

Becoming more transparent: Collecting and presenting data on biomedical Ph.D. alumniChristopher L. Pickett^{a,b,1} and Shirley Tilghman^{a,b}

^aRescuing Biomedical Research
1200 New York Avenue, Room 767
Washington, DC 20005

^bPrinceton University
Lewis-Sigler Institute
Princeton, NJ 08544

¹Corresponding author:
Christopher L. Pickett
Email: clp3@princeton.edu

Abstract

For more than 20 years, panels of experts have recommended that universities collect and publish data on the career outcomes of Ph.D. students. However, little progress has been made. Over the past few years, a handful of universities, including those in the National Institutes of Health's Broadening Experiences in Scientific Training consortium, and organizations, including the Association of American Universities and the Association of American Medical Colleges, launched projects to collect and publish data on biomedical Ph.D. alumni. Here, we describe the outcome of a meeting, convened by Rescuing Biomedical Research, of universities and associations working to improve the transparency of career outcomes data. We were able to achieve consensus on a set of common methods for alumni data collection and a unified taxonomy to describe the career trajectories of biomedical Ph.D.s. These materials can be used by any institution, with little or no modification, to begin data collection efforts on their Ph.D. alumni. These efforts represent an important step forward in addressing a recommendation that has been made for decades that will improve the ability of trainees to better plan for their careers and for universities to better tailor their training programs.

Background

Graduate training in the biomedical sciences prepares young scientists for a variety of careers, and Ph.D. graduates have secured and excelled in jobs in academia, government, and the private sector for decades. While tracking a Ph.D.'s path from graduate school to a faculty position is relatively straightforward, following Ph.D.s once they leave the academy can be difficult due to a lack of obvious tracking mechanisms. Instituting such a tracking mechanism would be beneficial for undergraduate students considering enrolling in graduate school, current graduate students exploring their career options, and departments looking to better evaluate and tailor their training programs.

Recommendations for improved tracking of the career paths of biomedical Ph.D.s are not new. In 1998, a National Academies report recommended that all departments "receiving federal funding for research or training should provide to its prospective graduate students specific information regarding" the career outcomes of its Ph.D. alumni (National Research Council, 1998). In the early to mid-2010s, reports from a variety of studies reiterated the call to collect and report on the career trajectories of biomedical Ph.D. alumni (Pickett et al., 2015). In 2015, the National Institute of General Medical Sciences responded by issuing a strong but non-binding directive to all graduate programs that receive federal training funds to collect and provide this information.

To date over two dozen universities have published career outcomes data for their Ph.D. alumni on their websites, and have thus acquired substantial experience in the collection and presentation of these data using both surveys and social media "sleuthing." On September 19, 2017 the Chief Academic Officers of the Association of American Universities, an organization that includes the 62 most research intensive universities in the U.S and Canada, "called on all Ph.D. granting universities and their respective Ph.D. granting colleges, schools, and departments, to make a commitment to providing prospective and current students with easily accessible information. This should include data on matters such as student demographics, time to degree, financial support, and career paths and outcomes within and outside of academia. AAU institutions should commit to developing the infrastructure and institutional policies required to uniformly capture and make public such data" (Association of American Universities, 2017).

Interviews conducted by Rescuing Biomedical Research of departments and institutions that had not yet collected or published career outcomes data revealed common roadblocks faced by institutions: (1) they are unsure of what data should be collected and how it should be presented, (2) they are unsure of the cost in staff time and financial outlays needed to pursue such a project and (3) they are concerned that publication of career outcomes would adversely affect graduate student recruitment. On the other hand, interviews of institutions that had collected and published data on their biomedical Ph.D. alumni suggested that many of these roadblocks could be overcome with efficient data collection and presentation programs, and that their commitment to transparency actually enhanced graduate student recruiting rather than diminishing it.

On Aug. 7, 2017, Rescuing Biomedical Research sponsored a meeting at the AAAS in Washington, D.C., to compile information about successful data collection efforts with the goal of disseminating the findings to institutions that had not yet begun this process. Attending the meeting were representatives of the Association of American Universities, the Association of American Medical Colleges, the National Institutes of Health and a number of universities including those that were part of the NIH's Broadening Experiences in Scientific Training (BEST) consortium (see Appendix A for attendee list).

