Inaccuracies have been reported in pulse oximetry measurements taken from people who identified as Black. Here, we identify substantial ethnic disparities in the population numbers within 12 pulse oximetry databases, which may affect the testing of new oximetry devices and impact patient outcomes.

There has been an unprecedented demand for pulse oximetry—a method of determining the oxygen saturation (SpO2) of the blood—during the COVID-19 pandemic to aid in medical decision-making. Pulse oximetry data are also widely used for medical research and algorithm development. The measurement of SpO2 involves shining light onto tissue at two separate wavelengths and derives the oxygen saturation from the relative changes in light absorption with each heartbeat. This pulsatile component is independent of skin pigmentation; however, other factors, such as the specific properties of the light source and the algorithms used by the product manufacturer, can produce variations that depend on skin pigmentation.

A recent study revealed the clinical importance of racial disparities in pulse oximetry readings. Specifically, when compared to measurements of arterial oxygen saturation, the pulse oximetry algorithms in these devices were found to produce systematically higher saturation values in Black patients compared to white patients. Such systematic racial biases could adversely affect clinical decision-making, such as triage for supplemental oxygen, due to the pulse oximetry readings of Black patients appearing artificially higher. Such inaccuracies may disproportionately increase the risk of unrecognized low oxygen saturation levels in Black patients under certain circumstances, for example in people with COVID-19.

Recent advancements in artificial intelligence (AI) have relied on using public databases to undertake feature extraction with pulse oximetry signals to assess hypertension, estimate lung function, and validate algorithms developed for monitoring patients with COVID-19. To prevent potential disparities in the calibration and accuracy of pulse oximetry devices and their algorithms, the pulse oximeter signals within such public databases need to be representative of the diverse populations on which these devices are used.

Since inaccuracies in pulse oximetry readings have been attributed to differences in skin pigmentation and skin pigmentation varies with race and ethnicity, it is essential to clarify these terms as used in this article. We have predominantly chosen to use the term ethnicity because that is the specific term used in the public datasets and is also the term that is identified by the patients themselves. For the purpose of discussing health disparities, while both race and ethnicity are social constructs, ethnicity has emerged as the preferred one since it encompasses cultural aspects of social identity that extend beyond the more simplistic view of race that is primarily based on shared skin pigmentation or physical characteristics. While members of a
given ethnicity can express a range of skin pigmentation, it is generally agreed that those patients who self-identify as Black generally have a darker skin pigmentation than other ethnic groups.

In order to investigate the proportion of individual ethnicities represented in publicly available pulse oximetry databases, we conducted a comprehensive assessment of accessible databases from 1st January 2012–1st January 2022 using PubMed consisting of Medical Subject Headings (MeSH) terms and Title/Abstract keywords. Applying the inclusion and exclusion criteria defined in Fig. 1 resulted in 12 research articles describing 12 publicly-available datasets to assess different medical conditions using pulse oximeter data.

In total, as of January 28th 2022, these databases have been used to produce over 6214 citations according to Google Scholar including 3544 citations for Medical Information Mart for Intensive Care (MIMIC III)9; 1049 citations for MIMIC II10; 531 citations for IEEEPPG Dataset11; 243 citations for Multiparameter Intelligent Monitoring in Intensive Care I (MIMIC I)12; 239 citations for WESAD13; 215 citations for Vortal Dataset14; 102 citations for the CapnoBase Dataset15; 87 citations for the University of Queensland Vital Signs Dataset16; 86 citations for PPG-DaLiA17; 63 citations for PPG-BP Dataset18; 50 citations for Wrist PPG Signals Recorded during Exercise19; and 5 citations for Medical Information Mart for Intensive Care IV (MIMIC-IV)20. We evaluated the existence of potential disparities in ethnicity based on the existing patient records as reported in the publicly available databases. In the absence of such information, the numbers of subjects of each category were inferred and quantified based on the authors’ research institutions’ locations or where the data was collected, as shown in Table 1.

To avoid any uncertainty in the results of ethnic disparity analysis for a given population, databases with inferred ethnicity information were excluded from the statistical analysis. Four databases for which data for ethnicity was clearly stated, MIMIC, MIMIC-II, MIMIC-III and MIMIC-IV, were included in the statistical analysis. The distribution of ethnic groups in the four databases is shown in Fig. 2.

