Using Linked Administrative Data to Follow Postdoctoral Researchers through COVID Disruptions [Outline]

Matthew VanEseltine[1]

[1] Institute for Research on Innovation and Science, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States of America

Email: matvan@umich.edu

Abstract

COVID-19 thoroughly disrupted the United States academic research enterprise, causing the closure and reduction of university operations, halts and adaptations to fieldwork, changes to meetings, interruptions in work time, and increased stress and distress. Postdoctoral scholars, already working in positions characterized by job insecurity and low support, faced especially difficult prospects when the pandemic began (Ahmed et al., 2020). Data gathering and integration over the past four years since the beginning of the pandemic allow us to begin to look more comprehensively at the institutional disruptions and career consequences for postdoctoral scholars at the time. Linking together UMETRICS administrative data from 14 U.S. research universities, Steppingblocks data on employment, OpenAlex publication data, and federal grant records, this paper investigates 11,366 postdoctoral researchers paid on projects funded by NIH and/or NSF, who each began their position between 2015 and 2021. This study follows these cohorts through 2022, observing career patterns of (1) departing postdoc employment, (2) entering faculty employment, (3) publishing papers, and (4) securing federal grants. Additionally, I investigate conditions of the research environment that may have mitigated or exacerbated the impacts of institutional COVID disruptions on postdoctoral researchers: the material environment of the research, characteristics of the research team, and the age and gender of the scholar.

1. Introduction

[This section will be prepared following analyses; see Timetable for Completion, below.]

2. Background

2a. Postdoctoral study and COVID-19

Concerns were soon raised that those with postdoctoral fellowships, typically underpaid and undersupported positions with annual contracts, were lacking the immediate resources and attention being more actively directed toward both students and faculty in the immediate rush to react to the many issues raised by COVID-19 at universities (Woolston, 2020). Disruptions to postdoctoral scholars threaten to derail career trajectories, potentially either cutting short or unnecessarily lengthening the precarious role, pushing postdocs out of academe or away from faculty positions, and generally interfering with scientific productivity.

2b. Contexts of COVID-19 disruptions

People working in science have been affected differently by COVID disruptions, with burdens falling most heavily on women (Caldarulo et al., 2022; NASEM, 2021; Viglione, 2020); marginalized and minoritized researchers (Douglas et al., 2022); parents of young children (Gao et al., 2021; Morgan et al., 2021). Indeed, some researchers reported increases in productivity (Barber et al., 2021) and many could report at least some positive consequence of working amid the pandemic (Johnson et al., 2021).

In addition to personal circumstances, the environment of scientific work may play an underexamined role in mitigating or exacerbating the impact of COVID-19. Researchers conducting their work as part of larger teams may benefit from greater institutional support and inertia than those working on smaller projects. The material environment of science (Latour & Woolgar, 1986) has been underappreciated in this context: very different kinds of physical and

personal actions and activities are involved under the broad umbrella of academic research. Scientific labor is *emplaced*: "it happens somewhere and involves material stuff" (Gieryn, 2000; Henke & Gieryn, 2008). The science conducted in finance research may be much more easily transferred to working from home than that of a research physiologist (Stenson et al., 2022). Some postdocs were therefore able to move to "desk-based activities" more easily than others (Beverstock & Pickersgill, 2022).

2c. Present study

This study represents an extensive look at a large population of postdoctoral scholars working at major American research universities before, during, and after the widespread impacts of COVID-19 beginning in March 2020. Drawing together linked data from a variety of sources, I compare cohorts of postdocs to assess the issues faced by postdocs during COVID-19 from several angles:

- Were postdoctoral positions either cut short or lengthened as a result of COVID-19?
- Were these postdocs prompted to more frequently leave academic work?
- Have these postdocs been less likely to transition into a professorship?
- Are these postdocs equally "productive," in the conventional senses of securing grant funding and publishing scientific papers?

Moreover, to draw in contexts:

- Do gender and age condition these patterns across postdoctoral scholars?
- Did the size of a postdoc's team environment influence these patterns?
- Did an element of the material environment of science—the rate of grant spending on purchasing supplies—mitigate or exacerbate the impacts of COVID?

4. Data

4a. UMETRICS

This study draws on university administrative records contributed and compiled by the Institute for Research on Innovation and Science (IRIS) as part of the upcoming 2024 Research Release of the UMETRICS (Universities measuring the EffecTs of Research on Innovation, Competitiveness, and Science) dataset.

