Mapping the U.S. Scientific Training Ecosystem: New Data, Methods, and Evidence*

Dror Shvadron[†] University of Toronto

Hansen Zhang[‡] Duke University

Lee Fleming[§] UC Berkeley

Daniel P. Gross[¶‖] Duke University and NBER

April 2025

Abstract:

How many PhDs does the U.S. produce in key fields, where are they trained, and who pays for their training? Many U.S. PhD graduates are foreign-born—how many remain in the U.S. after 5, 10, or 15 years? What happens when funding for PhD training declines? In this paper, we develop new data and methods for evaluating questions such as these about the U.S. scientific training ecosystem. We demonstrate that it is possible to do so using non-governmental and commercial data sources—especially information from PhD dissertations, later-in-life career outcomes, and natural language processing. Using these data, we explore how many PhDs are produced in specific critical technology areas, and by which universities; who funds doctoral education, and the effects of changes in public investment on PhD production; and how many graduates subsequently move abroad, where, and what the consequences may be for the U.S. innovation system. These methods complement existing data collection efforts by national statistics agencies, producing information which is otherwise costly to collect or not systematically observed.

JEL Classification: I23, I28, O31, O32, O33, O38

Keywords: Science policy, Scientific workforce, Doctoral training, Immigrant scientists

*We thank Bhaven Sampat for helpful conversations, as well as audiences at the 2024 ICSSI annual conference, 2024 Summer School on Data and Algorithms for Science, Technology & Innovation Studies conference, and 2025 NBER Investments in Early Career Scientists meeting for comments. We also thank Michelle Qiu and Max Murakami-Moses for valuable research assistance and the Duke University Fuqua School of Business and Alfred P. Sloan Foundation for financial support. Data described in this paper are being prepared for public release. This draft will be updated with a link to a data repository when available. Please contact the corresponding author if you are interested in these data and wish to be notified when they are posted. All errors are our own.

[†]University of Toronto, Rotman School of Management, Toronto, Canada; email: dror.shvadron@utoronto.edu.

[‡]Duke University, Fuqua School of Business, Durham, NC, USA; email: hansen.zhang@duke.edu.

[§]UC Berkeley, Haas School of Business, Berkeley, CA, USA; email: lfleming@berkeley.edu.

[¶]Duke University, Fuqua School of Business, Durham, NC, USA; email: daniel.gross@duke.edu.

[‖]Corresponding author