Fig. 1 Flow chart of study identification, inclusion and exclusion criteria. Presentation of the literature search and selection procedure together with the numbers of records at each stage sourced from PubMed from 1 January 2012 to 1 January 2022 using the following keywords: (‘critical care’[MeSH Terms] OR ‘critical’[All Fields] AND “care”[All Fields]) OR “critical care”[All Fields] OR (“oxygen”[MeSH Terms] OR “oxygen”[All Fields] AND “saturation”[All Fields]) OR “oxygen saturation”[All Fields]) OR (“o2”[All Fields] AND “saturation”[All Fields]) OR “o2 saturation”[All Fields]) OR “PPG”[All Fields] OR (“photoplethysmograms”[All Fields]) OR (“photoplethysmography”[MeSH Terms] OR “photoplethysmography”[All Fields])) AND (“publicly”[All Fields] AND (“availability”[All Fields] OR “available”[All Fields])) OR (“freely”[All Fields] AND (“access”[All Fields] OR “accessed”[All Fields] OR “accesses”[All Fields] OR “accessibilities”[All Fields] OR “accessibility”[All Fields] OR “accessible”[All Fields] OR “accessing”[All Fields]))) AND ((humans[Filter]) AND (english[Filter])). Papers that met the inclusion criteria, discussing the development and publication of original pulse oximetry datasets, were selected for analysis. This resulted in 12 research papers representing 12 publicly available datasets. Here n refers to the number of studies, where m refers to the number of publicly available databases.
We tested the statistical significance among all the subjects in the four databases considering a p-value < 0.05 as statistically significant and analyzed the variance using a one-way ANOVA followed by post hoc test to provide simultaneous two-way interactions using the Tukey’s honest significant difference criterion. The results indicated that there was a significant difference between the mean distributions of all racial groups; Asian and Black (p = 0.021), Asian and white (p = 4.10 × 10⁻⁹²), and Black and white (p = 5.01 × 10⁻⁶⁷). The same trend was observed between Other and Asian (p = 9.43 × 10⁻⁵⁹), Other and Black.

| Title                                                                 | Authors (year of publication) | Name of database | Number of subjects | Country of data location | Ethnicity is clearly stated or inferred | Ethnicity | Number of subjects based on ethnic group |
|----------------------------------------------------------------------|-------------------------------|-----------------|---------------------|--------------------------|----------------------------------------|-----------|------------------------------------------|
| A database to support development and evaluation of intelligent intensive care monitoring | Moody and Mark¹² (1996)       | MIMIC I         | 93                  | US                       | Clearly stated                        | 2.4% Asian | 2 Asian, 8 Black, 18 Other, 65 White       |
| CapnoBase: signal database and tools to collect, share and annotate respiratory signals | Karlen et al.¹⁵ (2010)         | CapnoBase       | 42                  | Canada                   | Inferred                              | 4.10 × 10⁻⁹² | 4 Asian, 1 Black, 6 Other, 31 White         |
| Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database | Saeed et al.¹⁰ (2011)          | MIMIC-II        | 32000               | US                       | Clearly stated                        | 2.5% Asian | 800 Asian, 3392 Black, 5664 Other, 22144 White |
| University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research | Liu et al.¹⁶ (2012)           | University of Queensland Vital Signs | 32                  | Australia                 | Inferred                              | 3.1% Asian | 1 Asian, 0 Black, 9 Other, 22 White         |
| TROIKA: a general framework for heart rate monitoring using wrist-type photoplethysmographic (PPG) signals during intensive physical exercise | Zhang et al.¹¹ (2015)         | IEEEPGG         | 12                  | China                     | Inferred                              | 100.0% Asian | 12 Asian, 0 Black, 0 Other, 0 White       |
| MIMIC-III, a freely accessible critical care database                | Johnson et al.⁹ (2016)        | MIMIC-III       | 53423               | US                       | Clearly stated                        | 2.4% Asian | 1282 Asian, 4113 Black, 9937 Other, 38091 White |
| An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram | Chariton et al.¹⁴ (2016)       | Vortal          | 45                  | UK                       | Inferred                              | 7.5% Asian | 4 Asian, 2 Black, 3 Other, 36 White         |
| Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion | Jarchi and Casson¹⁹ (2016)     | Wrist PPG Signals Recorded during Exercise | 8                  | UK                       | Inferred                              | 100% Other | 1 Asian, 1 Black, 0 Other, 0 White         |
| Introducing WESAD, a multimodal dataset for wearable stress and affect detection | Schmidt et al.¹³ (2018)       | WESAD           | 15                  | Germany                   | Inferred                              | 2.5% Asian | 0 Asian, 0 Black, 2 Other, 13 White         |
| A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China | Liang et al.¹⁸ (2018)          | PPG-BP          | 219                 | China                     | Inferred                              | 88.2% White | 203 Asian, 0 Black, 16 Other, 0 White      |
| Deep PPG: large-scale heart rate estimation with convolutional neural networks | Reiss et al.¹⁷ (2019)          | PPG-DaLiA       | 15                  | Germany                   | Inferred                              | 2.5% Asian | 0 Asian, 0 Black, 0 Other, 2 Other         |
| MIMIC-IV                                                             | Johnson et al.²⁰ (2021)       | MIMIC-IV        | 60000               | US                       | Clearly stated                        | 72.9% White | 13 White, 1800 Asian, 6000 Black, 6000 Other, 46200 White |

