Evaluating Diversity in Open Photoplethysmography (PPG) Datasets: a protocol for systematic review

Vedha Penmetcha, Lekaashree Rambabu, Brandon G Smith, Orla Mantle, Tom Edmiston, Laura Hobbs, Shobhana Nagraj, Peter H Charlton, Tom Bashford

Submitted to: JMIR Research Protocols
on: February 28, 2025

Table of Contents

Evaluating Diversity in Open Photoplethysmography (PPG) Datasets: a protocol for systematic review

Vedha Penmetcha[1, 2*] BA; Lekaashree Rambabu[1, 3, 4*] BSc, MBChB; Brandon G Smith[1, 5] BSc, PhD; Orla Mantle[1, 6] MBBS, MPhil; Tom Edmiston[1, 7] MBBS, BA; Laura Hobbs[1, 4] MBBS; Shobhana Nagraj[1, 5, 8] DPhil, MBBS, MPhil, BSc; Peter H Charlton[8, 9] PhD, MEng; Tom Bashford[1, 5, 4, 10] PhD, MBBS, MBiochem

[1]International Health Systems Group Department of Engineering University of Cambridge Cambridge GB
[2]6100 Main St Wiess School of Natural Sciences Rice University Houston US
[3]Accelerate Programme for Scientific Discovery Department of Computer Science and Technology University of Cambridge Cambridge GB
[4] Department of Anaesthesia Cambridge University Hospitals NHS Foundation Trust Cambridge GB
[5] Cambridge Public Health Interdisciplinary Research Centre University of Cambridge Cambridge GB
[6] GKT School of Medical Education King's College London London GB
[7] Department of Medicine University of Cambridge Cambridge GB
[8] Department of Public Health and Primary Care University of Cambridge Cambridge GB
[9] UCL Biomedical Research Centre London GB
[10] NIHR Global Health Research Group on Acquired Brain and Spine Injury University of Cambridge Cambridge GB
[*]these authors contributed equally

Corresponding Author:
Vedha Penmetcha BA
6100 Main St
Wiess School of Natural Sciences
Rice University
6360 Main St
Rice University
Houston
US

Abstract

Background: Photoplethysmography (PPG) is an optical method for measuring blood volume changes in microcirculation, through non-invasive photodetection. It has become a widespread and essential clinical tool, used in pulse oximeters and wearable devices. However, technical aspects of PPG make it susceptible to intrinsic bias, with the potential to adversely affect particular patient and consumer populations. Developments in PPG technology, increasingly driven by existing datasets as opposed to de novo experimentation, have the potential to help monitor an array of physiological variables. However, some populations may be under-represented in PPG datasets. We describe a protocol for a systematic review to assess the biases within open-access PPG datasets.

Objective: The aim of this review is to critically appraise diversity within PPG datasets. The review will evaluate the demographic characteristics present in open access PPG datasets to evaluate what data is collected in studies and utilised in developing and training new medical technology such as wearables and pulse oximeters. By evaluating the structural components of PPG datasets, we can elucidate current gaps and areas for improvement to reduce systemic bias in PPG based device development.

Methods: This review will be reported in accordance with the standard Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We will include primary studies that mention PPG and specifically reference openly accessible datasets since 2000. The datasets must contain physiological parameters such as heart rate (HR), blood pressure (BP), or respiratory rate (RR) as well as the PPG waveform data, collected from humans. Searches will be conducted in literature databases and data repositories.
Studies will be evaluated in accordance with the Standing Together Initiative recommendations, which are urging for healthcare technologies supported by representative data.

Results: All included studies and datasets will be described by their dataset characteristics and study outcomes using summary

statistics and statistical tests. We will analyse the dataset diversity and the structural basis of PPG datasets, and critically evaluate the different variables included in the datasets. We predict that there will be heterogeneity in the data reporting quality and the terminology used to signify specific variables. By utilising statistical test fit for nominal variable comparisons it will be possible to evaluate the frequencies of dataset characteristics.