4b. Steppingblocks

To augment these university records, IRIS has partnered with the data analytics provider Steppingblocks to link many of the university employees found in UMETRICS to more complete employment histories. Steppingblocks compiles employment information—dates, employers, and job titles—from a variety of sources, including "online profiles, resumes, job postings, government sources, such as the Bureau of Labor Statistics, public filings, company websites, public salary databases, and company firmographic databases" (Steppingblocks, 2024). Steppingblocks employment timelines are available for 5,580 of the 11,366 postdoctoral researchers in this sample, a match rate of 49.1%.

4c. OpenAlex/MAG

Following Microsoft's 2021 cessation of Microsoft Academic Graph, the OpenAlex project was launched as an open-source replacement (Priem et al., 2022). It currently includes nearly 250 million disseminated works (publications, books, theses, etc.) from 90 million authors. IRIS has developed a global disambiguation of authors in the final MAG dataset and is currently finalizing linkage of these authors to UMETRICS employees. This employee-author linkage allows the most complete available publication records to be attached to postdoctoral researchers in this study.

4d. NIH funding

Public records of projects funded by the National Institutes of Health, publicly available via NIH ExPORTER, include a persistent principal investigator (PI) identifier. Each UMETRICS research release includes name matching between UMETRICS employees and NIH PIs; this linkage for the 2024 Research Release is currently in progress and will be soon finalized by the IRIS research team. Analyses involving the measures derived from NIH data use only the subset of postdocs who received NIH funding (n = 8,758).

5. Methodology

5a. Sample construction

A subset of 14 UMETRICS universities have contributed consistent employee transaction data covering 2015 through 2022 in their entirety. Within these universities, 11,366 postdoctoral scholars paid on NIH or NSF projects began their postdoc employment between 2016 and 2021. These years and universities were chosen to capture multiple uninterrupted years of data with a range of experiences: the most recent two cohorts include those affected by COVID-19 every year, while most of those in the earliest cohort (2016) concluded their position prior to the pandemic. Following each cohort through 2022 yields 51,604 person-years of data.

5b. Outcome measurement

Each outcome is measured from annual data. The UMETRICS-derived conclusion of the postdoctoral position and leaving the postdoctoral university represent one-time status changes; other measures are annual binary or count values depending on the current status of the person.

Measure	Primary source	Description
Concluded postdoc	UMETRICS	Ended their postdoc position (no longer paid as a postdoc in university sponsored project records). One-time status change.

Left postdoc university	UMETRICS	Left the university of their postdoc (no longer listed as paid on any sponsored project). One-time status change.
Outside academe	Steppingblocks	Employers in longitudinal employment data do not include higher education employers. Annual binary.
Holding professorship	Steppingblocks	Job titles in longitudinal employment data include a professor position (excludes adjunct). Annual binary.
NIH awards	NIH ExPORTER	Number of funded NIH projects as PI (using subsample of NIH-funded postdocs only). Annual count. *[This outcome is not yet available; see Timetable for Completion below.]*
Publications	OpenAlex	Number of publications as author. Annual count. *[This outcome is not yet available; see Timetable for Completion below.]*

5c. Context measurement

This study highlights three factors that may mitigate or exacerbate the institutional disruptions of COVID-19 on the postdoctoral researcher. Each of these are measured during the first year of the postdoc's position in the UMETRICS data.

Measure	**Description**
Age	Birth month and year information is part of UMETRICS data collection; age is measured in whole years in March 2020.
Gender	Gender is imputed from name using Genni (Torvik & Agarwal, 2016) *[A pending improved measure will use an LLM method; see Timetable for Completion below.]*
Team size	Total count of employees paid on research awards that paid the postdoc during the first year. Two dichotomous measures created as small team size (lowest quintile) and large team size (highest quintile).
Reliance on supplies	Scientific environments requiring personal presence and interaction proxied by proportion of research vendor spending going to supplies (rather than travel, services, equipment, etc.) among projects also paying the postdoc during the first year. Two dichotomous measures created as low supplies spending (lowest quintile) and high supplies spending

	(highest quintile). Due to incomplete coding of spending categories, this measure is only available for a subset of cases (Zhang et al., 2022).

