

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No software was used for data collection

Data analysis Analyses were conducted using R Studio version 2021.09.1. For propensity score computation and assessment of covariate balance, we used the following packages: weightit (v 0.14.2) and cobalt (v 4.5.1). Documentation for these packages can be found at: <https://cran.r-project.org/web/packages/cobalt/index.html> and <https://cran.r-project.org/web/packages/WeightIt/index.html>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Public-use data (Survey of Doctorate Recipients 2019) are available at <https://ncesdata.nsf.gov/datadownload/>
A separate data availability statement section was added to the manuscript

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research.](#)

Reporting on sex and gender	N/A (this is not a study using identifiable data on human research participants)
Population characteristics	N/A (this is not a study using identifiable data on human research participants)
Recruitment	N/A (this is not a study using identifiable data on human research participants)
Ethics oversight	N/A (this is not a study using identifiable data on human research participants)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative analyses using cross-sectional data from the 2019 Survey of Doctorate Recipients.
Research sample	This study involved an existing dataset (the source is described in the Data section). This survey is conducted by the National Science Foundation every two years, and datasets are uploaded to the institution's publicly available data repository. The SDR 2019 dataset contains 80,882 observations, and the research sample was comprised by individuals that obtained a U.S. research doctorate degree between 1973 and 2017 in a science, engineering, or health field, and that were not 76 years of age or older, institutionalized, or terminally ill as of February 1st, 2019.
Sampling strategy	The survey relies on a fixed panel design. Subjects who graduated between 1973 and 2017 were selected from the Doctorate Records File (DRF), a yearly census of all doctorate recipients in the U.S, regardless of whether they had dropped out of the labor market or not. For the 2019 cycle, all 2015 and 2017 sampled members who remained age eligible were retained, and sampled members who did not respond in both the 2015 and 2017 surveys were dropped from the 2019 sample. A total of 10,000 new graduates who earned their degrees from 1 July 2015 to 30 June 2017 were added to the study sample, for a total of 120,000 participants, and 80,882 respondents (response rate: 69%), resulting in a weighted total of 1,148,817 doctorate recipients. The SDR utilizes survey weights that account for differential sampling rates, adjustments for nonresponse, unknown eligibility, and for aligning the sample with the DRF distribution on gender, race/ethnicity, location, degree field, and year. This revised sample stratification design strengthens reporting for small fields/cell sizes for fields of study, gender, and underrepresented minority indicator. Note to Editors: no specific sampling procedures were used by the NSF, as they attempted to include all eligible participants from the Doctorate Records File (DRF).
Data collection	The SDR used a trimodal data collection approach: self-administered online survey (Web), self-administered paper questionnaire (via mail), and computer-assisted telephone interview (CATI). Individuals in the sample generally were started in the Web mode depending on available contact information. After an initial survey invitation via postal mail and e-mail, the data collection protocol included sequential contacts by postal mail, telephone, and e-mail that ran throughout the data collection period. At any time during data collection, sample members could choose to complete the survey using any of the three modes. Nonrespondents to the initial survey invitation received follow-up with alternate survey modes. Note to Editors: the survey was completely remotely, and no blinding was used.
Timing	Data collection lasted six months (approximately from October 2019 to March 2020). The online survey closed 3 April 2020 and reception of hard-copy questionnaires ended on 20 April 2020.
Data exclusions	Our primary analyses were restricted to SDR participants from 2019 currently employed in a STEM field, living in the U.S., and working for an employer based in the U.S. at the time of the survey (N = 704,013). Therefore, 68.8% of participants were excluded from the primary analysis. Our rationale is that since we are analyzing differences in salaries, data on unemployed doctorate recipients or those living outside the U.S. could not be used for the analysis.

Non-participation	For the Survey of Doctorate Recipients 2019, the National Science Foundation reported a response rate of 69%.
Randomization	No randomization was used

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging