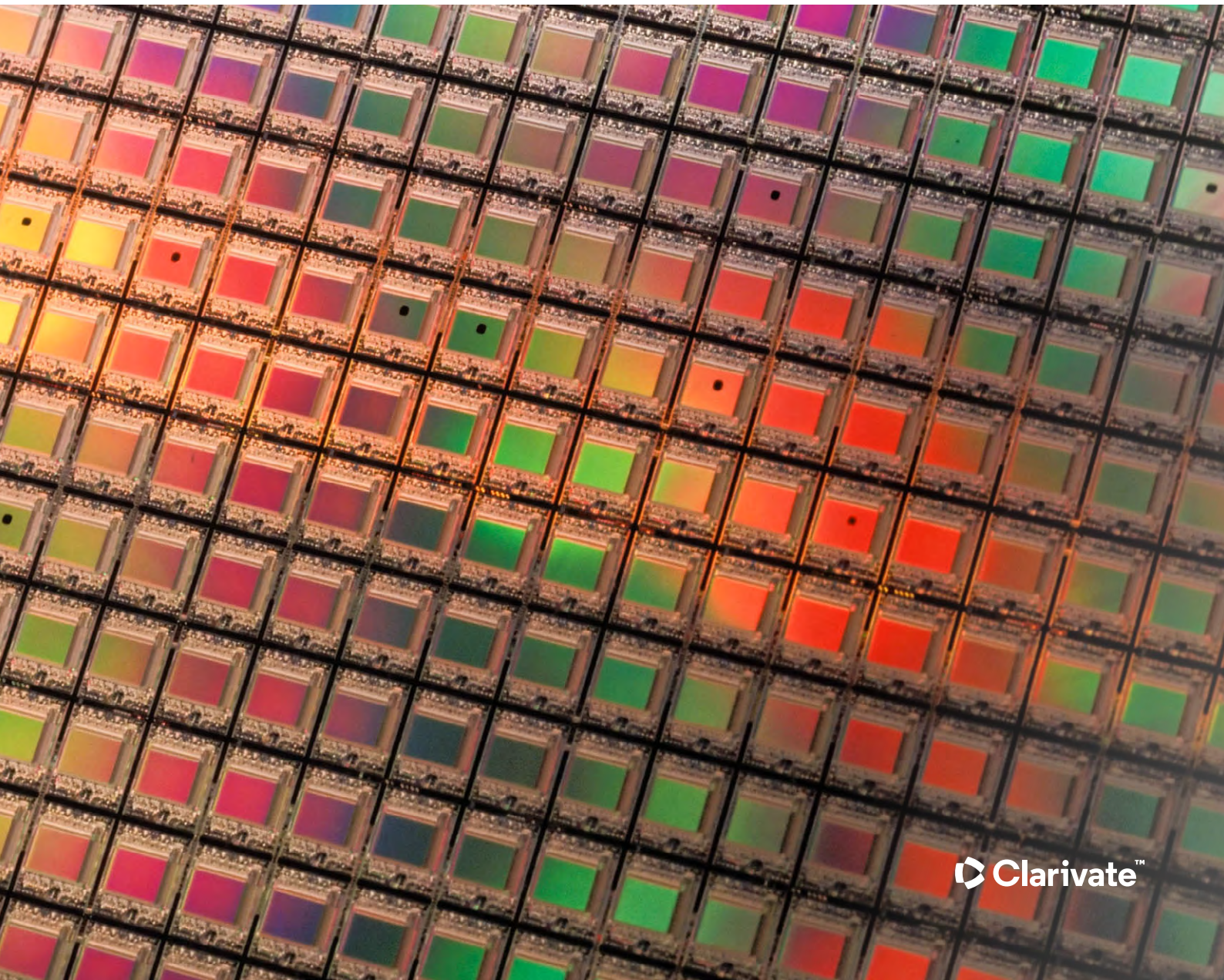
Institute for Scientific Information

Insights: Ethnic diversity in STEM in the United States

A bibliometric approach to identify under-represented minorities in scientific publications 2010 – 2020

Ryan Beardsley, Gali Halevi



Clarivate™

Author biographies

Ryan Beardsley is a Senior Consultant with the Institute for Scientific Information™. Ryan contributes to research and development initiatives, thought leadership and academic publications. In the decade prior to working at Clarivate™, Ryan was a condensed matter physicist, working in universities around the globe, where he was the principle investigator on multiple funded awards.
Web of Science ResearcherID: AAE-5094-2022

Gali Halevi is Director at the Institute for Scientific Information. Previously she was Associate Dean of Libraries & Information Sciences at the Icahn School of Medicine at Mount Sinai and the Mount Sinai Health System. She has a background in bibliometrics and has worked in academic publishing and as an information specialist for pharmaceutical and financial companies.
ORCID ID: 0000-0003-1967-4147
Web of Science ResearcherID: AAD-2257-2021

Foundational past, visionary future

About the Institute for Scientific Information

The Institute for Scientific Information at Clarivate has pioneered the organization of the world's research information for more than half a century. Today it remains committed to promoting integrity in research while improving the retrieval, interpretation and utility of scientific information.