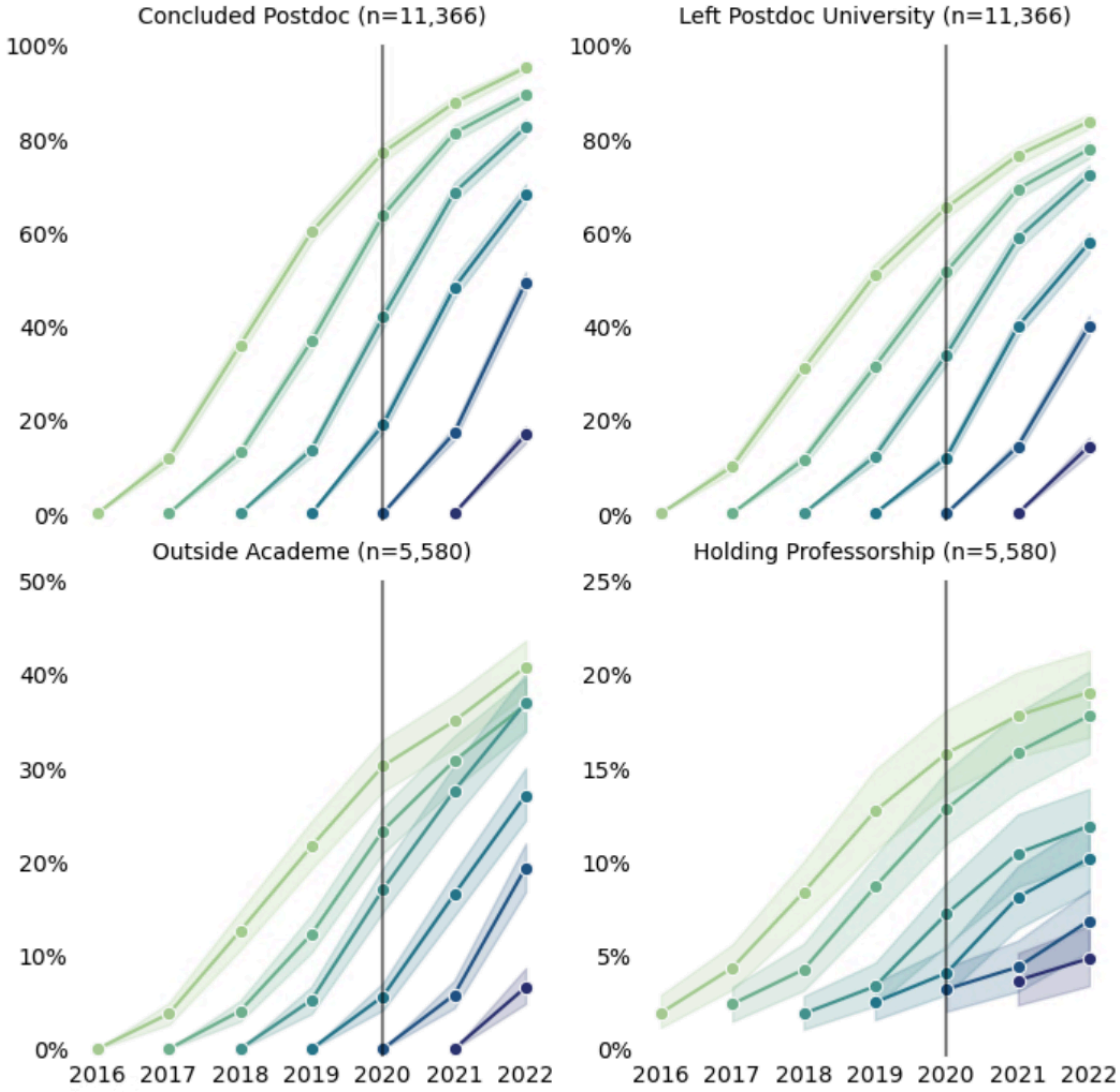
5d. Descriptive statistics

Variable	Mean	S.D.	Min	Max	N
Person-level					
Cohort			2016	2021	11,366
Gender					
Female	26.9%		0	1	11,366
Male	34.9%		0	1	11,366
Unknown	38.2%		0	1	11,366
Age in March 2020	33.2	5.0	19	73	10,157
Team size					
Count	7.3	15.3	1	450	11,366
Low (4 or fewer)	25.1%		0	1	11,366
High (17 or more)	20.3%		0	1	11,366
Reliance on supplies					
Proportion of vendor spending	0.74	0.3	0.0	1.0	5,855
Low (36% or less)	20.0%		0	1	5,855
High (100%)	19.2%		0	1	5,855
NIH-funded	77.1%		0	1	11,366
NSF-funded	27.2%		0	1	11,366
Time-varying					
Year	2019.9	1.7	2016	2022	51,604
Concluded postdoc	40.0%		0	1	51,604
Left postdoc university	34.0%		0	1	51,604
Outside academe	16.0%		0	1	26,354
Holding professorship	8.5%		0	1	26,354

