An Advancing the Science of "Science Funding Impact" White Paper

On

Best Practices for Funding Early Careers of Scientists: Evidence and Unanswered Questions

By

Pierre Azoulay
Kirk Doran
Megan MacGarvie

January 26, 2018

Recently there has been a surge of interest in funding early career scientists (i.e., scientists at the PhD, post-doc, and early tenure track stages of their careers). Changing demographics, the funding and regulatory landscape, and developing research needs have together made the choices and success of early career scientists a topic of interest for both funders and social scientists.

For example, social scientists have recently shown that the decline in birth rates over the last several decades and the end of mandatory retirement for university scientists since 1994 have combined to hasten the "graying" of the scientific workforce (Blau and Weinberg 2017). This graying could affect our future research output. In particular, if scientific creativity is greatest at younger ages, then such a shift in the age distribution could slow future scientific advances (Jones and Weinberg 2011). In this context, increases in early career funding could encourage a more balanced age distribution among investigators and therefore improved scientific output.

Furthermore, research suggests that early scientific careers may be more fragile than later ones, with the likelihood of leaving a career in science peaking at early career stages and decreasing as careers progress; this could mean that the most effective time to encourage people to remain in science would be at the early career stage, meaning that refocusing dollars towards the support of early careers could have a big impact on keeping the best minds in science. Another reason that a focus on early careers may make sense is that early career scientists may have the most flexibility to change their research agenda to focus on current needs and opportunities. Packalen and Bhattacharya (2017) show that younger scientists are more likely to build on new ideas in their work. During their early careers, scientists' research interests and methodologies may be most capable of being changed because they have not yet become entrenched in a long-term agenda. As a result, early career funding might be a particularly effective way to promote research on new ideas and new fields.

A final reason to focus on funding early careers is that established scientists often already have multiple ongoing sources of financial support. It is thus possible that creating new funding opportunities for established scientists will merely replace other sources of funding that these scientists would obtain anyway, this producing no changes in outcomes. But if funding levels for early career scientists are currently below optimal levels, then creating new funding opportunities for these scientists may be filling in a real gap and will therefore be more effective at producing real changes in outcomes.

In spite of these reasons for funding early scientific careers, there are important tradeoffs to consider. On the one hand, funders of early stage careers might improve scientific outcomes by providing early career researchers at the peak of their creativity with more autonomy relative to their outside options (which may be merely subsidiary roles in established researchers' existing funding streams). On the other hand, funders of later-career-stage researchers might do a better job picking the "right" people to fund by exploiting the information that success in the competitive funding environment provides. Because observations of graduate GPA, early career publications, and letters of recommendation are noisy measures of true potential, it may take many years before it is clear which researchers and which research programs are most and least likely to produce important results. Furthermore, it may be that aspects of scientific pedigree that are unrelated to the fundamental quality of the researchers or research programs themselves will matter more during early careers than they will after information revelation has occurred.

While individual funders are likely to have substantial insight into some of these questions, there is to date little systematic analysis of comprehensive data on early career funding that potential funders can rely on to inform their decision making. Given the substantial uncertainty surrounding these tradeoffs, we feel such a systematic analysis is warranted. This paper will discuss some of the prior empirical research on early career scientists, and will lay out some of what we view as the most pressing unanswered questions on this topic.

I. Review of the Nascent Literature

Despite the increasing importance of private sources of early career funding for scientific research, relatively little is known about the how foundation grants affect the careers of researchers and the science produced as a result. In addition to being an important source of funding, the particular form of governance of private funding organizations means that they may be relatively free to experiment with different approaches to grant making. This provides a unique opportunity for researchers and funders to evaluate foundation grant programs and learn what approaches have been most successful in furthering funders' objectives and contributing to the progress of science.

Evaluation is important because programs may have unintended effects that are unseen unless measures are defined to evaluate progress towards objectives. An example of this can be seen in the case of the Fields Medal, which is awarded by International Congress of Mathematicians (ICM) every four years to mathematicians under 40. The founder of the medal, John Charles Fields, wanted to encourage the recipients of the Medal to pursue further contributions in pure mathematics after age 40. Instead, data analyzed by Borjas and Doran (2015) shows that recipients of the medal become more likely to explore applied fields after winning the medal, relative to otherwise comparable non-recipients. This analysis thus suggests that the Fields medal has an effect on mathematicians' careers which runs counter to the original intentions of the funder.

