

Analysis

Is there a place for ethnicity in precision health?

Shoba Poduval¹
 Caroline Modeste²
 Namratha R. Kandula³
 R.M. Anjana⁴
 Amitava Banerjee^{1,5}

¹ Institute of Health Informatics, University College London, London, UK

² Patient and public author, London, UK

³ Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA

⁴ Dr Mohan's Diabetes Specialities Centre and Madras Diabetes Research Foundation, Chennai, India

⁵ Department of Cardiology, Barts Health NHS Trust

Correspondence to:

Full name: Amitava Banerjee

Mailing address: UCL Institute of Health Informatics, 222 Euston Road, London NW1 2DA

Email: ami.banerjee@ucl.ac.uk

Phone: [02035495449](tel:02035495449)

Word count: 2001

References: 21

KEY MESSAGES

- **Recording of ethnicity is inadequate to properly inform research and routine care.**
- **Inadequate ethnicity data may worsen health inequalities, particularly if genetic and biologic differences are overemphasized at the expense of social determinants of health.**
- **A precision health approach would better capture and integrate individual characteristics including ethnicity, environment and lifestyle.**

Contributors and sources

SP is an academic GP and Clinical Lecturer with an interest in ethnic health inequalities. CM is a patient representative. NRK is a specialist in Internal Medicine and Professor of Medicine and Preventative Medicine. RMA is a Consultant Physician and Managing Director of Dr Mohan's Diabetes Specialities Centre, Chennai and President of the Madras Diabetes Research Foundation. AB is a Consultant Cardiologist and Professor of Clinical Data Science. The article was conceived through discussions between SP and AB on collection

and use of ethnicity data in clinical research and practice. NRK and RMA joined the authorship team to add subject matter expertise and provide an international perspective. All authors contributed to drafting and editing of the article. AB is the guarantor.

Patient involvement

CM is a patient author with lived experience of long-term health conditions and took part in a series of Patient and Public Involvement workshops led by SP during planning of this article. The workshops were funded by a University College London Hospital Biomedical Research Centre Innovation grant (award number BRC1254/PPI/SP/104990) and involved discussion and co-design of research questions exploring use of ethnicity data in clinical research and practice to tackle ethnic inequalities in people with long-term health conditions.

Conflicts of Interest

We have read and understood [BMJ policy on declaration of interests](#) and have no interests to declare.

Licence

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd ("BMJ"), and its Licensees to permit this article (if accepted) to be published in The BMJ's editions and any other BMJ products and to exploit all subsidiary rights, as set out in [The BMJ's licence](#).

Is there a place for ethnicity in precision health?

Shoba Poduval and colleagues argue that the current use of ethnicity data in research and practice is inadequate. An approach which more accurately considers social and behavioural determinants of health is needed to tackle health inequalities.

Before, during and after the Covid-19 pandemic, ethnic inequalities in health and healthcare are well-documented and the focus of research and quality improvement. “Ethnicity” is “the social group a person belongs to, and either identifies with or is identified with by others, as a result of a mix of cultural and other factors including language, diet, religion, and geographical and ancestral origins”¹. Data-driven precision medicine is tailoring of diagnosis and treatment through analysis of genomic, environmental and lifestyle data²; rising in prominence due to increasing data from health care records, research, pathology, imaging and wearables³. We explore potential data-driven approaches to ethnic inequalities, after discussing current UK ethnicity data by broad categories, which may generalise across generations, communities and geographies, and potentially exacerbate health inequalities. Next, we describe how precision health links precision medicine with public health via socioeconomic, behavioural, environmental and cultural factors in individual and population-level interventions^{3,4}. Finally, we make recommendations to improve ethnicity data; enhance public trust and standardise ethnicity data categories; and integrate data on social determinants of health(SDOH)^{4, 5}.

Data on social and behavioural determinants of health is vital to individualise care

The NHS Personalised Medicine Strategy(2015) stated, “By bringing together technologies such as genome sequencing, personalised data and wearable technology, the NHS hopes to enter an era of truly personalised care.....embedded into mainstream healthcare.”¹.

Studying a person’s genome alone is insufficient to improve diagnosis and treatment since social and behavioural data are required to individualise care. Precision health is a broader concept, linking precision medicine with public health, directly targeting social, behavioural, genetic and biological determinants for health promotion and disease prevention³.

