BMJ Open

Reporting of data on participant ethnicity and socioeconomic status in high-impact medical journals: a targeted literature review

Sara C Buttery,[1,2] Keir E J Philip [ID],[1,2] Saeed M Alghamdi,[1,2]
Parris J Williams [ID],[1,2] Jennifer K Quint [ID],[1,2] Nicholas S Hopkinson[1,2]

ABSTRACT

Objectives To assess the frequency of reporting of ethnicity (or 'race') and socioeconomic status (SES) indicators in high-impact journals.

Design Targeted literature review.

Data sources The 10 highest ranked general medical journals using Google scholar h5 index.

Eligibility criteria Inclusion criteria were, human research, reporting participant level data. Exclusion criteria were non-research article, animal/other non-human participant/subject or no participant characteristics reported.

Data extraction and synthesis Working backwards from 19 April 2021 in each journal, two independent reviewers selected the 10 most recent articles meeting inclusion/exclusion criteria, to create a sample of 100 articles. Data on the frequency of reporting of ethnicity (or 'race') and SES indicators were extracted and presented using descriptive statistics.

Results Of 100 research articles included, 35 reported ethnicity and 13 SES. By contrast, 99 reported age, and 97 reported sex or gender. Among the articles not reporting ethnicity, only 3 (5%) highlighted this as a limitation, and only 6 (7%) where SES data were missing. Median number of articles reporting ethnicity per journal was 2.5/10 (range 0 to 9). Only two journals explicitly requested reporting of ethnicity (or race), and one requested SES.

Conclusions The majority of research published in high-impact medical journals does not include data on the ethnicity and SES of participants, and this omission is rarely acknowledged as a limitation. This situation persists despite the well-established importance of this issue and International Committee of Medical Journal Editors recommendations to include relevant demographic variables to ensure representative samples. Standardised explicit minimum standards are required.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study included recent studies from a range of the highest impact general medical journals.

⇒ Different inclusion/exclusion criteria for articles could be justifiably used, which may have produced different results.

⇒ We identified high-impact journals using the google scholar h5 index, however various other equally valid impact metrics exist, which could change the journals considered.

⇒ Our analysis focused on *if* ethnicity and/or race was reported, but not *how* they are reported which is an important and related area for discussion and research to that covered in this study.

INTRODUCTION

Information about the ethnicity and socioeconomic status (SES) of participants in clinical research is needed for the interpretation, generalisability and pooling of data as well as to inform discussion around health inequalities. The relevance of ethnicity and SES to health and biomedical research is well established but has been emphasised by the COVID-19 pandemic, during which specific ethnic groups and poorer individuals have been disproportionately affected.[1] The causal pathways driving health disparities are complex and multifactorial, however under-reporting of participant characteristics has been identified as a potential contributory factor.[2–4]

The International Committee of Medical Journal Editors recommendations,[5] and some journal instructions to authors promote inclusion of these data.[6 7] Previous studies have identified that reporting is frequently incomplete with limited progress made over the last three decades.[8–13] Recent years have seen an increased focus on ethnicity and SES in medicine, however there is a lack of research as to whether this has resulted in better reporting.

To evaluate the current situation in this area, we assessed the frequency of reporting of ethnicity (or 'race') and SES indicators in a sample of research articles published in high impact general medical journals in Spring 2021.

METHODS

We identified the 10 highest ranked journals as per Google scholar 'Health and Medical (general)' category up to April 2021. At the time of data collection, these were *The New England Journal of Medicine (NEJM), The Lancet, the Journal of the American Medical Association,*[7] *Proceedings of the National Academy of Sciences of the United States of America (PNAS), Nature Medicine, Public Library of Science One (PLOS One), The British Medical Journal (BMJ), Cochrane, Cell Metabolism and Science Translational Medicine. PNAS* and *PLOS One* include a wide range of subject areas therefore the subsections 'Biological Sciences, Medical Science' and 'Clinical Medicine' were used respectively. From each of these 10 journals, using the journals own websites, we worked backwards from 19 April 2021, selecting the 10 most recent journal articles that met inclusion/exclusion criteria. Inclusion criteria were as follows: research articles, reporting participant level data. Articles were excluded if they were not research (eg, editorial, news, images, etc.), animal/other non-human participant/subject or no participant characteristics reported. Laboratory studies using human-derived tissues or cells were included if donor information was provided. Journal reporting guidance and requirements were also assessed by evaluating author guidelines, websites and contacting the respective editorial/publishing teams. Data were collected on which participant level characteristics were reported and how. Data were also collected on if the absence of reporting these variables was noted as a limitation. The journals' accessible policies and guidance on reporting these variables was also reviewed. Data collection and analysis was conducted by SCB, KEJP, SMA and PJW. All journals were reviewed and articles selected by at least two researchers independently, who then came together to discuss any inconsistencies with a third researcher.

