PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Taube JC, Merritt A, Bansal S. 2025 Twenty years of infectious disease dynamics research: gender and race imbalances in publication and citation practices. *Proc. R. Soc. B* **292**: 20250960.
https://doi.org/10.1098/rspb.2025.0960

Received: 9 April 2025
Accepted: 2 June 2025

Subject Category:
Ecology

Subject Areas:
computational biology, health and disease and epidemiology

Keywords:
citation analysis, science of science, infectious disease dynamics

Author for correspondence:
Shweta Bansal
e-mail: shweta.bansal@georgetown.edu

†Joint first authors.

Special Feature: Research culture in infectious disease dynamics. Guest edited by Shweta Bansal, Paula Christen, Anne Cori and Amy Wesolowski.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7900913.

THE ROYAL SOCIETY
PUBLISHING

Twenty years of infectious disease dynamics research: gender and race imbalances in publication and citation practices

Juliana C. Taube†, Alexes Merritt† and Shweta Bansal

Department of Biology, Georgetown University, Washington, DC, USA

JCT, 0000-0002-0100-7648; AM, 0000-0003-1835-2074; SB, 0000-0002-1740-5421

Publication practices accumulate to affect credibility and career advancement. Understanding authorship and citation practices is critical to addressing inequities. While citation bias has been demonstrated in several fields, it remains uncharacterized in infectious disease dynamics (IDD), a quantitative, interdisciplinary domain highly visible during the COVID-19 pandemic. We analyse IDD articles and their bibliographies from 2000 to 2019 using machine-learning algorithms to infer the gender and race/ethnicity of each article's lead and senior authors. We examine authorship and citation patterns by gender and racial group across geographic scales, including characterizing the author composition of each article's bibliography relative to the field. Our analysis reveals persistent gender and race imbalances in IDD research. Man-authored and White-authored publications dominate the field, with little progress in racial diversification of US and UK publications over the last two decades. Woman-authored articles have the most representative citation practices but are undercited, especially when women are senior authors. In the USA and UK, most citations feature White lead and senior authors, even when citing articles have lead or senior authors of colour. These findings underscore the urgent need for more inclusive IDD research practices. We discuss possible mechanisms and solutions to create opportunities for researchers from underrepresented groups.

1. Introduction

Diverse teams produce better science [1,2] yet numerous barriers prevent underrepresented individuals from entering, flourishing and staying in scientific research (e.g. [3–5]). As it stands, White men generally dominate scientific fields: while women compose 51% of the United States (US) population, for example, they only make up 35% of the US STEM workforce [6]. The gender disparity in STEM fields follows a scissor-shaped curve: women slightly outnumber men at the undergraduate level, but in later career stages (PhD and beyond), men dominate, widening the gap between genders at a particularly visible professional level [7]. There are similar disparities in racial and ethnic representation within STEM fields [6,8]. Identifying these inequities within each academic field is necessary to address field-specific equity issues and, ultimately, to improve national and international scientific output and individual research experiences.

Publication and citation practices are two critical markers of scientific success that are particularly vulnerable to bias and could be contributing to these disparities. These metrics are integral to visibility, credibility, career advancement and attaining leadership positions but are not always objectively constructed and may be amplifying existing inequities [9]. Characterizing how publication rates and citation practices vary within the scientific ecosystem is one way to evaluate the diversity and equity within a scientific field. This work has been conducted across disciplines [10–12] and in several specific fields including political science [13], international relations [14,15], neuroscience [16–18], physics [19], astronomy [20] and medicine [21]. For example, Bertolero and colleagues found that White neuroscience authors preferentially cite other White authors and that neuroscience authors of colour are being increasingly undercited despite authoring an increasing proportion of articles [17]. Likewise, Caplar and colleagues found that women are lead-authoring an increasing proportion of astronomy publications, but their work is undercited compared to men [20]. Within infectious disease research, a recent study on editors and authorship in 40 infectious disease journals found that women are underrepresented in senior author and editor positions and suggested that recruiting more women editors could help increase the publication rate for women [22]. The field of infectious disease dynamics (IDD) is distinct from most infectious disease research due to its highly quantitative nature and is highly visible to the public due to its ability to provide actionable insights during health crises. These distinctions make the field both prone to gender and race disparities and a priority for ensuring equitable research methods. However, there have been no comprehensive analyses of publication and citation practices in IDD to date.

Here, we address this gap by examining and quantifying the author composition and citation practices by gender and race/ethnicity within the interdisciplinary and impactful field of IDD. We compile a dataset of over 10 000 articles in IDD and infer the gender and race/ethnicity of lead (first) and senior (last) authors of these articles using validated machine-learning packages. We test three hypotheses: (i) men and White (lead/senior) authors are overrepresented in articles in the field (especially in the Global North); (ii) even after accounting for the composition of the field, articles with women (lead/senior) authors are undercited compared to articles authored by men authors; and (iii) articles by women and authors of colour exhibit more representative citation practices than articles with men and White (lead/senior) authors. Our findings characterize the degree of inequity in IDD publication and citation practices and suggest potential ways to remediate these issues.

2. Methods

(a) Data collection

IDD is an interdisciplinary field that publishes in a broad range of disciplinary and general scientific journals. Therefore, it is challenging to define the boundaries of the published literature in IDD by selecting journals or conducting keyword searches. Similarly, because no IDD international professional organization exists, membership in the field is not feasible to identify. For our analysis, we thus identified authors in the IDD field by identifying the articles that cite a set of influential primary research and review IDD articles before 2020. To identify this set of influential research, we searched the Web of Science Core Collection (WoSCC) and selected articles that developed or discussed dynamical models of population-scale disease transmission and had accrued an average of 50 or more citations per year since publication. This resulted in 23 research articles published between 1990 and 2015. (Additional details and the complete list of articles can be found in the electronic supplementary material.) To characterize authorship patterns in IDD, we define the *authorship dataset* as the set of articles published in 2000–2019 in English in WoSCC that cite these influential articles as of 25 November 2024 and are cited by another authorship dataset article by 25 November 2024.