“Precision medicine” and “personalised medicine” are often used synonymously, despite distinct meanings, or with different definitions in different policies, including the NHS Long Term Plan⁴. Precision health is not integrated in these strategies, in terms of data on individual characteristics and risk, with personalised care-planning, perpetuating “research-clinical

practice” gaps. In the US, this research-clinical practice gap is being tackled with research programs including the “All of Us” program which gathers rich data from a combination of physical measurements, survey responses, electronic healthcare records(EHRs), wearables and genomic data ¹. An application is the study of SDOH and genetics influence risk of common diseases such as coronary heart disease(CHD)(Figure 1).

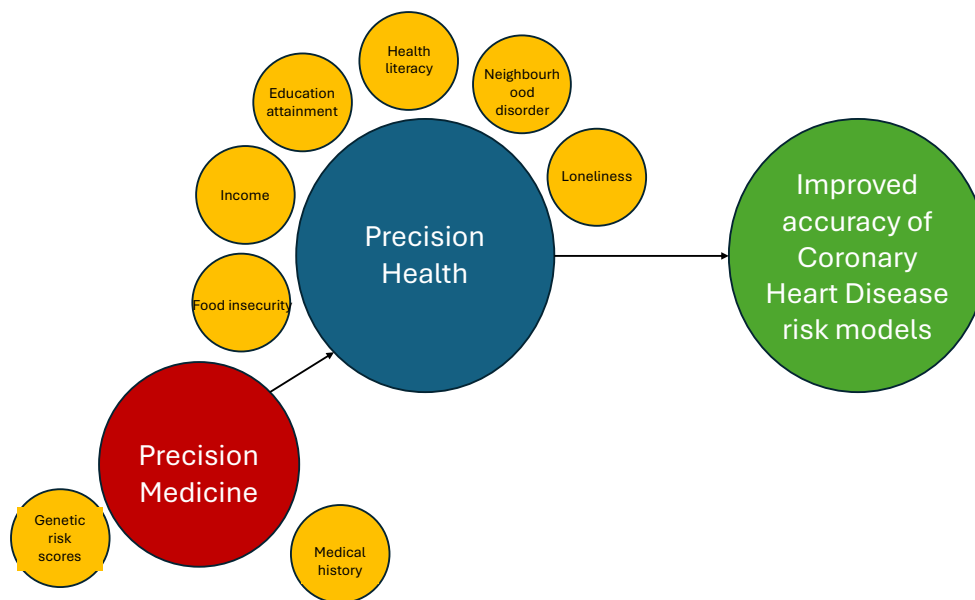


Figure 1. Precision Medicine, Precision Health and potential application.

Limitations to ethnicity data: In UK research and clinical practice, ethnicity data, predominantly self-reported, lack granularity, accuracy and completeness⁶. Among 60 million individuals in English primary care, 10% lacked recorded ethnicity data ⁶. Most studies have investigated ethnicity as five categories(e.g., Asian, Black/African/Caribbean, White, Mixed, Other)⁷, although more than 250 patient-identified ethnicity sub-groups⁶ exist. Unethical practices in data collection and use threaten public trust in data sharing⁸, e.g. flawed implementation of the UK’s care.data project, which aimed to link and provide access to health and social care information from different settings, but risked selling personal data to drug companies and non-healthcare industries⁹. Some countries(e.g. France) do not collect ethnicity data due to principles of secularism and universalism, with some arguing that ethnic inequalities diminish after controlling for social, demographic, familial and economic characteristics, and diagnostic stage and type of therapy². The counter-argument is that

limited data to evidence health disparities between ethnic minority groups, leads to inadequate evidence to inform decision-making on key structural and SDOH³. Similarly, in Australia, there is reluctance to report race/ethnicity data, and variables that measure cultural and language diversity (collective cultural groupings), based on an individual's country of birth and language spoken at home are emphasised². Relying on cultural diversity data can lead to poor ethnicity data mapping and underestimation of health disparities among racial and ethnic groups².

Lack of ethnicity data prevents researchers and policy-makers from identifying higher disease risk in marginalised groups, e.g. during the Covid-19 pandemic¹⁰. After the UK's Covid-19 vaccination rollout in December 2020, Black and ethnic minority groups were less likely to be vaccinated, but appeared to have higher Covid-19 mortality risk, attributed to biological differences and higher risk factor incidence, including type 2 diabetes. For example, people from a Bangladeshi background are more than 60% more likely to have a long-term health condition with increased infection risk, potentially explaining excess mortality¹². More accurate and complete ethnicity data could provide better understanding of individual and population diversity; vital for precision health and personalised care and addressing health inequalities⁶.

