Toward Responsible Collection and Use of Demographic Information in Scholarly Publishing

Beryne Odeny

The growing emphasis on demographic data collection among publishers and journals underlines the crucial need for responsible data management in scholarly publishing. As a follow-up to CSE's webinar by Dr Beryne Odeny and Julia Robinson, this synopsis provides insights into dilemmas of, and considerations for, collecting and using demographic information in scholarly publishing. We focus on the value of diversity, equity, inclusion, and accessibility (DEIA) at various levels of scholarly communication.[1] We delve into the importance of collecting demographic data as a tool for achieving DEIA, considering the vulnerabilities of those who share their data and potential harms, and harm mitigation strategies, including acknowledgement of the positionality of those who collect and use the data.[2] Responsible data management should prioritize conscientious collection, analysis, interpretation, and dissemination of demographic data.

The Importance of Diversity in Research

In 2020, the Royal Society of Chemistry (RSC) convened over 50 global publishers to endorse a joint commitment to DEIA in scholarly publishing.[3] As part of this commitment, publishers are actively collecting data on sex, gender, race, and ethnicity to assess DEIA in their editorial and publication processes.[4] There is growing evidence that promoting diversity in research and publishing strengthens the generalizability of scientific conclusions to broader populations.[5] Designing research for widespread

Beryne Odeny, MD, MPH, PhD, Department of Surgery, Washington University, St Louis, MO.

dissemination enhances its accessibility, societal relevance, and scientific merit.[6]

Promoting diversity among editors, peer reviewers, and authors is crucial for broadening perspectives and ensuring fairness and inclusivity in scientific research. In fact, studies have demonstrated that diversified research teams produce more diverse, nuanced, and impactful research. For example, Nielsen and colleagues,[7] showed that teams with diverse backgrounds and perspectives are more likely to generate novel ideas, approach problems creatively, and produce research that has broader societal impacts. Conversely, less diverse editorial boards are less likely to publish diverse and generalizable research.[8] Considering this, publishers have an ethical imperative to uphold equity and justice alongside assessing scientific merit in publication processes. The Belmont Report's justice principle on equitable sharing of burdens and benefits in research should extend to scholarly publishing.[9] Equitable dissemination of research from diverse authors and contexts is essential to upholding this principle.

Demographic Data Is Vital for Addressing the Lack of Diversity

Unfortunately, the current lack of diversity in scholarly publishing poses a threat to equitable and inclusive editorial practices. Demographic data can provide valuable insights into the diversity and inclusivity of editorial boards and authors, as well as the representation of diverse populations in research.[10] Existing research on the composition of editorial boards, authorship groups, and peer reviewer panels has revealed a concerning lack of diversity, particularly underrepresentation of women and researchers from low- and middle-income countries (LMICs).[11-15] Issues like parachute science—where researchers from high-resource settings conduct research in low-resource communities without meaningful community engagement and capacity building—undermine efforts to promote DEIA.[16] To address

CONTINUED

demographic disparities in publishing, it is crucial to monitor unjust publication practices, collect demographic data, and develop robust metrics for evaluating DEIA initiatives' impact on scientific knowledge generation and dissemination.[13,17]

Dilemmas and Potential Harms of Collecting Demographic Data

Although collecting and using demographic data offers valuable insights, publishers must gather this information with caution and consideration of potential harms to respondents sharing their information.[9] Publishers can learn from the experiences of researchers who, when collecting and using demographic data for research purposes, have encountered unintended consequences—largely borne by research participants. To further illustrate this dilemma for researchers, Call and colleagues,[18] aptly explores how the collection of demographic data unintentionally reinforces harmful stereotypes and stigma, and further marginalizes underserved and underrepresented groups, especially within the context of existing societal structures and forces.[19-20] For instance, maintaining sample homogeneity by excluding groups based on their characteristics can exacerbate underrepresentation of minoritized populations and lead to disparities in scientific advances across varying contexts.[21] In light of this, publishers need to anticipate potential harms associated with the collection and use of demographic data, such as highlighting an author's underrepresented status inadvertently, and introducing bias into the peer review process.

Other dilemmas around the use of demographic data for research provide critical lessons for publishers. For instance, although the temptation for researchers to prioritize homogeneity of demographic characteristics is convenient for analytical and interpretation purposes, it can inadvertently hinder DEIA efforts. In some disciplines, heterogeneity may be seen as a nuisance that needs to be controlled for to uphold internal validity.[21] Due to personal preference or expertise, editors and peer reviewers may prioritize manuscripts that conform to conventional research methodologies, which may exclude studies from marginalized communities or use more nuanced approaches like implementation science.

Another potential dilemma and harmful practice in research is the use of demographic variables as proxies for sociocultural factors or experiences driven by structural factors. This can lead to misclassification of individuals, inaccurate inferences, and development of inappropriate or unacceptable interventions for certain contexts.[22] For example, an affluent Black respondent may be subject to irrelevant classification and intervention, whereas a more socioeconomically disadvantaged White individual

may be misclassified and overlooked when dispensing interventions.

While the collection of demographic data from authors and peer reviewers is a relatively recent practice, emerging knowledge from researchers indicates that publishers need to proactively monitor and mitigate potential unintended consequences for respondents.[9] This involves ensuring that demographic data is collected and used ethically, and responsibly, avoiding the perpetuation of stereotypes, stigma and inequities.

Considerations for Demographic Data Collection

Publishers and editors who are recipients and users of demographic data need to be mindful of their positionality as they review and draw conclusions from these data. In research, positionality refers to the way personal background, experiences, and social identity influence how we see and interpret information.[2] In scholarly publishing, it acknowledges that editors and publishers are not neutral observers but bring their own perspectives and biases to their interpretation of data. Those analyzing and interpreting these demographic data can acknowledge their positionality and indicate measures to minimize potential biases. This will help to build trust with respondents sharing their data and avoid the perpetuation of biases.

The standardized RSC demographic questionnaire[1] is not universally relevant and accessible. There are concerns about selection bias in its design and the voices represented given the categories, which are primarily US-centric and tailored more toward high-income settings. The relevance, appropriateness, and acceptability of standardized demographic questionnaires to diverse global contexts has not been fully explored and should be revisited. While relevant to some contexts, the questionnaire does not fully capture the nuances of diverse identities and experiences of authors (such as how early in a career), publishers, and editorial boards in LMICs.

Collecting demographic data is fraught with challenges—some beyond publishers' control—associated with inflexible paradigms for representing the spectrum of demographic diversity. There are concerns that demographic questionnaires constrain individuals to predetermined categories, which undermines efforts to uphold diversity.[23,24] For example, the category "Black/African American" may not accurately reflect the experiences or identities of individuals from Africa who do not identify as being Black nor being an African American but identify as having "brown skin" rather than "black skin." Various approaches to address these narrow classification paradigms have been proposed.[23-25] One approach is to give respondents freedom

CONTINUED

to qualitatively describe themselves and code these data; however, this presents challenges for quantitative analysis and comparison across settings. To address this, questionnaires with closed-ended questions and follow-up questions that provide opportunities for individuals to self-describe their identity can help to mitigate feelings of invalidation and ensure a more comprehensive understanding of diverse identities. The RSC's standardized questionnaire[1] provides this flexibility and has been adopted by several journals, such as the *BMJ*.

Analysis and reporting of demographic data also require special attention and deliberation. Very specific or unique demographic characteristics can risk the confidentiality of respondents if their numbers are small. Conversely, grouping small data cells into larger categories to minimize identification can be harmful if the categories mask characteristics that are understudied and underrepresented, limiting a deeper understanding of minoritized groups. This can have downstream effects on data analyses, potentially failing to expose the real issues underlying disparities and contributing to a vicious cycle of undue stigmatization and stereotyping of individuals.[26-27] To address these potential harms, processes for protecting privacy and confidentiality must be clear and transparent to respondents and users of demographic data. Those who share their demographic data need to be informed about how their data will be stored, analyzed, and used, and how potential harms will be mitigated—akin to consenting processes in a research setting.

Recommendations for Responsible Data Collection and Use

In response to these dilemmas and considerations, we provide 7 theory-driven recommendations founded on Call and colleagues'[18] ethics and social justice framework for collecting and using demographic data in the context of scholarly publishing.

1. Publishers can clearly define their goals for collecting demographic data, ensuring transparency and implementing robust privacy measures. Part of this means disclosing their positionality and implementing strategies to mitigate biases when using demographic data in their custody.

2. Publishers can engage diverse scientific communities in revising or refining data collection questionnaires to be contextually relevant, in addition to developing strategies for diversifying authors, reviewers, and editorial boards.

3. Editors can consider blinded peer review processes where applicable, to protect marginalized groups of authors, and strategically select reviewers who are

familiar with the research context. In some cases, as an equity measure, financial compensation may be considered to enhance the engagement of reviewers from underrepresented groups who possess specialized knowledge of the research context.

4. Publishers need to give greater visibility to the work of underrepresented researchers (e.g., special issues and collections) as a strategy for actively promoting inclusivity in curated journal content.

5. Publishers should actively consult with, and seek feedback, from underrepresented communities about how to interpret and use demographic data to benefit those at highest risk of exclusion.

6. Publishers should institute and communicate mechanisms to eliminate or mitigate the anticipated harms that can arise from demographic data collection and use. This will further demonstrate their commitment to fostering trust with respondents and readers.

7. Publishers can commit to share demographic reports for accountability to the scientific community. Publishers should consider reporting de-identified data summaries and outlining how the data will inform efforts to improve author and reviewer diversity and editorial composition. Responsible data collection should culminate in actionable strategies to reduce inequities, as merely collecting data without a clear plan for utilization falls short of responsible data management.

A Call for Collective Action

This synopsis is applicable beyond publishing and can be applied to funding agencies, as the effects of bias and inequity reverberate throughout the research lifecycle, from funding to publication and implementation. While the success of DEIA efforts hinges on evaluating and acting upon demographic data, more is needed to safeguard those sharing their data and the data itself. The approach by which demographic data are collected, handled, and reported must adhere to the Belmont Report to avoid propagating biases and disparities in scholarly publishing. A justice-oriented approach to demographic data collection and use within scholarly publishing, research, and funding agencies can significantly advance DEIA efforts by prioritizing ethics, transparency, and accountability. This is a call for thoughtful collection, use, and dissemination of demographic data to create a more diverse and equitable scientific ecosystem.

References and Links

1. https://www.rsc.org/policy-evidence-campaigns/inclusion-diversity/joint-commitment-for-action-inclusion-and-diversity-in-publishing/diversity-data-collection-in-scholarly-publishing/

CONTINUED

2. Jacobs K. Positionality and reflexivity statement. Work. 2023;76:895–896. https://doi.org/10.3233/wor-236010.

3. https://www.rsc.org/policy-evidence-campaigns/inclusion-diversity/joint-commitment-for-action-inclusion-and-diversity-in-publishing/

4. Lewis EF, Beaty C, Boltze J, Breathett K, Clair WK, de las Fuentes L, Essien UR, Goodell H, Hinson HE, Kershaw KN, et al. Creation of the American Heart Association Journals' Equity, Diversity, and Inclusion Editorial Board: the next step to achieving the 2024 impact goal. Circulation. 2022;146:721–723. https://doi.org/10.1161/circulationaha.122.061450.

5. Eldridge L, Ginsburg O, Gopal S, Odeny TA. Equity, diversity, and inclusion in global cancer clinical trial authorship. Cancer. 2023;129;2771–2772. https://doi.org/10.1002/cncr.34940.

6. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CH. The academic, economic and societal impacts of open access: an evidence-based review. F1000Research. 2016;5:632. https://doi.org/10.12688/f1000research.8460.3.

7. Nielsen MW, Alegria S, Börjeson L, Etzkowitz H, Falk-Krzesinski HJ. Gender diversity leads to better science. Proc Natl Acad Sci U S A. 2017;114:1740–1742. https://doi.org/10.1073/pnas.1700616114.

8. Roberts SO, Bareket-Shavit C, Dollins FA, Goldie PD, Mortenson E. Racial inequality in psychological research: trends of the past and recommendations for the future. Perspect Psychol Sci. 2020;15:1295–1309. https://doi.org/10.1177/1745691620927709.

9. https://irb.wisc.edu/regulatory-information/belmont-report/

10. https://www.councilscienceeditors.org/2-7-diversity--equity--and-inclusion-in-scholarly-publishing

11. *Manan MR, Nawaz I, Rahman S, Razzaq A, Zafar F, Qazi A, Liblik K. Diversity, equity, and inclusion on editorial boards of global health journals. Asian Bioeth Rev. https://doi.org/10.1007/s41649-023-00243-8.*

12. Skrivankova VW, Hossmann S, Cornell M, Ballif M, Dupont C, Huwa J, Seintaridis K, Kalua T, Wandeler G, Kassanjee R, et al. Authorship inequalities in global health research: the IeDEA Southern Africa collaboration. BMJ Glob Health. 2023;8:e013316–e013316. https://doi.org/10.1136/bmjgh-2023-013316.

13. Shah SGS, Dam R, Milano MJ, Edmunds LD, Henderson LR, Hartley CR, Coxall O, Ovseiko PV, Buchan AM, Kiparoglou V. Gender parity in scientific authorship in a National Institute for Health Research Biomedical Research Centre: a bibliometric analysis. BMJ Open. 2021;11:e037935. https://doi.org/10.1136/bmjopen-2020-037935.

14. Gallifant J, Zhang J, Whebell S, Quion JM, Escobar B, Gichoya JW, Herrera K, Jina R, Chidambaram S, Mehndiratta A, et al. A new tool for evaluating health equity in academic journals; the Diversity Factor. PLOS Glob Public Health. 2023;3:e0002252–e0002252. https://doi.org/10.1371/journal.pgph.0002252.

15. Hedt-Gauthier BL, Jeufack HM, Neufeld NH, Alem A, Sauer S, Odhiambo J, Boum Y, Shuchman M, Volmink J. Stuck in the middle: a systematic review of authorship in collaborative health research in Africa, 2014–2016. BMJ Glob Health. 2019;4:e001853. https://doi.org/10.1136/bmjgh-2019-001853.

16. Odeny B, Bosurgi R. Time to end parachute science. PLOS Med. 2022;19:e1004099. https://doi.org/10.1371/journal.pmed.1004099.

17. Jack L Jr, Olson PJ, Baskin PK, Iwuchukwu OF. Building diversity, equity, inclusion, and accessibility capacity: resources to promote best practices among professionals in scholarly publishing. Prev Chronic Dis. 2023;20:E105. https://doi.org/10.5888/pcd20.230332.

18. Call CC, Eckstrand KL, Kasparek SW, Boness CL, Blatt L, Jamal-Orozco N, Novacek DM, Foti D. An ethics and social-justice approach to collecting and using demographic data for psychological researchers. Perspect Psychol Sci. 2022;18:174569162211373. https://doi.org/10.1177/17456916221137350.

19. Fernandez T, Godwin A, Doyle J, Verdin D, Boone H, Kirn A, Benson L, Potvin G. More comprehensive and inclusive approaches to demographic data collection. Paper presented at 2016 ASEE Annual Conference & Exposition, New Orleans, LA, June 2016. https://doi.org/10.18260/p.25751.

20. Castillo W, Gillborn D. How to "QuantCrit:" practices and questions for education data researchers and users. EdWorkingPaper: 22-546. Annenberg Institute at Brown University; 2022. https://doi.org/10.26300/v5kh-dd65.

21. Lewis NA. What counts as good science? How the battle for methodological legitimacy affects public psychology. Am Psychol. 2021;76:1323–1333. https://doi.org/10.1037/amp0000870.

22. Byrd DA, Rivera-Mindt MG. Neuropsychology's race problem does not begin or end with demographically adjusted norms. Nat Rev Neurol. 2022;18:125–126. https://doi.org/10.1038/s41582-021-00607-4.

23. Strunk KK, Hoover PD. Quantitative methods for social justice and equity: theoretical and practical considerations. In: Strunk KK, Locke LA, editors. Research methods for social justice and equity in education. Palgrave Macmillan; 2019, p. 191–201. https://doi.org/10.1007/978-3-030-05900-2.

24. Flanagin A, Frey T, Christiansen SL, Bauchner H. The reporting of race and ethnicity in medical and science journals: comments invited. JAMA. 2021;325:1049–1052. https://doi.org/10.1001/jama.2021.2104.

25. Hughes JL, Camden AA, Yangchen T. Rethinking and updating demographic questions: guidance to improve descriptions of research samples. Psi Chi J Psychol Res. 2016;21:138–151. https://doi.org/10.24839/2164-8204.jn21.3.138.

26. Ross PT, Hart-Johnson T, Santen SA. Zaidi NLB. Considerations for using race and ethnicity as quantitative variables in medical education research. Perspect Med Educ. 2020;9:318–323. https://doi.org/10.1007/s40037-020-00602-3.

27. Noroña-Zhou A, Bush NR. Considerations regarding the responsible use of categorical race/ethnicity within health research. PsyArXive. 2021. https://doi.org/10.31234/osf.io/kfa57.