How might Fields have avoided this outcome? The first step would have been to explicitly and publicly define an objective measure of success, e.g. continued contributions in pure math as opposed to applied fields. The second step would be continual evaluation of progress towards this goal; in the case of the Fields Medal, such evaluation would have alerted the ICM of a marked tendency towards divergence of the program's impact from its goals at an earlier stage. Finally, post-award management can often play an important role as well. The Fields Medal is a one-time award, but for other ongoing grants careful post-award management can bring straying scientists back into the fold.

Measuring the ultimate impacts of science funding is difficult. Scientific discoveries often take place many years after grants are made, draw on prior contributions from many different scientists, and can have direct as well as indirect effects on a wide range of economic activities, many of which are not foreseen by funders. Knowing what metrics to track in order to evaluate progress requires imagination and forward thinking, and it may not always be possible to predict the specific impacts of a grant ex ante. Moreover, to capture the full impact of a grant or program, one would ideally be able to see the entire scientific ecosystem of which the grant or program is part. This is because grants may substitute for each other.

For example, Jacob and Lefgren (2011) find only a small impact of the receipt of National Institutes of Health (NIH) funding on publications: recipients of grants publish 7% more articles over the five years following the grant than non-recipients receiving approximately the same score in the NIH grant review process. However, most scientists not funded by the NIH ultimately obtained funding elsewhere, so the effect of getting an NIH grant cannot be measured in isolation. This observation has several implications. One is that researchers may make more progress by focusing on variation in the *manner* in which grants are made, rather than trying to measure the causal impact of a particular agency like the NIH. For example, Azoulay et al. (2011) compares grants from the Howard Hughes Medical Institute (HHMI) to grants from NIH, and shows that the longer time horizons and greater independence of HHMI grants contributed to more risk-taking and higher-impact publications by scientists funded under that program.

Another lesson from this example is that research on science funding will be more productive if it can see the whole ecosystem of science funding. A grant from a specific funder may increase the productivity of the researcher who receives the grant relative to what would have happened if that researcher had not received the grant. But this does not mean that the overall amount of scientific output in the whole field increased because of this grant. For a variety of reasons, some other researcher may produce less precisely when the first researcher produces more. In that case the field overall may produce the same amount as it would have had the grant never existed. Moreover, the measurement of impact of a grant on a particular research area may overstate impact if government funding agencies reduce their funding of an area in response to the perception that the area is already well supported by foundations. For all of these reasons, in order to obtain an accurate measure of impact, one would need data on grants and output from all sources in a scientific domain.

In what follows, we discuss several design choices or "parameters" faced by funders of scientific research. In most cases, there is little evidence to date on how these choices affect outcomes, and in every case the right choice will depend on the objectives of the funder. More research is needed on how the design of funding programs affects the outcomes of the funding.

II. Internal, External, and Dynamic Consistency in Program Design

There is much more to setting up a funding program for early careers that simply setting a budget and a target number of young scientists to support. Rather, such programs need to be carefully designed with respect to each stage (solicitation and application process; evaluation; ongoing management; renewal or wind-down; post-award data collection). Moreover, design parameters should not be chosen independently of each other.[1] Rather, program designers should choose them as a bundle, so that different design elements for the program reinforce, rather than undermine each other. This is what we call *internal consistency*.

Funders should also consider whether the design of their program is *externally consistent*, i.e., whether its features position it to tackle challenges in the funding environment at large. For instance, how similar will the mode of selection, and targeted population be to that of other funders active in the area? If the goal is to target early scientists, will the potential recipient of funds be selected at the stage where scientists are particularly at risk of dropping out of science?

Another guiding principle for the design of funding programs is that of *dynamic consistency*: Will the program's effectiveness be robust to internally- and externally-driven changes? For example, will design features pertaining to solicitation, selection, and post-award management scale as the program expands its reach? What would be the impact of large swings in government funding for the program's likely effectiveness?

