Demographic reporting in biosignal datasets: a comprehensive analysis of the PhysioNet open access database

Sarah Jiang, Perisa Ashar*, Md Mobashir Hasan Shandhi, Jessilyn Dunn*

The PhysioNet open access database (PND) is one of the world's largest and most comprehensive repositories of biosignal data and is widely used by researchers to develop, train, and validate algorithms. To contextualise the results of such algorithms, understanding the underlying demographic distribution of the data is crucial—specifically, the race, ethnicity, sex or gender, and age of study participants. We sought to understand the underlying reporting patterns and characteristics of the demographic data of the datasets available on PND. Of the 181 unique datasets present in the PND as of July 6, 2023, 175 involved human participants, with less than 7% of studies reporting on all four of the key demographic variables. Furthermore, we found a higher rate of reporting sex or gender and age than race and ethnicity. In the studies that did include participant sex or gender, the samples were mostly male. Additionally, we found that most studies were done in North America, particularly in the USA. These imbalances and poor reporting of representation raise concerns regarding potential embedded biases in the algorithms that rely on these datasets. They also underscore the need for universal and comprehensive reporting practices to ensure equitable development and deployment of artificial intelligence and machine learning tools in medicine.

Lancet Digit Health 2024

Published Online
October 1, 2024
https://doi.org/10.1016/S2589-7500(24)00170-5

*Contributed equally

Department of Biomedical Engineering (S Jiang, P Ashar, M M H Shandhi, J Dunn), **Department of Computer Science** (S Jiang), **and Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA** (J Dunn); **Duke Clinical Research Institute, Duke University, Durham, NC, USA** (J Dunn)

Correspondence to:
Dr Jessilyn Dunn, Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA
jessilyn.dunn@duke.edu

Introduction

Combined advances in physiologic sensing technology and computing have led to health-care applications of artificial intelligence (AI), such as wearable devices that monitor people with chronic and infectious diseases. Such successes include automated monitoring for irregular heart rhythms,[1,2] sleep health,[3,4] and stress.[5,6] The AI-based biosignal algorithms underlying these technologies, however, face challenges that also exist in AI more generally, eg, reduced accuracy under particular circumstances or for some populations.[7–9] Most concerningly, performance inadequacies tend to disproportionately affect marginalised groups, often on the basis of demographic attributes such as age and race, leading to bias.[10] There have been several high-profile examples of bias in AI, including in algorithms used for hiring, the US justice system, credit scoring, health applications, and facial recognition.[11] In the realm of facial recognition, AI algorithms are less accurate in identifying individuals with darker skin tones, with one study showing error rates of 0·8% for light-skinned men and 34·7% for dark-skinned women.[12] In health care, the use of AI might amplify existing inequities. For example, a tool using health-care cost as a proxy for need resulted in Black patients receiving worse care than White counterparts.[13]

Such performance inadequacies are more likely to occur when the data used to train and test AI have poor diversity. Such data fail to represent a wide range of populations and conditions and can result in algorithm bias, where a machine learning model produces unfair or inaccurate outcomes due to these imbalances. Additionally, a machine learning model is likely to have poor real-world performance if there is a mismatch between the data used to build the model and the context of the model's real-world deployment, or if those data contain unwanted societal biases. Reporting standards for artificial intelligence and machine learning (AI–ML) data have been proposed, including 'datasheets for datasets';[14–20] however, no singular and widespread consensus process for documenting the data used for model building exists. One challenge is that key metadata that should be reported is often dependent on the field, application, and context, and therefore there is no single solution.

In health-related applications, the complications created by poor reporting on data characteristics are pronounced. A 2020 review of 164 articles applying machine learning to improve clinical decision making with use of electronic health records data revealed that race and ethnicity were not reported in 64% of studies, sex and gender were missing from 24%, age from 21%, and socioeconomic status from 92%.[19] Whether demographic variables were used as model inputs was rarely clearly reported. A 2021 National Institutes of Health (NIH) guideline aiming to address this challenge mandated that grantees must report individual-level study participant data on sex or gender, race, ethnicity, and age in annual progress reports.[20]

Our Viewpoint is specifically concerned with reporting on characteristics of biosignal data, which is essential to the design, evaluation, and comparison of biosignal algorithms. The PhysioNet open access meta-database (PND) is one of the largest compilations of biosignal databases. It consists of an archive of well characterised digital recordings of physiological signals and related data for use by the biomedical research community. PND is one of the world's largest, most comprehensive, and most widely used repositories of biosignal data.