We also define an alternative definition for the IDD authorship dataset by identifying articles that have cited three seminal IDD books: *Infectious Diseases of Humans: Dynamics and Control* [23], *Modeling Infectious Diseases in Humans and Animals* [24] and *Mathematical Tools for Understanding Infectious Disease Dynamics* [25]. These articles were extracted on 13 April 2023, from WoSCC. We present our findings based on this alternative definition in electronic supplementary material, figure S1.

To characterize citation patterns in IDD, we define the *citation dataset* as the set of articles cited in the articles of the authorship dataset. To exclude irrelevant citations, citation dataset articles must either be in the authorship dataset or have been cited sufficiently based on their date of publication (from more than 2 times for 2019, to more than 5 times for 2000, determined using the 90th percentile of the number of citations for other articles published the same year). We perform sensitivity analyses on this threshold for the citation dataset (electronic supplementary material, figures S19 and S20).

(b) Inferring identities

We inferred the gender and race/ethnicity of the lead and senior authors of each article in the authorship and citation datasets using the machine-learning algorithms implemented in `genderize.io` [26,27] and `ethnicolr` [28], respectively. Genderize.io (https://genderize.io/) is designed to infer the gender of a first name using a proprietary algorithm. The response is either male, female, or none, as well as a confidence probability, representing the number of data entries used to calculate the response and the proportion of names with the gender returned in the response. The underlying data (of over a billion individuals) are collected from social networks across 79 countries and 89 languages. Ethnicolr (https://github.com/appeler/ethnicolr) is a Python package designed to predict race and ethnicity based on the sequence of characters in a person's name. The package uses character-level recurrent neural networks (RNNs) to make these predictions using either first and last name or just last name. The model is trained on multiple data sources, including US Census data (2000 and 2010), Florida voting registration data and Wikipedia data. The predictions are broad race/ethnicity categories based on US definitions (White, Black, Asian, Hispanic), as well as a probability score for each potential category. Only identities classified with over a 70% confidence were included; we performed sensitivity analyses on this threshold (electronic supplementary material, figures S15–S18). We also assessed the sensitivity of our results to the choice of algorithm by additionally using `namsor` [29] for classification (electronic

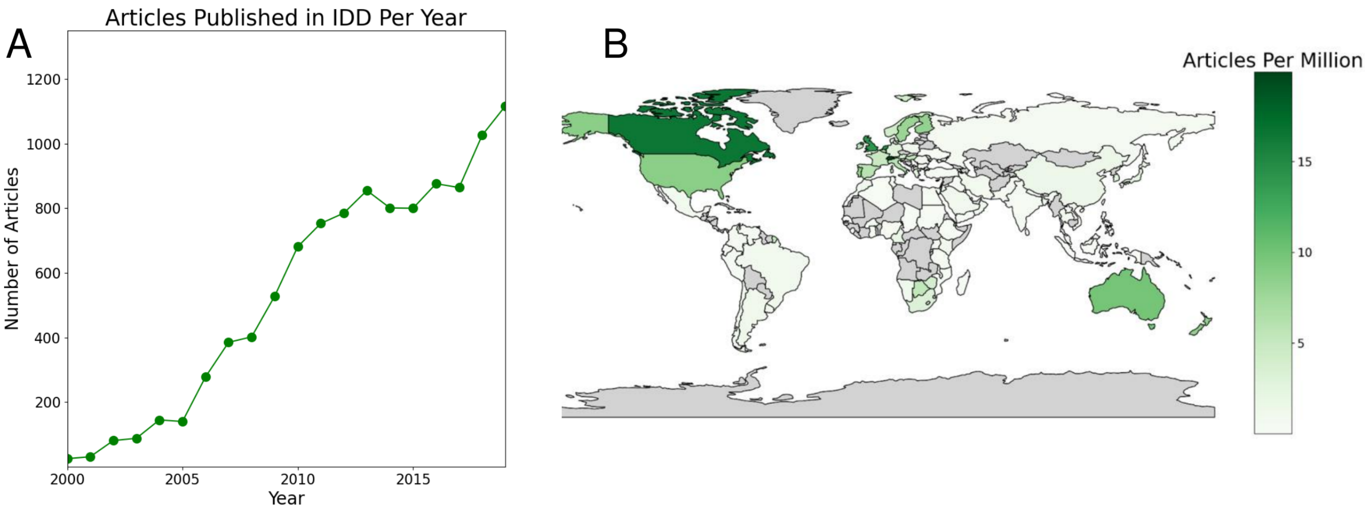

Figure 1. Infectious disease dynamics (IDD) publications have rapidly increased in the last two decades, but most senior authors are concentrated in the Global North. (A) Number of IDD publications per year from 2000 to 2019. (B) Distribution of articles published per million (general) population by country from 2000 to 2019, based on the affiliations of the senior authors.

supplementary material, figures S13 and S14). The accuracy of all three algorithms has been validated and these algorithms are regularly used in academic research [30–36]. We additionally use the affiliation for the senior author to assign a home country to each article, allowing us to specify our analysis to certain countries (e.g. USA, UK) or the Global North (which is broadly comprised of Northern America, Europe, Israel, Japan, South Korea, Australia and New Zealand). All geographically restricted analyses exclude articles that are missing senior author country affiliations. For articles with missing author first names or non-delineated first and last names, we pulled additional author name information from the CrossRef API, if available. To disambiguate authors who are recognized by different names across articles or with only initials for first names (e.g. E. Halloran versus Elizabeth Halloran), we used the algorithm in [16] to identify authors in the dataset with first and last names matching the initials and last names and replaced initials with the inferred full name. If there were different full author names that match a set of initials, then we did not replace the name (e.g. Elizabeth Halloran and Edward Halloran).

(c) Analyses

To understand authorship patterns, we measured the proportion of articles in the authorship dataset authored by each gender and race/ethnicity group across time and geography. When analysing mentorship patterns, we removed any articles with only a single author. We performed bootstrapping (1000 draws) to address small sample sizes. These proportions are compared relative to the gender and race/ethnicity representation of the general population and STEM workforce, where possible, in the respective geography (e.g. global, Global North, USA, UK).