Conclusions: This review will provide insight into the potential gaps of existing open-access PPG datasets, and the limitations of studies performed using these datasets. It will inform future collection and design of medical devices including wearables to avoid perpetuating biases, allowing for application in diverse clinical settings. Clinical Trial: PROSPERO [CRD42024564759]

(JMIR Preprints 28/02/2025:73040)
DOI: https://doi.org/10.2196/preprints.73040

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Paper Type: Protocol

Title: Evaluating Diversity in Open Photoplethysmography (PPG) Datasets: a protocol for systematic review

Authors: Vedha Penmetcha*[1,5], Lekaashree Rambabu*[2,3,6], Brandon G. Smith[5,9], Orla Mantle[5,7], Tom Edmiston[2,5], Laura Hobbs[5,6], Shobhana Nagraj[4,5,9], Peter H. Charlton[4,11], Tom Bashford[5,6,8,9]

* Vedha Penmetcha and Lekaashree Rambabu contributed equally as co-first authors

1. Department of Kinesiology, Rice University, Houston, Texas
2. Department of Medicine, University of Cambridge, U.K.
3. Accelerate Programme for Scientific Discovery, Department of Computer Science and Technology, University of Cambridge
4. Department of Public Health and Primary Care, University of Cambridge, U.K
5. International Health Systems Group, Department of Engineering, University of Cambridge, U.K.
6. Department of Anaesthesia, Cambridge University Hospitals NHS Foundation Trust, Cambridge, U.K.
7. GKT School of Medical Education, King's College London, London, U.K.
8. NIHR Global Health Research Group on Acquired Brain and Spine Injury, University of Cambridge, U.K.
9. Cambridge Public Health Interdisciplinary Research Centre, University of Cambridge, U.K.
10. East London NHS Foundation Trust, London, UK
11. Research Centre of Biomedical Engineering, City, University of London, London, UK

Corresponding Author: Vedha Penmetcha
Address: 14428 Goddard St, Overland Park, KS< 66221
Phone Number: +1 913-304-9350
Email: vedhapenmetcha@gmail.com

Keywords: photoplethysmography; ppg; data; diversity; skin tone; bias; equity; fairness

Abstract:

Background

Photoplethysmography (PPG) is an optical method for measuring blood volume changes in microcirculation, through non-invasive photodetection. It has become a widespread and essential clinical tool, used in pulse oximeters and wearable devices. However, technical aspects of PPG make it susceptible to intrinsic bias, with the potential to adversely affect particular patient and consumer populations. Developments in PPG technology, increasingly driven by existing datasets as opposed to *de novo* experimentation, have the potential to help monitor an array of physiological variables.

However, some populations may be under-represented in PPG datasets. We describe a protocol for a systematic review to assess the biases within open-access PPG datasets.

Objective

The aim of this review is to critically appraise diversity within PPG datasets. The review will evaluate the demographic characteristics present in open access PPG datasets to evaluate what data is collected in studies and utilised in developing and training new medical technology such as wearables and pulse oximeters. By evaluating the structural components of PPG datasets, we can elucidate current gaps and areas for improvement to reduce systemic bias in PPG based device development.

Methods

This review will be reported in accordance with the standard Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We will include primary studies that mention PPG and specifically reference openly accessible datasets since 2000. The datasets must contain physiological parameters such as heart rate (HR), blood pressure (BP), or respiratory rate (RR) as well as the PPG waveform data, collected from humans. Searches will be conducted in literature databases and data repositories.

Studies will be evaluated in accordance with the Standing Together Initiative recommendations, which are urging for healthcare technologies supported by representative data.

Outcomes

All included studies and datasets will be described by their dataset characteristics and study outcomes using summary statistics and statistical tests. We will analyse the dataset diversity and the structural basis of PPG datasets, and critically evaluate the different variables included in the datasets. We predict that there will be heterogeneity in the data reporting quality and the terminology used to signify specific variables. By utilising statistical test fit for nominal variable comparisons it will be possible to evaluate the frequencies of dataset characteristics.

