The FASEB Journal • Essay

Is 6 Years Too Long to Get a Ph.D. in Biomedical Science?

Henry R. Bourne[1]
University of California, San Francisco

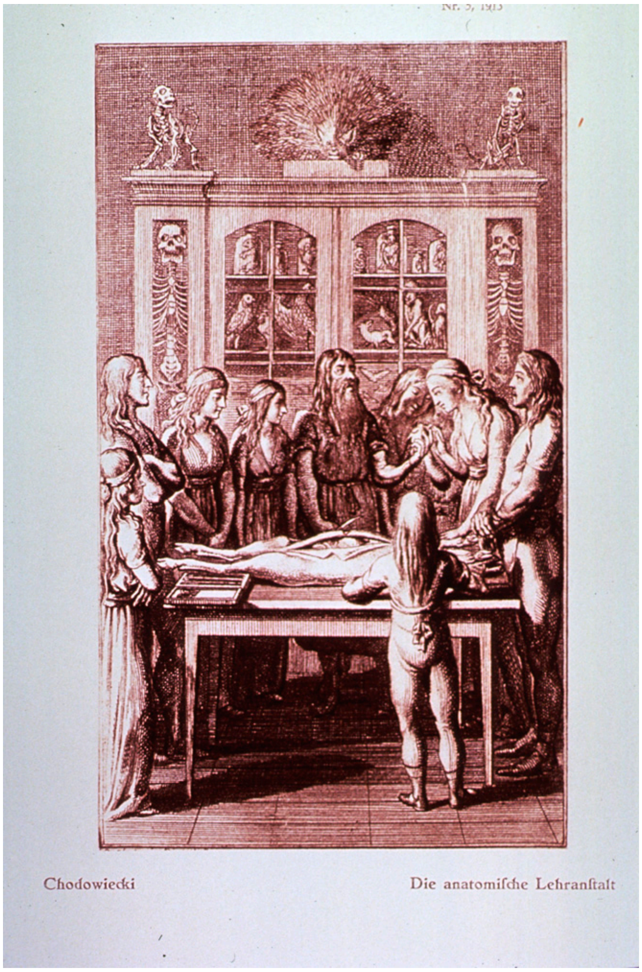
BACK IN THE 1960S, I USED my MD degree to sneak into experimental biology laboratories and spent 4 decades learning how to perform and think about experiments under the guidance of scientists with Ph.D. degrees, sometimes helping Ph.D. students to learn the same skills. Those experiments prompted me to ask 2 questions. First, must the average biomedical Ph.D. student take more than 6 years to get her degree? Second, what must a Ph.D. student learn? In this essay, I propose an experiment to answer the first question and attempt to answer the second.

A THOUGHT EXPERIMENT

To begin, let's try out this thought experiment: magically reduce the average time to degree (TTD) for U.S. biomedical science Ph.D. students from its present duration of 6.5 to 4.7 years (1) and specify that any new training regime does not decrease its graduates' ability to do research. If this magic works, it could produce far-reaching consequences:

1. Beginning scientists would assume permanent research positions 2 years earlier.
2. Almost $1 billion (B) of the U.S. National Institutes of Health (NIH) budget could be used directly to fund research rather than graduate student stipends and tuition (see below).
3. Composition and organization of U.S. academic laboratories would change dramatically, because 1) if yearly production of Ph.D. graduates remains constant, the shorter TTD will by itself reduce the number of NIH-funded Ph.D.-earning trainees in U.S. laboratories by ~ 31%; 2) the shorter TTD will require principal investigators (PIs) to focus their attention more on teaching prospective Ph.D. candidates how to conduct scientific research rather than produce blockbuster papers.

Of these results, the first would surely prove favorable by furnishing earlier independence to young scientists during their most creative years. In view of severe budgetary constraints on federal research funding, the second result could also be positive if the freed-up money is used to fund research grants awarded to individual investigators—but only with the proviso, perhaps naïve, that the NIH and Congress spend the $1B wisely. The third result would alarm some academic scientists and delight others; one group fears what might follow any change in the composition of academic laboratories, whereas the second group anticipates quite different effects and expects that these will be salutary.