Although PND is one of the largest archives of well characterised digital recordings of physiological signals and is widely used for developing and validating novel biosignal algorithms,[21–23] there is no standard reporting

For more on the **PhysioNet open access meta-database** see https://physionet.org/about/database/

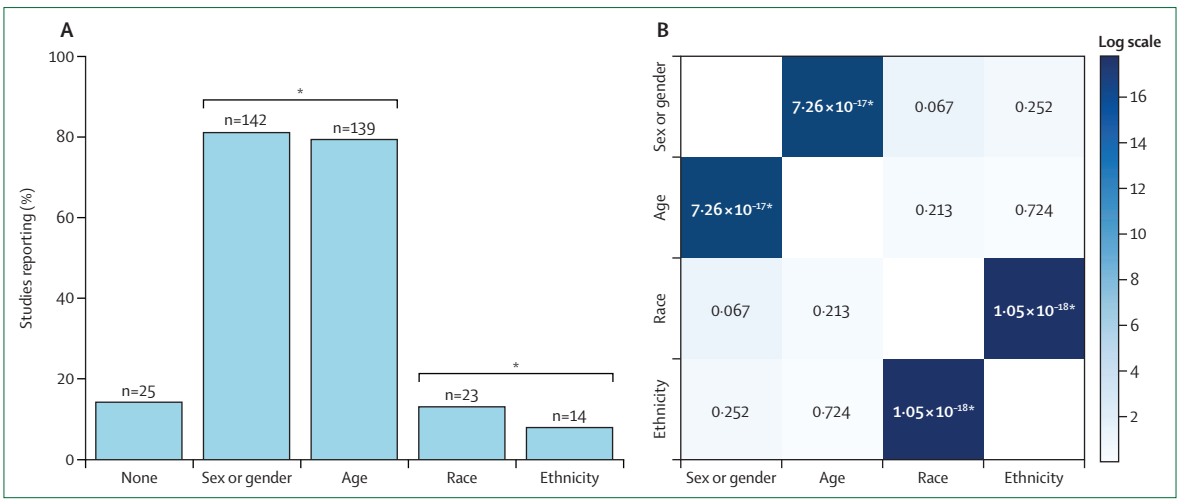

Figure 1: **Demographic reporting frequency**
(A) The proportion of studies on PhysioNet involving people that report each of the four demographic variables of interest: race, ethnicity, sex or gender, and age. Raw counts are reported above each bar (N=175), and pairs of variables that are associated significantly are represented by a bracket and asterisk (indicating a significant p value result from χ^2 tests of independence). (B) Relationships between the demographic variables, with numerical values representing p values from the paired χ^2 tests of independence, significant p values are indicated by an asterisk. p values less than 0·0083 are coloured according to the logarithmic colour scale on the right; values at or above 0·0083 are shaded in pale blue. The self-tests (diagonal) are shaded in white.

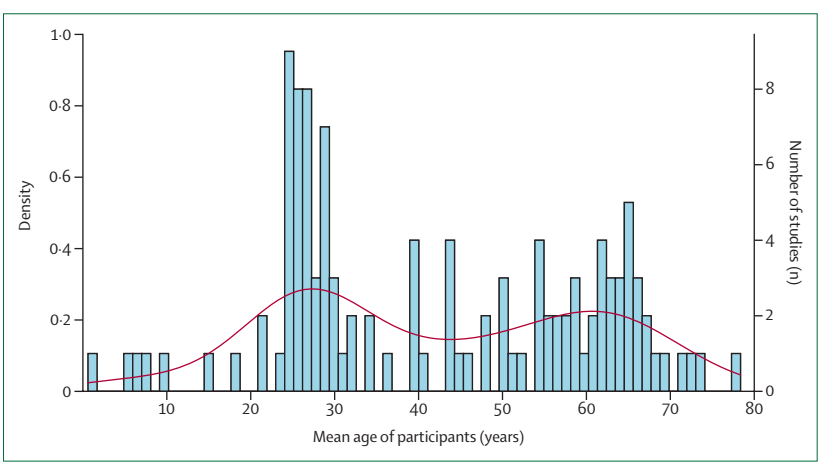

Figure 2: **Age distribution of participants**
The distribution of the mean age of each study's participants for the 114 studies reporting either each participant's age or the mean age of all participants. Left-hand y-axis pertains to the density graph (red), right-hand y-axis pertains to the histogram.

See Online for appendix

about the people from whom the data originated. In this Viewpoint, we have explored the demographic information available on PND to understand whether and how demographic information is reported and to uncover any common trends or biases that might be discerned from the available information. We further explored the role of various study-related factors (eg, study size, location, and biosignal type) in the reporting on and the characteristics of the study demographics (appendix p 2). This information brings clarity to if and how such datasets should be leveraged for developing and validating biosignal algorithms that are generalisable and equitable.