It maintains the knowledge corpus upon which the Web of Science™ index and related information and analytical content and services are built. It disseminates that knowledge externally through events, conferences and publications while conducting primary research to sustain, extend and improve the knowledge base. For more information, please visit www.clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/.

About ISI Insights

ISI Insights are a new series of analyses from the Institute for Scientific Information (ISI) which offer a concise and informative analysis of topical research trends, using best-in-class citation data and analytics from Clarivate.

This, our first ISI Insights paper, combines self-identification data from the U.S Census with the uniquely structured and curated Web of Science data to examine the issue of diversity in authorship of scientific publishing in the U.S.

Throughout this ISI Insights paper, we use the same ethnic terms used by the U.S Census.

DOI Number: https://dx.doi.org/10.14322/ISI.insight.1

Cover image: Silicon wafer

In electronics, a wafer (also called a slice or substrate) is a thin slice of semiconductor, such as a crystalline silicon (c-Si), used for the fabrication of integrated circuits and, in photovoltaics, to manufacture solar cells. The wafer serves as the substrate for microelectronic devices built in and upon the wafer. It undergoes many microfabrication processes, such as doping, ion implantation, etching, thin-film deposition of various materials, and photolithographic patterning.

ISI Insights

1 Most areas of study that we examined in the United States exhibit a sustained underrepresentation of specific minority groups, most notably *Black Only* and *Hispanic* authors. Moreover, the situation has changed little across the 10-year period we looked at. **We see a need to develop mentorship, collaborations and partnerships that will encourage authorship from all underrepresented minorities.**

2 In all areas we looked at, publication rates from *Asian/Pacific Islands Only* authors are vastly higher than might be expected in the context of the U.S. population. For example, only 5% of the U.S. population self-identify with this category, but in the field of computer science – artificial intelligence **they represent 33% of authorships, more than six times the greater population.** This is an upward trending fraction in that field, where *Asian/Pacific Islands Only* authorship appears to have exceeded that of *White Only* (64% U.S. population) since 2017. No other minority group has a comparable upward trend.

3 Of the fields we studied, **medical research has the highest representation of *Black Only* authors (5.5%)**, but this is still less than half of the same group's representation within the U.S. population of 12%.

4 Math is dominated by *White* authorship but sees a **steady growth of *Asian/ Pacific Islands Only* authors.** However, there are low authorship levels of *Black* and *Hispanic* authorship in the field – and hardly any growth.

5 ***Black Only* and *Hispanic* authors do not show significant growth or decrease,** and participation is mostly low and stagnant.

6 ***Native American/Alaska Native* authorship is very low across the studied disciplines** and in some cases, such as computer science, sees decrease in participation.

Data behind this analysis

The Web of Science organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is the world's largest publisher-neutral citation index and research intelligence platform

Learn more:

https://clarivate.com/webofsciencegroup/solutions/web-of-science/

Introduction

Diversity in the workplace has been found to have a myriad of benefits, from accelerated innovation and discovery to better decision-making.

A range of perspectives and backgrounds can improve productivity (which may result in stronger financial results in business settings) and problem solving. In academic settings[2], improved diversity - whether that includes diversity of ethnicity, gender, physical ability or neurodiversity - may also improve outcomes, broaden scope and foster innovation in research.

This ISI Insights paper will focus on diversity of ethnicity in authorship of some STEM research subjects in the United States, which has been a topic of much public discussion in the past few years (Diversity et al., 2020; Fry et al., 2021; Miriti, 2020). Other previous work has focused on publishing in the area of diversity itself. A 2019 survey of 469 faculty members in ecology and evolutionary biology across the U.S. found that the vast majority of faculty engaging in activities pertaining to diversity and inclusion are those identifying themselves as non-white, non-male, first generation college attendees and the majority of respondents also felt that diversity and inclusion work did not matter in tenure decisions. The most prevalent reasons for lack of diversity and inclusion work in higher education was attributed to insufficient time, funding and knowledge of best practices.

Publications focusing on ethnic inequality in STEM often concentrate on the access to and participation in quality education as a pathway to a more diverse workforce, which is seen as the main barrier for underrepresented minorities in the research arena (Anjur, 2021). Many of the articles published on this topic address issues of unequal higher education access, and call for higher education leadership to promote and support underrepresented minorities in their pursuit of degrees in STEM (Bilimoria & Singer, 2019; Griffin et al., 2020). While these are important signals of increased awareness, they are mostly based on institutional surveys or employment records that shed a light on the topic but do not provide much insight into the research landscape as a whole.

In this Insights paper we sought to find a way to identify the ethnicity of authors of STEM research articles. The purpose was to not only be able to identify gaps in ethnic diversity in research but also to discover the trends underpinning participation and inclusivity of authorship across disciplines. In addition, we aimed to discover if and how the research landscape is changing, and whether there are changes in the levels of authorship of underrepresented minorities.

This ISI Insights paper will focus on diversity of ethnicity in authorship of some STEM research subjects in the United States, which has been a topic of much public discussion in the past few years

[2] https://clarivate.com/blog/annual-g20-scorecard-the-link-between-covid-19-response-and-research-subject-diversity/

Methodology

To track ethnic authorships trends on a larger scale and across science, technology, engineering and mathematics (STEM) disciplines we used bibliographic authors' last names retrieved from articles indexed in the Web of Science and used the U.S. Census last names database, which includes self-identifying ethnic identification to approximate the ethnic identification of authors using their last names.

