee (months 1-9)	■	■	■	■	■	■	■	■	■																$ 43,500
Bibliometric Analysis (60 awardees) (Months 1-9)	■	■	■	■	■	■	■	■	■																$ 36,600
Pilot expert review (months 11-12)										■	■														$ 20,000
Expert Review for 60 awardees (Months 110-15)										■	■	■	■	■	■										$ 78,000
Case Studies (10-14 case studies) (Months 10-18)										■	■	■	■	■	■	■	■	■							$ 118,000
Write report (Months 19-24)																			■	■	■	■	■	■	$ 142,500
Monthly meetings and udpates	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	$ 45,600
																									$ 590,200

Timeline: Option 1

Review and finalize study design (Months 1-2) [cost: $27,500]

- Develop study questions
- Update literature review
- Finalize study design

Survey and Expert Reviewer Protocols (Months 1-14)

Develop survey and expert protocols & submit OMB Clearance Package [cost: $52,000]

- Pilot survey and expert protocols. Submit revised protocols to OMB, if needed
- Obtain OMB Clearance
- Conduct survey ($87,000)
- Pull top 3 publications & 3 reviewers recommended by awardee for expert review

Select comparison group candidates (ESI R01s) (Months 1-5) [cost: $26,000]

- Choose characteristics for sample strata
- Stratify sample and select 61 ESI R01s
- Run descriptive statistics and compare to NIA population

Collect other data (Months 1-9) [cost: $47,000]

- Collect, clean, and analyze input data from NIH databases and other sources
- Tabulate and analyze results

Pilot Expert Review (~9 interviews for 2 awardees): (Months 13-18) [cost: $13,000]

- Select and invite expert reviewers
- Create Expert Review packages (3 publications, instructions, protocol)
- Summarize and code expert reviewer answers
- Assess if protocol or other changes need to be made & if so, make changes.

Conduct Expert Review (~37reviewers will review ~10 packages each, for a total of 366 reviews): (Months 13-18) [cost: $170,000]

- Select and invite expert reviewers
- Create Expert Review packages (3 publications, instructions, protocol)
- Summarize and code expert reviewer answers

Bibliometric Analysis (122 awardees) (Months 1-18) [cost: $55,000]

- Pull data from Web of Science for each awardee
- Conduct bibliometric analysis using standard and emerging indicators (see literature review) on NIA only and all publications
- Review and analyze results

Case Studies (equal amount of NIA and R01 awardees) (Months 10-18) [cost: $201,000]

- Review and write-up initial case study using awardee progress reports and collected data
- Conduct discussions with awardees
- Summarize and code expert reviewer answers

Write report (Months 19-24) [cost: $150,000]

- Summarize findings from each type of analysis
- Prepare report for NIH review. Finalize report
- Monthly meetings [cost: $45,00]

Estimated Costs

For option 1, study would involve the full set of 61 NIA awardees and 61 R01 ESIs awardees for 2007-2008. It would include all the components outlined above. The estimated cost would be $875,000.

For option 2, the size of the population would be reduced to 30 NIA awardees and 30 R01 ESIs. The number of case studies would also be reduced to 5 to 7 case studies for each program for a total of 10 to 14. Finally, a survey would not be conducted, so that an OMB clearance is not needed. OMB does not require clearance for expert review.[47] The estimated cost would be $590K.

[47] OMB asked us to remove the Expert Review protocol from the OMB Clearance package for the NDPA outcome evaluation.

References

Amabile, Teresa M. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43 (5):997-1013.

Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby, and Michael Herron. 1996. Assessing the Work Environment for Creativity. *The Academy of Management Journal* 39 (5):1154-1184.

Austin, Faye C. 2008. High-Risk High-Reward Research Demonstration Project. Paper read at NIH Council of Councils, November 20, 2008, at Bethesda, MD.

Azoulay, Pierre, Waverly Ding, and Toby Stuart. 2006. The Impact of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output. 57 (4):637-676.

Azoulay, Pierre, Joshua Graff Zivin, and Gustavo Manso. 2009. Incentives and Creativity: Evidence from the Academic Life Sciences. *NBER Working Paper*.

Bergstrom, Carl T. 2010. Use ranking to help search. *Nature* 465 (17 June):870.

Burrell, Quentin L. 2007. Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics* 73 (1):19-28.

Burt, Ronald S. 2004. Structural Holes and Good Ideas. *The American Journal of Sociology* 110 (2):349-399.

Colwell, Rita R. 2003. Keynote Address to the International Life Sciences Summit of Georgetown University. Paper read at International Life Sciences Summit of Georgetown University, October 20, 2003, at Washington, D.C.

Cronin, Blaise, and Lokman Meho. 2006. Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology* 57 (9):1275-1278.