Exploring the PhysioNet open access meta-database

There were 181 unique datasets in PND as of July 6, 2023, of which 175 involved people. The datasets for each study contained various biosignals (eg, electrocardiogram, photoplethysmogram, and accelerometry) representing clinical areas ranging from cardiac electrophysiology to physical movement (appendix pp 3–8). These datasets were published on PhysioNet between Aug 3, 1999, and Jan 18, 2023. We grouped studies by their biosignal types from the categories adapted from PND.[24] Sex or gender was treated as a binary variable, and 34 studies that did not specifically state the number of male and female participants were removed from that analysis. When information was not available on PhysioNet, we checked the original publications. For the χ^2 tests of independence, the assumption of independence of observations was violated as several datasets were derived from the same overarching database. Thus, we randomly selected one study from each of those that were derived from the same dataset, resulting in the inclusion of 163 of the 175 studies involving people. The world heat map to explore the geographical origin of the datasets was generated with Rworldmap 1.3–6 (R version 4.3.1). Pandas 2.0.3, NumPy 1.25.0, SciPy 1.10.1, and Seaborn 0.12.2 (Python version 3.9) were used for all other analyses.

Reporting of demographic factors

Of the 175 studies analysed, only 12 (6·9%) reported all four key demographic variables of age, sex or gender, race, and ethnicity. 25 studies (14·3%) did not report any demographic information (figure 1). Fewer studies reported race and ethnicity than age and sex or gender (p<0·0083,

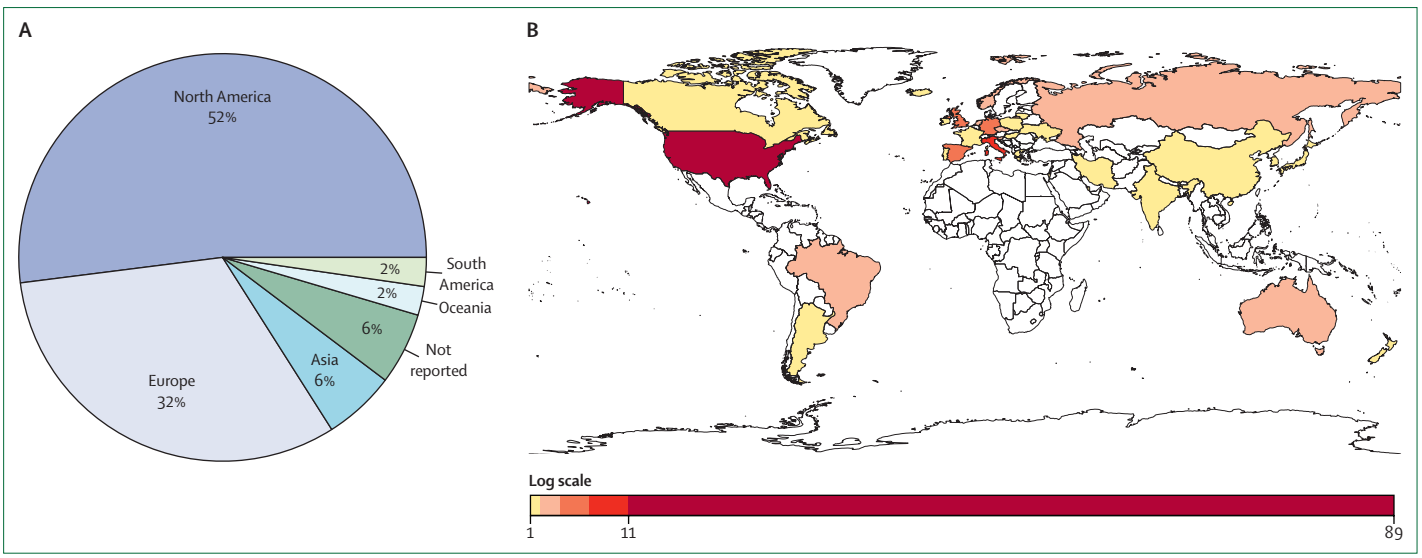

Figure 3: **Geographical distribution of PhysioNet studies**
(A) The geographical distribution of studies by continent, dominated by North America. The Not reported section indicates studies where the location could not be determined based on the descriptions in the PhysioNet database or the lack of affiliations of the authors, as well as studies where participants hailed from multiple different countries within a single study. (B) The country sources of studies on the PhysioNet open database. Due to the significant discrepancy between the number of studies published by the USA *vs* other countries, the heat map had tailored intervals within the figure legend to properly visualise the difference of studies published by various countries of the world.

χ^2 test). Only 23 studies (13·1%) reported race, and only 14 reported ethnicity (8·0%), whereas 142 (81·1%) reported sex or gender and 139 (79·4%) reported age (figure 1A). Race and ethnicity had varying reporting methods: two studies reported ethnicity as race, and five vice versa, potentially due to the definitions of these concepts varying among different countries and cultures.[25–31] The only ethnicities reported were Hispanic or Latino, non-Hispanic or non-Latino, and unknown (appendix pp 9–10). We found that reporting of race was not independent of reporting ethnicity and reporting of age was not independent of reporting sex or gender ($p<0·0001$, χ^2 tests of independence; figure 1B). We found that several studies on PhysioNet used overlapping data, in that smaller studies drew their data from larger studies. Of note, how these related substudies reported demographic information varied. Importantly, the availability of demographic information in the larger datasets dictated which information could be reported by its substudies, highlighting the influence of larger datasets in shaping overall demographic reporting.

Demographic distributions within and across datasets

Of the 150 studies that reported at least one demographic factor, eight did not report sex or gender (or a related variable similar to gender). Excluding the 16 studies with only female participants that investigated pregnancy, 51·6% of participants identified as male, and 41·3% identified as female, indicating an overall slight skewness towards male participants (appendix p 11). Sex or gender was either unknown or not reported for the remaining percentage of participants.

Of the 139 studies that reported age, the methods of age reporting varied widely, from the minimum and maximum age of subgroups or all participants, to individual ages of participants, to summary statistics of the entire participant pool or more general age descriptions (ie, college students). Of the 114 studies reporting mean or participant-level age, the average age across studies was 42·0 years (SD 18·4) and the median was 39·9 years (IQR 26·4–59·4; figure 2). The distribution of average participant age by study appears to be roughly bimodal, with a more prominent peak for ages 24–36 years (n=43 studies) and a subtler peak for ages 60–70 years (n=23 studies). These peak ranges were chosen based on buckets from the histogram that had reasonably more studies, and were confirmed from the average ages reported in the table in the appendix. Of the 43 studies in the younger peak, 26 (60·5%) studied healthy individuals (n=41) and athletes (n=2), and were focused on comparing and validating biosignals or performing exploratory biosignal analyses in healthy populations. Several of the studies in the younger group were focused on biosignals during pregnancy (n=9). Of the 23 studies in the older peak, 12 (52·2%) were focused on health conditions including heart failure, hypertension, and Parkinson's disease, as well as patients recieving ambulatory or intensive care. Overall, there were few (n=2) studies on healthy participants of an older age; the predominant focus in these studies was on disease conditions.[32,33]

Geographic distribution of datasets

Demographic reporting behaviours could vary by geographical location. In particular, we hypothesised that