There were two primary outcomes from this meeting that would allow any university to launch its own data collection and presentation efforts. First, we developed a standard set of methods to collect data on Ph.D. alumni. Second, we developed a single, unified taxonomy to classify career outcomes. These tools will help you and your university launch a successful data collection effort on biomedical Ph.D. alumni.

Part 1: Strategy for career outcome data collection

Identify time frame of data collection

Determine the time frame for which to gather data on Ph.D. alumni, for example going back 15 years, and then identify all alumni who matriculated and graduated within this period. Several offices that may have such a list include:

- Office of the graduate school
- Registrar's office
- Alumni office
- Human resources
- Umbrella graduate programs
- Departmental lists of Ph.D. graduates

To ensure complete coverage, you may want to collect lists of alumni from multiple sources.

Data collection – university information

The next step in creating a database of Ph.D. alumni is to collect as much alumni information as possible from university sources. First, choose a spreadsheet or program that satisfies the needs of your data collection efforts and is, ideally, compatible with university sources of alumni information. When developing your database, consider how to best collect longitudinal data so that updating alumni information does not necessarily overwrite critical data.

It is unlikely university data sources will have complete information on career trajectories of Ph.D. alumni so it is important to collect as much information as possible to simplify and enable your efforts to reach out to alumni.

- Essential metrics for collection from university sources
 - Last and first names
 - Last known email address
 - Year of matriculation
 - Year of graduation
 - Degree conferred
 - Program/Department of thesis work
- Additional data that may be of use for contacting alumni
 - Alternate email addresses
 - Other last name
 - Gender
 - Race/ethnicity
 - Citizenship
 - PI
 - Last known job title
 - Last known employer
 - LinkedIn or other online profile
 - A notes field could provide a history of outreach attempts or explain an ambiguous outcome

- Date of data collection/update
- A “Collected by” field should be used if you anticipate that multiple people will collect data

Identify primary contact information

The next step is to find contact information, likely an email address, for each of the alumni. If current contact information is not included in the above data collection effort, you may be able to find contact information in a variety of places:

- Permanent email addresses may be catalogued in exit surveys.
- LinkedIn or other social media profiles may bear contact information; however, this information is sometimes restricted to those connected with the alum.
- Former PIs may have current contact information.

Data collection – alumni outreach

Alumni are the best sources of information on their career trajectory. Consider what other university offices are already conducting surveys or data collection on the population you are interested in. Adding your data collection needs into an existing platform may accelerate your work.

If you will be reaching out to alumni, consider using a mail-merge document that can take advantage of the database that you have already created. This will allow you to use a boilerplate inquiry letter so that personal email messages are streamlined and amenable to a mail merge. If possible, more recognizable or higher ranking university staff members are more likely to get a response from alumni. Consider having the most senior person reach out for information.

For those who don't respond to emails or for whom an email address was not found, internet sleuthing can be highly effective in tracking alumni. Data collection via internet sleuthing can be done by permanent, temporary or student workers who are committed to data quality and maximum data recovery. LinkedIn searches are likely to be the most fruitful. However, web searches are not guaranteed to turn up current, or even correct, information and data should be verified via other websites such as ResearchGate and other social media platforms. In addition, while this kind of data collection can be done by student workers, classification and binning of Ph.D. alumni career trajectories should be done by someone familiar with the landscape of biomedical career possibilities.

Funding and time required

Funding these efforts is highly variable depending on the quality of data and the number of people devoted to collecting and curating the data. This effort could be led by a single paid staff member or split among a team. Additional time needed for the project include website development and data presentation. However, the costs are almost all salaries, and expenses beyond salaries are minimal.

There are many different aspects to the data collection effort, but most estimates put the work time for identifying alumni, collecting contact information and collecting data on alumni at approximately 120 hours for a 15-year retrospective analysis of roughly 1,200 alumni.

Tracking current students

Beginning the tracking process when people are still enrolled is the best way to ensure maximal career trajectory data recovery. Consider encouraging all graduate students to provide a permanent email address at matriculation and graduation. Another way to keep track is to set up an alumni LinkedIn or Facebook page and ask all current students to become a member of that page.

Part 2: Career classification scheme and aggregate data

Nearly two dozen universities have already published data on their biomedical Ph.D. alumni, and each had to create their own taxonomies to classify the types of jobs held by their alumni. This limits comparison of career outcomes among programs and universities because the taxonomies were not always compatible.

