The wrong word for the job? The ethics of collecting data on 'race' in academic publishing

John McMillan ![ORCID],[1] Brian D Earp,[2] Wing May Kong,[3] Mehrunisha Suleman ![ORCID],[4] Arianne Shahvisi ![ORCID] [5]

Socially responsible publishers, such as the BMJ Publishing Group, have demonstrated a commitment to health equity and working towards rectifying the structural racism that exists both in healthcare and in medical publishing.[1] The commitment of academic publishers to collecting information relevant to promoting equity and diversity is important and commendable where it leads to that result.[2] However, collecting sensitive demographic data is not a morally neutral activity. Rather, it carries with it both known and potential risks. Among these are issues around privacy or data misuse, as well as more basic concerns about how, when or why people should be categorised in certain ways and/or prompted to conceive of themselves or their identities in certain terms.[3] If such data are to be collected, therefore, their effectiveness in achieving the stated ends must have a sufficiently compelling evidence base so as to justify the various risks involved. And where possible, these risks must also be identified and minimised. As Varcoe *et al*[4] argue,

> While most leaders and healthcare workers and some patients [in their study] envisioned potential benefits associated with *having* ethnicity data, these benefits were seen as largely contingent upon action being taken to [actually] ameliorate inequities. Overwhelmingly, however, leaders from ethno-cultural communities and patients of diverse identities anticipated potential harm arising both from having ethnicity data and the process of collection. The analysis illustrates that in today's sociopolitical context, collecting ethnicity data in clinical contexts may

engender considerable harm, particularly for racialized, vulnerable patients. (p1569)

Varcoe *et al* refer to 'ethnicity' data, which as they note is an ambiguous concept, potentially encompassing such diverse notions as ancestry, language, religion or culture. It is also a term that is in some contexts—for example, the USA—often used interchangeably with a different, more highly charged term: 'race'.

Data on race (or ethnicity when used as a synonym or euphemism) are often collected with the clearly worthy goals of documenting and attempting to ameliorate *racism*. In other words, although racism is often correlated with and related to other forms of invidious discrimination (eg, discrimination based on language, immigration status or religion), it is not reducible to those other constructs.

Racism itself has been conceptualised, operationalised and measured in different ways depending on the field of study and the practical or theoretical aims of undertaking such an (inevitably value-laden) inquiry.[5] On one conception, it involves the unjust differential treatment of people who are differently 'racialised' (that is, socially regarded or perceived to be of different recent ancestry, usually based on stereotypical features). This mistreatment may occur at an individual level, as with overt acts of racism, or at a more structural or systems level, due to historical or ongoing practices, laws or policies that have long-lasting downstream effects by way of their cultural and institutional implications.[6]

In order to deal with such a serious moral and political problem as racism, the precise nature, manifestation and extent of the problem must be understood. This is often pursued—appropriately and necessarily—through some form of data collection and analysis. However, the type of data, the manner in which it is collected and the uses to which it is put can be more or less suitable or warranted.

To address racism in medical publishing, it might seem obvious that specific data

on stakeholders' (eg, authors' or peer reviewers') racialised identities would need to be routinely collected: for example, to establish whether and to what extent there are potentially worrisome differences between racialised groups in certain areas (eg, relative rates of article submission vs acceptance, membership on editorial boards). The causes of these differences could then be critically investigated and, where relevant and appropriate, evidence-based remedies developed and pursued.

However, in this editorial, we suggest that simply asking how a person identifies in terms of race—including by asking them to choose among a preset list of potentially culturally or historically relative 'racial' identities—can be problematic. This is especially true in the context of an increasingly globalised scientific and research industry that nevertheless remains dominated by a select group of powerful publishers in the Global North. The term race, we suggest, as well as particular purported racial identities, can have quite different and even divisive connotations across diverse cultural contexts.

Before proceeding with our argument, we wish to reiterate that there are strong *prima facie* reasons in favour of collecting information that can be used to promote equity, diversity and inclusiveness, not only in academic publishing but in healthcare more generally. As Varcoe *et al* note, while it has been routine for medical organisations and researchers to collect self-reported data on people's racialised identities in the USA or UK, it might not be routine in all countries.[4] Nevertheless, such data can be invaluable for identifying and rectifying health inequities. In particular, if collected accurately, respectfully and in a culturally sensitive manner, such data may help to flag possible inequities that are not primarily due to other, often correlated issues, such as socioeconomic status and/or various aspects of ethnicity, but rather to racism (as defined above) as such.[7]

In a paper published in the *JME*, Schmidt, Roberts and Eneanya explore the ethics of ventilator allocation in intensive care units (ICU) during the COVID-19 pandemic.[8] They note how 'colourblind' ICU triage protocols may systematically disadvantage people racialised as black in the USA, and they suggest a number of ways of correcting this health inequity that has deep historical roots. Similarly, Bruce and Tallman[7] argued that triage protocol adjustments based only on socioeconomic factors may be 'problematic because racial minorities suffering from [racism-related] health disparities do not always live in

