

SIDE BAR**Publication Output by Underrepresented Groups and Impact on R&D Careers**

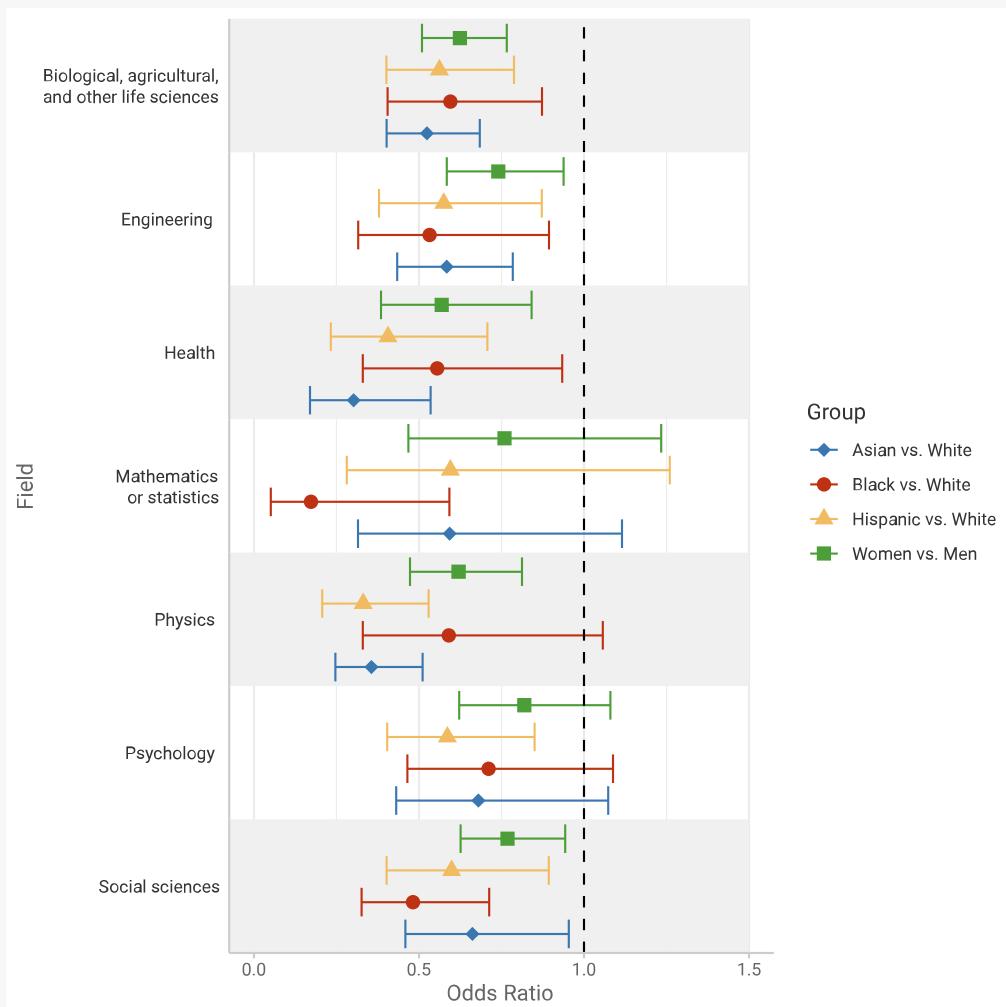
The National Science Board stated in its Vision 2030 report that “women and underrepresented minorities remain inadequately represented in S&E relative to their proportions in the U.S. population” (NSB 2020). These disparities have also been found in the publication of peer-reviewed articles (Hopkins et al. 2013). The National Center for Science and Engineering Statistics (NCSES) has undertaken research to examine linkages between publication output and careers in research (Chang, White, and Sugimoto forthcoming).