Lack of granularity in ethnicity recording could be due to confusion about reasons for ethnicity data collection, concerns about discrimination and belief that ethnic categories are unrepresentative⁸. We should ensure adequate capture of diversity within broad heterogeneous ethnic populations. Cultural history, immigrant status, language, socioeconomic status, and experiences of structural and interpersonal racism all influence health outcomes. These nuances are lost once data are aggregated to broad ethnicity categories. Accurate and complete data regarding ethnicity and protected characteristics could provide better understanding of individual and population diversity; facilitating precision health and personalised care and addressing health inequalities⁶.

As well as data, how ethnicity and associated data are used in health policy is important. Despite historical conflation with genetics, ethnicity and nationality are examples of culturally and politically constructed, social categories⁹. Biological reductionism implies that health disparities between ethnic groups rely entirely or predominantly upon underlying genetic differences, rather than structural and SDOH, including exclusionary policies, unhealthy environments, poverty, and education¹⁰. Moreover, people may be placed in ethnic categories unreflective of their social and behavioural characteristics. Therefore, population-level interventions may not work, with resources directed away from those most at need.

Even when accurately collected, ethnicity data are increasingly inadequate, with evidence that health outcomes are a result of complex interplay of many factors across the life course, including place of birth, social deprivation, environment, health-related behaviours and migration. An intersectionality lens posits that socially constructed identities(e.g. gender, ethnicity), social categories(e.g. social class) and social processes(e.g. discrimination) exist together, are additive and mutually reinforcing, and fundamental to how society structures power and resources to produce social and health inequality¹³. Epidemiological studies emphasising one dimension(e.g. social class, ethnicity), rather than intersectionality, fail to capture complexity of lived experiences of health inequality among marginalised groups, or identify implementable solutions. Men living in deprived areas of the UK live shorter lives, highlighting socioeconomic inequality and varying impact by gender identity¹³. Intersectional action is needed to address health inequalities across diverse populations.

Public health interventions need to go beyond narrow definitions of ethnicity

After the start of the Covid-19 vaccine rollout in the UK in December 2020, emerging data showed that people from Black and ethnic minority groups were more likely to die from Covid-19 but less likely to be vaccinated. Vulnerability to Covid-19 mortality was attributed to biological differences and higher incidence of risk factors, including type 2 diabetes. However, further studies showed that living in a multi-generational household contributes to excess risk, particularly in elderly South Asian women¹⁴, prompting calls to direct public health interventions to communities where multi-generational households are highly prevalent, including their prioritisation for vaccination. However, interventions continued to be targeted at ethnic minority groups as a whole, or one of the five large sub-groups, without attention to intersections of household composition, family structure and health of older adults¹⁵.

South Asians have traditionally been considered a high-risk ethnic group by type 2 diabetes susceptibility,¹⁶ with higher intra-abdominal and hepatic fat levels and lower insulin sensitivity, even at lower BMI levels. However, “South Asians” could include individuals of Indian, Pakistani, Bangladeshi, Sri Lankan, Bhutanese, Maldivian or Nepalese origin who, despite purported common genetic origins, differ widely in socioeconomic and environmental determinants of health and disease, and cultural and behavioural practices, resulting in differential type 2 diabetes susceptibility and complications. Even within this apparently homogenous ethnic group, factors such as family size, age and gender might play more important roles than shared genetics in susceptibility to disease, access to healthcare, response to treatment and implementation of effective management strategies. For example,

changes in dietary and culinary practices should target whole families rather than individuals attending a clinic or practice for those ethnic South Asians living in large multigenerational families. It may be inaccurate to categorise all individuals of Asian origin residing in countries such as the UK, as “South Asians”. Precision medicine approaches must go beyond narrow definitions of ethnicity for the greatest population benefit.