To characterize gendered citation practices, we calculated the cumulative fraction of articles of each lead and senior authorship gender type (man–man (MM), woman–woman (WW), woman–man (WM), man–woman (MW)) across time; this distribution represents the pool of articles available to cite in a given bibliography. We then compared the makeup of each bibliography with that of the field one month prior to publication to assess which author types were being over- or undercited. We included citations of articles written before 2000 by assuming that the composition of the field had not changed before 2000. To understand whether the identity of the citing authors affected their citation practices, we also calculated the proportion of each gender pairing in each bibliography for articles with men lead and senior authors versus articles with a woman lead and/or senior author. We bootstrapped both of these estimates by randomly selecting articles from the authorship dataset 1000 times and analysing their bibliographies. We repeat the same analyses for race/ethnicity with a focus on White–White and lead and/or senior person of colour (PoC)-authored articles due to low sample size of mixed-race authorship. We restricted this analysis to years with 25 or more citations. In all our analyses of citation practices, we exclude self-citations to measure broader scholarly impact, reduce sub-disciplinary biases [37] and prevent inflation of diversity biases [38]. In our supplementary analyses, we consider the sensitivity of our findings to this decision (electronic supplementary material, figures S21 and S22).

We further analysed citation practices using a linear regression model to understand how lead and senior author identity predict citation rate (defined as the ratio of the number of total citations of an article and the number of years since publication). We exclude self-citations and control for 2023 Journal Impact Factor (JIF) and whether an article was published by authors in the Global North. (See electronic supplementary material for sensitivity analyses and diagnostics; articles missing senior author affiliation country are excluded.) The model was structured as

$$\begin{aligned}&\log_{10}\text{citation rate}_i\\&\sim \beta_0+\beta_1\mathrm{MW}_i+\beta_2\mathrm{WM}_i+\beta_3\mathrm{WW}_i+\beta_4\mathrm{HC}_i+\beta_5\mathrm{CH}_i+\beta_6\mathrm{CC}_i+\beta_7\mathrm{J}_i+\beta_8 1_{\mathrm{GlobalNorth}_i}+\beta_9\mathrm{MW}_i\mathrm{J}_i+\beta_{10}\mathrm{WM}_i\mathrm{J}_i\\&\qquad+\beta_{11}\mathrm{WW}_i\mathrm{J}_i+\beta_{12}\mathrm{HC}_i\mathrm{J}_i+\beta_{13}\mathrm{CH}_i\mathrm{J}_i+\beta_{14}\mathrm{CC}_i\mathrm{J}_i+\epsilon_i,\end{aligned}$$

where M stands for man, W for woman, C for PoC, H for White, and J for the square root of 2023 JIF, and $\varepsilon \sim N\left(0,\sigma^2\right)$.

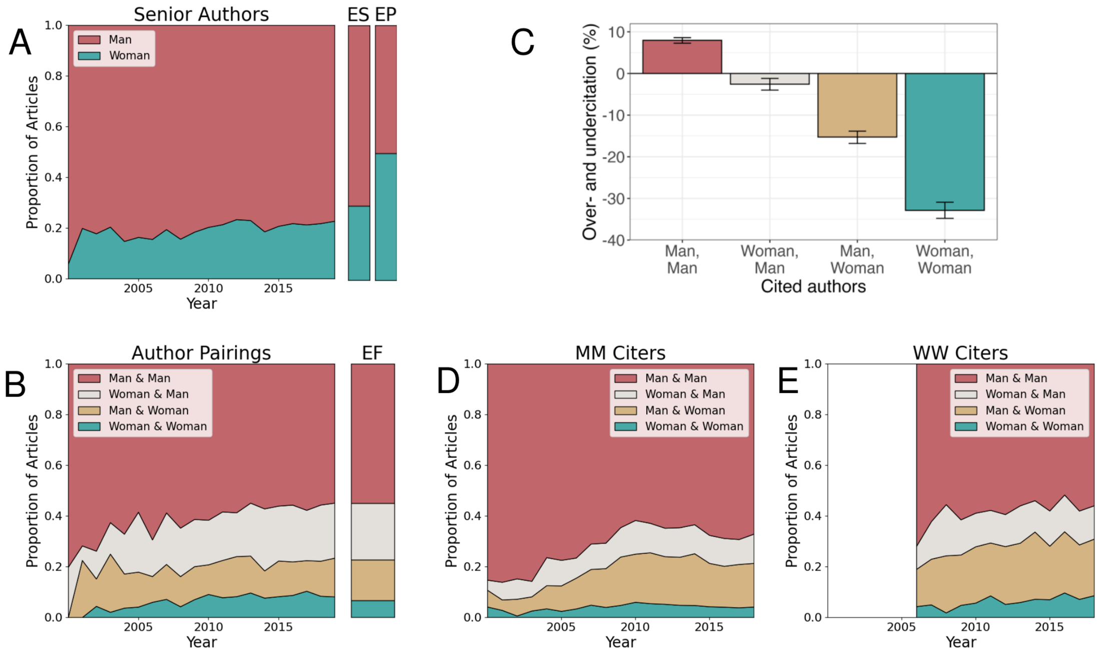

Figure 2. Globally, infectious disease dynamics publications are predominantly authored by men, but mentorship practices across genders do not appear biased. Women are undercited, particularly by man–man-authored articles. (A) Proportion of articles senior-authored by each gender from 2000 to 2019. Expected proportions based on the global STEM workforce (ES) in 2023 [39] and the (global) population (EP) in 2019 are shown in the colour bars at right. (B) Proportion of articles authored by a given (lead and senior author) gender pair from 2000 to 2019. Expected proportions based on the field (EF) in 2019 are shown in the colour bar at right; these expectations are based on the observed proportions in electronic supplementary material, figure S5. (C) Rates of over- and undercitation by article lead and senior author gender based on authorship gender composition prior to publication date. (D) Authorship gender breakdown of articles cited in articles by men lead and senior authors (MM) from 2000 to 2019. (E) Authorship gender breakdown of articles cited in articles written by women lead and senior authors (WW) from 2006 to 2019. From 2000 to 2005, there were not enough citations for this analysis.

3. Results

We identify 10 660 articles in the authorship dataset and 16 195 unique articles for the citation dataset cited in 220 570 citations across the bibliographies of the authorship dataset. The published literature of the IDD field grew rapidly and consistently across the last two decades (figure 1A) driven primarily by articles lead- and senior-authored by men and White individuals (in the USA and UK) (electronic supplementary material, figures S6–S9). The authorship dataset spans 101 countries of affiliation, with the majority of articles (65%) published between 2000 and 2019 written by senior authors from the Global North (figure 1B). The two countries in the Global North with the most published articles were the UK with 938 articles (14 per million people) and the USA with 2855 articles (8.70 per million people) (electronic supplementary material, figure S2). We were able to infer senior author gender for 85% of authorship articles and 86% of unique citation articles and senior author race for 63% of authorship articles and 53% of unique citation articles.

(a) Infectious disease dynamics publication authorship is dominated by men while woman-authored articles are undercited

We find that men lead-authored 61% of articles and senior-authored 67% of the articles published in IDD from 2000 to 2019, despite men composing only 50.4% of the global population. The ratio of man-authored to woman-authored articles has remained relatively constant since the early 2000s for both lead and senior authors (figure 2A and electronic supplementary material, figure S5). Concerning mentorship, 43% of articles have men lead and senior authors while only 6% of articles have women lead and senior authors. These ratios are heavily skewed compared to the global population and equal gender representation within the field but do not indicate unequal mentoring practices after taking into account the field's disproportionate number of men authors (figure 2B). These findings are consistent across geographic scales, gender-inference approaches, field definitions and after accounting for uncertainty in our sample (electronic supplementary material, figures S10, S11, S13 and S3).

In the context of citation practices, women lead- and senior-authored articles are undercited and the woman's authorship position affects the degree of undercitation. WM articles are undercited by 3%, MW by 15% and WW by 33% (figure 2C). These findings are robust to geography and gender inference package (electronic supplementary material, figures S10 and S13) but are sensitive to the definition of the IDD field (electronic supplementary material, figure S11); the rate of undercitation of WW-authored articles is not different from zero in the alternative definition of the field that has a smaller sample size. In 2019, 66% of references in MM articles are also MM-authored publications and this trend has been fairly consistent over time (figure 2D). WW articles have had increasingly gender-diverse bibliographies with only 55% of their citations corresponding to MM articles in 2019 (figure 2E). MM articles consistently cite articles with women as lead and/or senior authors at a lower rate than WW articles cite articles with women as lead and/or senior authors (statistically significant for each year with a *t*-test). However,

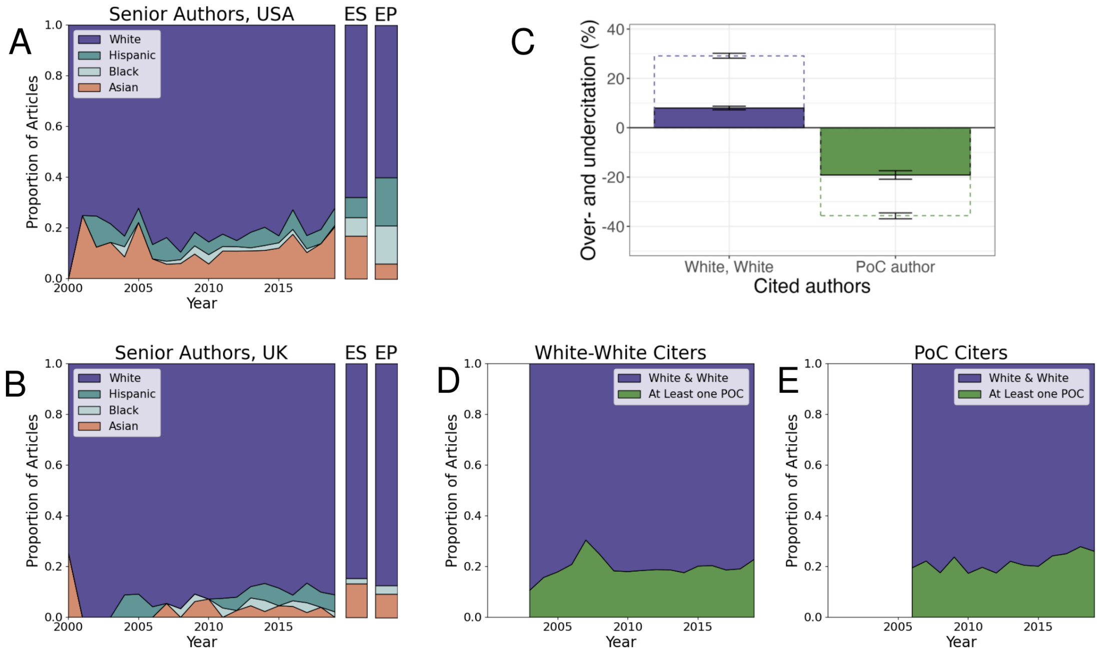

Figure 3. Authorship in infectious disease dynamics is not racially diverse and White-authored articles are overcited compared to articles that have a non-White lead and/or senior author. (A) Proportion of articles senior-authored by each race/ethnicity group in the USA from 2000 to 2019. Expected values (ES) based on the national STEM workforce based on 2019 National Science Foundation reporting [40] and expected values (EP) based on 2019 Census data [41] are shown in the colour bars to the right. (B) Proportion of articles senior-authored by each race/ethnicity group in the UK from 2000 to 2019. Expected values (ES) based on the national STEM academic staff [42] in 2019, expected values (EP) based on 2019 Office of National Statistics Data [43] are shown in the colour bars to the right. (C) Rates of over- and undercitation by article lead and senior author race/ethnicity based on authorship race/ethnicity composition prior to publication date. PoC author denotes that the lead and/or senior author was a person of colour; sample size was too small to disaggregate by author position for PoC authors. Filled boxes denote results for articles published by authors with affiliations within the Global North citing other work from the Global North. Dashed boxes denote results for articles published by authors with affiliations within the Global North citing articles with any affiliation location. (D) Racial/ethnic composition of articles authored in the Global North from 2003 to 2019 cited in articles written by White lead and senior authors in the Global North. From 2000 to 2002, there were not enough citations for this analysis. (E) Racial/ethnic composition of articles authored in the Global North from 2006 to 2019 cited in articles written by a lead and/or senior author of colour in the Global North. From 2000 to 2005, there were not enough citations for this analysis.

MM articles are increasingly citing articles with women as lead and/or senior authors at a slightly higher rate compared to the rate at which WW articles cite articles with women lead and/or senior authors (slope = 0.011, intercept = 0.21 for MM citers and slope = 0.006, intercept = 0.360 for WW citers with a Mann–Kendall test with $p < 0.05$). In electronic supplementary material, figure S21, we test the sensitivity of these findings to the inclusion of self-citations and find the results to be qualitatively robust.

(b) Race/ethnicity disparities are present in authorship and citation practices

Because race is a social construct that is experienced and treated differently across contexts, we focus on the race/ethnicity of authors in the Global North, where race is understood similarly. We restrict our analysis to four race/ethnicity categories in the USA and UK (White, Black, Hispanic and Asian) so that we can compare authorship composition with the general population and STEM workforce of each country. We find that most US publications have White authors with growing Asian authorship, which is less diverse than the general population (figure 3A and electronic supplementary material, figure S8). Similarly, nearly all UK publications have White authors, but authorship more closely reflects the general population (figure 3B and electronic supplementary material, figure S8). These findings are consistent across definitions of the IDD field, race/ethnicity inference methods and after accounting for uncertainty in our sample (electronic supplementary material, figures S12, S14, and S4).

Likewise, citation practices show significant biases. Articles with a lead and/or senior author of colour are undercited, while articles with White lead and senior authors are overcited in the Global North (figure 3C). The rates of over- and undercitation are more extreme when author's bibliographies are expected to reflect global race/ethnicity authorship (dashed boxes in figure 3C). Articles authored by lead and/or senior PoC generally cite articles authored by lead and/or senior PoC at a higher rate than articles with White lead and senior authors cite lead and/or senior PoC authored articles (8 of 14 years show statistical significance with a t-test, $p < 0.005$). Additionally, both groups of authors are increasingly citing articles with authors of colour as lead and/or senior authors at similar rates (White–White citers: slope = 0.0031, intercept = 0.15; PoC citers: slope = 0.0060, intercept = 0.15 with a Mann–Kendall test with $p < 0.05$) (figure 3D,E). In electronic supplementary material, figure S22, we test the sensitivity of these findings to the inclusion of self-citations and find the results to be qualitatively robust.

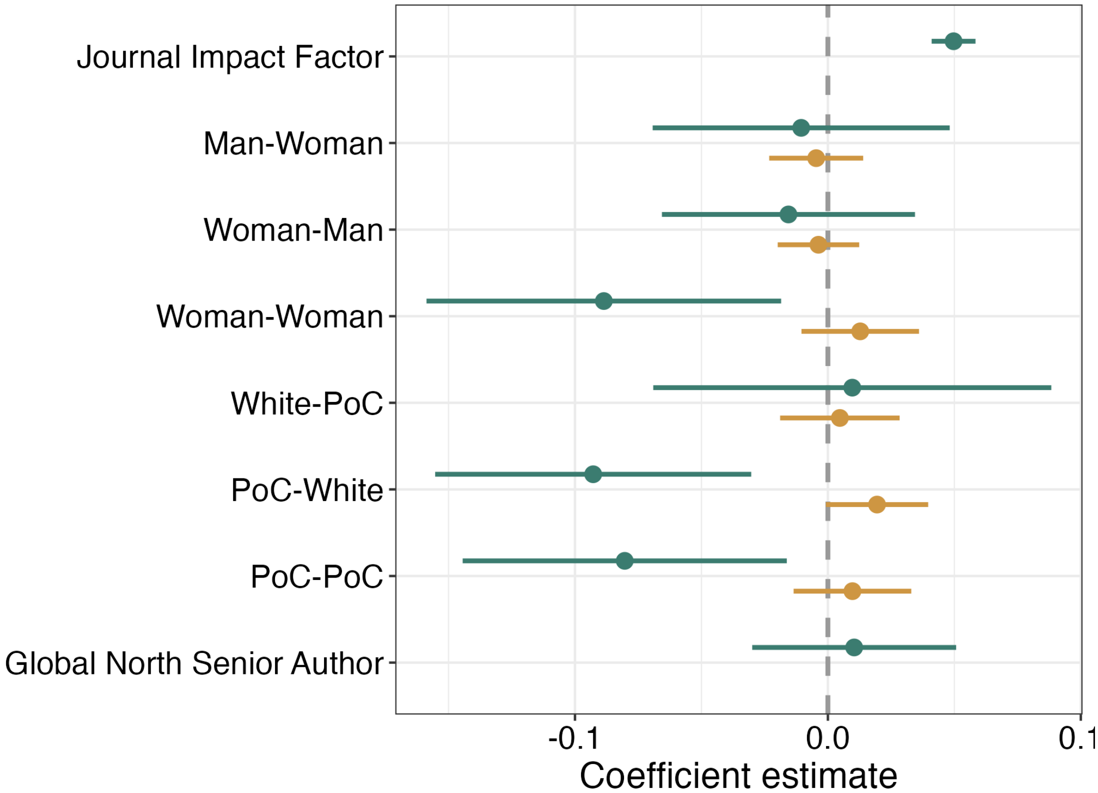

Figure 4. Citation rates in infectious disease dynamics are gender- and race/ethnicity-biased. Coefficient estimates are shown from a linear model predicting citation rate (no. citations/years since publication) based on article and author information. Teal point-ranges show the main effects (coefficient estimate and 95% CI) for the independent impact of the predictor on citation rates, and gold point-ranges show the interaction effects for the impact of the predictor on citation rates for different values of Journal Impact Factor. Coefficients represent the difference from citation of man–man and White–White authored articles with non-Global North senior authors and average Journal Impact Factor. Woman–woman, PoC–White and PoC–PoC authored articles are significantly less cited, while articles in high-impact journals are significantly more cited. Journal Impact Factor is square root transformed to ensure normality. PoC stands for person of colour.

(c) Author race and gender influence citation rate after controlling for publication venue

We found that author gender pairing and race pairing are significantly associated with article citation rate, even after accounting for the impact factor of the journal where the article was published (figure 4). WW, PoC–White and PoC–PoC articles had significantly lower citation rates, while JIF significantly increased citation rate. However, WW, PoC–White and PoC–PoC publications in high-impact-factor journals had higher citation rates relative to MM and WW publications, though these interactions were not statistically significant. Sensitivity analyses varying geography, self-citation inclusion and citation thresholds for inclusion can be found in electronic supplementary material, figures S25–S32.