We limited our Web of Science[3] dataset to publications from U.S. institutions and U.S. authors only, as we were using comparative data from the U.S. Census. We excluded publications that had international affiliations. For all publications identified as U.S. publications only, we extracted the authors' last names from the Web of Science and extracted the U.S. Census data from the government website[4]. Using an exact match between authors' last names and U.S. Census data we were able to match, depending on the discipline, 75%-80% of authors' last names to the U.S. Census ethnic identification. Throughout this ISI Insights paper, we use the same ethnic terms used by the U.S Census. According to the 2010 Census, the overwhelming majority (97%) of the total U.S. population reported only one race in 2010. The U.S. Census data used in this study is comprised of people identifying as:

- *White Only* (64%)
- *Black Only* (12%)
- *Asian/Pacific Islands Only* (5%)
- *Native American/Alaska Native* (1%)
- *Two or More Races* (2%)
- *Hispanic* (16%)
- *Unknown* (<1%)

We then calculated the probability of a certain last name to be identified as (1) *White Only*, (2) *Black Only*, (3) *Hispanic*, (4) *Asian/Pacific Islands Only*, (5) *Native American/Alaska Native*, (6) *Two or more races* and (7) *Unknown*. We assumed that each name takes fractions indicated in the Census data. Therefore, an author is not assigned to one ethnic group but rather each author's last name is assigned the fractional probability of the respective ethnicities. For example, if the probability of a last name is being 90% *White Only*, 5% *Black Only*, 2% *Asian/Pacific Islands Only* and 3% *Two or More Races*, we retain all of these probabilities; we do so for each authorship in our dataset. By averaging probabilities for each ethnic group, we obtain the mean contribution of each group to the set, which was over a 10-year time window and partitioned across several disciplines.

While comparing the overall population's racial distributions and authorships trends, one must remember that the scientific authorship does not necessarily represent the overall population. While racial groups may be distributed in a certain way in the population, they might not be in the research community. Therefore, in our analysis we look at both authorships trends over time and in the context of the U.S. population.

Areas of research

For this study we selected four areas of research, representing diverse disciplines in STEM:

1. Biochemistry

2. Mathematics

3. Medical research

4. Computer science

[3] https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[4] https://www.census.gov/data/developers/data-sets/decennial-census.html

Biochemistry

The biochemistry dataset included 205,481 U.S. exclusive publications from 2010-2020 and 128,571 unique last names. Figure 1 shows the authorships and the U.S. population as described by the U.S. Census data. *Asian/Pacific Island Only* authorships exceed the U.S. background by a substantial margin while *Hispanic* authorships are dramatically underrepresented compared to the background.

As can be seen in Figure 1, the average author names identified as *White Only* have the majority in the field. White-identifying academics represent slightly more than 44% of all authors and remain at this level through the years, up to 2018. In 2018 and 2019 the authorship identified as *White Only* decreases to 42.5% and 41.9% respectively with a slight increase seen in 2020.

The level of *Asian/Pacific Islands Only* authorship remains at the 23%-24% average, showing slight increase in 2018, 2019 and 2020. However, *Black Only* authorship remains stagnant at 5.5%. *Hispanic* authorship shows a slight increase from 4.1% in 2010 to 5.2% in 2019 with a slight decrease in 2020 to 4.8%. *Native American/ Alaska Natives'* authorship in biochemistry is significantly lower at 0.33% with no evident increase.