[1] For instance, prize-like programs are not aligned with the funders' objectives when they seek to use the funding as an inducement to change some entrenched behavior, such as lack of risk taking. An opportunity to compete for renewing the funding provide a mechanism to hold grantees accountable.

III. Design Parameters for Early Career Funding Programs

In what follows, we describe design parameters at each of three phases of the funding process: the solicitation and application phase, the application phase, and post-award management. We argue that there should also be a fourth phase of funding programs: the evaluation stage. Programs should be designed with evaluation in mind, and we conclude with recommendations related to designing for evaluation.

i. Solicitation and Application Phase

Program designers face a basic tradeoff between breadth and focus, with no single obviously correct choice. The NSF CAREER Awards can be awarded to young scientists in all areas of science in which NSF funding take place; the Packard and Sloan Foundations solicit applicants in a relatively broad cross-section of scientific fields. In contrast, HHMI, the Pew and Searle Scholarship funds, and the Damon Runyon Cancer Research Foundation only select applicants working in a narrow set of biological subfields.

Research by Aghion, Dewatripont and Stein (2008) highlights the role of information asymmetry between funders and scientists which suggests it may be optimal to allow scientists substantial autonomy in the choice of topic and how to structure research in early-stage basic science. However, when scientists' incentives are not perfectly aligned with funders' goals, funders may want to attract attention to specific fields. Recent research by Myers (2017), comparing grants awarded by the NIH in open contests with Requests for Applications (RFAs) finds that RFA applications are 18% less similar to prior work by the applicant than open-contest applications, suggesting that the RFA approach appears to be redirecting scientist effort into new fields.

A broad scope for the applicant population might make sense if the funder is uniquely focused on excellence and identifying the "best and brightest." The challenge of breadth is felt most acutely at the evaluation stage, since relatively few (or maybe even no) scientists have the ability to judge the scientific worth of research that falls outside of their areas of expertise. As a result, topically broad programs often either piggy-back on evaluation done for other processes (e.g., NSF), or do not have solicited application processes and instead identify potential awardees through nominations from already-established scientists (e.g., the MacArthur “Genius” awards).

More targeted programs are those which, after considering the extent to which they cohere in shaping a funding program that attracts the most desirable set of applicants, select from that set the “best” awardees (where what is “best” will of course depend on program goals), and then ensure that the agenda of funding recipients is enabled.

Another choice relates to whether to solicit applications from the EC scientist, or her mentor, of the mentor/mentee pair. Focusing on the EC scientist may improve career outcomes for that individual, due to increased autonomy and bargaining power. However, it may result in reduced investments from the mentor, who may have less incentive to contribute relative to if the mentor were directly responsible for the grant. Funding the mentor instead (for example with a training grant) holds mentors accountable for the quality of the training, but comes at the cost of a loss of autonomy for the early-career researcher. To date there is little evidence to date on the impact of funding “portability,” though Blume-Kohout and Adhikari (2017) find that PhDs who were funded primarily as research assistants are significantly more likely to take research-focused jobs in the U.S. scientific workforce after they graduate, as compared to PhDs who were primarily supported as trainees or fellows.

A final choice has to do with whether the funded area is already relatively well populated by prominent researchers, or is a white space toward which funders wish to direct scientific attention. Crowdedness may be an indicator of scientific potential, while white space may exist because the barriers to making discoveries are high. However, one risk of funding a crowded area is that grants may go to more marginal projects; another is that other funding agencies may reduce their commitment to this area in response to the influx of additional resources. Moreover, the existence of “white space” may reflect high fixed costs that a potential new funder would have to incur in order to support the heretofore neglected area. One bold funder willing to overcome these barriers might be all that is needed to enable a flowering of research that was previously not feasible. One example of the latter type of project is the

Sloan Digital Sky Survey, a three-dimensional map of the universe funded by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and by participating institutions. This project has resulted in the publication of thousands of scientific papers on a wide variety of astronomy topics.

ii. Selection phase

Once a pool of applicants has been collected, funders must evaluate applications. This raises questions about the best way to aggregate evaluator sentiment. A very common metric is an average, which is a good measure of the typical rating of evaluators, but does not incorporate the intensity of sentiment and how it varies across evaluators. It is possible that diversity of opinion might itself be a marker of creative potential, in which case funders should look closely at grants with a high variance in evaluator scores. NIH grant applicants often complain that one bad review is enough to torpedo a proposal – however, the most original projects may be more likely to garner negative reviews because they do not fit with established views of what is likely to lead to progress.

Another approach, similar to one used by the Gates Foundation, is to issue reviewers a limited supply of "gold stars," where in theory, all proposals with one gold star get funded. This forces reviewers to think carefully about how to allocate their stars across projects (Azoulay et al. 2017). One could also issue evaluators a limited number of "rotten tomatoes" which have the capacity to sink a proposal.

There are also questions about the degree of discretion to give program administrators to overrule evaluators' opinions. On the one hand, administrators are better informed because they may "see the whole deck" rather than just a handful of proposals. Some program administrators admit that their decisions are based largely on "gut," and that additional information about the potential of a scientist (beyond what is available in the CV and letters) can be obtained from face-to-face interviews.

The question of how much discretion to allow program officers or evaluators in choosing projects is an important one. In a study of the NIH R01 selection process, Li (2017) finds that grant evaluators are biased in favor of projects related to their own research, but their evaluations of the quality of related projects are more accurate. A recent evaluation of Department of Energy program ARPA-E by the National Academies of Sciences, Engineering, and Medicine suggests that the exercise of discretion may be working well to select the right projects. The review showed that program managers have access to peer-reviewed scores but in many cases overrule them to promote some projects and demote others. However, there is no apparent difference in outcomes between projects upgraded by managers and those awarded funds based purely on the score (Goldstein and Kearney 2017). This seems to suggest that it may be optimal to allow program managers some scope to use information beyond what is captured by the reviewers' scores to select projects.

In contrast, Hoffman, Kahn and Li (2017) show that the exercise of managerial discretion may reduce the quality of hiring relative to a score-based metric, and a literature in behavioral economics growing out of the work of Daniel Kahneman and Amos Tversky also shows that overruling objective measures with judgement is often a bad decision. To the extent that administrators are allowed to use discretion, funders need mechanisms to evaluate and hold administrators accountable for the exercise of discretion.

When numeric scores with cut-offs for funding are used, design parameters may depend on perceptions of how evaluators add value. The main role of evaluators may be in distinguishing meritorious applicants from those with little or no potential, or it may be in distinguishing between fine gradations among applicants with high potential. If the former, then it would make sense to use evaluators to determine the pool of finalists, and then decide on the winners by lottery. This approach has the additional benefit of introducing randomization, which makes robust evaluation possible. If the latter, then a single score in the context of a fixed award pool can provide a path to evaluation.

A number of other questions remain about how to conduct evaluations. Should review panels be composed exclusively of researchers in related fields, or should non-traditional evaluators be included? Should cutoffs for funding be rigidly determined by whether a numeric score is above some cutoff, or based on more fluid criteria? Should deliberation among evaluators be face-to-face or decentralized?

iii. Post-award management phase

Funders also need to decide to what extent they want ongoing involvement with funded researchers. The pure prize approach, which rewards investigators for past success, requires little to no post-award management. For example, the HHMI Early Career Scientist Program selects researchers who have led independent laboratories for two to six years and gives them six years of non-renewable funding for full salary and research support. The program is designed to encourage investigators to explore and take risks. In 2015, HHMI together with the Bill & Melinda Gates Foundation and the Simons Foundation introduced the Faculty Scholars Program which offers funding for five years. Like the Early Career Scientist Program, the Faculty Scholars Program encourages investigators to explore, take risks, and potentially even change direction in their research as long as it remains on a topic that fits the mandate of the program.

By contrast, an incentive program with renewals -- such as the program used by HHMI, with its longer-term funding (five years followed by renewal for another five years without necessarily requiring substantive results for renewal), and its focus on "people not projects" -- is costlier to manage, but provides a lever for the funder to influence the direction of scientific effort. This also can enable funders to deepen their financial commitment for the more successful grantees. However, this works best if the funder has a clear way to measure the types of practices or behaviors they want to influence. For example, one of the goals of HHMI is to encourage risk-taking, so the first renewal is based on whether the investigator took risks (Azoulay et al. 2011). Funders that can do this can design renewals to help shepherd the project along to make sure it is meeting the funder's objectives. This works less well if the funder's idea of "success" is an ultimate impact (e.g. the cure for cancer) rather than an intermediate outcome.