Ethnicity and race are related yet different constructs and arguably the latter term should be abandoned.[14] However, given the frequent lack of standardisation in the literature and that the terms are in practice often used interchangeably we accepted the use of either term. For the purpose of this study, ethnicity (or race) was defined as variables explicitly stated by the authors as 'ethnicity', 'ethnic group' or 'race', 'racial group'. Similarly, regarding reporting of SES indicators, various often inconsistent methods are used, therefore we opted to assess both direct measures such as the Index of Multiple Deprivation, but also measures from which SES could be inferred such as educational attainment and job role. The focus being if, rather than how, such measures are reported. Variables were considered to be indicators of SES if they were explicitly stated as being included for this purpose in the studies reporting them, or if not explicitly stated in the study itself, variables that might be considered SES indicators were discussed between researchers and included or excluded based on consensus opinion. Given the potential degree of subjectivity related to this approach, we have provided the specific terms used by included studies in the results section below. The agreed
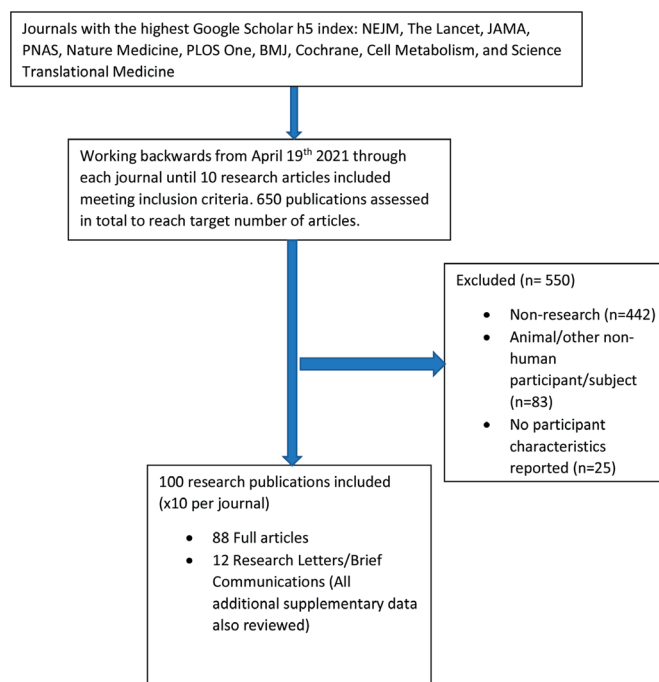
approach was to take a more inclusive approach, so that if these variables were found to be infrequently reported, such findings would not be dismissed as relating to overly stringent inclusion criteria.

Patient and public involvement
None.

RESULTS
650 publications were assessed to identify 100 meeting inclusion criteria (see figure 1 and online supplemental tables 1–3). Of 100 research articles included, 35 reported ethnicity (or race) and 13 reported SES. By contrast, 99 reported age, and 97 reported sex or gender (table 1).

Among the articles not reporting ethnicity, only 3 (5%) highlighted this as a limitation, and only 6 (7%) highlighted where SES data were missing. Median number of articles reporting ethnicity per journal was 2.5/10 (range 0/10 (*PLOS One*) to 9/10.[7] Only two journals explicitly requested reporting of participant ethnicity (or race), and one requested SES. Types of research included—interventional studies (n=30), cohort studies (n=35), case–control studies (n=3), systematic reviews and meta-analyses (n=16), epidemiological and surveys (n=3) and other (n=13). Twenty of the 100 were laboratory studies (either observational or involving interventional manipulation of samples) using human samples, of which four reported ethnicities of sample donors (of others, none mentioned as a limitation), and none reported SES.