Discussion

This review will provide insight into the potential gaps of existing open-access PPG datasets, and the limitations of studies performed using these datasets. It will inform future collection and design of medical devices including wearables to avoid perpetuating biases, allowing for application in diverse clinical settings.

Registration

PROSPERO [CRD42024564759]

Intro

Photoplethysmography (PPG) is an optical method for measuring blood volume changes in microcirculation, through non-invasive photodetection. It has become a widespread and essential tool, providing the sensing technology for pulse oximeters in clinical use, and being widely used in consumer wearables such as smartwatches for heart rate monitoring. However, technical aspects of PPG make it susceptible to intrinsic bias, with the potential to adversely affect particular patient and consumer populations. Developments in PPG technology, increasingly driven by existing datasets as opposed to *de novo* experimentation, have the potential to help monitor an array of physiological variables with both clinical and wellness applications. However, these may disenfranchise those populations who are under-represented in the existing training datasets.

PPG waveforms depict blood volume changes due to changes in absorption and reflection patterns detected by the photodetector (1). PPG measurements are usually obtained by placing the device on the extremities of the body where the vascular bed is shallow and widespread (2). These locations can include the wrist, fingers, toes, and/or ears. Through signal processing, and in some cases the use of different wavelengths of light, PPG can provide information on respiratory rate, blood pressure, sleep patterns, arterial stiffness, and blood oxygen saturation (3). In addition, the periodicity of electrocardiogram (ECG) and PPG waves can be correlated in terms of heart rate variability and arrhythmias, as the waveforms correlate to specific points within the cardiac cycle (4).

The two primary components of a PPG waveform consist of the pulsatile (AC) and the lower frequency (DC) portions. The AC component directly correlates to the volumetric changes associated with the heart rate (5). The DC component, on the other hand, consists of the light absorbed by the tissue, veins, and blood, which correlates to volume capacity (5). The optical measurements are usually taken at red or near infrared (NIR) wavelengths (625nm-750 nm and 750nm-1400nm) or green wavelengths (500-565nm) (6). These are commonly used because the red or NIR wavelengths penetrate up to 2cm beyond the skin surface whereas the green penetrates up to 0.3 mm and allow for the differentiation between oxyhaemoglobin and deoxyhaemoglobin (7-9). However, some PPG methods include the use of multiple wavelengths (MW PPG), including blue, green, and/or yellow light to improve the signal-to-noise ratio (10,11).

PPG, along with pulse oximetry, has limitations attributed to motion, ambient light, sensor contact, pulse pressure reduction, and disturbances in contact between the sensor and the skin. One key limitation which has recently been highlighted is that of inaccuracy across different skin tones (7,12,13). *Sjoding et. al* recently reported that black patients had almost three times the frequency of occult hypoxia – falsely elevated oxygen saturation levels – compared to their white counterparts (14). Occult hypoxia can lead to higher risk of mortality and organ failure for darker-skin toned patients and has been shown to reduce the provision of life saving treatment modalities like supplemental oxygen and medications. This is clearly of concern, both in terms of clinical devices but also for the possible biases within commercial technology which systematically underserve select populations. Furthermore, as clinical monitoring and the availability of wearable technology increases globally without correctional strategies, this bias will only worsen.

While the original experiments in pulse oximetry involved permissive hypoxia of test subjects, a number of current PPG developments are largely driven by training on available datasets. However, the biases within openly available datasets have not been described. This makes it difficult to understand their possible impact on both clinical and consumer technology. We describe a protocol for the first systematic review of which we are aware designed to assess the biases within currently open-access PPG datasets. This has immediate value for those seeking to derive physiological data from PPG technology using training datasets. By focusing on openly accessible datasets and repositories we will provide insight into the demographic characteristics of the data that is commonly used for a majority of existing PPG technology including wearables.