At the post-award stage, funders may also have varying degrees of involvement in scientific choices of the awardee. An "active management" approach is hands-on regarding the choice of collaborators, choice of topics, milestones, with appropriate ongoing monitoring and possible early termination of the project. This type of management may be less relevant for early-career funding (where mentors may play the role of the active managers). For example, in the ARPA-E program, the program officer has significant influence on project selection and how projects are conducted, including the assembling of the team and the definition of milestones. In contrast, as described above, HHMI allows investigators substantial discretion over how they conduct their research and even the topic of research.

iv. Designing for Evaluation

The first question to ask when evaluating a grants program is: what are the objectives of the program, and how can progress towards these objectives be measured? It can be useful to distinguish between outcomes (clearly defined and measurable products of an initiative, specified in advance) and impacts (broader, possibly longer-term results). For example, whereas the overall desired impact of a program may be to improve life expectancy for patients with a particular health condition, it may be easier to measure narrower and more intermediate outcomes such as the number of research articles published in a given field. When measurable outcomes are clearly defined from the outset, outcome evaluation lends itself well to rigorous quantitative analysis.

The approaches used by private funders to evaluate their grants vary. At the Sloan foundation, for every grant that is made, program officers identify the ultimate objective of the grant and three or four metrics that can be used to measure progress towards that goal. These metrics are then followed over time and the objective periodically revisited to evaluate how well the grant is performing.

IV. Requirements for Quantitative Analysis

Quantitative analysis requires data, and here funders have a particular advantage. The grant process itself generates data about applications, grants and outcomes that, if collected and managed appropriately, can provide insights to help improve the performance of the grants. Given this potential, grant applications should be designed with future evaluation in mind, so that the information supplied by applicants can be used for data analysis.[2]

Clearly, measuring progress towards an objective will be easier in some cases than in others, with tracking publication counts on the easy end of the spectrum, and measuring the impact on human welfare of basic science research at the other end. However, if funders can articulate goals and measure progress towards them using high-quality data and rigorous analysis, funders may gain insight into the process of grant-making that will help them better achieve their goals.

i. Intermediate Outcomes

Given that the time lag between initial funding and the outcomes of that funding can be extremely long, funders can benefit from intermediate indicators of progress. However, it is an open question which intermediate outcomes best predict ultimate success. In some cases, the scientists with the biggest eventual impacts took a very long time to realize their full potential, and would likely not have appeared promising to funders evaluating potential grants to EC scientists.[3] Funders seeking to encourage projects with high risk but potentially high return must be prepared for some funded projects to fail. However, how can funders tell the difference between grants that failed for the "right" reasons (they funded inherently risky but still promising science) and those that failed because the project was ill-conceived from the outset? Moreover, how can the scientific community share information about failures as well as successes in order to learn from failure and avoid duplication of effort?

---

[2] The increasing internationalization of science and mobility of scientists across borders means it will be necessary to track scientists even after they leave the country. Most available datasets are currently limited to scientists who remain in the home country.

[3] During his early career, Nobel-prize winning chemist Stefan Hell worked in relative isolation and obscurity. He has said that "If I had worked with a group of other people initially, I would not have started to do what I did. Actually, many people told me that it would not work. Sometimes, being naive and a bit ignorant about the difficulties of a subject helps to see it from a radically different angle." ("The Rotarian Conversation with Nobel laureate Stefan Hell," from the August 2016 issue of *The Rotarian* https://www.rotary.org/en/rotarian-conversation-nobel-laureate-stefan-hell)

V. **Conclusions**

Informed by the potential for unintended consequences and missed opportunities in the funding of science, as well as the critical importance of understanding the best way to fund early career scientists, we propose that an initiative be established to collect comprehensive data on the ecosystem of private scientific grants and scientific outcomes. Tracking research inputs and research outputs across subfields and over time will allow funders and foundation staff to tailor their strategies to their aims. Early career scientists will receive funding more closely aligned with their career goal of producing ground-breaking science. Reductions in wasteful spending will free up resources to pursue more and better science, leading to a more dynamic and successful scientific future.