4. Discussion

IDD is a rapidly growing area of interdisciplinary research. As the field reflects on lessons learned from the early stages of the COVID-19 pandemic and identifies areas of future work, it is paramount to consider which voices are underrepresented and underrecognized in the field. This introspection is essential for promoting inclusivity, but also because publications and citations are key ways that academic success, credibility and potential are measured, and public dissemination is achieved. We address this pressing issue by comprehensively analysing the authorship and bibliographic landscape of IDD publications. We overcome challenges to define the field by curating a list of influential articles and analyse the relevant articles citing these influential works. We use validated machine-learning packages [32] to infer the gender and race/ethnicity identity of the authors of each article and conduct sensitivity analyses on this identity inference, our definition of the field, and using varying spatial scales to demonstrate that our results are robust.

We find that publication and citation practices in the field of IDD are notably biased. Most publications are authored by men or White authors which does not reflect the composition of the general population (see intersectional results in electronic supplementary material, figure S33). Work authored by men or White scientists is overcited even after accounting for their disproportionate composition of the field. In contrast, work authored by women or people of colour is heavily undercited, and this amplifies with gender or race assortativity (i.e. mentorship by authors of the same identities). Author position also affects the rate of undercitation, with greater bias when women are in more senior authorship positions. The same trend may apply for authors of colour, but we did not have sufficient sample size to detect it. Our regression analysis shows that these issues persist even after accounting for JIF. However, we do learn that citation rates of the articles by women or people of colour in high-impact journals are higher relative to MM and WW authored articles, respectively. We speculate that this might reflect the phenomenon of overperformance for underrecognized individuals, in which systemically disadvantaged individuals are driven to and held to a higher standard for reaching the top of their field [44].

These gender results are consistent across spatial scales, but our race/ethnicity findings show greater geographic variability. The USA and the UK produced the most articles in the Global North, but authorship within these countries is not representative of the racial/ethnic makeup of the general population. Citation practices by race/ethnicity are similarly biased. Articles authored by White lead and senior authors are overcited while articles with at least one author of colour are undercited by more than 30%. This difference cannot be explained by White lead and senior authors citing fewer authors of colour (figure 3D,E).

We recognize that our interpretation of gender as binary and race/ethnicity fitting into four classifications is limiting and prevents analysis of some of the most vulnerable scientists, including, but not limited to, nonbinary individuals and Indigenous populations. Furthermore, race/ethnicity categorization can meaningfully differ across countries. Using names to infer gender and race can be inaccurate and ignores physical forms of identity expression [45]. However, names and affiliations are the key identifying information that editors, reviewers and authors have when choosing to read, accept, review or cite a given article. Thus, inaccuracies in gender or race/ethnicity inference may well reflect how an unknown author is perceived by others. We focus on authors listed first and last, ignoring shared authorship positions and the identity makeup of the full authorship list. We acknowledge that our sample may be missing some IDD publications or include non-IDD work, but we must start to address these issues of inequity even in the absence of perfect data. Despite these limitations, our findings align with those from other disciplines, such as neuroscience, physics, astronomy and medicine [16,19–21]. In particular, similar to the fields of mathematics and computer science, we find an increasing percentage of articles authored by women, but not at a sufficient rate to equal the rate at which men are authoring articles [46,47]. Likewise, our findings are nearly identical to those in software engineering and computer science, more broadly, where man-authored articles are overcited and woman-authored articles are undercited, with variation by woman author position [48,49]. However, WW-authored articles in software engineering and computer science overcite other WW-authored articles to a greater degree than in our study. Analyses of citation practices by race/ethnicity in quantitative fields are limited, though across published US authors of any field, Freeman and Huang observed assortativity in collaborations based on race/ethnicity [50]. Thus, the field of IDD appears to have demographic composition and citation practices similar to those of other quantitative disciplines.

As the field of IDD moves forward with additional urgency and scrutiny due to the COVID-19 pandemic, it is imperative to take action to increase the diversity and equity in the field [51]. We must also recognize how the pandemic has likely only exacerbated many of the disparities we have documented here [52]. For example, women had lower research productivity than men [53], including submitting fewer manuscripts during the early months of the pandemic [54]. We advocate for increased journal editorial board and reviewer diversity [22,55–57] to leverage the effect of homophily. Additionally, bibliographic transparency [19], in which editors and authors ensure that the composition of submitted bibliographies reflects the gender and race/ethnicity composition of the field, is critical. To facilitate this process, we've created a tool so that authors can check their own bibliographies relative to expected compositions for the field of IDD [58]. Additionally, individuals should consider citing both older initial work and recent applications by more junior and diverse authors in their manuscripts, a practice that can be aided by the removal of reference length limits by journals. Finally, academia should consider broadening the definitions of scientific contributions to reflect the variety of work that advances the field but may not be traditionally valued, for example, by acknowledging the value of dashboard, app, and software package creation, science communication and policy engagement. To move forward, the field must assess and examine our current composition and practices so that we can measure future progress; we begin such an effort here and call on others to join us in acknowledging and correcting the inequities within our field.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This project utilizes a commercial dataset provided by Clarivate and the raw data cannot be republished. For further information, contact Clarivate. The data and code used in our analysis are provided for reproducibility [59] and permanently on Dryad [60].

Electronic supplementary material is available online [61].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. J.C.T.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft, writing—review and editing; A.M.: formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; S.B.: conceptualization, data curation, investigation, methodology, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. We appreciate data sharing and valuable technical support provided by Ann Beynon and Rob Pritchett of Clarivate. We are grateful to Simon Frost for providing statistical expertise and assistance. We also thank Eva Rest and Andrew Tiu for their technical assistance and Juliet Pulliam, Anne Cori, Amy Wesolowski for their valuable feedback on this work.