Among the 24 papers describing clinical trials, 50% reported ethnicity, with none highlighting the absence of these data as a limitation; 12.5% of trials reported an

Journals with the highest Google Scholar h5 index: NEJM, The Lancet, JAMA, PNAS, Nature Medicine, PLOS One, BMJ, Cochrane, Cell Metabolism, and Science Translational Medicine

Working backwards from April 19th 2021 through each journal until 10 research articles included meeting inclusion criteria. 650 publications assessed in total to reach target number of articles.

Excluded (n= 550)
- Non-research (n=442)
- Animal/other non-human participant/subject (n=83)
- No participant characteristics reported (n=25)

100 research publications included (x10 per journal)
- 88 Full articles
- 12 Research Letters/Brief Communications (All additional supplementary data also reviewed)

Figure 1 Flow diagram of study inclusion/exclusion.

Table 1 Reporting of ethnicity and/or race, and socioeconomic status indicators in research articles

Report participant level characteristics	N 100	Additional notes
Report ethnicity and/or race	35/100 report 65 not report	Range per journal: JAMA 9/10, with clear guidance that this information is expected.
Noted in limitations	62 of the 65 do not state this as a limitation 3 do highlight this as a limitation.	Some studies identify race and ethnicity as being relevant to the research focus, yet did not provide relevant data on their study participants or highlight this a limitation of their study, for example, ▶ *in the case of DNA-based mutation testing, poor sensitivity in detecting mutations in infants from ethnic and racial minority groups (DOI: 10.1126/ scitranslmed.abd8109)* ▶ *peripheral oxygen saturation can substantially differ from the Sao2 under certain conditions and may be less accurate in Black patients than in White patients (DOI: 10.1056/NEJMoa2032510)*
Report socioeconomic status indicator	13/100 report at a measure of SES (six direct measure, for example, Index of Multiple Deprivation, Poverty income ratio; seven measures from which SES can be inferred, for example, educational attainment, job role) 87/100 did not report any indication of SES	
Noted in limitations	6/87 identified this as a limitation	
Age reported	99/100	
Sex or gender reported	97/100	

Percentages not given as most results have 100 as the denominator.
SES, socioeconomic study.

indicator of SES, with one of the 21 not reporting SES highlighting this absence as a limitation.

Of note, two of the research articles included in our sample identified ethnicity as being relevant to their research topic, yet did not provide relevant data on their study participants or highlight the lack of this data as a limitation of their study *in the case of DNA-based mutation testing, poor sensitivity in detecting mutations in infants from ethnic and racial minority groups,* and *peripheral oxygen saturation can substantially differ from the SaO$_2$ under certain conditions and may be less accurate in Black patients than in White patients.*[15]

DISCUSSION

The majority of research published in high-impact medical journals does not include data on the ethnicity and SES of participants, and this omission is rarely acknowledged as a limitation. This finding echoes related historical research,[8–13] but its persistence is of concern and is surprising given current awareness of such issues.[16 17]

These findings have important implications for the interpretation and application of research findings, both within academia and beyond, with the ongoing omission no longer justifiable as simple oversight. As highlighted by Baker *et al*,[18] in relation to data relating to LGBTQI+ communities, but equally relevant here, *Data are fundamentally political: decisions about which data are collected and which are overlooked both reflect and shape policy and programme priorities.*

Our results could have multiple contributory factors. For some research including secondary data analyses, ethnicity and SES data may not have been available to the researchers, but given the lack of explanation, it remains unclear if these data were unavailable, or available but not included in publications. The low level of reporting in controlled clinical trials suggests issues beyond unavailability of data, as in these studies, such data would be simple to collect. Additionally, given research successfully reporting these data, the justification for these omissions remains unexplained. Non-reporting of ethnicity (or race) and SES data may also result from explicit or implicit racism, or other forms of discrimination such as that based on SES, which could include failing to appreciate the relevance of these factors to the generalisability of findings.