For each cohort, for each year 2016–2022, outcomes are presented in Figure 1 *[for information on other outcomes see Timetable for Completion section below]*. These initial

descriptive figures suggest that postdoc timelines were, in the aggregate, largely unaffected by COVID-19's arrival in 2020. The top half of the figures shows that (left) postdoctoral scholars completed their positions and (right) postdoctoral scholars left the universities of their focal position with approximately the same timings regardless of cohort. The lower two figures, however, suggest an aggregate difference in careers following the postdoc, with both higher rates of leaving academe and lower rates of professorships, particularly among the cohorts of 2018 and later—most of whom were still postdocs in the spring of 2020.

Figure 1. Annual outcome prevalence by cohort, 2016–2022.



5e. Modeling approach

I fit a series of time-varying regression models for each outcome to estimate the differences between cohorts in their career progression (with Stata 17, *xtlogit* for dichotomous and *xtpoisson* for count outcomes). Time is marked annually by cohort, such that each cohort's t_0 is the first year observed with the postdoctoral position, as high as t_6 for the 2016 cohort in 2022. All models include university-level fixed effects and are estimated using robust standard errors. The marginal effects of differences in outcomes between pre- and post-pandemic years will provide aggregate estimates of the degrees to which COVID-19 disrupted specific outcomes. Interaction terms with contexts measures will demonstrate whether disruptions substantially varied across personal and scientific circumstances, or whether—for these measures, and within these populations—the disruptions were more evenly distributed.

6. Results

[The Results, Discussion, and Conclusion sections will be prepared following complete data integration; see Timetable for Completion, below.]

7. Discussion

8. Conclusion

Conflict of Interest Statement

The author is employed by the Institute for Research on Innovation and Science, which produces the UMETRICS dataset and linkages used in this study.

Timetable for Completion

The data tranche for the UMETRICS 2024 Research Release has been selected, and work at IRIS currently proceeds toward an anticipated release to the IRIS research community in the spring. Certain data elements for this study are not complete, but they are being finalized and will be available for inclusion in this study. This estimated timeline is for my own progress on this paper, not the IRIS team's work.

Jan 26	Annual publications outcomes will be added via new OpenAlex/MAG linkage, currently being finalized by Dr. Jinseok Kim as part of an NSF contract at IRIS.
Feb 2	Annual NIH PI outcomes will be added via updated UMETRICS linkage, and expanded and improved gender imputation for 2024 Research Release will be included. Both elements are currently being completed by the IRIS Research Team, led by Dr. Natsuko Nicholls.
Mar 1	Complete draft prepared for private feedback.
Mar 29	Paper revised, prepared, and available for dissemination in advance of conference.

References

ACD Working Group on Re-envisioning NIH-Supported Postdoctoral Training. (2023). *Draft Report to the NIH Advisory Committee to the Director*. https://acd.od.nih.gov/working-groups/postdocs.html

Ahmed, M. A., Behbahani, A. H., Brückner, A., Charpentier, C. J., Morais, L. H., Mallory, S., & Pool, A.-H. (2020). The precarious position of postdocs during COVID-19. *Science*, *368*(6494), 957–958. https://doi.org/10.1126/science.abc5143

Barber, B. M., Jiang, W., Morse, A., Puri, M., Tookes, H., & Werner, I. M. (2021). What Explains Differences in Finance Research Productivity during the Pandemic? *The Journal of Finance*, *76*(4), 1655–1697. https://doi.org/10.1111/jofi.13028

Beverstock, J., & Pickersgill, M. (2022). Producing knowledge in a pandemic: Accounts from UK-based postdoctoral biomedical scientists of undertaking research during the COVID-19 pandemic. *Humanities and Social Sciences Communications*, *9*(1), Article 1. https://doi.org/10.1057/s41599-022-01160-1

Caldarulo, M., Olsen, J., Frandell, A., Islam, S., Johnson, T. P., Feeney, M. K., Michalegko, L., & Welch, E. W. (2022). COVID-19 and gender inequity in science: Consistent harm over time. *PLOS ONE*, *17*(7), e0271089. https://doi.org/10.1371/journal.pone.0271089