Daniel Nicklaus Chodowiecki (1726–1801) "*Die anatomische Lehranstalt.*" (How Anatomy is Taught), ca. 1775. Image from the *History of Medicine* (U.S. National Library of Medicine).

This short TTD scenario may be more than a magical possibility. Over the past 15 years, the Watson School at Cold Spring Harbor (CSH) has awarded its Ph.D. degrees at an average TTD of 4.7 years and claims that its graduates get academic research jobs at rates comparable to graduates of leading Ph.D. programs with long TTDs (2, 3). Why has no other biomedical research institution tried to

---

[1] Correspondence: Department of Cellular and Molecular Pharmacology, School of Medicine, Genentech Hall, N212F, 600 16th St., University of California, San Francisco, CA 94158, USA. E-mail: henry.bourne@ucsf.edu

doi: 10.1096/fj.15-0201ufm

0892-6638/15/0029-0357 © FASEB

reduce its TTD to this same degree? Perhaps many PIs believe that "experience shows that first-rate Ph.D. [graduates] cannot be produced in a short time," as one of my colleagues avers. To the contrary, I believe the 6.5 year period is surely long enough and probably longer than necessary.

More critically, we do not yet know whether shorter TTDs can be achieved for Ph.D. graduates produced in good academic environments outside CSH. Here I propose an experiment to test whether the CSH result can be achieved elsewhere. If the answer is yes, the experiment will have lasted long enough to allow careful consideration of the potential consequences of translating it into practice. We shall return to this issue later.

THE TRAINING EXPERIMENT

Funded by the NIH, the experiment will be performed at 5 institutions in 3 of the following steps:

1. *Assess results and practice at the Watson School, CSH* (year 1). A small committee of respected biomedical scientists and educators first reviews the CSH experience by visiting CSH to interview the program's leader, faculty, students, and graduates. After judging the validity of the CSH results, it recommends or disapproves a controlled experiment to ask whether and how short TTDs can be achieved in other academic settings. A positive recommendation leads to step 2.
2. *Experimentally test whether an average TTD of 4.5 years can be achieved across a spectrum of academic research institutions and assess quality of graduates* (begin in year 2 and extend through year 11). Each of the 5 new subprograms in 5 research centers admits 8 Ph.D. students per year. Within 10 years, 5 classes of "experimental" students (~200 total or 8 Ph.D. students per experimental subprogram per year for the experiment's past 5 years) will graduate. Equal numbers of control students will be trained in established programs at the same center by PIs who may also train students in the experimental programs. To make the control groups valid and avoid increasing total number of graduate students at each location, the 8 students admitted into each experimental subprogram every year must not increase the total number of NIH-supported Ph.D. students admitted to the research center. Instead, both control and experimental students will be randomly chosen from pools of students admitted into existing graduate programs who agree (before matriculation) to be randomly assigned to control or short TTD arms of the experiment. (Experimental students may be matriculants to several separate graduate programs at the institution, as long as the number of informed-consenting students from each existing program allows assignment of 2 or more students to each arm of the trial.) Students in the experimental program are funded to the same extent from training and fellowship grants (TGs and FGs, respectively), university funds, etc., as are students in traditional programs (including the control students) at that center, with the exception that, for experimental students, the "extra" funds—ordinarily paid from research project grants (RPGs) in later years of Ph.D. training—must come from NIH as part of this experiment (see below), independent of grants awarded to the student's PI. (PIs will also be asked to provide their own informed consent to participate in this experiment. Only "consenting faculty" will be available to serve as PIs for students who give their own informed consent and are randomized into the experimental arm.)
3. *Research center applications to participate in the experiment.* Specific plans for educating experimental students include explicit goals of training (TTD and otherwise); admission criteria, including previous courses and research experience; curriculum requirements; protocols for supervising and monitoring student progress; and criteria for awarding the Ph.D. degree. Each center will closely monitor the admissions process (comparing short TTD matriculants *vs.* controls who also gave their consent to be randomized and are assigned to traditional programs) and record each student's progress and faculty deliberations/decisions with respect to each student. It is crucial that no experimental subprogram "copy" the CSH approach to achieve a short TTD; instead, each devises its own approach, which may incorporate some features of the CSH program, but will be based on principles and strategies analyzed and articulated by the experimental subprogram itself. Moreover, the TTD goal is an average, not a lock-step "maximum sentence" for every student; no student's training will be curtailed arbitrarily. Each subprogram will, however, focus its effort on an approach that is likely to be "scalable" to larger or more specialized graduate programs.
4. *Initial evaluation of results* (years 2–11). Each experimental subprogram will gather requisite information about its students and controls (students admitted in the same year and randomized, as described above). For both experimental and control students, data will include TTD; students' preadmission qualifications; student evaluations and outcomes of both sets of students during training (publications, evaluations, and examinations); positive or negative graduation decisions (by the faculty, student, or both); student trajectories after graduation, including processes and results of searching for postdoctoral training and later permanent positions; publications; career choices; the students' own evaluation of their Ph.D. training; and faculty evaluation of individual students and the program itself, *vis-à-vis* established programs. In addition, experimental programs will be evaluated with respect to feasibility and achievement of the TTD goal, relative success of trainees during and after graduation, and costs per graduate (*vs.* those for control graduates in established programs). It will also be especially important in evaluating "successful" subprograms to assess the program's apparent scalability (see above) and aspects of training that correlate with trainee success during and after their graduate school.