Figure 1. Ethnic distribution of authorship in biochemistry

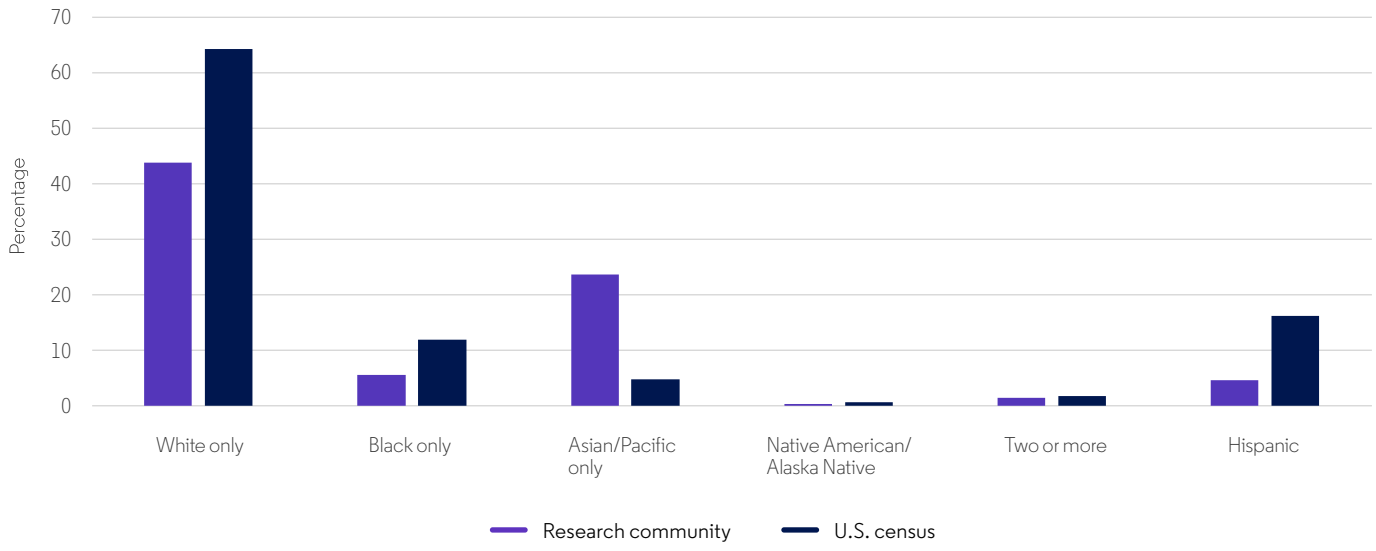
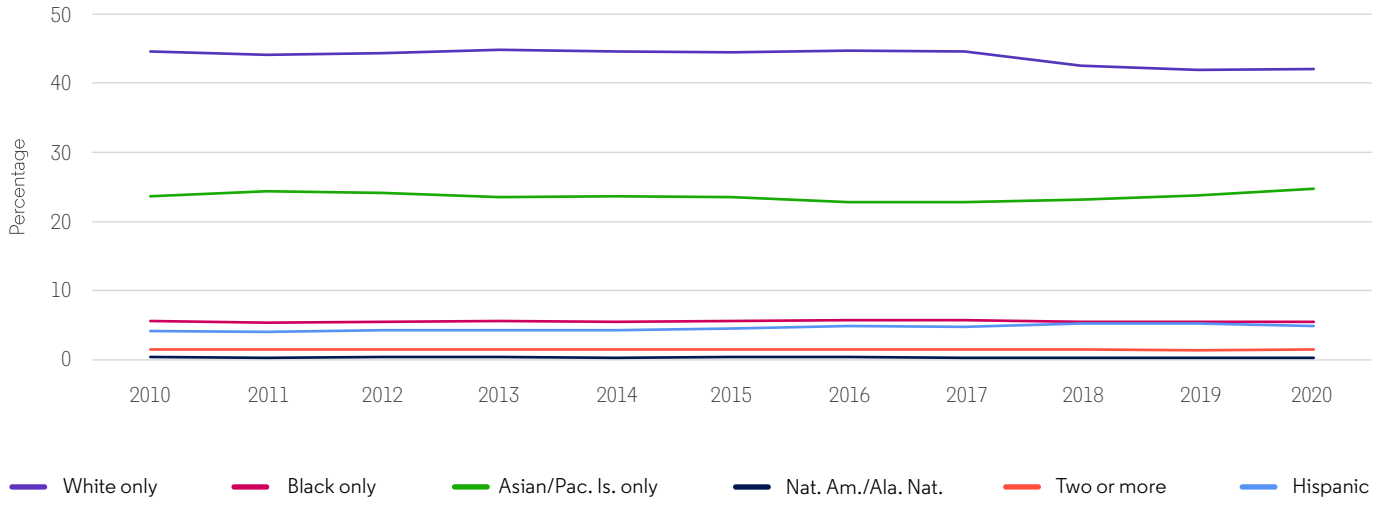


Figure 2. Change in authorship distribution over time in biochemistry



Mathematics

The mathematics dataset contained 48,080 U.S. exclusive publications from 2010-2020 and 17,909 unique names. Figure 3 indicates an excess of *Asian/Pacific Island Only* authors compared to the U.S. background in mathematics. The gap is smaller here than for other disciplines, but *Hispanic* and *Black Only* authorship is equally underrepresented as elsewhere.

As can be seen in Figure 4, the field has a majority of authors identified as *White Only*. *White Only* authorship is approximately 47% but sees a steady decrease from 2010 onward to 45% in 2020. In contrast to biochemistry, in mathematics the gap between *White Only* authorship and other ethnic groups is larger.

Asian/Pacific Islands Only authorship is estimated at 15%-19% through the years, showing a very slight increase in 2020. *Black Only* authorship is stagnant at 4.6%, while *Hispanic* authorship is stagnant at approximately 3.8%. The level of *Native American/Alaska Native* authors is extremely low at 0.28%.