Douglas, H. M., Settles, I. H., Cech, E. A., Montgomery, G. M., Nadolsky, L. R., Hawkins, A. K., Ma, G., Davis, T. M., Elliott, K. C., & Cheruvelil, K. S. (2022). Disproportionate impacts of COVID-19 on marginalized and minoritized early-career academic scientists. *PLOS ONE*, *17*(9), e0274278. https://doi.org/10.1371/journal.pone.0274278

Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R., & Wang, D. (2021). Potentially long-lasting effects of the pandemic on scientists. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-26428-z

Gieryn, T. F. (2000). A Space for Place in Sociology. *Annual Review of Sociology*, *26*, 463–496.

Henke, C. R., & Gieryn, T. F. (2008). Sites of scientific practice: The enduring importance of place. In E. J. Hackett & Society for Social Studies of Science (Eds.), *The handbook of science and technology studies* (3rd ed). MIT Press.

Johnson, T. P., Feeney, M. K., Jung, H., Frandell, A., Caldarulo, M., Michalegko, L., Islam, S., & Welch, E. W. (2021). COVID-19 and the academy: Opinions and experiences of university-based scientists in the U.S. *Humanities and Social Sciences Communications*, *8*(1), 1–7. https://doi.org/10.1057/s41599-021-00823-9

Jones, K. (2023). Precarity of post doctorate career breaks: Does gender matter? *Studies in Higher Education*, *48*(10), 1576–1594. https://doi.org/10.1080/03075079.2023.2245849

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Lokhtina, I. A., Castelló, M., Lambrechts, A. A., Löfström, E., McGinn, M. K., Skakni, I., & van der Weijden, I. (2022). The impact of the COVID-19 pandemic on early career researcher activity, development, career, and well-being: The state of the art. *Studies in Graduate and Postdoctoral Education*, *13*(3), 245–265. https://doi.org/10.1108/SGPE-10-2021-0076

Kabanova, A., Gavriilaki, E., Pelzer, B. W., Brunetti, L., & Maiques-Diaz, A. (2020). Effect of the COVID-19 Pandemic on Laboratory and Clinical Research: A Testimony and a Call

to Action From Researchers. *HemaSphere*, *4*(6), e499. https://doi.org/10.1097/HS9.0000000000000499

Morgan, A. C., Way, S. F., Hoefer, M. J. D., Larremore, D. B., Galesic, M., & Clauset, A. (2021). The unequal impact of parenthood in academia. *Science Advances*, *7*(9), eabd1996. https://doi.org/10.1126/sciadv.abd1996

National Academies of Sciences, Engineering, and Medicine. (2021). *The impact of COVID-19 on the careers of women in academic sciences, engineering, and medicine*. https://doi.org/10.17226/26061

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (arXiv:2205.01833). arXiv. https://doi.org/10.48550/arXiv.2205.01833

Torvik, V. I., & Agarwal, S. (2016). *Ethnea—An instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database*. University of Illinois at Urbana-Champaign. https://hdl.handle.net/2142/88927

Stenson, M. C., Fleming, J. K., Johnson, S. L., Caputo, J. L., Spillios, K. E., & Mel, A. E. (2022). Impact of COVID-19 on access to laboratories and human participants: Exercise science faculty perspectives. *Advances in Physiology Education*, *46*(2), 211–218. https://doi.org/10.1152/advan.00146.2021

Sims, D., Nicholas, D., Tenopir, C., Allard, S., & Watkinson, A. (2023). Pandemic Impact on Early Career Researchers in the United States. *SAGE Open*, *13*(3), 21582440231194394. https://doi.org/10.1177/21582440231194394

Steppingblocks. (2024). *Our data*. https://www.steppingblocks.com/our-data. Accessed January 2, 2024.

Viglione, G. (2020). Are women publishing less during the pandemic? Here's what the data say. *Nature*, *581*(7809), 365–366. https://doi.org/10.1038/d41586-020-01294-9

Woolston, C. (2020). Pandemic darkens postdocs' work and career hopes. *Nature*, *585*(7824), 309–312. https://doi.org/10.1038/d41586-020-02548-2

Zhang, Z., Nicholls, N., VanEseltine, M., Brown, C., & Owen-Smith, J. (2022). *Object Code Classification for the IRIS UMETRICS Fall 2022 Supplementary Data Release*. https://doi.org/10.21987/V9ZP-C742