The increased frequency of reporting ethnicity, compared with SES, may indicate differences between the perceived relevance of these variables. This would be in keeping journal author guidelines and ICMJE recommendations that encourage the inclusion of relevant

demographic variables to ensure representative samples,[5] more often explicitly stating race and/or ethnicity, than SES. The relevance of these factors may not have been apparent to authors and editorial teams, however ICMJE Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly work in Medical Journals[5] states *Because the relevance of such variables as age, sex or ethnicity is not always known at the time of study design, researchers should aim for inclusion of representative populations into all study types and at a minimum provide descriptive data for these and other relevant demographic variables.* Of note, not all of the journals in our sample state that they follow the ICMJE recommendations.[19] However, whether or not the journal states they follow guidance or not, this has no impact on the relevance of these data and the importance of reporting them. Additionally, Maduka *et al*[20] found no difference between journals stating they follow ICMJE recommendations, and those that do not, in the frequency of reporting race and ethnicity in a sample of surgical research publications in 2019.

Certain considerations and limitations require highlighting. First, different approaches to selecting research papers may alter findings. Second, we identified high-impact journals using the google scholar h5 index but acknowledge various other equally valid methods exist. Third, our analysis focused on if ethnicity and/or race was reported, but we acknowledge that these are not synonymous terms. In addition to *if* these variables are reported, *how* they are reported is also an important area for discussion and research. The choice to analyse 100 papers was somewhat arbitrary. We wanted to include an adequate number of articles from the selected journals to provide a representative sample of their original research papers. Furthermore, given the substantial differences in the number of original research papers published between journals, keeping to 10 per journal ensured all included papers were published within a 4-month window. If we had included 100 papers per journal, the sample from some journals might be 2 months, while others nearer 2 years, which could complicate interpretation given the potential for changing levels of reporting over time. The widespread omissions identified by this research suggests a structural problem. Indeed, we the authors have published research which would have met the inclusion criteria and failed to report these specific characteristics. Our intention is to highlight an issue and suggest approaches to address it.

Given that inadequate reporting persists despite research highlighting the issue, author and ICMJE recommendations, and the current sociopolitical climate, there is a clear need for more explicit requirements that are adhered to in practice. This is likely best achieved if steps are integrated into each stage of the research process, from protocol to publication. For example, Fain *et al*[21] compared reporting of race and ethnicity on ClinicalTrials.gov before and after the requirement to report these data (if collected), was introduced, finding that this was associated with an increase from 42% to 92%. Similar explicit requirements could be taken in Enhancing the QUAlity and Transparency Of health Research (EQUATOR) guidelines,[22] and research ethic applications. From our sample, the journal *JAMA* had the most explicit guidance for reporting race and ethnicity, and this variable was reported in 9/10 of the articles we reviewed. Of note, from 2022, the *New England Journal of Medicine* will be requiring authors of research articles to provide data on the representativeness of the sample including race or ethnic group,[23] though it is unclear if SES indicators will also be required. Much of the recent literature appears to focus on ethnicity reporting, likely due to the COVID-19 pandemic exposing its disproportionate effects on some ethnic groups.[24] One recent publication in *Nature medicine*[24] suggested that it would require changes at policy level as well as engaging with professionals, patients and the public to communicate the importance of this issue in understanding inequalities. Barriers suggested include problems collecting ethnicity data, whether this be reported by a healthcare professional or self-reported, and in defining ethnic groups where categorisation is inconsistent.[24 25] This is reflected in the diverse terms used to report ethnicity in the papers we reviewed (online supplemental table 3). Future research would be useful investigating changing in reporting overtime, especially in relation to specific actions taken to improve this issue, which could inform research reporting guidelines.

CONCLUSION

The reporting of ethnicity and socioeconomic status in high-impact medical research remains poor, despite a consensus on its importance. Omission of these participant characteristics limits the interpretation, generalisability and pooling of data that are required to facilitated informed discussion around health inequalities. Guidance and encouragement have so far proven insufficient to change practice in this area. Standardised, explicit, minimum standards are required.