Exploring the role of ethnicity in clinical algorithms

Ethnicity is used in clinical algorithms in a variety of clinical areas, including cardiovascular disease and maternity care¹⁷. Proponents argue that ethnicity can act as a proxy for biological factors to identify individuals at higher risk of disease¹⁷, but for reasons described above, use of ethnicity in clinical decision-making is being increasingly questioned. Additionally, use of ethnicity in risk stratification is particularly inaccurate for people from cohorts that were not included in original datasets that contributed to development of risk calculators, including people of Hispanic or Asian descent. Patient awareness and perspectives vary regarding consideration of race in clinical algorithms and decision-making, with many being unaware of how race is used or concerned about racial discrimination associated when providing ethnicity data or making treatment decisions⁴. There have been mitigation efforts by working with communities to design appropriate ethnicity categories and explaining rationale for data collection⁵, and to eliminate ethnicity correction from clinical algorithms, particularly when adjusting by ethnicity might worsen existing ethnic inequity. Notable examples include revised vaginal birth after caesarean section calculator, estimated glomerular filtration rate and prediction tool for atherosclerotic cardiovascular disease in the US¹⁸. A new PREVENT risk calculator estimates 10- and 30-year risk of cardiovascular disease using ethnicity-free risk calculations¹⁹. Measuring proximal risk factors like blood pressure, diabetes status and cholesterol accurately predicted risk without using ethnicity. Raised blood pressure and cholesterol show associations with SDOH, e.g. living in segregated neighbourhoods or experiencing racism, which accumulate over time. We should recognise ethnicity as a sociopolitical, rather than biological, construct, affecting people over their lifecourse. Research provides a first step towards determining causal and social factors underpinning ethnic differences in CVD risks and outcomes.

Improved data capture, linkage and transparency are needed

We suggest three actions for research and practice. Firstly, linkage, consistency and minimum standards across primary and secondary care data are required. The Office for National Statistics(ONS) sets the gold standard for ethnicity recording in England and Wales, but Scotland has its own coding, and primary and secondary care and disease-specific registries

categorise using different census results²⁰. Interoperability and comparability could be facilitated if datasets complied to same standards⁶. A minimum ethnicity coding standard (by ONS definitions: currently 19 categories under 5 groups in the 2021 census) could be implemented across all routinely collected clinical data and disease registries. Enhanced efficiency and transferability would help to motivate researchers and policymakers.

Secondly, improved capture of interplay of ethnicity with health, and relevant social factors mediating this relationship, is needed, potentially through health data linkage with ONS and local authority data on social factors like employment and housing. These data could be made available to healthcare commissioners and to GPs during consultations. Successful examples include use of health and local authority data to explore factors relevant to multiple long-term conditions among working-age adults in a deprived borough of East London²¹, finding strong association between household tenure and MLTCs to guide related policy. More accurate and complete ethnicity data could target resources where they are most needed, which could persuade policymakers to take this action to address health disparities.

Thirdly, we recommend emphasis on training, culture and standards for researchers and health systems in ethnicity coding, and public education and engagement. Focus on how and why ethnicity is used in studies, and critical thinking about how this might perpetuate harmful biases, could improve recording⁸. and design of algorithms to improve ethnic health inequalities¹⁷. Barriers to change can be overcome by showing impact of poor coding.

Healthcare decisions based on inaccurate or unavailable data can lead to groups at risk of poorer health outcomes receiving inappropriate or no care. Precision health approaches, combining data on health and social factors, promote a more accurate picture of individual characteristics, informing intervention development to target factors which reinforce inequality.