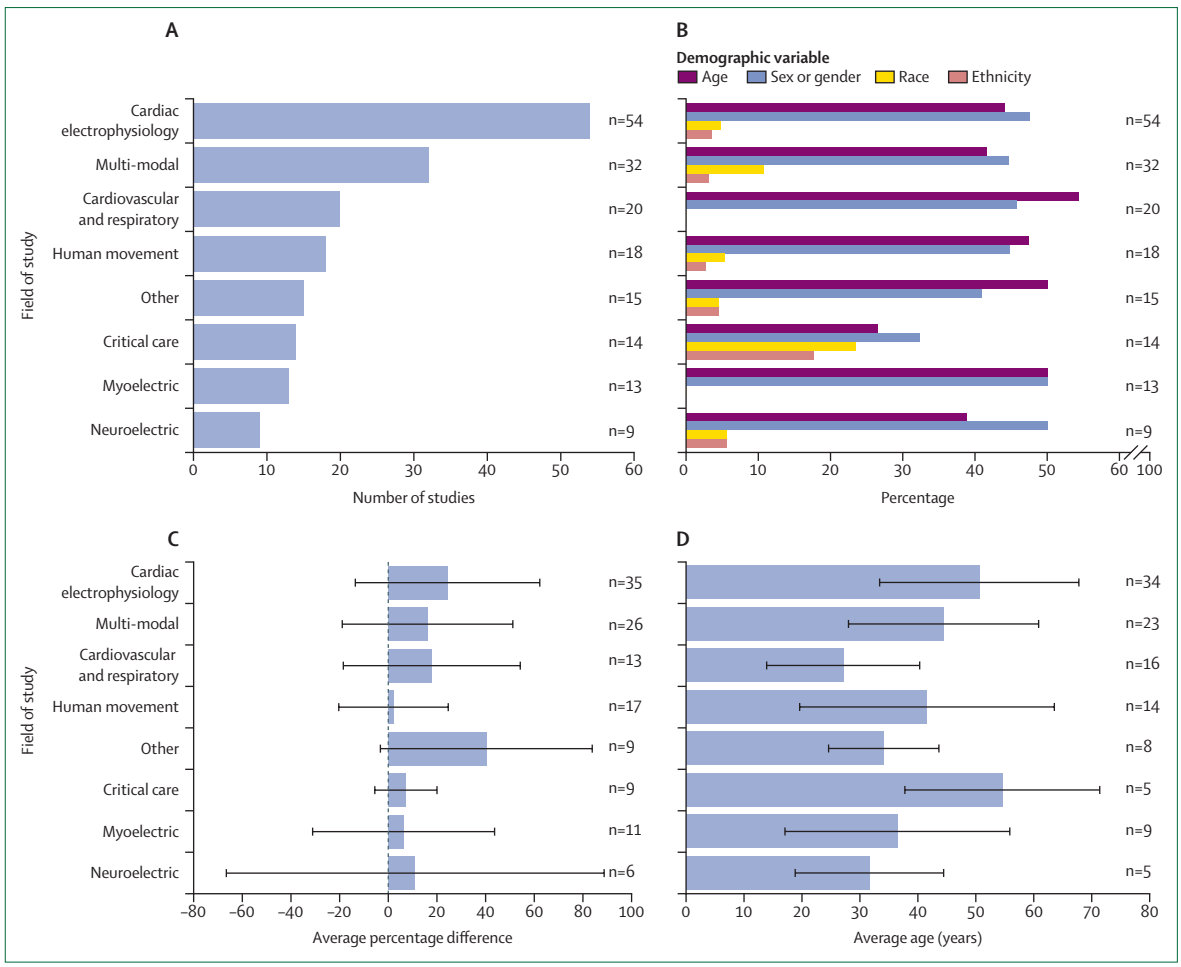

Figure 4: **Biosignal types and demographics reporting**
(A) The number of studies categorised into different biosignal focus areas. (B) The percentage of the demographic variables (sex or gender, age, race, and ethnicity) reported by studies (of the total number of studies categorised by different biosignal types). (C) The mean percentage differences of male to female participants among studies across different biosignal types. Studies that could only analyse one sex or gender due to certain biological conditions only being present in one sex or gender were removed (n=16 studies on pregnancy). (D) Graph depicts the mean value of average ages of patients in studies across different biosignal focus areas.

in regions with more homogeneous populations by race and ethnicity, the practice of reporting on these demographic variables might be less common. Of the total 175 studies analysed, 164 reported the geographical location of the data collection, representing 31 unique countries and five continents (figure 3). We found that the majority of the datasets (n=91) were generated in North America, with nearly 98% (n=89) of those from the USA (figure 3A, appendix pp 12–13). The USA alone accounted for about 51% of the total studies on PND. Grouping by continent to gain deeper insights into the geographical distribution of biosignal studies, we found that Africa, surprisingly, and Antarctica, as expected, were the only continents not represented (figure 3B, appendix p 14). Such an absence of representation is probably due both to PhysioNet's origin in the USA, and lower rates of biosignal research overall in some regions. Only five of the 31 countries represented had studies reporting race (appendix pp 12–13). Race was reported in 17 (19·1%) of 89 studies published in the USA, two (66·7%) of three from the Netherlands, one (20·0%) of five from Spain, two (66·7%) of three from Brazil, and one (25·0%) of four from Russia Additionally, only two countries, the USA and Brazil, had studies reporting ethnicity. Of the studies published in the USA, 13 (14·6%) reported ethnicity, and one (33·0%) of the three studies from Brazil reported ethnicity. Taken together, race and ethnicity are clearly infrequently reported upon in studies worldwide. All 31 countries had at least one study that reported sex or gender. Of the USA studies, 71·9% (n=64) reported sex or gender. 135 studies from 30 countries reported age, 65 (48·1%) of which were done in the USA.

Demographic reporting by biosignal type

Many biosignal studies focus on a specific measurement modality and clinical or physiological area of interest. The most prevalent biosignal type in PND was cardiac

electrophysiology (n=54; figure 4A, B). The higher proportion of male-to-female participants held true across biosignal types (figure 4C). For example, on average, cardiac electrophysiology had 24% more male than female participants. This disparity is shown, but to a lesser extent, for studies related to critical care, with 7% more male than female participants, and those related to human movement, with 2% more male participants. Notably, no study type had more female than male participants after excluding the pregnancy studies from this analysis. None of the observed differences in representation by sex or gender were significant ($p=0·1447$, χ^2 test).

We next explored how a study's focus area might be related to the age of its study participants. We analysed 139 studies that reported both age and biosignal type. Notably, critical care (n=5 studies) had the highest average age at 54·6 years (SD 16·8), and cardiac electrophysiology (n=34) had the second highest average age of 50·6 years (SD 17·2; figure 4D). The Kruskal–Wallis test revealed differences in the mean ages of participants across biosignal types ($p=0·0022$), suggesting that age distributions vary notably between different fields of study.

Discussion

Overall, we found that 14% of studies in the PND did not report any study participant demographics. Furthermore, race and ethnicity were reported 7× and 10× less frequently than both age and sex or gender, which were reported in about 80% of studies. Of no doubt are the challenges to adequate reporting of race and ethnicity; such socially constructed bulk characterisations, which are often self-reported, could miss key subgroups and are unlikely to be as related to underlying physiology as ancestral or genetic information. However, genetic information is both more difficult to come by and also plagued by privacy challenges. Future work should establish how best to characterise and report on race and ethnicity, including when genetic or ancestral information is needed, such that the level of reporting is adequate for relating outcomes of interest to true underlying physiology and appropriately accounting for racial or ethnic confounders.