Egghe, Leo. 2006. Theory and practise of the g-index. *Scientometrics* 69 (1):131-152.

Gaughan, M. 2009. Using the curriculum vitae for policy research: an evaluation of National Institutes of Health center and training support on career trajectories. *Research Evaluation* 18:117-124.

Grant, J. and Allen, L. 1999. Evaluating high risk research: an assessment of the Wellcome Trust's Sir Henry Wellcome Commemorative Awards for Innovative Research. *Research Evaluation* 8:201-204.

Heinze, Thomas, and Gerrit Bauer. 2007. Characterizing creative scientists in nano-S&T: Productivity, multidisciplinarity, and network brokerage in a longitudinal perspective. *Scientometrics* 70 (3):811-830.

Heinze, Thomas, Philip Shapira, Jacqueline Senker, and Stefan Kuhlmann. 2007. Identifying creative research accomplishments: Methodology and results for nanotechnology and human genetics. *Scientometrics* 70 (1):125-152.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102 (46):16569-16572.

Jin, Bihui. 2007. The AR-index: complementing the h-index. *ISSI Newsletter* 3 (1):6.

Kaplan D., Lacetera N., and C. Kaplan. 2008. Sample Size and Precision in NIH Peer Review. *PLoS ONE* 3 (7):e2761.

Lal, B., M. Hughes, E. Lee, A. Marshall, and S. Shipp. 2011. NIH Director's Pioneer Award Outcome Evaluation for FY 2004 – 2005. IDA Science & Technology Policy Institute.

Leydesdorff, Loet and Jung C. Shin. 2010. How to evaluate universities in terms of their relative citation impacts: fractional counting of citations and the normalization of differences among disciplines. In *arXiv*.

Mavis, Brian, and Michael Katz. 2003. Evaluation of a Program Supporting Scholarly Productivity for New Investigators. *Academic Medicine* 78 (7):757-765.

National Academy of Sciences. Committee on Science, Engineering, and Public Policy. 1999. *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. Washington, D.C.: National Academies Press.

NRC. 2006. *Evaluation of the Markey Scholars Program*. Washington, D.C.: The National Academies Press.

———. 2006. *Evaluation of the Markey Scholars Program*. Washington DC: The National Academies Press.

Ochse, R. 1990. *Before the gates of excellence: the determinants of creative genius*. New York: Cambridge University Press.

Pion, Georgine M., and David S. Cordray. 2008. The Burroughs Wellcome Career Award in the Biomedical Sciences. *Evaluation & the Health Professions* 31 (4):335-369.

Pion, Georgine M., and Charles B. Hammond. 2005. The American Association of Obstetricians and Gynecologists Foundation Scholars Program: Additional data on research-related outcomes. *American Journal of Obstetrics and Gynecology* 193 (5):1733-1739.

Porter, Alan, Alex Cohen, J. David Roessner, and Marty Perreault. 2007. Measuring researcher interdisciplinarity. *Scientometrics* 72 (1):117-147.

Rosenbaum, P. R., & Rubin, D.B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freemman. 1993. *Evaluation: A Systematic Approach*. Newburyt Park, CA: Sage Publications, Inc.

Saha, Somnath, Sanjay Saint, and Dimitri A. Christakis. 2003. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association* 91 (1):42-46.

Seglen, Per O. 1997. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314:498-502.

Simonton, Dean Keith. 1997. Creative Productivity: A Predictive and Explanatory Model of Career Trajectories and Landmarks. *Psychological Review* 104 (1):66-89.

———. 2003. *Creativity in science: chance, logic, genius, and Zeitgeist*. New York, NY: Cambridge University Press.

———. 2003. Scientific Creativity as Constrained Stochastic Behavior: The Integration of Product, Person, and Process Perspectives. *Psychological Bulliten* 129 (4):475-494.

Stephan, Paula E., Shiferaw Gurmu, Albert J. Sumell, and Grant Black. 2007. Who's Patenting in the University? Evidence from the Surevery of Doctorate Recipients *Economics of Innovation and New Technology* 16 (2):71-99.

Sternberg, R. 1990. *Wisdom: Its nature, origins, and development*. New York: Cambridge University Press.

Wagner, Caroline S., J. David Roessner, Kamau Bobb, Julie Thompson Klein, Kevin W. Boyack, Joann Keyton, Ismael Rafols, and Katy Börner. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* In Press, Corrected Proof.

Westat. 2007. Feasibility, Design, and Planning Study for Evaluating the NIH Career Development Awards.

Williard, R, and E O'Neil. 1998. Trends among biomedical investigators at top-tier research institutions: a study of the Pew Scholars. *Academic Medicine* 73 (7):783-789.