References

1. Yang Y, Tian TY, Woodruff TK, Jones BF, Uzzi B. 2022 Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proc. Natl Acad. Sci. USA* **119**, e2200841119. (doi:10.1073/pnas.2200841119)
2. AlShebli BK, Rahwan T, Woon WL. 2018 The preeminence of ethnic diversity in scientific collaboration. *Nat. Commun.* **9**, 5163. (doi:10.1038/s41467-018-07634-8)

3. Spoon K, LaBerge N, Wapman KH, Zhang S, Morgan AC, Galesic M, Fosdick BK, Larremore DB, Clauset A. 2023 Gender and retention patterns among U.S. faculty. *Sci. Adv*. **9**, i2205. (doi:10.1126/sciadv.adi2205)
4. Cori A. 2024 SIR… or MADAM? The impact of privilege on careers in epidemic modelling. *Epidemics* **47**, 100769. (doi:10.1016/j.epidem.2024.100769)
5. Ross MB, Glennon BM, Murciano-Goroff R, Berkes EG, Weinberg BA, Lane JI. 2022 Women are credited less in science than men. *Nature* **608**, 135–145. (doi:10.1038/s41586-022-04966-w)
6. Docter-Loeb H. 2023 White men still hold majority of US science and engineering jobs. *Nature* (doi:10.1038/d41586-023-00865-w)
7. Joyce JA, Masina S, Michalik L, Pot C, Sempoux C, Amati F. 2024 Closing the scissor-shaped curve: strategies to promote gender equality in academia. *Cell* **187**, 1335–1342. (doi:10.1016/j.cell.2024.01.050)
8. National Science Foundation. 2023 *Diversity and STEM: women, minorities, and persons with disabilities 2023*. Alexandria, VA: National Center for Science and Engineering Statistics.
9. Zurn P, Teich EG, Simon SC, Kim JZ, Bassett DS. 2022 Supporting academic equity in physics through citation diversity. *Commun. Phys.* **5**, 240. (doi:10.1038/s42005-022-00999-9)
10. Liu F, Rahwan T, AlShebli B. 2023 Non-White scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. *Proc. Natl Acad. Sci. USA* **120**, e2215324120. (doi:10.1073/pnas.2215324120)
11. Kozlowski D, Larivière V, Sugimoto CR, Monroe-White T. 2022 Intersectional inequalities in science. *Proc. Natl Acad. Sci. USA* **119**, e2113067119. (doi:10.1073/pnas.2113067119)
12. Lerman K, Yu Y, Morstatter F, Pujara J. 2022 Gendered citation patterns among the scientific elite. *Proc. Natl Acad. Sci. USA* **119**, e2206070119. (doi:10.1073/pnas.2206070119)
13. Dion ML, Sumner JL, Mitchell SM. 2018 Gendered citation patterns across political science and social science methodology fields. *Polit. Anal*. **26**, 312–327. (doi:10.1017/pan.2018.12)
14. Maliniak D, Powers R, Walter BF. 2013 The gender citation gap in international relations. *Int. Organ*. **67**, 889–922. (doi:10.1017/s0020818313000209)
15. Mitchell SM, Lange S, Brus H. 2013 Gendered citation patterns in international relations journals. *Int. Stud. Perspect*. **14**, 485–492. (doi:10.1111/insp.12026)
16. Dworkin JD, Linn KA, Teich EG, Zurn P, Shinohara RT, Bassett DS. 2020 The extent and drivers of gender imbalance in neuroscience reference lists. *Nat. Neurosci*. **23**, 918–926. (doi:10.1038/s41593-020-0658-y)
17. Bertolero MA, Dworkin JD, David SU, Lloreda CL, Srivastava P, Stiso J. 2020 Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. *bioRxiv.* (doi:10.1101/2020.10.12.336230)
18. Palser ER, Lazerwitz M, Fotopoulou A. 2022 Gender and geographical disparity in editorial boards of journals in psychology and neuroscience. *Nat. Neurosci*. **25**, 272–279. (doi:10.1038/s41593-022-01012-w)
19. Teich EG *et al*. 2022 Citation inequity and gendered citation practices in contemporary physics. *Nat. Phys*. **18**, 1161–1170. (doi:10.1038/s41567-022-01770-1)
20. Caplar N, Tacchella S, Birrer S. 2017 Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron*. **1**, 1. (doi:10.1038/s41550-017-0141)
21. Chatterjee P, Werner RM. 2021 Gender disparity in citations in high-impact journal articles. *JAMA Netw. Open* **4**, e2114509. (doi:10.1001/jamanetworkopen.2021.14509)
22. Last K, Hübsch L, Cevik M, Wolkewitz M, Müller SE, Huttner A, Papan C. 2022 Association between women's authorship and women's editorship in infectious diseases journals: a cross-sectional study. *Lancet Infect. Dis*. **22**, 1455–1464. (doi:10.1016/s1473-3099(22)00367-x)
23. Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
24. Keeling MJ, Rohani P. 2011 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press.
25. Diekmann O, Heesterbeek H, Britton T. 2013 *Mathematical tools for understanding infectious disease dynamics*. Princeton, NJ: Princeton University Press.
26. Demografix APS. Find gender from a name—accurate gender classification API. See https://genderize.io/.
27. Wais K. 2016 Gender prediction methods based on first names with genderizeR. *R J*. **8**, 17. (doi:10.32614/rj-2016-002)
28. Laohaprapanon S, Sood G, Naji B. 2022 Ethnicolr: predict race and ethnicity from name. See https://github.com/appeler/ethnicolr.
29. Namsor. Namsor | name checker for gender, origin and ethnicity determination. See https://namsor.app/.
30. Chumthong T, Jitkajornwanich K, Kraishan O, Kee KF, Narabin A. 2024 Leveraging race prediction algorithms to enhance team composition in big data science teams. In *2024 IEEE Int. Conf. on Big Data (BigData), Washington, DC, USA, 15–18 December 2024*, pp. 3105–3113. (doi:10.1109/BigData62323.2024.10825329)
31. Xie F. 2022 rethnicity: an R package for predicting ethnicity from names. *SoftwareX* **17**, 100965. (doi:10.1016/j.softx.2021.100965)
32. VanHelene AD, Khatri I, Hilton CB, Mishra S, Gamsiz Uzun ED, Warner JL. 2024 Inferring gender from first names: comparing the accuracy of Genderize, Gender API, and the gender R package on authors of diverse nationality. *PLoS Digit. Health* **3**, e0000456. (doi:10.1371/journal.pdig.0000456)
33. Sebo P. 2021 Using genderize.io to infer the gender of first names: how to improve the accuracy of the inference. *J. Med. Libr. Assoc*. **109**, 609–612. (doi:10.5195/jmla.2021.1252)
34. Sebo P. 2021 Performance of gender detection tools: a comparative study of name-to-gender inference services. *J. Med. Libr. Assoc*. **109**, 414–421. (doi:10.5195/jmla.2021.1185)
35. Sebo P. 2023 How well does NamSor perform in predicting the country of origin and ethnicity of individuals based on their first and last names? *PLoS One* **18**, e0294562. (doi:10.1371/journal.pone.0294562)
36. Bursztyn L, Chaney T, Hassan TA, Rao A. 2024 The immigrant next door. *Am. Econ. Rev*. **114**, 348–384. (doi:10.1257/aer.20220376)
37. Patience GS, Patience CA, Blais B, Bertrand F. 2017 Citation analysis of scientific categories. *Heliyon* **3**, e00300. (doi:10.1016/j.heliyon.2017.e00300)
38. King MM, Bergstrom CT, Correll SJ, Jacquet J, West JD. 2017 Men set their own cites high: gender and self-citation across fields and over time. *Socius* **3**, 237802311773890. (doi:10.1177/2378023117738903)
39. World Economic Forum. 2023 *Global Gender Gap Report 2023*. See https://www.weforum.org/publications/global-gender-gap-report-2023/digest/ (accessed April 2025).
40. National Science Board. 2021 *The STEM labor force of today: scientists, engineers, and skilled technical workers*. National Science Foundation. See https://ncses.nsf.gov/pubs/nsb20212/participation-of-demographic-groups-in-stem (accessed March 2025).
41. US Census Bureau. 2025 *National population by characteristics: 2010-2019*. See https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-detail.html#par_textimage_1537638156 (accessed March 2025).
42. Royal Society. 2024 Ethnicity and disability STEM data for academic staff in higher education. See https://royalsociety.org/-/media/about-us/who-we-are/diversity/ethnicity-and-disability-stem-data-for-academic-staff-in-higher-education.pdf (accessed 19 March 2025).
43. Office for National Statistics. 2021 Population estimates by ethnic group and religion, England and Wales. See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/populationestimatesbyethnicgroupandreligionenglandandwales/2019 (accessed 19 March 2025).
44. Basson I, Ni C, Badia G, Tufenkji N, Sugimoto CR, Larivière V. 2023 Gender differences in submission behavior exacerbate publication disparities in elite journals. *eLife* **12**, RP90049. (doi:10.7554/eLife.90049.1)
45. Lockhart JW, King MM, Munsch C. 2023 Name-based demographic inference and the unequal distribution of misrecognition. *Nat. Hum. Behav*. **7**, 1084–1095. (doi:10.1038/s41562-023-01587-9)
46. Mihaljević-Brandt H, Santamaría L, Tullney M. 2016 The effect of gender in the publication patterns in mathematics. *PLoS One* **11**, e0165367. (doi:10.1371/journal.pone.0165367)