The coalition present at the RBR meeting created a single, unified taxonomy of biomedical Ph.D. career trajectories by merging the taxonomies in use by UCSF and the BEST consortium. The resulting taxonomy is a 3-tier description of career paths. Sector—academia, government, industry, etc.—is the first level of classification and is the broadest categorization of career description. Career Type—primarily research, science related, further training, etc.—is the second level and represents a coarse description of job duties and activities. Job Functions is the third level of classification and is a specific descriptor of the job the alumnus holds.

With regard to the representation of the data, it was agreed that a basic representation—a spreadsheet with N values and percentages of alumni—would be ideal for ease of understanding the data and for comparisons across institutions. More sophisticated representations of the data, in graphical form for example, would be at the discretion of universities.

Widespread adoption of this taxonomy is critical to improve transparency around career outcomes for biomedical Ph.D.s. A set of definitions and example job types are included to help speed your classification of Ph.D. alumni.

In addition to the taxonomy, there is a growing interest in the presentation of demographic information and other aggregate data on graduate students and Ph.D. alumni. Information such as gender, citizenship, race/ethnicity, completion rate and time to degree should be collected and presented alongside career trajectory information.

Tier 1: Sector

	Definition	Coding clarifications
Academia	Any academic institution including K-12 institutions, colleges, universities, some medical centers, or free-standing research institutions where training occurs.	This does not include VA hospitals, but does include teaching, for-profit, and other types of hospitals.
Government	Any organization operated by federal, state, local or foreign governments.	Includes VA hospitals
For-Profit	Any organization that operates to make a profit, including some industry research.	
Nonprofit	Any non-governmental organization that does not operate to make a profit.	
Other	Individuals who are unemployed, full-time caretaker or parent, on extended medical leave or employed at an organization not included in other options.	
Unknown	Unknown	

Tier 2: Career type

	Definition	Coding clarifications
Primarily research	The primary, although not necessarily the only, focus is the conduct or oversight of scientific research.	Includes academic faculty titles at R1-R3 institutions as identified through Carnegie classifications.
Primarily teaching	The primary, although not necessarily the only, focus is education and teaching.	Includes academic faculty at all other institutions
Science-related	Career that is relevant to the conduct of scientific research, but does not directly conduct or oversee research activities	
Not related to science	Career that is not directly relevant to the conduct of scientific research	
Further training or education	Temporary training position	Postdoctoral research, completing medical residency, or pursuing an additional degree.
Unknown	Unknown	

Tier 3: Job function

	Definition	Coding clarifications
Administration	Administrative-intensive roles.	Faculty affairs, graduate program administrators, human resources, academic admissions, career development offices, grant and contracts management, research development, PhD-level program development
Business development, consulting, and strategic alliances	Role that involves the development, execution, management, or analysis of a business. Role may include relationship management, refinement of operational efficiency, or fee-based advisory services.	Management consultant, business development professional, market researcher, investment analyst, venture capitalist
Clinical research management	Role that is responsible for the oversight, management, or design of clinical research trials.	Clinical research project/trials manager or coordinator.
Clinical services	Role that involves that administration of clinical services or research	Genetics counselor, testing specialist, clinical laboratory staff
Data science, analytics, and software engineering	Role that may combine programming, analytics, advanced statistics, data communication, and/or software development.	
Entrepreneurship	Founder, co-founder, CEO or other role that develops, manages, and provides/obtains capital to initiate a	