References

1. NHS England. Improving Outcomes through Personalised Medicine 2016.
2. Data Saves Lives. What is Health Data? 2024 [Available from: <https://datasaveslives.eu/what-is-health-data>].
3. Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision Health: The Role of the Social and Behavioral Sciences in Advancing the Vision. *Annals of Behavioral Medicine*. 2020;54(11):805-26.10.1093/abm/kaaa018
4. Department of Health and Social Care. Major conditions strategy: case for change and our strategic framework. 2023.
5. North Carolina Precision Health Collaborative. What is Precision Health? : North Carolina Biotechnology Center; 2024 [cited 2024 03 Sep]. Available from: <https://www.ncbiotech.org/transforming-life-sciences/sectors-attention/precision-health#:~:text=Precision%20Health%20combines%20data%20about,personalize%20health%20and%20wellness%20plans>.
6. Pineda-Moncusí M, Allery F, Delmestri A, Bolton T, Nolan J, Thygesen JH, et al. Ethnicity data resource in population-wide health records: completeness, coverage and granularity of diversity. *Sci Data*. 2024;11(1):221.10.1038/s41597-024-02958-1
7. Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *Lancet*. 2021;397(10286):1711-24.10.1016/s0140-6736(21)00634-6
8. Bignall T, Phillips J. Improving the Recording of Ethnicity in Health Datasets. London, UK.; 2022.
9. Juengst ET, McGowan ML. Why Does the Shift from "Personalized Medicine" to "Precision Health" and "Wellness Genomics" Matter? *AMA J Ethics*. 2018;20(9):E881-90.10.1001/amajethics.2018.881
10. Sankar P, Cho MK, Condit CM, Hunt LM, Koenig B, Marshall P, et al. Genetic Research and Health Disparities. *JAMA*. 2004;291(24):2985-9.10.1001/jama.291.24.2985
11. Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG) 2019. <https://www.abs.gov.au/statistics/classifications/australian-standard-classification-cultural-and-ethnic-groups-ascceg/latest-release>
12. Raleigh V. The health of people from ethnic minority groups in England 2023 [Available from: <https://www.kingsfund.org.uk/publications/health-people-ethnic-minority-groups-england#diabetes>].
13. Bauer GR, Scheim AI. Methods for analytic intercategorical intersectionality in quantitative research: discrimination as a mediator of health inequalities. *Social Science & Medicine*. 2019;226:236-45 [PubMed](#) .
14. Nafilyan V, Islam N, Ayoubkhani D, Gilles C, Katikireddi SV, Mathur R, et al. Ethnicity, household composition and COVID-19 mortality: a national linked data study. *Journal of the Royal Society of Medicine*. 2021;114(4):182-211.10.1177/0141076821999973
15. Paudyal P, Skinner E, Majeed-Hajaj S, Hughes LJ, Magar NK, Keeling DI, et al. COVID-19 health information needs of older adults from ethnic minority groups in the UK: a qualitative study. *BMJ Open*. 2022;12(6): [PubMed](#) e059844.10.1136/bmjopen-2021-059844
16. Patel SA, Shivashankar R, Ali MK, Anjana RM, Deepa M, Kapoor D, et al. Is the "South Asian Phenotype" Unique to South Asians?: Comparing Cardiometabolic Risk Factors in the CARRS and NHANES Studies. *Glob Heart*. 2016;11(1):89 [PubMed](#) - 96.e3.10.1016/j.gheart.2015.12.010

17. El-Azab, S., & Nong, P. (2023). Clinical algorithms, racism, and “fairness” in healthcare: A case of bounded justice. *Big Data & Society*, 10(2).
<https://doi.org/10.1177/20539517231213820> (Original work published 2023)
18. Vyas Darshali A, Eisenstein Leo G, Jones David S. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine*. 2020;383(9):874-82.10.1056/NEJMms2004740
19. Mancini GBJ, Ryomoto A. Adoption of the PREVENT (Predicting Risk of Cardiovascular Disease EVENTS) Risk Algorithm: Potential International Implications. *JACC: Advances*. 2024;3(8):101122 [PubMed](https://doi.org/10.1016/j.jacadv.2024.101122) .<https://doi.org/10.1016/j.jacadv.2024.101122>
20. UK Government. Ethnicity facts and figures. <https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups/#:~:text=recording%20people's%20ethnicity,-,The%20ethnic%20groups%20were:,Any%20other> (Accessed 22/3/25)
21. Ingram E, Gomes M, Hogarth S, McDonald HI, Osborn D, Sheringham J. Household Tenure and Its Associations with Multiple Long-Term Conditions amongst Working-Age Adults in East London: A Cross-Sectional Analysis Using Linked Primary Care and Local Government Records. *International Journal of Environmental Research and Public Health* [Internet]. 2022; 19(7). 10.3390/ijerph19074155

1. Bick AG, Metcalf GA, Mayo KR, Lichtenstein L, Rura S, Carroll RJ, et al. Genomic data in the All of Us Research Program. *Nature*. 2024;627(8003):340-6.10.1038/s41586-023-06957-x
2. Renzaho AMN. The Lack of Race and Ethnicity Data in Australia-A Threat to Achieving Health Equity. *Int J Environ Res Public Health*. 2023;20(8).10.3390/ijerph20085530
3. Meudec M, Affun-Adegbulu C, Cosaert T. Review of health research and data on/with racially minoritised groups: Implications for addressing racism and racial disparities in public health practice and policies in Europe: a study protocol. *F1000Res*. 2023;12:57.10.12688/f1000research.128331.1
4. Schmidt IM, Shohet M, Serrano M, Yadati P, Menn-Josephy H, Ilori T, et al. Patients' Perspectives on Race and the Use of Race-Based Algorithms in Clinical Decision-Making: a Qualitative Study. *J Gen Intern Med*. 2023;38(9):2045-51.10.1007/s11606-023-08035-4
5. Bignall T, Phillips J. Improving the Recording of Ethnicity in Health Datasets. London, UK.; 2022.