The absence of comprehensive race and ethnicity reporting could be a result of racially or ethnically homogeneous populations in some regions, but importantly, a poor level of such reporting in countries with large majority populations could mask performance issues in small subpopulations. Furthermore, the absence of universally defined concepts of race and ethnicity in different parts of the world might also contribute to the inconsistent reporting of race and ethnicity. Missing demographic information could pose a larger challenge than imbalanced, but reported, demographics because encoded biases might be less obvious and therefore more likely to evade detection. This potential bias is concerning considering that PND data are widely used for developing and validating AI–ML algorithms.[21,34] Often, the algorithms trained on these biased datasets are proprietary, making the extent of the potential problem difficult to assess.[35–37] However, their potential commercial usage, which could include an effect on life-saving decisions such as detecting hypoxia or cardiovascular events, poses serious concerns.

Standardised demographic reporting guidelines including comprehensive, participant-level demographic data could be implemented in biosignal databases to address insufficient demographic reporting (panel). These guidelines could be derived from the Office of Management and Budget's *Statistical Policy Directive (No 15:) on Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity*, or from the NIH policy and guidelines for human participants study demographics, which require participant-level reporting of race, ethnicity, and sex or gender in human studies.[39] Expanding upon the NIH's *Inclusion Across the Lifespan*[40] policy, we also recommend that individual-level age be reported, with aggregation to protect participant privacy on the basis of the risk level of the data (eg, age >89 years reported as a group following the Health Insurance Portability and Accountability Act privacy rule known as the Safe Harbor method).[41] Likewise, PhysioNet and other biosignal repositories should either adapt or develop a standardised demographic reporting template and enforce it for future dataset release. This template could require curation of detailed participant-level data, including race, ethnicity, sex, gender, age, and relevant socioeconomic factors such as income and education level, and a summative score (eg, from 0 to 4 based on the four primary demographic variables of interest required by the NIH) to indicate the comprehensiveness

***Panel*: Proposed guidelines for equitable representation and demographic reporting**

Proposed guidelines are crucial for ethical and responsible artificial intelligence and machine learning (AI–ML) algorithm design and implementation in health care.

Standardised guidelines for demographic reporting:

- Addresses potential biases in AI–ML algorithms due to unconsidered or unreported demographics.
- Must include participant-level reporting of race, ethnicity, and sex.[38]
- Could be derived from the Office of Management and Budget's *Statistical Policy Directive (No 15:) on Race and Ethnicity* data standards or National Institutes of Health policy and guidelines for human study demographics.[38–40]

Diverse participant populations:

- Essential for representative data and algorithm outputs.
- Consider race, ethnicity, gender, age, and geography.
- Transparently share demographic data among researchers.

Search strategy and selection criteria

We explored the PhysioNet open access database, which contained 181 studies as of January, 2024, of which 175 involved human participants. Our data was collected on July 6, 2023. All 175 studies involving human participants were included in this analysis. Any resources on policy guidance referenced were the most up-to-date policies as of January, 2024; only open access datasets from PhysioNet were analysed. Additional references included in this Viewpoint are provided as supporting evidence.

of the demographic data. This score could also include information on the quality of the reporting (eg, the granularity of the variable levels reported). However, with these more transparent and comprehensive data reporting practices comes the risk of overstepping a participant's autonomy and privacy; thus, finding a balance between transparent practices and preserving participant anonymity is paramount. Such policies should be revisited as population demographics shift, and reporting standards should be regularly updated to reflect the evolving nature of biosignal research and its applications. Researchers and authors also should discuss their rationale for choosing a specific sample for their research objective and address potential biases that could result from their choices.

Our findings also point to a need for more diversity in biosignal data, ensuring wider representation of demographic and geographical variables. Study populations were predominantly White and US-based, and some fields had disproportionate over-representation of male participants. This lack of diverse representation raises concerns about the applicability of research findings to female and non-binary individuals, and people of colour in the USA and other Western countries whose populations are predominantly White. The absence of diversity in study populations could lead to biosignal-related technology innovations that do not work equally well for everyone. Our study is limited by the inconsistent manner of reporting of demographic data in PND, requiring inference by the study team to enable comparison across studies. This variability in data reporting highlights the need for standardised demographic data documentation in biosignal research.

Increasing transparency and balance in biosignal data demographics could be achieved through a multifaceted approach involving clear and standardised demographic reporting on biosignal datasets, and more diverse participant sampling. First, the development of demographic reporting standards is paramount. Second, participant-level demographics should be accessible to those using biosignal data to develop AI-based models to ensure that the models perform equally well for everyone. Finally, the demographics of any data used for model development should also be disclosed. However, sufficient attention should be given to the participant's privacy while sharing participant-level demographic information to ensure that it does not lead to identification of study participants. Not reporting demographic information greatly reduces the real-world potential of biosignal algorithms and increases the chances of developing biased models and applications. However, over-reporting could also be problematic as it could lead to identification of study participants. Thus, in addition to sufficient demographic reporting, it is also crucial to preserve participant anonymity and ensure that data are properly de-identified and cannot be traced back to their original source.