	business or enterprise. This function does not include staff at a start-up business.	
Faculty: nontenure track	Leading an academic research team and ineligible for tenure.	Research assistant professor, research associate professor, research professor
Faculty: tenured/tenure track	Leading an academic research team and eligible for or already tenured.	Assistant professor, associate professor, professor
Faculty: tenure track unclear or not applicable	Leading an academic research team at an institution where tenure is not granted or tenure status is unknown.	For those tracking down alumni and binning them into job functions, whether someone is or is not on a tenure track is often not clear and should be sorted here.
Group leader (research)	Leading a research team in a nonacademic setting.	Anyone working in industry, non-profit or government who is running a somewhat independent research group. This includes those with "Faculty" titles at VA hospitals and other government research institutions.
Healthcare provider	Role where the primary responsibility is providing healthcare	Doctor, nurse, medical resident, veterinarian
Full-time teaching staff	Full-time institutionalized teaching position with no research responsibilities.	Instructor, Lecturer. Distinct from "Primarily teaching, faculty," these are people teaching at a single university without a faculty appointment.
Intellectual property and law	Role that involves the curation, management, implementation or protection of intelligence and creation, including trademarks, copyrights, patents, or trade secrets.	Patent agent, patent attorney, technology transfer specialist.
Part-time teaching staff	Contingent teaching role that is contracted on a single-semester, short-term, or non-permanent basis with no research responsibilities.	Instructor, Lecturer. Distinct from "Primarily teaching, faculty," these could include people teaching at multiple universities, indicating contingent status.
Postdoctoral	Temporary mentored training position following completion of doctoral degree.	
Regulatory affairs	Role that involves controlling or evaluating the safety and efficacy of products in areas including pharmaceuticals, medicines, and devices	Institutional regulatory affairs professional, quality control specialist, compliance officer
Research staff or technical director	Role that directly involves performing or managing research	Research staff, staff scientists, lab/core managers, directors of research facilities, public health analyst, and epidemiologists.
Sales and marketing	Non-technical role that is related to the sales or marketing of a science-related product or service	Medical science liaison, technical sales representative, marketing specialist
Science education and outreach	Role that involves K-12 teaching or public outreach at primary/secondary schools, science museum, scientific society, or similar	High school teacher, museum curriculum development, outreach program administrator

Science policy and government affairs	Role that involves policy or program development and review, including analysis, advisory, or advocacy	Program officer, public affairs or government affairs staff at scientific societies, foundations, government entities, or think tanks
Science writing and communication	Role that involves the communication of science-related topics	Science, medical, or technical writer, journalist, science editor, science publisher
Technical support and product development	Role that requires specialized technical knowledge of a science-related product	Technical support specialist, field application specialist, product development scientist or engineer
Other	Role that does not require scientific training or involve the direct implementation or communication of science	Full-time homemaker, care-taker, chef, food or hospitality services, some types of military service or mission work, or currently unemployed
Completing further education	Pursuing additional education that usually results in graduation with conferment of a degree or certificate; this does <i>not</i> include postdoctoral research	Pursuing an additional degree in medicine, law, business, or other area.
Deceased/retired	Deceased or retired	
Unknown	Unknown	

Aggregate data

	N	%
Total students matriculated		
Alumni with outcomes identified		
Gender	N	%
Male		
Female		
Other/not reported		
Citizenship	N	%
U.S. citizen or permanent resident		
Temporary visa holder		
Unknown citizenship		
Race/ethnicity	N	%
American Indian or Alaska Native		
Asian		
Black or African American		

Hispanic or Latino
White
More than one race
Other race

Completion rate	N	%
Finish without degree		
Completion: Masters		
Completion: PhD		

Time to completion	Time in program
Finish without degree	
Completion: Masters	
Completion: PhD	

Conclusion

In summary, we have created a concise set of methods and options that departments and universities can use to track the demographics and career outcomes of their biomedical Ph.D. alumni. Furthermore, we agreed on a singular taxonomy to describe biomedical Ph.D. career outcomes. The effort by a variety of universities and organizations described here represents an important step forward in fulfilling the 20-year old recommendation to improve data collection efforts on those graduating with a biomedical Ph.D.

Acknowledgements

We thank the attendees of the RBR meeting for their tireless work and efforts to reach consensus on these critical issues. We also thank Jeremy Berg, Mary Sue Coleman, Jessica Polka, Gary McDowell, Joan Reede and other members of Rescuing Biomedical Research for their comments on previous versions of this white paper. Rescuing Biomedical Research is supported, in part, by the Open Philanthropy Project and the Rita Allen Foundation.

References

- Association of American Universities. 2017. AAU Chief Academic Officers Call for Greater Doctoral Education Data Transparency. Available at [https://www.aau.edu/newsroom/press-releases/aau-chief-academic-officers-call-greater-doctoral-education-data_\(accessed 20 October 2017\)](https://www.aau.edu/newsroom/press-releases/aau-chief-academic-officers-call-greater-doctoral-education-data_(accessed 20 October 2017)).
- National Research Council-Committee on Dimensions, Causes, and Implications of Recent Trends in Careers of Life Scientists. 1998. *Trends in the Early Careers of Life Scientists* Washington, D.C.: National Academy Press.
- Pickett CL, Corb BW, Matthews CR, Sundquist WI, Berg JM. 2015. Toward a sustainable biomedical research enterprise: Finding consensus and implementing recommendations. *Proc Natl Acad Sci U.S.A.* 112(35):10832–10836. doi: 10.1073/pnas.1509901112.