Methods

This review will be reported in accordance with the standard Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The study has been registered with PROSPERO [CRD42024564759] (https://www.crd.york.ac.uk/prospero/) prior to the initial literature search. The full planned search strategy is provided in the Supplementary Material, which has been developed in close concordance with a medical librarian (IK and EE). This search strategy includes the databases and data repositories: MedLine OVID, IEEE, SCOPUS, Physionet, National Sleep

Research Resource, UC Irvine Machine Learning Repository, and Borealis.

Eligibility Criteria

We will include any peer-reviewed primary studies – observational and non-observational – which utilise PPG datasets. This includes PPG datasets in published randomised controlled trials, non-randomised trials, cross-sectional studies, cohort studies, and algorithmic models.

We will include studies published since 2000 to ensure the datasets are relevant to current technology developments. We plan to include studies that mention PPG and specifically reference openly accessible datasets. The datasets must contain physiological parameters such as heart rate (HR), blood pressure (BP), or respiratory rate (RR) as well as data describing the waveform, such as pulse wave or slope transit time. These terms have been chosen as they are indicators of parameters routinely collected in PPG datasets.. The inclusion and exclusion criteria is outlined in Table 1. Included studies will be evaluated for the extent to which they include demographic characteristics and consider these characteristics.

Table 1: Summary of inclusion and exclusion criteria

Inclusion	Exclusion
● Openly accessible datasets ● Literature containing datasets years 2000 - present ● Must contain PPG signal and physiological parameters such as HR, BP, or RR. ● Human studies	● Animal studies ● Lack of PPG signals ● Primary literature without reference to datasets ● Conference papers and abstracts

Outcomes

We predict that there will be heterogeneity in the data reporting quality and the terminology used to signify specific variables, therefore it is important to utilise broad inclusion criteria to capture relevant articles.

We also anticipate significant heterogeneity of demographic characteristics. *Jiang et. al* report that less than seven percent of physiological datasets, specifically those available on PhysioNet, include all 4 key demographic variables – age, sex, race, and ethnicity (15). Consequently, it is important to not include these variables in our inclusion or exclusion criteria to truly understand the level to which PPG datasets consider demographic aspects.

Data Analysis

Following the initial literature search, all studies will be reviewed independently by two authors using systematic review software, Rayyan.ai (https://www.rayyan.ai/). Studies will initially be screened by title and abstract. The full-text articles will then undergo screening. Disagreements will be discussed with and resolved by a third author, a senior author with expertise in the topic. At each step, studies will be retrieved only if they meet the eligibility criteria. The studies will be

assessed using the Standing Together Initiative guidelines as a framework for evaluating the datasets, as there are no widely accepted tools for assessing the risk of bias of datasets (16,17). The Standing Together Initiative provides guidelines that address dataset diversity and inclusion as a means of assessing the quality of training datasets and their potential to contribute to health inequity (15). Though these recommendations do not make up a definitive risk of bias tool, they provide content areas for analysing the datasets.

After dataset extraction a summary of all included studies and datasets will be provided by describing their dataset characteristics and study outcomes. Due to the heterogeneity of the studies and datasets included, the summary will provide an overview of the current state of PPG based dataset repositories. Aspects such as the devices, biosignals collected, clinical parameters (sample size), demographic characteristics (age, sex, race/ethnicity), and geographical information (country, healthcare setting, dates) will provide insight into the dataset characteristics.