Twitter Keir E J Philip @keirphilip, Parris J Williams @ParrisWilliams1 and Nicholas S Hopkinson @COPDdoc

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. All data used in this study are publicly available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Keir E J Philip http://orcid.org/0000-0001-9614-3580
Parris J Williams http://orcid.org/0000-0001-8027-1879
Jennifer K Quint http://orcid.org/0000-0003-0149-4869

REFERENCES

1. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.10.1038/s41586-020-2521-4
2. Chastain DB, Osae SP, Henao-Martínez AF, et al. Racial Disproportionality in Covid clinical trials. *N Engl J Med Overseas Ed* 2020;383:e59.
3. Krieger N, Waterman PD, Chen JT, et al. Missing again: US racial and ethnic data for COVID-19 vaccination. *The Lancet* 2021;397:1259–60.
4. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA* 2020;323:2466–7.
5. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals 2019.
6. The Lancet. Information for authors, 2021. Available: https://marlin-prod.literatumonline.com/pb-assets/Lancet/authors/tl-info-for-authors.pdf [Accessed 02 Sep 2021].
7. JAMA. Instructions for authors, 2021. Available: https://jamanetwork.com/journals/jama/pages/instructions-for-authors [Accessed 02 Sep 2021].
8. Geller SE, Koch A, Pellettieri B, et al. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? *J Womens Health* 2011;20:315–20.
9. Brahan D, Bauchner H. Changes in reporting of race/ethnicity, socioeconomic status, gender, and age over 10 years. *Pediatrics* 2005;115:e163–6.
10. Kanakamedala P, Haga SB. Characterization of clinical study populations by race and ethnicity in biomedical literature. *Ethn Dis* 2012;22:96–101.
11. Bokor-Billmann T, Langan EA, Billmann F. The reporting of race and/or ethnicity in the medical literature: a retrospective bibliometric analysis confirmed room for improvement. *J Clin Epidemiol* 2020;119:1–6.
12. Lee SJ, Kavanaugh A. A need for greater reporting of socioeconomic status and race in clinical trials. *Ann Rheum Dis* 2004;63:1700–1.
13. Ma IWY, Khan NA, Kang A, et al. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *J Clin Epidemiol* 2007;60:572–8.
14. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the use of race in medical research. *JAMA* 2021;325:623–4.
15. Schjørring OL, Klitgaard TL, Perner A, et al. Lower or higher oxygenation targets for acute hypoxemic respiratory failure. *N Engl J Med* 2021;384:1301–11.
16. Flanagin A, Frey T, Christiansen SL, et al. The reporting of race and ethnicity in medical and science journals. *JAMA* 2021;325:1049–52.
17. The New England Journal of Medicine. Race and medicine 2021.
18. Baker KE, Streed CG, Durso LE. Ensuring that LGBTQI+ people count — collecting data on sexual orientation, gender identity, and intersex status. *N Engl J Med* 2021;384:1184–6.
19. International Committee of Medical Journal Editors. Journals stating that they follow the ICMJE recommendations, 2021. Available: http://www.icmje.org/journals-following-the-icmje-recommendations/ [Accessed 09 Sep 2021].
20. Maduka RC, Broderick M, White EM, et al. The reporting of race and ethnicity in surgery literature. *JAMA Surg* 2021;156:1036.
21. Fain KM, Nelson JT, Tse T, et al. Race and ethnicity reporting for clinical trials in ClinicalTrials.gov and publications. *Contemp Clin Trials* 2021;101:106237.
22. EQUATOR network. reporting guidelines 2021, 2021. Available: https://www.equator-network.org/reporting-guidelines/ [Accessed 04/10/2021].
23. , Rubin E, Editors. Striving for diversity in research studies. *N Engl J Med* 2021;385:1429–30.
24. Routen A, Akbari A, Banerjee A, et al. Strategies to record and use ethnicity information in routine health data. *Nat Med* 2022;28:1338–42.
25. Khunti K, Routen A, Banerjee A, et al. The need for improved collection and coding of ethnicity in health research. *J Public Health* 2021;43:e270–2.