[1]Bioethics Centre, University of Otago, Dunedin, New Zealand
[2]Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK
[3]Chair of Trustees, Institute of Medical Ethics, London, UK
[4]Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[5]Ethics, Brighton and Sussex Medical School, Brighton, UK

Correspondence to Professor John McMillan, Bioethics Centre, University of Otago, Dunedin, 9016, New Zealand; john.r.mcmillan68@gmail.com

disadvantaged communities' (p209). Or as Varma *et al* have stated:

> … socioeconomic variables are not [adequate] proxies for the ill effects of racism on health. Research suggests that racial disparities in health persist even after controlling for socioeconomic status. This is because living in a society that assigns value based on the social interpretation of how one looks (which is what we call 'race') results in differential opportunities, exposures, resources and risks by race and ethnicity.[9]

Such observations do suggest a need for collecting and appropriately analysing data on the variable impact of the pandemic—or any other health threat—on people from different racialised communities within a given sociohistorical context (ie, not just on people currently facing different levels of deprivation). But it does not tell us exactly how or when such data should be collected, analysed or used. For example, asking people as they enter the ICU to fill out a form that requires them to self-identify as one of a limited number of 'races' would likely not be among the most appropriate methods.

At the same time, it should be noted that the collection and analysis of data on race (or even ethnicity) are not always relevant or helpful for addressing health inequities. In a recently published commentary, for instance, Saylor and Martschenko[10] criticise the tendency to use race or ethnicity when discussing diagnostic equity in genomics:

> …the identification of patients for genomic variant reclassification using socially constructed race or ethnicity runs the risk of perpetuating the harmful conflation of race and genetic ancestry. Race is a socially constructed idea tied to concepts of inferiority and superiority. When the term was first applied to human populations, it was used to argue that there exist inborn biological differences between groups of humans that justified the social order. (p821)

For this reason, they suggest that for the sake of both diagnostic accuracy, and what they call 'diagnostic equity', we should move towards reclassifying genomic variants via genetic similarity—not race or ethnicity. More broadly, as Suleman and Qureshi[11] observe in a recent *JME* editorial:

> …racial and ethnic categories as data collection tools carry inherent imprecision, potentially failing to accurately encapsulate an individual's identity or cultural heritage. Consequently, this imprecision can engender misclassification, underreporting, or oversimplification, thereby laying the foundation for research findings that may lack precision and accuracy… here exists a lurking potential for racial and ethnic data to inadvertently bolster stereotypes, thereby fostering stigmatisation and discrimination. (p725)

Needless to say, if a data collection tool ends up reinforcing stereotypes and increasing stigmatisation and discrimination, it will have produced a serious harm. Given that these tools usually aim at addressing health inequities, if they cause harm, that undermines their rationale.

The devil is in the details. Recently, a number of academic publishers have joined forces and created a standardised data collection tool. The Royal Society for Chemistry (RSC) and Elsevier are leading an initiative that at least 54 publishers have agreed to join. The RSC's website hosts the *Joint Commitment for Action on Inclusion and Diversity in Publishing* group, which has produced a demographic questionnaire that academic publishers can send to editors, authors and reviewers so as to collect information relevant to equity and diversity.

The questionnaire includes items on gender (itself a fraught construct we will not be able to analyse here), as well as ethnic origins or ancestry, the latter of which may or may not be appropriate or useful depending on how these data are ultimately analysed and what is done with the results. However, the questionnaire also includes items that ask directly about race which, we suggest, may have certain drawbacks or unintended consequences in some populations.

The questions on race are as follows:

How would you identify yourself in terms of race?

Please select ALL the groups that apply to you:

- ▶ Asian or Pacific Islander
- ▶ Black
- ▶ Hispanic or Latino/a/x
- ▶ Indigenous (eg, North American Indian Navajo, South American Indian Quechua, Aboriginal or Torres Strait Islander)
- ▶ Middle Eastern or North African
- ▶ White
- ▶ Self describe* *[open text box]*
- ▶ Prefer not to disclose

In focussing on how respondents *identify*, the questionnaire attempts to avoid implying that these are biological or essentialist concepts of race. That is to the credit of the survey designers. Nevertheless, because of the historical uses and abuses of the term race and its associated conceptual baggage—as well as its different connotations in different cultures and groups—we suggest that this is a questionable way of framing data collection in this context.

Importantly, a great many of the publishers who make up the Joint Commitment are, like the BMJ, focused on publishing scientific and medical research. Given the prestige of these publishers, as well as the broader epistemic authority that is extended to science and medicine, their use of the term race might be taken as endorsement of uncritical uses of the term in other contexts, or might entrench the idea that the term has a straightforward biological reality. There is therefore particular reason for care in how the Commitment is worded.