The numbers of subjects and distribution of different ethnic groups as identified or inferred in the publicly available datasets. In case of the absence of patient ethnicity information, the location of the authors’ research institutions’ or the data collection location was used to infer ethnicity based on local statistics. For Vortal and Wrist PPG Signals Recorded during Exercise databases, we inferred the ethnicity of subjects based on the Institute of Race Relations (https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prod/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=PR&Code2=01&Data=Count&SearchText=canada&SearchType=Begin&SearchP3=01&Tab1=All&TABID=1). As an example, the Black population accounts for 3.5% of Canada’s total population (https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prod/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=PR&Code2=01&Data=Count&SearchText=canada&SearchType=Begin&SearchP3=01&Tab1=All&TABID=1).
examine the demographic statistics of each data set, based on
citly stated, the ethnic disparity is not known. However, if we
graphic groups are provided in Table 2.
results on the statistical separability tests for all pairs of demo-
ject appeared in all four MIMIC databases where the ethnicity
and ethnic biases in algorithms and devices, leading to possible
Our findings highlight clear disparities in pulse oximetry
databases. As these biased databases would be used during the
development of new pulse oximeter databases with more
racial biases in algorithms and devices, leading to possible
Since last year, the US Food and Drug Administration has
led to an increased number of data-
approach to the matter of constructing reference databases,
mainly when applied to vascular optical measurement technology
that can be influenced by skin color characteristics. White sub-
jects appeared in all four MIMIC databases where the ethnicity
was clearly stated, consisting of an average of 73.19% of the total
population. However, Black subjects only accounted for an
average of 9.29% of the sample population. In addition, Asian
subjects comprised an average of 2.67% of the total population
and ethnic biases in algorithms and devices, leading to possible
challenges in their wider application in medicine.

Our findings highlight clear disparities in pulse oximetry
databases. As these biased databases would be used during the
premarket phase to adjust pulse oximeter accuracy and to develop
algorithms for oxygen saturation determination, they place sub-
jects with darker skin pigmentation at increased risk of unrec-
ognized health conditions. Such health inequalities necessitate
the development of new pulse oximeter databases with more
racial biases in algorithms and devices, leading to possible
challenges in their wider application in medicine.

Data availability
Supplementary Data 1 contains source data for the main figures in this manuscript. Pulse
oximetry databases can be accessed via the following links: MIMIC-I (https://www.
physionet.org/content/mimicibs1.0.0/); CapnoBase (https://dataverse.harvard.edu/
info/dataverse/capnobase---text=The%20CapnoBase%20benchmark%20dataset%20
contains,that%20may%20arise%20during%20anesthesia); MIMIC-II (https://archive.
physionet.org/physiobank/database/mimicii/); MIMIC-III (https://physionet.org/physiobank/database/mimiciii/); University of Queensland Vital Signs
(https://oufbox.east.uq.edu.au/ouflj100s/oufitalsgndataset/index.html#:~:text=Introduciton%40the%20Royal%20Adelaide%20Hospital.); IEEEPPG (https://zenodo.
org/record/3932710?persistent=1); MIMIC-IV (https://physionet.org/content/
physionet.org/content/mimiciv/1.0/).