New methodologies for leveraging imbalanced demographic data for algorithm development should also be explored via collaborative efforts across various disciplines, including health care, statistics, AI, ethics, and social sciences. While we hope this Viewpoint will help to foster more balanced demographics in new biosignal datasets, approaches such as transfer learning, oversampling under-represented groups, or use of synthetic data might enable the fair use of existing datasets that have imbalanced demographics. Finally, the absence of geographical diversity in biosignal data is concerning. Balanced representation locally and globally could be achieved through targeted funding efforts,[42] establishing collaborations between biosignal researchers in different regions, and encouraging the submission of biosignal studies taking place in diverse geographical locations to major open-source databases. This increased reach will lead to advancements that could improve human health for everyone.

Conclusions

Substantial evidence has shown that bias in biosignal algorithms can have severe consequences (eg, inferior performance of pulse oximeter-based blood oxygen saturation measurement tools in people with darker skin tone).[43] With this study, we sought to characterise demographic reporting in one of the largest open-source biosignal archives. We specifically investigated demographic parity and reporting practices in these data and its relation to factors such as study size, location, and clinical area. The findings from this study bring clarity to if and how these data should be leveraged for developing and validating biosignal algorithms that are generalisable and equitable and point to areas for improvement on reporting of data characteristics.

Contributors

SJ: literature and database review, data collection, statistical analysis, code development and synthesis, writing, figure and table generation, references, and final formatting. PA: literature and database review, data collection, statistical analysis, code development, writing, and figure and table generation. MMHS: project administration, supervision, validation, writing, review, and editing. JD: conceptualisation, funding acquisition, project administration, supervision, validation, writing, review, and editing.

Declaration of interests

We declare no competing interests.

Data sharing

All collected demographic and reporting data from 175 databases on PhysioNet Open Databases, including the study protocol, collected data, statistical analysis plan, and analytic code, are available on GitHub to anyone who wishes to access the data for any purpose.

Acknowledgments

The authors of this Viewpoint received funding from the National Science Foundation CAREER Award Number: 2339669.

References

1 Melillo P, Izzo R, Orrico A, et al. Smart health for assessing the risk of events via ECG database. *PhysioNet* 2015; published online May 19. https://doi.org/10.13026/C2RC7C.

2 Melillo P, Izzo R, Orrico A, et al. Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. *PLoS One* 2015; **10:** e0118504.

3 Alvarez-Estevez D, Rijsman R. Haaglanden Medisch Centrum sleep staging database. *PhysioNet* 2002; published online March 18. https://doi.org/10.13026/t79q-fr32.

4 Alvarez-Estevez D, Rijsman RM. Inter-database validation of a deep learning approach for automatic sleep scoring. *PLoS One* 2021; **16:** e0256111.

5 Amin MR, Wickramasuriya D, Faghih RT. A wearable exam stress dataset for predicting cognitive performance in real-world settings. *PhysioNet* 2002; published online May 26. https://doi.org/10.13026/kvkb-aj90.

6 Rafiul Amin M, Wickramasuriya DS, Faghih RT. A wearable exam stress dataset for predicting grades using physiological signals. 2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); March 10–11, 2022 (30–36).

7 Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digit Health* 2021; **3:** e51–66.

8 Wu J, Lui X, Li M, et al. Clinical text datasets for medical artificial intelligence and large language models—a systematic review. *NEJM AI* 2024; published online May 17. https://doi.org/10.1056/AIra2400012.

9 Wen D, Khan SM, Ji Xu A, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. *Lancet Digit Health* 2022; **4:** e64–74.

10 Charpignon ML, Carrel A, Jiang Y. Going beyond the means: exploring the role of bias from digital determinants of health in technologies. *PLOS Digit Health* 2023; published online Oct 12. https://doi.org/10.1371/journal.pdig.0000244.

11 Kordzadeh N, Ghasemaghaei M. Algorithmic bias: review, synthesis, and future research directions. *Eur J Inf Syst* 2021; **31:** 1–22.

12 Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proc Mach Learn Res* 2018; **81:** 77–91.

13 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; **366:** 447–53.

14 Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets. *arXiv* 2021, published online Dec 1. http://arxiv.org/abs/1803.09010 (preprint).

15 Holm S, Petersen E, Ganz M, Feragen A. Bias in context: what to do when complete bias removal is not an option. *Proc Natl Acad Sci USA* 2023; **120:** e2304710120.

16 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385:** e078378.

17 Rostamzadeh N, Mincu D, Roy S, et al. Healthsheet: development of a transparency artifact for health datasets. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; June 21–24, 2022 (1943–61).

18 The STANDING Together Collaboration. Recommendations for diversity, inclusivity, and generalisability in artificial intelligence health technologies and health datasets. *Zenodo* 2023; published online Oct 30. https://doi.org/10.5281/zenodo.10048356.