47. Wang LL, Stanovsky G, Weihs L, Etzioni O. 2021 Gender trends in computer science authorship. *Commun. ACM* **64**, 78–84. (doi:10.1145/3430803)
48. Tchilinguirova K, Belle AB, Mahamud G. 2024 Biases in gendered citation practices: an exploratory study and some reflections on the Matthew and Matilda effects. See https://arxiv.org/abs/2410.02801.
49. Nakajima K, Sasaki Y, Tokuno S, Fletcher G. 2024 Quantifying gendered citation imbalance in computer science conferences. In *Proc. AAAI/ACM Conf. on AI, Ethics, and Society*, vol. **7**, pp. 1011–1022, (doi:10.1609/aies.v7i1.31699)
50. Freeman RB, Huang W. 2015 Collaborating with people like me: ethnic coauthorship within the United States. *J. Labor Econ*. **33**, S289–S318. (doi:10.1086/678973)
51. Swartz TH, Palermo AGS, Masur SK, Aberg JA. 2019 The science and value of diversity: closing the gaps in our understanding of inclusion and diversity. *J. Infect. Dis*. **220**, S33–S41. (doi:10.1093/infdis/jiz174)
52. Carreri A, Naldini M, Tuselli A. 2024 Inequalities in academic work during COVID-19: the intersection of gender, class, and individuals' life-course stage. *Soc. Sci*. **13**, 162. (doi:10.3390/socsci13030162)
53. Lee KG, Mennerat A, Lukas D, Dugdale HL, Culina A. 2023 The effect of the COVID-19 pandemic on the gender gap in research productivity within academia. *eLife* **12**, e85427. (doi:10.7554/elife.85427)
54. Squazzoni F, Bravo G, Grimaldo F, García-Costa D, Farjam M, Mehmani B. 2021 Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals. *PLoS One* **16**, e0257919. (doi:10.1371/journal.pone.0257919)
55. Liu F, Holme P, Chiesa M, AlShebli B, Rahwan T. 2023 Gender inequality and self-publication are common among academic editors. *Nat. Hum. Behav*. **7**, 353–364. (doi:10.1038/s41562-022-01498-1)
56. Helmer M, Schottdorf M, Neef A, Battaglia D. 2017 Gender bias in scholarly peer review. *eLife* **6**, e21718. (doi:10.7554/elife.21718)
57. Murray D, Siler K, Larivière V, Chan WM, Collings AM, Raymond J. 2018 Author-reviewer homophily in peer review. *bioRxiv*. (doi:10.1101/400515)
58. Taube J. 2024 Infectious disease dynamics citation analysis tool. See http://bansallab.com/iddcitationanalysis.
59. Taube J. 2025 The extent of gender and race/ethnicity imbalance in infectious disease dynamics research. GitHub. See https://github.com/bansallab/IDD_imbalance.
60. Taube J, Merritt A, Bansal S. 2025 Data for: Twenty years of infectious disease dynamics research: gender and race imbalances in publication and citation practices. Dryad Digital Repository. (doi:10.5061/dryad.djh9w0w98)
61. Taube JC, Merritt A, Bansal S. 2025 Supplementary material from: Twenty years of infectious disease dynamics research: gender and race imbalances in publication and citation practices. Figshare (doi:10.6084/m9.figshare.c.7900913)