Matching publication output data to demographic survey data provides a key to understanding publication output in conjunction with authors’ demographic, training, and career information. Prior researchers have attempted to add author demographics using various methods, such as sex and race disambiguation algorithms (e.g., NamSor, Ginni, Ethnicolr, OriginsInfo), that estimate the probability of race or sex from given names (or, in the case of Face⁺⁺, from images). The accuracy of these matches varies dramatically by country and field; sex disambiguation algorithms perform better for western countries and poorly for other countries, specifically in Asia and South America (Karimi et al. 2016). In addition, some scientific fields, such as astronomy and astrophysics, generally use initials rather than given names. Despite these limitations, researchers have observed sex and race disparities in publication output (Hopkins et al. 2013; Larivière et al. 2013; Marschke et al. 2018; and **NSB Indicators 2018: S&E Publication Patterns, by Gender**).

The limitations associated with the earlier approaches can be overcome using data directly collected from the authors. One such source is the NCSES Survey of Doctorate Recipients (SDR),^{*} which provides demographic, education, and career history information from a sample of individuals with a U.S. research doctoral degree in a science, engineering, or health field (NCSES 2021). Clarivate, the architect of Web of Science (WoS),[†] matched SDR respondents to publication records in the WoS publication output database. The results provide demographic information, such as sex and race or ethnicity of publication authors.

These data shed light on publication output differences between groups defined by race or ethnicity and sex, by discipline, and by impacts to R&D career paths (Chang, White, and Sugimoto forthcoming).[‡] The point estimates in Figure PBS-A show the odds of pre-doctorate student publishing by ethnic group or sex relative to White students (or men, for the sex comparison) while the error bars show the confidence around that point estimate (95% confidence interval). The confidence interval is closely linked to the size of the sample. In the SDR-WoS data, the number of minorities and women receiving degrees in the population influences the sample size—and, consequently, the ability to measure odds ratios. For example, there are 3,750 women who received mathematics or statistics PhDs compared to 10,450 men (Table SPBS-32). A similar issue arises for mathematics or statistics PhDs by race or ethnicity (Table SPBS-33). Overall, compared to White graduates, Asian, Black, or Hispanic graduates are less likely to publish before their doctorate in biological, agricultural, and other life sciences; engineering; health sciences; and social sciences.

Figure PBS-A

S&E pre-doctorate publishing odds ratio, by sex and selected race or ethnicity: 1995–2006**Note(s):**

S&E doctorates include science, engineering, and health PhD candidates at U.S. research doctorate institutions. Computer sciences is not included in the figure because the odds ratio and confidence interval show no conclusive results for any demographic group or sex. Table shows the estimated odds ratios of publishing at least one article or conference proceeding during the five years before receiving a doctorate in the combined Web of Science and Survey of Doctorate Recipients database. For more detail, see Table SPBS-32 and Table SPBS-33.

Source(s):

National Center for Science and Engineering Statistics, Survey of Doctorate Recipients; Clarivate, Web of Science.

Science and Engineering Indicators

Compared to men, women are less likely to publish before graduation in the biological sciences, agriculture, engineering, health sciences, physics, and social sciences. Pre-doctorate publications appear to factor into obtaining a job in which research is the primary activity.[§] For those with at least one pre-PhD publication, 56% reported that their first job has research as its primary activity compared to 37% of those without a publication (Chang, White, and Sugimoto forthcoming).

* A machine learning approach matches the SDR respondents to the authors of publications indexed by the Web of Science (WoS). The matching algorithm incorporates name commonality, research field, education, employment affiliations, coauthorship network, and self-citations to predict matches from the SDR respondents to the WoS.

[†] WoS is a bibliometric database of conference proceedings and peer-reviewed literature with English-language titles and abstracts.

[‡] To predict pre-doctorate publishing propensity, separate models were fitted for each doctoral field, and the following factors from the NCSES's Survey of Earned Doctorates were controlled: doctorate award year, type of PhD-awarding institution, source of primary support, community college experience, U.S. citizenship status at the time of degree award, level of parental education, marital status, dependents under 18 years old, disability, graduate debt, and name commonality.

[§] The model controls for critical factors, such as the PhD institutions ranking as a high research institution, year of graduation, citizenship, parental degree, and student debt. The model does not measure article submissions or rejections.