19 Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. *J Am Med Inform Assoc* 2020; **27:** 1878–84.

20 NIH Staff. What are the requirements for reporting on human study participant age, sex/gender, and race/ethnicity? 2021. https://nexus.od.nih.gov/all/2021/04/29/what-are-the-requirements-for-reporting-on-human-study-participant-age-sex-gender-and-race-ethnicity/ (accessed Jan 16, 2024).

21 Hong S, Zhang W, Sun C, Zhou Y, Li H. Practical lessons on 12-lead ECG classification: meta-analysis of methods from PhysioNet/computing in cardiology challenge 2020. *Front Physiol* 2022; **12:** 811661.

22 Sannino G, De Pietro G. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Gener Comput Syst* 2018; **86:** 446–55.

23 Reyna MA, Kiarashi Y, Elola A, et al. Heart murmur detection from phonocardiogram recordings: the George B Moody PhysioNet challenge 2022. *PLoS Digit Health* 2023; published online Sept 11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10495026/.

24 Golberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000; published online June 13. https://doi.org/10.1161/01.CIR.101.23.e215.

25 Johnson A, Pollard T, Badawi O, Raffa J. eICU collaborative research database demo (version 2.0.1). *PhysioNet* 2012; published online May 6. https://doi.org/10.13026/4mxk-na84.

26 Novak V, Quispe R. Cerebromicrovascular disease in elderly with diabetes (version 1.0.1). *PhysioNet* 2022; published online Aug 5. https://doi.org/10.13026/00bm-0x81.

27 Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; **3:** 160035.

28 Bennett A, Wiedekopf J, Ulrich H, Johnson A. MIMIC-IV clinical database demo on FHIR (version 2.0). *PhysioNet* 2022; published online June 7. https://doi.org/10.13026/2f5g-dh02.

29 Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV-ED Demo (version 2.2). *PhysioNet* 2023; published online Feb 8. https://doi.org/10.13026/jzz5-vs76.

30 Escalona-Vargas D, Rathinaswamy B, Govindan A, et al. MMG database. *PhysioNet* 2016; published online March 4. https://doi.org/10.13026/C2ZW95.

31 Logé C, Ross E, Dadey DYA, et al. Q-Pain: a question answering dataset to measure social bias in pain management. *arXiv* 2021; published online Aug 3. https://doi.org/10.48550/arXiv.2108.01764 (preprint).

32 Wayne PM, Manor B, Novak V, et al. A systems biology approach to studying Tai Chi, physiological complexity and healthy aging: design and rationale of a pragmatic randomized controlled trial. *Contemp Clin Trials* 2013; **34:** 21–34.

33 Stein P. Normal sinus rhythm RR interval database. *PhysioNet* 2023; published online March 3. https://doi.org/10.13026/C2S881.

34 Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000; published online June 13. https://doi.org/10.1161/01.CIR.101.23.e215.

35 Holdcroft A. Gender bias in research: how does it affect evidence based medicine? *J R Soc Med* 2007; **100:** 2–3.

36 No authors listed. Tackling biases in clinical trials to ensure diverse representation and effective outcomes. *Nat Commun* 2024; published online Feb 15. https://doi.org/10.1038/s41467-024-45718-w.

37 Upchurch M. Gender bias in research. In: Naples NA, ed. Companion to women's and gender studies. March 24, 2020. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119315063.ch7 (accessed Jan 16, 2024).

38 NIH Grants and Funding. Inclusion of women and minorities. Oct 11, 2022. https://grants.nih.gov/policy/inclusion/women-and-minorities.htm (accessed July 10, 2024).

39 Orvis K. OMB publishes revisions to statistical policy directive no. 15: standards for maintaining, collecting, and presenting federal data on race and ethnicity. 2024. https://www.whitehouse.gov/omb/briefing-room/2024/03/28/omb-publishes-revisions-to-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity/ (accessed April 4, 2024).

For **demographic and reporting data on GitHub** see https://github.com/Big-Ideas-Lab/PhysioNet-Open-Access-Database-Analysis

40 NIH Grants and Funding. Inclusion across the lifespan. Jan 10, 2023. https://grants.nih.gov/policy/inclusion/lifespan.htm (accessed Jan 16, 2024).
41 Office for Civil Rights. Guidance regarding methods for de-identification of protected health information in accordance with the health insurance portability and accountability act (HIPAA) privacy rule. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html (accessed March 28, 2024).
42 Cho PJ, Yi J, Ho E, et al. Demographic imbalances resulting from the bring-your-own-device study design. *JMIR Mhealth Uhealth* 2022; **10:** e29510.
43 Al-Halawani R, Charlton PH, Qassem M, Kyriacou PA. A review of the effect of skin pigmentation on pulse oximeter accuracy. *Physiol Meas* 2023; **44:** 05TR01.