Questions and Answers

What is the best way to fund early career scientists?

This paper has described some of the current empirical economics research on funding early career scientists. This research has shown that choices about how to organize funding programs can have unintended consequences. Furthermore, the research has left open a wide array of basic questions which can only be addressed with more comprehensive data.

Given the multidimensional and long-term nature of the benefits of science, how can collecting and analyzing data on the ecosystem of private grants and scientific outcomes capture the full impact of scientific funding?

It is impossible to predict in advance what the long-term impact of any scientific investment will be. However, there are many aspects of these impacts that are measurable and are leading indicators of richer long-term scientific outcomes. Nevertheless, even these potentially measurable impacts require a collaborative institutional commitment to collecting comprehensive data on scientific inputs and outputs.

Given the importance of collecting data on the ecosystem of private grants and scientific outcomes, why hasn't this data collection already occurred?

Numerous researchers in the science of science have attempted to collect data on such grants and such outcomes. But the lack of uniformity among different grant agencies and different databases of scientific outcomes makes this inherently difficult for a single research team to accomplish; coupled with privacy concerns, such data collection has been partial and temporally-limited. Granting agencies themselves have not pursued such data collection in part because they have not been fully aware of the dangers and missed opportunities arising from not doing so. One goal of this white paper is to raise awareness on this issue.

What is the best candidate for an institution that would collect data on the ecosystem of private grants and scientific outcomes while preserving confidentiality as necessary?

Because non-profit private grant agencies are not in competition with each other, the optimal institution to collect and manage this data need not be a third party; it could be a granting agency itself.

If analysis of the ecosystem of private funding of early career scientists reveals significant waste and low impacts, might funders be reluctant to continue funding early scientific careers?

Funders want to fund strategies that demonstrably achieve their goals. Analysis of the ecosystem of private funding of early career scientists will likely reveal unproductive funding strategies as well as useful and successful funding strategies. Given funders' desire to have an impact on science, they are more likely to shift resources towards successful strategies in response to this evidence than to give up entirely. Likewise, private foundations will have time to adjust their strategies in response to this evidence because they will be partnering to fund the institution which produces it. Finally, any analysis of the data would be cognizant of the limitations of the data; would take into account both qualitative and quantitative evidence; and would acknowledge the existence of both intermediate and long-run goals.

References

Aghion, Philippe, Mathias Dewatripont and Jeremy C. Stein "Academic Freedom, Private-Sector Focus, and the Process of Innovation," *The RAND Journal of Economics*, Vol. 39, No. 3 (Autumn, 2008), pp. 617-635

Azoulay, Pierre, Yuly Fuentes-Medel, Julian Kolev, and Fiona Murray (2017). "Evaluating Innovation: Identifying Promising Ideas at the Gates Foundation." Working paper, MIT.

Azoulay, Pierre, Joshua S. Graff Zivin, Gustavo Manso "Incentives and creativity: evidence from the academic life sciences", *The RAND Journal of Economics,* Volume 42, Issue 3, Fall 2011, Pages 527–554

Blume-Kohouta, Margaret E. and Dadhi Adhikari. "Training the Scientific Workforce: Does Funding Mechanism Matter?" *Research Policy* 45 (2016) pp. 1291-1303.

Borjas and Doran (2015). "Prizes and productivity how winning the Fields medal affects scientific output." Journal of Human Resources 50.3 pp. 728-758.

David M. Blau and Bruce A. Weinberg (2017). "Why the US science and engineering workforce is aging rapidly" *PNAS* 114 (15) 3879-3884

Goldstein, Anna P., and Michael Kearney. "Uncertainty and Individual Discretion in Allocating Research Funds." Working Paper, MIT Sloan.

Jacob, B. A., & Lefgren, L. (2011). The Impact of Research Grant Funding on Scientific Productivity. *Journal of Public Economics*, 95(9-10), 1168–1177. http://doi.org/10.1016/j.jpubeco.2011.05.005

Kahn, S. and D. Ginther (2017) "The impact of postdoctoral training on early careers in biomedicine" *Nature Biotechnology* 35, 90–94

Hoffman, Mitchell, Lisa B. Kahn, Danielle Li, "Discretion in Hiring," NBER Working Paper No. 21709