The Joint Commitment drew on the expertise of Professor Ann Morning of New York University, who conducts research on the use—and misuse—of racial and ethnic classifications in censuses. When writing in response to a critique of her account in *An Ugly Word: Rethinking Race in Italy and the United States*,[12] she says:

> The issue of race, prominently featured in the title of our book, remains a subject worthy of investigation, if only because it has been institutionalized in myriad ways across the globe as a sociopolitical category of considerable significance, and as such, it merits study as well as being the object of political action. If however it is a "floating signifier" [as has been argued] to which we want to apply a satisfactory degree of analytical precision, we must employ more clearly defined categories, whether or not this approach ultimately reaffirms the centrality of the hovering, metamorphic everyday concept of "race".[13] (p3)

Race continues to be used in many countries, although often in different and inconsistent ways. It is therefore ripe for sociological analysis, and Professor Morning's book is an interesting critique of the assumption that race tends to be more 'biological' in the USA and more 'cultural' in Europe. However, as the title of her book suggests, race is in many parts of the world an 'ugly' concept and one that, depending on its use, is likely to cause justified offence to some people. For example, this may be the case in countries such as New Zealand or the UK where the concept of ethnicity is more commonly used to pick out what have been called racialised identities. Such countries do not have a tradition of surveying people

about the race they identify with, not least because race is often seen in these contexts as a form of external classification, while ethnicity is associated with linguistic, cultural and historic worlds with which people tend to self-identify.

There is an argument to be had about whether, depending on the context and purpose of data collection, ethnicity or race is preferable. This has been discussed in a recently published student essay by Kawano.[14] He makes an argument in favour of race when collecting data about South Asian cardiovascular disease in the USA. However, he does not suggest that his argument should be extended beyond that case. Rather, he is focused on US-based research that deals with a very specific research population. In general, we suggest that whether or not race is the best concept depends on the particular audience, purpose and justification for the use of the term in a given context.

It is also important to reiterate that the terms race and ethnicity are often conflated, a confusion that is exacerbated by the use of ethnicity as a euphemism for race, and by the fact that 'racism' can also refer to instances of discrimination on the basis of ethnicity. The Joint Commitment's choice of terminology might further contribute to this unhelpful conflation.

The Joint Commitment on publishing aims to collect demographic information that can be used to investigate and rectify inequities in academic publishing. Given that aim and that academic publishing is global, the terminology should be chosen so as to minimise any risks of harm or offence. In our view, it would have made more sense to choose a less contested concept that has a less controversial history in this context than race. As the *JME* operates under a medical parent publisher and is co-owned by the Institute of Medical Ethics (which is a registered charity), we are especially conscious of the shameful histories of racism in our fields. Perhaps, the Joint Committee should have used ethnicity (although this choice, too, would need to be justified). Alternatively, they could have simply asked, 'How do you identify yourself?'

Check for updates

ORCID iDs
John McMillan http://orcid.org/0000-0002-7507-1861
Mehrunisha Suleman http://orcid.org/0000-0002-8819-0659
Arianne Shahvisi http://orcid.org/0000-0002-9347-7566

REFERENCES

1 Adebowale V, Rao M. It's time to act on racism in the NHS. *BMJ* 2020:m568.
2 Richards M. BMJ journals collect gender, race, and ethnicity data on submissions. *BMJ* 2022;379:2542.
3 Ali S, Karlsen S, Learner H, *et al*. UK policy response to female genital mutilation needs urgent rethink. *BMJ* 2023;383:e074751.
4 Varcoe C, Browne AJ, Wong S, *et al*. Harms and benefits: collecting ethnicity data in a clinical context. *Soc Sci Med* 2009;68:1659–66.
5 Faucher L. Racism. In: Alcoff I, Anderson L, eds. *The Routledge Companion to the Philosophy of Race*. Routledge, 2017: 405–22.
6 Fish M, Shahvisi A, Gwaambuka T, *et al*. A new Tuskegee? Unethical human experimentation and Western neocolonialism in the mass circumcision of African men. *Dev World Bioeth* 2021;21:211–26.
7 Bruce L, Tallman R. Promoting racial equity in COVID-19 resource allocation. *J Med Ethics* 2021;47:208–12.
8 Schmidt H, Roberts DE, Eneanya ND. Rationing, racism and justice: advancing the debate around "colourblind" COVID-19 ventilator allocation. *J Med Ethics* 2022;48:126–30.
9 Varma T, Jones CP, Oladele C, *et al*. Diversity in clinical research: public health and social justice imperatives. *J Med Ethics* 2023;49:200–3.
10 Saylor KW, Martschenko DO. Promoting diagnostic equity: specifying genetic similarity rather than race or ethnicity. *J Med Ethics* 2023;49:820–1.
11 Suleman M, Qureshi Z. Should medicine be colour blind? *J Med Ethics* 2023;49:725–6.
12 Maneri M, Morning A. *An Ugly Word: Rethinking Race in Italy and the United States*. Russel Sage Foundation, 2022.
13 Maneri M, Morning A. Repositioning, not replacing, race: the case for concepts of descent-based difference[online ahead of print]. *Ethnic and Racial Studies* 2023:1–9.
14 Kawano B. Constructing a South Asian cardiovascular disease: a qualitative analysis on how researchers study cardiovascular disease in South Asians. *J Med Ethics* 2024;50:70–4.