Additionally, summary statistics will be provided based on the characteristics of the datasets. Statistical analysis will be performed on the dataset characteristics based on parameters such as frequency and other quantitative aspects extracted from the studies (17). Raw counts will be visualised as percentages or proportions to depict the array of demographic characteristics reported in the datasets. The mentioning of these characteristics, such as age, sex, race, and ethnicity, will be analysed using Chi-Squared for Frequency and Independence (14). These tests will provide insight into the differences on both the frequencies and the co-occurrence of demographic characteristics, respectively. The Kruskal-Wallis test will provide information on whether the frequency of particular variables or demographic characteristics significantly differs across datasets. Further analysis of the literature will contextualize why such differences may or may not be statistically significant. Visualisations such as histograms and will include information on the mean ages, dataset sizes, and number of waveform characteristics reported. This analysis will provide insight into how datasets are structured and the amount of information reported in them to qualify potential reasons for statistically significant differences in dataset reporting. Additional information such as geographical or socioeconomic data will be displayed cartographically or in a heatmap to depict the sources of PPG studies and datasets.

Discussion

Wearables including smartwatches are commonly used to track personal health measures such as heart rate, sleep, and blood pressure. These commercially available medical devices often utilise PPG to take physiological measurements. However, PPG data collected by wearables have been observed to be affected by skin tone and higher adipose tissue due to obesity (12). In combination, differences of up to 61% in signal intensity have been observed (12). The Fitzpatrick scale is commonly used to assess for differences in skin tone, and while it provides an insight into the physiological signal values caused by greater melanin levels, it is subjective, and treats skin tone as a categorical, rather than a continuous, variable (18). This has limitations when considering the many underlying factors that result in the chromatic appearance of an individual's skin tone, particularly if complicated by clinical conditions such as jaundice or hypoxia. Exploring the extent to which variables such as this are captured within existing datasets may be instructive in understanding their inherent ability to account for overall skin colouration.

Clearly, this same need to account for diversity in skin tone and clinical condition applies to clinical PPG studies and their resultant datasets. In a clinical setting, having erroneous oxygen saturation values or heart rate measurements can result in poorer clinical outcomes. There is evidence that the obese and those with greater melanin content in their skin suffer from worse health outcomes, with inaccurate monitoring a clear potential contributor (18). Obesity can be attributed to

physiological changes such as the reduction in trans-epidermal water loss (TEWL) and an increase in skin thickness due to adipose tissue can influence the light scattering patterns needed to obtain a PPG reading (18). In addition to melanin content and obesity, other factors influencing perfusion and skin temperature can all affect PPG. When expanding PPG technology to low resource settings there is a need to account for diverse populations not only in terms of melanation but also factors such as obesity and conditions altering perfusion. Moreover, they may have a greater diversity of endemic clinical conditions including anaemia, parasitaemia, and malnutrition – whose effects on PPG have not been clearly described. Effective clinical PPG monitoring in these patients mandates an interpretation based on data from representative populations.

In this study, we will analyse the structural basis of PPG datasets and critically evaluate the different variables included in the datasets. By critically appraising the dataset diversity, we can highlight the gaps present in current PPG data to inform future collection and design of medical devices including wearables to be more inclusive to avoid perpetuating biases. Additionally, by looking at the structure of PPG datasets and their representativeness can also provide insight into the gaps that may be present in the context of PPG applications in low resource settings.

Acknowledgements:

We would like to acknowledge the medical librarians Isla Kuhn and Emma Etteridge for their help in developing the search strategy for this systematic review.

References

1. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron. 2018;4(4):195–202.

2. Park J, Seok HS, Kim SS, Shin H. Photoplethysmogram Analysis and Applications: An Integrative Review. Front Physiol [Internet]. 2022 Mar 1 [cited 2025 Jan 20];12. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.808451/full

3. Almarshad MA, Islam MS, Al-Ahmadi S, BaHammam AS. Diagnostic Features and Potential Applications of PPG Signal in Healthcare: A Systematic Review. Healthcare. 2022 Mar;10(3):547.

4. Tang Q, Chen Z, Guo Y, Liang Y, Ward R, Menon C, et al. Robust Reconstruction of Electrocardiogram Using Photoplethysmography: A Subject-Based Model. Front Physiol [Internet]. 2022 Apr 25 [cited 2025 Jan 20];13. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.859763/full

5. Loh HW, Xu S, Faust O, Ooi CP, Barua PD, Chakraborty S, et al. Application of photoplethysmography signals for healthcare systems: An in-depth review. Comput Methods Programs Biomed. 2022 Apr 1;216:106677.