Addendum

Since our meeting and subsequent report, several new developments indicate the effort to collect and publish data on Ph.D. career outcomes is broad and gaining momentum. In September, the provosts of Association of American Universities released a statement committing their institutions to improving transparency in the career outcomes of their Ph.D. alumni (1). This is a notable commitment as these institutions granted nearly 50 percent of all U.S. research doctorates in 2016.

In December, 10 institutions announced the formation of the Coalition for Next Generation Life Sciences (2). The coalition established a portal to visualize demographics and career outcomes data from each of the institutions. Their first release of data concerned graduate admission and enrollment, graduation rates, time-to-degree and demographics of their Ph.D. populations. Similar information about postdoctoral scholars and career outcomes of all trainees using the RBR taxonomy will be published in the coming months.

In February, the University of Toronto released its 10,000 Ph.D.s Project, which is an effort to collect and publish the career outcomes of the institution's Ph.D. alumni since 2000 (3). One wrinkle of the U of T study is that they collected all of their information through publicly available databases, and did not reach out to their alumni. Also in February, the Office of Fellows' Career Development at the National Institute for Environmental Health Sciences released a detailed examination of the career paths of NIEHS postdoc alumni (4). The NIEHS used a three-tier taxonomy, developed several years ago independently of the taxonomy described in this document, to describe career outcomes.

Finally in May, Wayne State University published a study of biomedical Ph.D. career outcomes over the past 15 years (5). The study uses the RBR taxonomy to examine career outcomes at intervals of 5, 10 and 15 years. WSU is a member of the NIH's BEST consortium and the consortium continues to implement and improve on the RBR taxonomy.

References

1. Association of American Universities. 2017. "AAU Chief Academic Officers Call for Greater Doctoral Education Data Transparency" Available at <https://www.aau.edu/newsroom/press-releases/aau-chief-academic-officers-call-greater-doctoral-education-data> (accessed 20 October 2017).
2. Blank R, et al. A new data effort to inform career choices in biomedicine. *Science* **358** (6369), 1388-1389 (2017).
3. School of Graduate Studies, University of Toronto. 2018. "10,000 PhDs Project" <http://www.sgs.utoronto.ca/about/Pages/10,000-PhDs-Project.aspx> (accessed 7 June 2018).
4. Xu H, et al. Visualizing detailed postdoctoral employment trends using a new career outcome taxonomy. *Nat Biotechnol* **36**, 197-202 (2018).
5. Mathur A, et al. Visualization of gender, race, citizenship and academic performance in association with career outcomes of 15-year biomedical doctoral alumni at a public

research university. *PLoS ONE* **13**(5): e0197473.
<https://doi.org/10.1371/journal.pone.0197473>

Appendix A: Meeting attendees

The following people attended an in-person meeting in Washington D.C. at the American Association for the Advancement of Science building on Aug. 7, 2017. Attendance should not be construed as an endorsement of this document or its recommendations.

Attendee	Affiliation	Working group
Patrick Brandt	University of North Carolina at Chapel Hill	Data collection
Deirdre Brekken	University of Texas Southwestern Medical Center	Data presentation
McKenzie Carlisle	University of Utah	Data collection
Roger Chalkley	Vanderbilt University	Data presentation
Mary Sue Coleman	Association of American Universities, RBR	University implementation
Peter Espenshade	Johns Hopkins University	--
Kenny Gibbs	National Institute for General Medical Sciences	--
Peter Hitchcock	University of Michigan	--
Ambika Mathur	Wayne State University	Data presentation
Gary McDowell	Future of Research, RBR	University implementation
Emily Miller	Association of American Universities	--
Jessica Polka	ASAPbio, RBR	Data collection
Phil Spector	Johns Hopkins University	--
Abby Stayart	University of Chicago	Data collection
Shirley Tilghman	Princeton University, RBR	University implementation
Renetta Tull	University of Maryland, Baltimore County	University implementation
Elizabeth Watkins	University of California, San Francisco	Data presentation
Jody Yellin	Association of American Medical Colleges	University implementation
Lillian Zwemer	Duke University	Data collection
Chris Pickett	RBR	--