Figure 3. Ethic distribution of authorship in mathematics

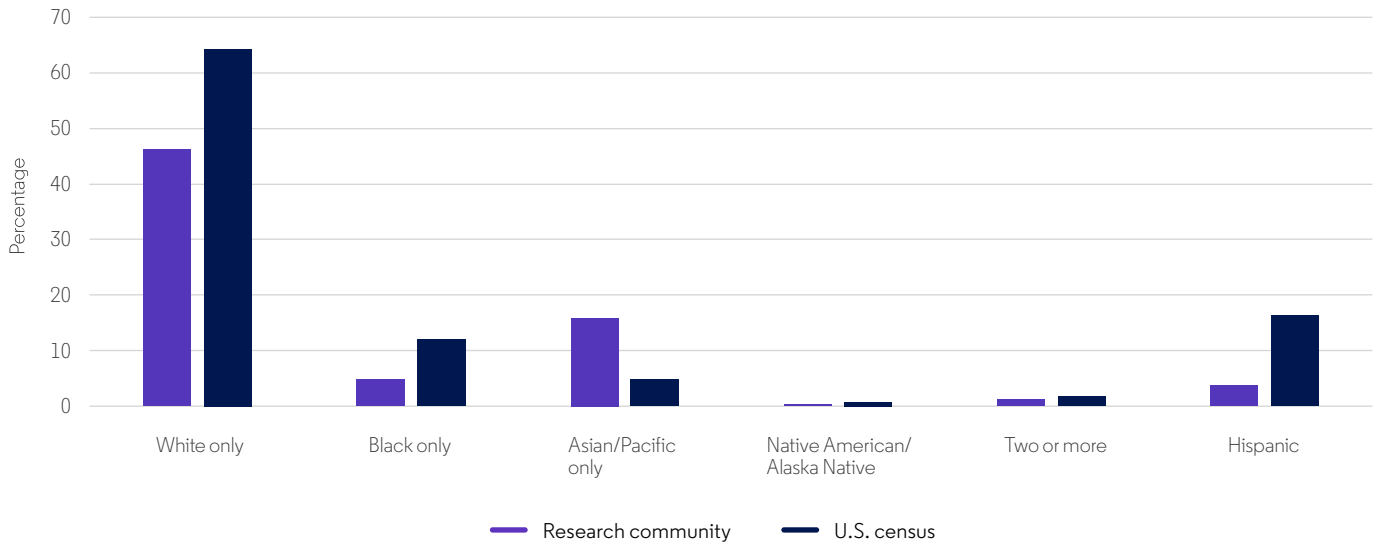
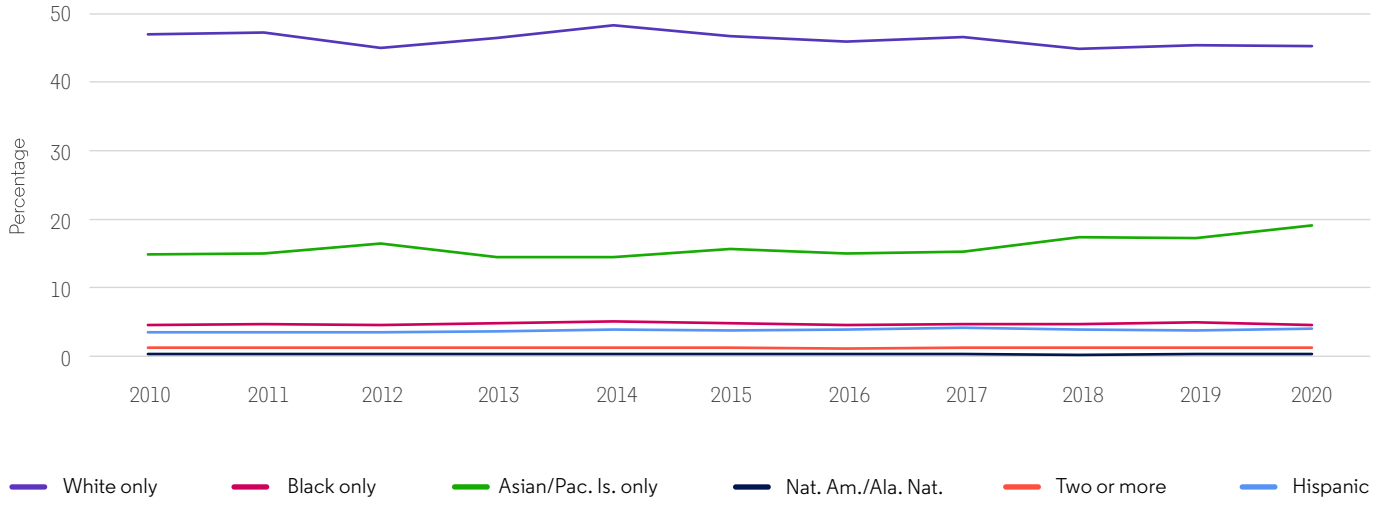


Figure 4. Change in authorship distribution over time in mathematics



Medical research

Our medical research dataset consisted of 102,209 U.S. exclusive publications from 2010-2020 and 90,145 unique names. Figure 5 reproduces the essential features of other fields, with high *Asian/Pacific Island Only* engagement but under-representation in other ethnicities.

As can be seen in Figure 6, although *White Only* authorship dominates, it is steadily decreasing over time from approximately 45% in 2010 to 42% in 2020, while *Asian/Pacific Islands Only* authorship is steadily increasing from approximately 21% in 2010 to 23% in 2020.

Black Only authorship is low but steady, at approximately 5.5%. *Hispanic* authorship is low but steadily increasing from approximately 4.3% in 2010 to 5.1% in 2020. *Native American/Alaska Native* authorship is extremely low at approximately 0.3% in 2010 and shows a very slight increase of 0.32% in 2020.

Figure 5. Ethnic distribution of authorship in medical research

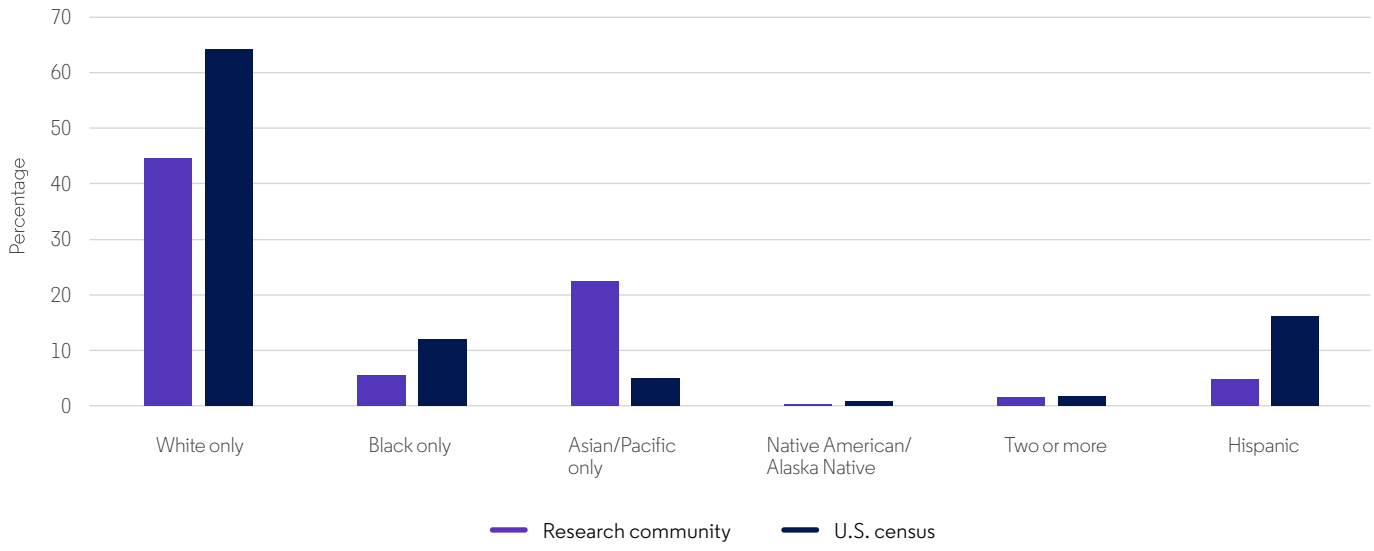
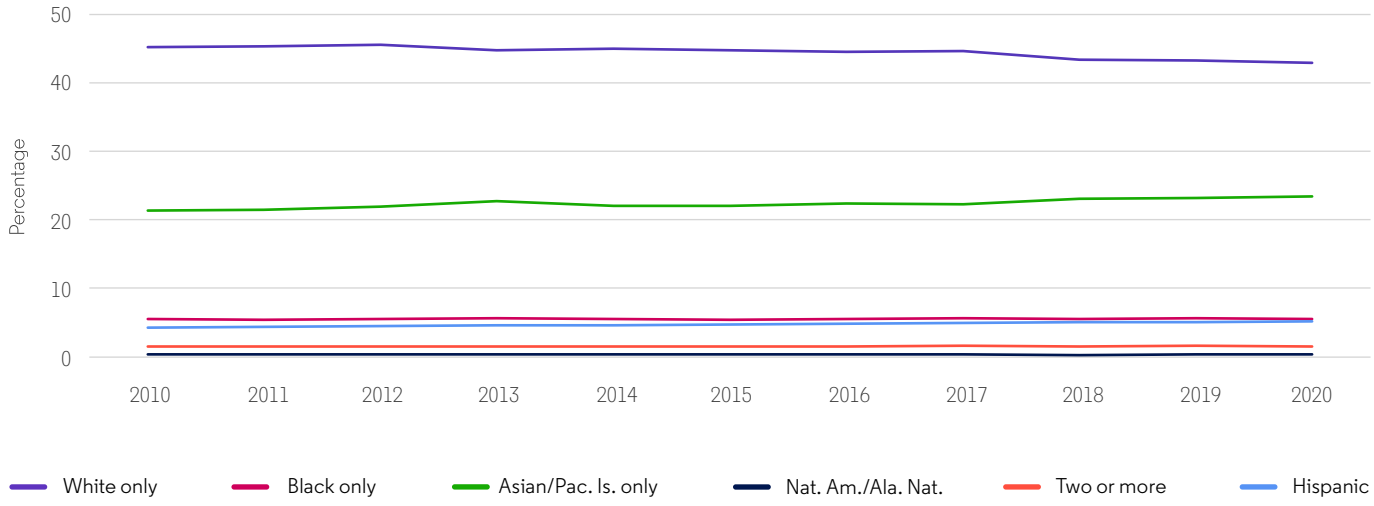


Figure 6. Change in authorship distribution over time in medical research



Computer science

The computer science dataset included 59,117 U.S. exclusive publications from 2010-2020 and 37,536 unique names. This area is seeing profound changes in authorship trends. Here we see relatively low representation of *White Only* authors at around half the U.S. population rate, with 37.7% in 2010, decreasing to approximately 30.7% in 2020.

However, *Asian/Pacific Islands Only* authorship increased substantially from 28.1% in 2010 to 38.1% in 2020, overtaking *White Only* authorship from 2017 onwards. *Black Only* authorship sees fluctuations with approximately 4.2% in 2010, a slight decrease in 2012-2013, an increase in 2016-2017 and a decrease from 2018 onwards to 3.8%.

Hispanic authorship is also decreasing slightly through the years from approximately 3.7% in 2010 to 3.2% in 2020. Finally, *Native American/ Alaska Native* authorship is lower than the other disciplines we examined with approximately 0.24% in 2010 decreasing to 0.23% in 2020.

Figure 7. Ethnic distribution of authorship in computer science – artificial intelligence

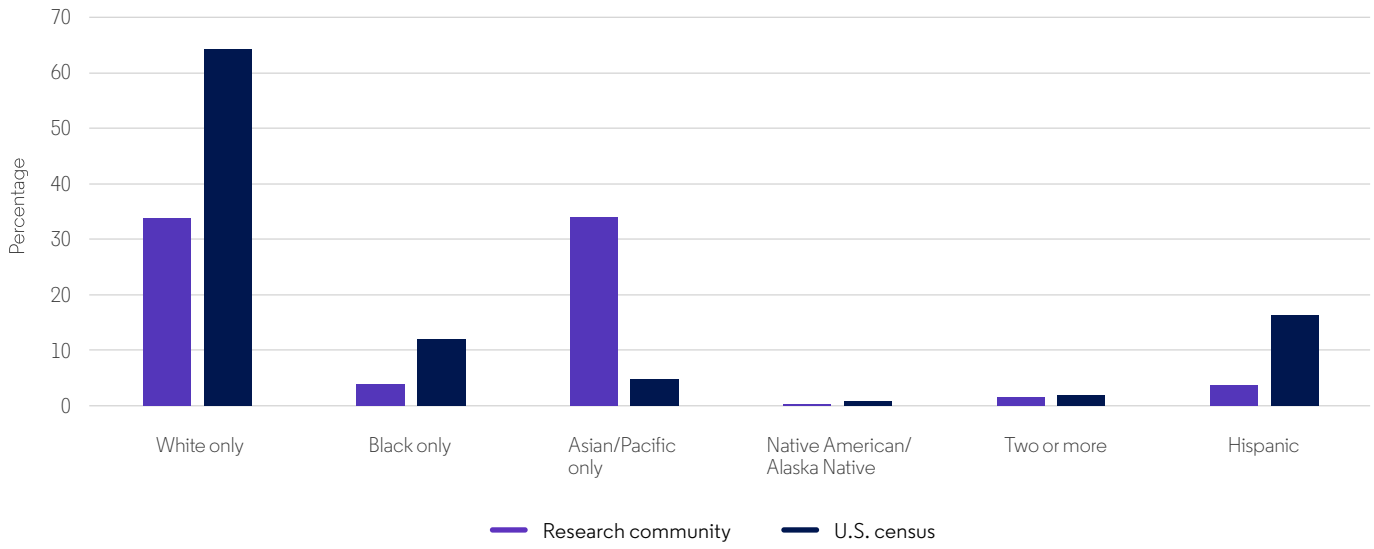
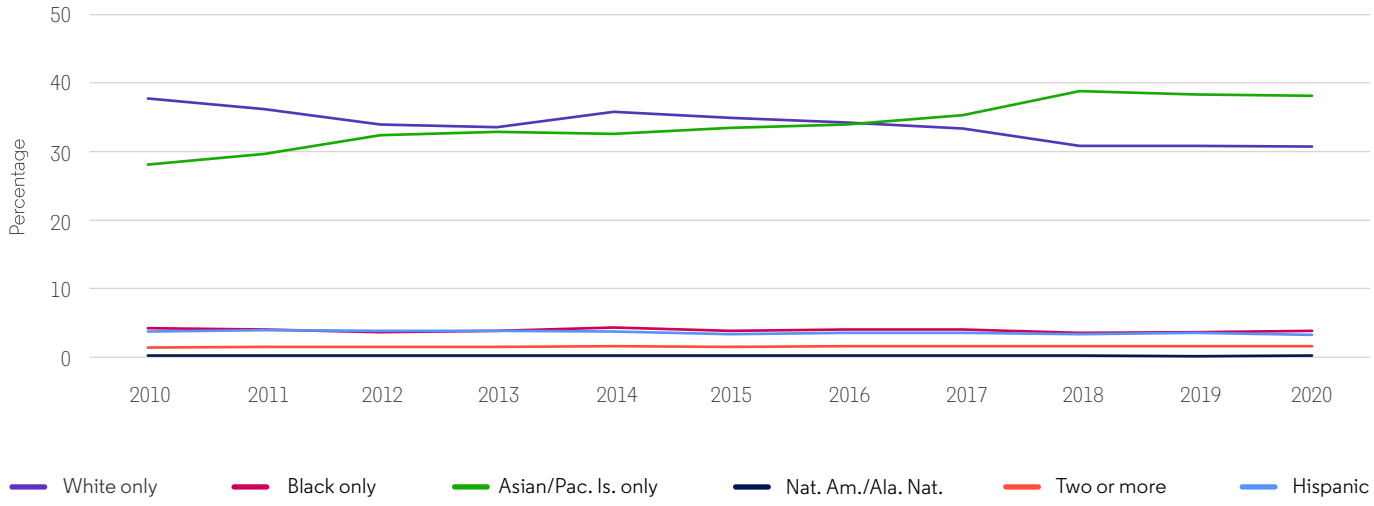


Figure 8. Change in authorship distribution over time in computer science – artificial intelligence



Conclusions

Our analysis shows that in most of the disciplines we chose to analyze, ethnicity of authorship hasn't changed much in the last 10 years, despite gradually increasing awareness of the importance of improved diversity.