THE BOTTOM LINE(S)

The experiment will cost the NIH \$6 million (M) per year for 10 years based on the following assumptions: 1) one biomedical Ph.D. student for 1 year costs \$50,000, including stipend, fees, and tuition (depending on the latter, the total may be higher); 2) to administer the program and record and analyze results, each subprogram will need an administrative assistant, a monitor/analyst, and a faculty director with partial support (\$195,000, including salaries + benefits for all 3 positions each year); 3) each participating research center ordinarily supplies (from its coffers, TGs, and FGs) at least 2.0 years of support for every Ph.D. student; 4) "experimental" TTDs across subprograms will average ≤5.0 years (longer than the desired goal of 4–4.5 years); 5) each subprogram admits 8 students each year so that at steady-state the subprogram will comprise ~40 students at any one time, with ~8 students leaving each year, for an average TTD of 5 years; 6) the institution and TGs/FGs together support 2 of the average 5 years of training, leaving (at steady state, with 40 total students) 24 students to be supported each year by non-RPG funds to be supplied by the NIH (over and above the TG/FG funds NIH already supplies to train students at the host institution). Once steady state is reached, each subprogram will cost (per year) \$1,395,000 [= (24 students × \$50,000) + (\$195,000 for the administrative assistant, monitor/analyst, and director]. Five simultaneous subprograms in separate centers will cost \$6.975M per steady-state year (5 × \$1.395M). Over 10 years, the total will come to ~\$60M or ~\$6M per year. (This total is less than 10 × \$6.975M per year because NIH dollars will support fewer than 24 experimental students in each subprogram in each of the 4 years before steady state is reached and because all students are supported in their first 2 years from institutional and TG/FG funding.)

Will this yearly expenditure be manageable for NIH? In 2013, NIH awarded ~\$734M in graduate TGs/FGs (Ruth L. Kirchstein Awards; see ref. 4). The NIH does not know how many graduate students are presently supported on NIH RPGs (1), making it difficult to calculate the total RPG dollars it expends for their support each year. I estimate that this RPG support comes to ~\$2.45B based on estimates that 1) 68% of all biomedical graduate students receive RPG support (5); 2) the number of such students in the U.S. is ~80,000 (1); and 3) the cost of each student per year is \$45,000 (close to the actual value at my home institution). If so, NIH spends ~\$3.18B per year for training biomedical Ph.D. students (= \$2.45B from RPGs + \$0.73B in TGs/TFs)—that is, slightly more than 10% of its total annual \$30.15B budget. Thus, NIH already pays part of the money required for the proposed experiment in stipends and fees for trainees in established programs and laboratories; indeed, almost every entering student in the experimental program would otherwise be supported by the NIH for 1 to 5 years in an established program (via TG/FGs or RPGs) because of the stipulation that "experimental" students not increase the total number of NIH-supported Ph.D. students admitted to any research center. Overall, the \$6M that NIH would pay for the experiment each year amounts to ~0.82% of the ~\$734M that NIH awards yearly (4) in graduate TGs/FGs and only ~0.18% of the \$3.18B it pays for graduate education via TGs/FGs in combination with RPGs. Compared with the money NIH already spends on graduate training, the cost of this experiment is trivial.