Table 2 Tukey simultaneous tests for differences of means.

| Difference of levels | Difference of means | SE of difference | 95% CI       | T-value | Adjusted P-value |
|----------------------|---------------------|-----------------|--------------|---------|------------------|
| Asian-black          | 6.77                | 2.00            | (0.84, 12.71)| 3.39    | <0.0001          |
| Asian-other          | 13.78               | 2.00            | (7.84, 19.71)| 6.89    | <0.0001          |
| Asian-white          | 69.38               | 2.00            | (63.44, 75.31)| 34.69  | <0.0001          |
| Black-other          | 7.00                | 2.00            | (1.06, 12.94)| 3.50    | 0.02             |
| Black-white          | 62.60               | 2.00            | (56.66, 68.54)| 31.30  | <0.0001          |
| Other-white          | -55.60              | 2.00            | (-61.54, -49.66)| -27.80| <0.0001          |

Fig. 2 Box plots of the ethnic makeup by proportion in all databases.
This figure combines all the databases used in the four publicly available
pulse oximeter databases that clearly reported the distribution of ethnic
groups. The data supports the hypothesis that disparities exist here.
Significant differences are evident between white and Black (p < 0.0001),
white and Asian (p < 0.0001), and Black and Asian populations
(p = 0.021). All pairs of groups were tested by using a simultaneous
pairwise Tukey test. The bottom and the top of the box are the 25th and
75th percentiles, and the line inside the box is the 50th percentile
(median). Whiskers from minimum to maximum are determined with a
95% confidence interval.
Received: 29 June 2021; Accepted: 3 May 2022; Published online: 27 May 2022

References
1. Colvonen, P. J. Response to: investigating sources of inaccuracy in wearable optical heart rate sensors. npj Digit. Med. 4(1), 1–2 (2021).
2. Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E. & Valley, T. S. Racial bias in pulse oximetry measurement. N. Engl. J. Med. 383, 2477–2478 (2020).
3. FDA. Pulse oximeter accuracy and limitations: FDA safety communication, https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication (2021).
4. Elgendi, M. et al. The use of photoplethysmography for assessing hypertension. npj Digit. Med. 2, https://doi.org/10.1038/s41746-019-0136-7 (2019).
5. Jiang, W. et al. A wearable tele-health system towards monitoring COVID-19 and chronic diseases. IEEE Rev. Biomed. Eng. 15, 61–84 (2021).
6. Luks, A. M. & Swenson, E. R. Pulse oximetry for monitoring patients with COVID-19 at home: potential pitfalls and practical guidance. Ann. Am. Thorac. Soc. 17(9), 1040–1046 (2020).
7. Ford, C. L. & Harawa, N. T. A new conceptualization of ethnicity for social epidemiologic and health equity research. Soc. Sci. Med. 71(2), 251–258 (2010).
8. Harawa, N. T. & Ford, C. L. The foundation of modern racial categories and implications for research on black/white disparities in health. Ethn. Dis. 19, 209–217 (2009).
9. Johnson, A. E. W. et al. MIMIC-III, a freely accessible critical care database. Sci. Data 3, 160035 (2016).
10. Saed, M. et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. Crit. Care Med. 39(5), 952 (2011).
11. Zhang, Z., Pi, Z. & Liu, B. TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. IEEE Trans. Biomed. Eng. 62, 522–531 (2015).
12. Moody, G. B. & Mark, R. G. A database to support development and evaluation of intelligent intensive care monitoring. Comput. Cardiol. 1996, 657–660 (1996).
13. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C. & Laerhoven, K. V. Introducing WISAD, a multimodal dataset for wearable stress and affect detection, in Proceedings of the 20th ACM International Conference on Multimodal Interaction 400–408 (Association for Computing Machinery, 2018).
14. Charlton, P. H. et al. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. Physiol. Meas. 37, 610–626 (2016).
15. Karlen, W., Turner, M., Cooke, E., Dumont, G. & Ansermino, J. M. CapnoBase: signal database and tools to collect, share and annotate respiratory signals. 2010 Annual Meeting of the Society for Technology in Anesthesia, West Palm Beach, Florida, January 13–16, 2010 (2010).
16. Liu, D. et al. University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. Anesth. Analg. 114(3), 584–589 (2012).
17. Reiss, A., Indlekofer, L., Schmidt, P. & Van Laerhoven, K. Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors 19, 3079 (2019).
18. Liang, Y., Chen, Z., Liu, G. & Elgendi, M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. Sci. Data 5, 180020 (2018).
19. Jarchi, D. & Casson, A. Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion. Data 2, 1 (2016).
20. Johnson, A. et al. MIMIC-IV (version 1.0). PhysioNet. https://doi.org/10.13036/sfn6-xd98 (2021).

Author contributions
M.E. designed and led this investigation. F.S., R.W., D.A., J.A., R.F., C.M. and M.E. conceived the study. All authors approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-022-00121-8.

Correspondence and requests for materials should be addressed to Mohamed Elgendi.

Peer review information Communications Medicine thanks Steve Greenwald and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.