6. Chen SH, Chuang YC, Chang CC. Development of a Portable All-Wavelength PPG Sensing Device for Robust Adaptive-Depth Measurement: A Spectrometer Approach with a Hydrostatic Measurement Example. Sensors. 2020 Nov 17;20(22):6556.

7. Feiner, John R., Severinghaus, John W.; Bickler, Philip E. Dark Skin Decreases the Accuracy of Pulse Oximeters at Low Oxygen Saturation: The Effects of Oximeter Probe Type and Gender | Anesthesia & Analgesia. [cited 2025 Jan 20]. Available from: https://journals.lww.com/anesthesia-analgesia/fulltext/2007/12001/dark_skin_decreases_the_accuracy_of_pulse.4.aspx

8. Abay TY, Kyriacou PA. Photoplethysmography for blood volumes and oxygenation changes during intermittent vascular occlusions. J Clin Monit Comput. 2018 Jun;32(3):447–55.

9. Moço AV, Stuijk S, de Haan G. Skin inhomogeneity as a source of error in remote PPG-imaging. Biomed Opt Express. 2016 Oct 26;7(11):4718–33.

10. Liu J, Yan BPY, Dai WX, Ding XR, Zhang YT, Zhao N. Multi-wavelength photoplethysmography method for skin arterial pulse extraction. Biomed Opt Express. 2016 Sep 27;7(10):4313–26.

11. McMurray JP, Branan KL, Hsiao CT, Idah-Oze S, Thrailkill A, Coté GL. Multiwavelength photoplethysmography signal analysis as a function of varied wrist contact pressure. In: Optical Diagnostics and Sensing XXIII: Toward Point-of-Care Diagnostics [Internet]. SPIE; 2023 [cited 2025 Jan 20]. p. 31–5. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12387/1238707/Multiwavelength-photoplethysmography-signal-analysis-as-a-function-of-varied-wrist/10.1117/12.2649531.full

11. Monte Carlo analysis of optical heart rate sensors in commercial wearables: the effect of skin tone and obesity on the photoplethysmography (PPG) signal | Biomedical Optics Express. 2021 Nov 10 [cited 2025 Jan 20]. Available from: https://opg.optica.org/boe/fulltext.cfm?uri=boe-12-12-7445&id=464758

13. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. N Engl J Med. 2020 Dec 16;383(25):2477–8.

14. Racial Bias in Pulse Oximetry Measurement | New England Journal of Medicine [Internet]. [cited 2025 Jan 20]. Available from: https://www-nejm-org.laneproxy.stanford.edu/doi/full/10.1056/NEJMc2029240

15. Jiang S, Ashar P, Shandhi MMH, Dunn J. Demographic reporting in biosignal datasets: a comprehensive analysis of the PhysioNet open access database. Lancet Digit Health. 2024 Nov 1;6(11):e871–8.

16. Alderman JE, Palmer J, Laws E, McCradden MD, Ordish J, Ghassemi M, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations. Lancet Digit Health. 2025 Jan 1;7(1):e64–88.

17. Alderman JE, Charalambides M, Sachdeva G, Laws E, Palmer J, Lee E, et al. Revealing transparency gaps in publicly available COVID-19 datasets used for medical artificial intelligence

development—a systematic review. Lancet Digit Health. 2024 Nov 1;6(11):e827–47.

18. Wound care and skin tone: signs, symptoms and terminology for all skin tones – Wounds Asia [Internet]. [cited 2025 Jan 20]. Available from: https://woundsasia.com/journal-articles/wound-care-and-skin-tone-signs-symptoms-and-terminology-for-all-skin-tones/

Supplementary Files

Multimedia Appendixes

Search Strategy.
URL: http://asset.jmir.pub/assets/918f96525d914a9ddd9f506167c5db44.docx