Finally, consider the amount of money that the NIH would have to spare if the average TTD were reduced by 1.8 years from 6.5 to 4.7. Almost all of the reduction would come from RPGs, because Ph.D. students typically spend their last 2 years in the laboratory. If NIH pays \$2.45B to support 68% of all graduate students through RPGs and the average TTD is 6.5 years, its RPG funds are covering 0.68 × 6.5 = 4.4 years of graduate training per student. Reducing this RPG cost by 1.8 years would release ~\$1.00B [= (1.8/4.4) × \$2.45B] for RPGs to spend directly for research rather than training.

DECISIONS BASED ON RESULTS

As the experiment proceeds, evaluators will review each year's progress or lack thereof. Several of the following criteria will define optimal success: all (or most) subexperiments in separate centers produce Ph.D. graduates who meet the required TTD goal; experimental students equal or surpass controls in traditional programs (*e.g.*, research skills acquired, satisfaction with training, postdoctoral fellows in excellent laboratories, and publications); high-faculty satisfaction with program and students; and subsequent postgraduate careers (assessed over only 1–5 years, a very brief period).

If such success is achieved and if one or more successful approaches are deemed scalable, institutions will consider changing their own graduate programs and will presumably attract many prospective Ph.D. students. Depending on its judgments with respect to the ultimate consequences of shorter TTDs, NIH may gradually and deliberately apply short (average) TTDs as a strong criterion (or even a requirement) for positive review of new or renewal Ph.D. training program applications.

The results could well be mixed, with different subprograms producing results that range from good to low "success." Less likely but still possible, all subprograms may fail miserably. On the basis of the results, NIH will assess which program features correlate with success and decide to redesign or terminate the experiment. At the very least, institutions, prospective students, and the academic biomedical research community will know whether the CSH results can be replicated and can think more clearly about possible consequences.

By the time this experiment begins, NIH and institutions will have improved their ability to track all students' careers after graduation. If possible, the data will be useful to assess accomplishments and careers of experimental *vs.* control students for 15–20 years to determine what Ph.D. graduates lose or gain by either the old or the new approach. I predict that the differences will be small or undetectable. Such long-term evaluation (not included in cost estimates above) is distinct from the modest goals of the proposed experiment, which can at best determine whether short TTDs are feasible and whether the resulting Ph.D. graduates can become good postdoctoral fellows over the short term. If both answers are positive, then it is likely—albeit not proven—that such graduates will enjoy successful long term careers.

DOES TTD MATTER?

Does shortening the TTD for Ph.D. students justify performing a costly experiment? As suggested earlier, the experiment is worth doing because translating a successful result into practice could put the brightest young scientists into laboratory positions 2 years earlier and NIH RPGs could devote ~$1B paid from RPGs for graduate training to support research more directly (see above). The experiment also deserves hard thinking, because these and other consequences of translation into practice would change both composition of academic laboratories and conduct of research in many ways, including:

- *PIs' conflict of interest between training and research productivity.* Because students contribute more to the progress of research once they have learned how to conduct the research, PIs are motivated to keep them working until they can publish a highly significant paper before graduation. Perhaps for this reason, the national average TTD has changed little since the 1970s (1); in contrast, the CSH graduate program's short TTDs are achieved in a program entirely supported by TGs, FGs, and CSH itself, with no fiscal contribution from the PIs' grants (2, 3).
- *Inadequate peer review and spotty quality of training.* Inefficiency and unacceptably variable quality of Ph.D. training in the U.S. is well documented (1, 2). PIs' conflicts of interest probably play a role, but it is also indisputable that peer review of RPGs—the principal support of graduate students in later years of training—has focused exclusively on the quality of research and not at all on what trainees learn. In institutions with few or no TGs or FGs, the Ph.D. students will receive little or no peer review aimed at judging quality of training. The solution would be to subject all training in the U.S. to rigorous peer review focused on learning. This will be an easier task if short TTDs reduce the years of RPG support required for graduate training.
- *Staff scientists in the workforce.* Most academic laboratories hire students and postdoctoral fellows rather than staff scientists with Ph.D. degrees, partly from habit but mainly because students and postdoctoral fellows cost approximately half as much. Still, a good staff scientist probably can more than double the research accomplishments of the average graduate student. Staff scientists could improve the stability and efficiency of academic laboratories, but institutions and NIH will need to devise modest incentives to hire them (1, 2, 6, 7).
- *Sustainability of biomedical research.* RPG funding creates new Ph.D. graduates in more or less direct proportion to RPG support of research and makes it impossible to scale numbers of new Ph.D. graduates in proportion to demand. The resulting positive feedback loop has brought serious consequences (1, 2, 6–8), including an ever-expanding postdoctoral "holding tank," hyper-competition among grant applicants and institutions, and a marked graying of the academic research professoriate. At least in part, these problems can be alleviated by subjecting graduate training to more rigorous peer review and funding training primarily (or, if possible, exclusively) via TGs and FGs rather than RPGs.

At the outset, I posed 2 questions. The experiment proposed here can answer the first question by showing whether or not biomedical Ph.D. students need take more than 6 years, on average, to earn their degrees. No feasible experiment can answer the second question—what should Ph.D. students learn? Instead, I suggest a provisional answer. Ph.D. students should learn how to identify a significant question; design and execute experiments that can answer the question; analyze experimental results, using quantitative skills and statistics; and discriminate results that must be pursued now from those best pursued later. None of these abilities necessarily involves publishing a blockbuster paper—a paper that will serve a PI's career but also create an inevitable conflict of interest. If the proposed experiment does no more than persuade PIs and research universities to reckon with this conflict, NIH dollars will have been well spent. FJ

REFERENCES

1. NIH. (2012) Biomedical research workforce working group report. Retrieved November 12, 2014, from http://acd.od.nih.gov/biomedical_research_wgreport.pdf
2. Bourne, H.R. (2013). A fair deal for PhD students and postdocs. *eLife* **2**:e0119
3. Data from Watson School. Cold Spring Harbor. Retrieved November 12, 2014 from http://www.cshl.edu/images/stories/wsbs/docs/WSBSstats.pdf
4. Department of Health and Human Services. Justification of Estimates for Appropriations Committees, Fiscal year 2015, vol. 1, NIH Overview, page ES-30. Retrieved November 12, 2014, from http://officeofbudget.od.nih.gov/pdfs/FY15/FY2015_Overview.pdf
5. National Science Foundation. Survey of Graduate Students and Postdoctorates in Science & Engineering, Table 40. Retrieved November 12, 2014 from http://www.nsf.gov/statistics/nsf13331/content.cfm?pub_id=4290&id=2
6. Alberts, B., Kirschner, M. W., Tilghman, S., and Varmus, H. (2014) Rescuing US biomedical research from its systemic flaws. *Proc. Natl. Acad. Sci. USA* **111**, 5773–5777DOI: 10.1073/pnas.1404402111
7. Bourne, H. R. (2012) NIH support for PhD training (TRR-IV). *Biomedwatch* Retrieved November 12, 2014 from http://biomedwatch.wordpress.com/2012/08/13/nih-support-for-phd-training-trr-iv/
8. Bourne, H. R. (2013) The writing on the wall. *eLife* **2**, e00642

The opinions expressed in editorials, essays, letters to the editor, and other articles comprising the Up Front section are those of the authors and do not necessarily reflect the opinions of FASEB or its constituent societies. The FASEB Journal *welcomes all points of view and many voices. We look forward to hearing these in the form of op-ed pieces and/or letters from its readers addressed to journals@faseb.org.*