The Sustainable Development Goals[5] from the UN stress the need for gender equality and reduced inequalities. As the scientific endeavor works towards solving these goals, it is particularly important that the workforce represents all countries, and that authors from diverse backgrounds are represented in scholarly publishing and journals too, in order to reduce inherent biases and enable a complete and global picture of the challenges we face. We recommend that more be done in terms of mentorship, development and education to encourage diversity in science that will eventually bear fruit through diverse authorship across disciplines. This must be a collaboration across the research landscape, from universities to funders, publishers and other organizations working in academia. It is an issue that cannot be solved by one group alone, but working together, progress can and should be made.

More must be done in terms of mentorship, development and education to encourage diversity in science that will eventually bear fruit through diverse authorship across disciplines.

Limitations of analysis

The purpose of this analysis was to introduce a methodology to explore measures of diversity through a combination of U.S. Census data and Clarivate bibliometric data. Some limitations of the study relate to the range of data used. We restricted ourselves to publications with exclusively U.S. based authors, as we wanted to be consistent with the Census data we had access to. As a pilot study, we also took a small sample of disciplines to investigate. A fuller range of research areas would be an interesting further work, as well as investigation cutting across institutions or geographical regions. Further to this, our methodology uses name matching to establish ethnic probabilities, which inevitably leads to some authors being excluded as their name is unknown to the U.S. Census. These unknown names could be handled differently in a more extensive work.

Are you interested in assessing diversity at your institution?

Contact us to learn how Clarivate can help:

ISI@clarivate.com

[5] https://www.undp.org/sustainable-development-goals

References

Abramo, G., & D'Angelo, C. A. (2015). Accounting for gender research performance differences in ranking universities. Current Science, 109(10), 1783–1789. . https://doi.org/10.18520/v109/i10/1783-1789

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. Journal of Informetrics, 7(4), 811–822. https://doi.org/10.1016/j.joi.2013.07.002

Anjur, S. (2021). Diversity and Inclusion in the STEM Classroom.

Bilimoria, D., & Singer, L. T. (2019). Institutions Developing Excellence in Academic Leadership (IDEAL): A partnership to advance gender equity, diversity, and inclusion in academic STEM. Equality, Diversity and Inclusion: An International Journal.

Diversity, A.-P. P. G. (APPG) on, Inclusion in Science, E., & Maths (STEM). (2020). Inquiry into Equity in the STEM Workforce.

Frandsen, T. F., Jacobsen, R. H., Wallin, J. A., Brixen, K., & Ousager, J. (2015). Gender differences in scientific performance: A bibliometric matching analysis of Danish health sciences Graduates. Journal of Informetrics, 9(4), 1007–1017.

Fry, R., Kennedy, B., & Funk, C. (2021, April 1). STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity. Pew Research Center Science & Society. https://www.pewresearch.org/science/2021/04/01/stem-jobs-see-uneven-progress-in-increasing-gender-racial-and-ethnic-diversity/

Griffin, K., Mabe, A., & Bennett, J. (2020). A Guidebook for a Campus Self-assessment of Successes and Challenges in STEM Faculty Diversity and Inclusion (2020).

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. Nature, 504(7479), 211–213. https://doi.org/10.1038/504211a

Miriti, M. N. (2020). The elephant in the room: Race and STEM diversity. BioScience, 70(3), 237–242.

Office, U. C. B. P. I. (n.d.). 2010 Census Shows America's Diversity—2010 Census—Newsroom—U.S. Census Bureau. Retrieved November 12, 2021, from https://www.census.gov/newsroom/releases/archives/2010_census/cb11-cn125.html

Professional services playbook: Adapt today, thrive tomorrow | Explore Our Thinking | Plante Moran. (n.d.). Retrieved November 10, 2021, from https://www.plantemoran.com/explore-our-thinking/insight/2020/06/professional-services-playbook-adapt-today-thrive-tomorrow

About Clarivate

Clarivate™ is a global leader in providing solutions to accelerate the pace of innovation. Our bold mission is to help customers solve some of the world's most complex problems by providing actionable information and insights that reduce the time from new ideas to life-changing inventions in the areas of Academia & Government, Life Sciences & Healthcare, Professional Services and Consumer Goods, Manufacturing & Technology. We help customers discover, protect and commercialize their inventions using our trusted subscription and technology-based solutions coupled with deep domain expertise. For more information, please visit clarivate.com.

The Web of Science™ is the world's largest publisher-neutral citation index and research intelligence platform. It organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research.

Download here:
www.clarivate.com/isi

Need to evaluate research at your organization?
Contact us to find out how Clarivate can help.

clarivate.com/contact-us/

WS813442335 / 05