https://doi.org/10.1038/s42256-023-00614-8

An extension to the FDA approval process is needed to achieve AI equity

More than thirty years have passed since evidence regarding differential pulse oximetry accuracy afflicting racial and ethnic minoritized groups emerged[1]. Yet, testimony before the US Food and Drug Administration (FDA) in November 2022 has made clear that new generations of pulse oximeters continue to perpetuate the problem, generating flawed readings that "most likely contributed to the several-fold greater number of deaths in COVID-19 in ethnic minority patients than in white patients," according to Dr. Amal Jubran, a pulmonary critical care doctor at Loyola Medicine in Chicago[1]. As of 2022, the FDA's approval process for medical devices and software does not consider race or ethnicity, which results in harmful, even fatal, consequences for marginalized groups[2]. The patient populations that the FDA currently considers to assess medical devices before they are brought to market are primarily white men and, to a lesser extent, white women[2]. In this current form, the FDA's approval process lacks a health equity focus, as demonstrated by the manifold disparities that arise from medical imaging devices, pulse oximeters and infrared thermometers, which have been implemented without being tested on a representative cohort of patients and have consequently resulted in readout fallacies[3–5]. Despite these well-documented disparities, the most recent FDA guidelines to improve trial diversity lack concrete accountability measures[2]. To achieve health equity, the FDA should mandate that manufacturers test their medical devices and software on diverse patient populations. In addition, companies should provide information on the composition of patients who participated in the design and calibration of the device or software.

A recent study found that artificial intelligence (AI) deep learning models can identify the race of a patient purely based on chest images from X-rays (area under the receiver operating characteristics curve (AUC): 0.91–0.99), CT machines (AUC: 0.87–0.96) and mammograms (AUC: 0.81)[3]. The risk inherent in such imaging devices is that AI models can conceivably leverage race as a feature in informing decisions through algorithms that use medical images as input, thereby perpetuating or exacerbating existing racial disparities. Particularly for medical imaging, this risk is compounded by the fact that human experts cannot identify racial identity from medical images, which means that human oversight of AI models is of limited use in mitigating this problem[3]. This issue creates an enormous risk for all model deployments in medical imaging: if an AI model relies on its ability to detect racial identity to make medical decisions but in doing so produces race-specific errors that clinical radiologists — who do not typically have access to racial demographic information — cannot account for, such factors may unduly influence care decision processes.

Similarly, a recent study on pulse oximeters found that, compared with white patients, Black, Hispanic and Asian patients treated in intensive care units have more significant discrepancies between blood saturation levels that are detected using pulse oximeters

Table 1 | Suggested extensions of FDA premarket approval pathway requirements to advance health equity

Additional requirements for FDA technical section	Example
Racial and ethnic distributions of the patient cohort should reflect the general or disease-specific population.	The racial composition of clinical trials for the COVID-19 vaccines reflected the general US population: • Pfizer: 81.9% white, 9.8% Black, 4.4% Asian, 0.6% American Indian/Alaska Native, 0.2% Native Hawaiian or Other Pacific Islander, 26.2% Hispanic[5] • Moderna: 79.4% white, 9.7% Black, 4.7% Asian, 0.8% American Indian/Alaska Native, 0.2% Native Hawaiian or Other Pacific Islander, 20.0% Hispanic[5] As people of colour were disproportionately affected by COVID-19, Pfizer and Moderna both partnered with historically Black colleges and universities and minority-serving institutions[5].
Medical device and software approvals should require a detailed implementation and market strategy plan that elucidates how the product will be made accessible to marginalized populations.	Recent medical devices and software offer templates for how manufacturers may be able negotiate with insurers to secure reimbursement, thereby allowing for wider accessibility for marginalized patients. • IDx-DR (Digital Diagnostics), an AI diagnostic system which detects diabetic retinopathy without clinician oversight, is the first autonomous software to be reimbursed by Medicare under the Medicare Physician Fee Schedule and the Outpatient Prospective Payment System[6]. • The Viz LVO software for stroke detection and Caption Guidance software for cardiac ultrasound imaging acquisition are reimbursable by Medicare under the New Technology Add-on Payments in the Inpatient Prospective Payment System[6].
New medical devices and software should rely on something other than prior measurements known to be error-prone, to prevent existing disparities from being magnified downstream.	AI algorithms that rely on medical imaging devices, pulse oximeters and infrared thermometers may amplify inherent biases, causing physicians to deliver disparate care to patients, such as prescribing or terminating the use of supplemental oxygen because of artificially high blood oxygen readouts for Black patients[1].
Clinicians should be trained in the use of newly approved medical devices and software, especially those known to cause racial disparities.	On average, Black patients have more dense breasts than white patients. As dense breast tissue manifests as white areas in mammograms, whereas fatty tissue appears as dark areas, mammograms are less sensitive in women with dense breasts. As a result, physicians may miss potential tumours in Black patients and should therefore be trained to accurately detect dark areas of dense breast tissue on mammograms in order to prevent racial disparities[7].

and blood tests[1]. Falsely elevated readings heighten the risk for hidden hypoxemia, which occurs at higher incidence among racial and ethnic minority groups and is associated with higher mortality rates[1]. Additionally, many clinical AI models rely on oxygen saturation readings to determine diagnoses, which makes downstream errors challenging to detect. Finally, infrared thermometers also mirror this same trend of fallacy[4]. A recent study from Emory University showed that — because of skin colour discrepancies — forehead thermometers had a 26% lower chance of detecting fever in Black patients than oral thermometers, leading to missed fevers, delayed diagnoses and antibiotic treatment, and increased death rates compared with white patients[4]. Medical devices and software ultimately generate racial bias because manufacturers do not design and calibrate them on diverse patient cohorts before deploying them in clinical settings. As such, it is imperative to create approval processes that ensure a broader representation of patients during clinical trials.

Based on its recent 2022 report entitled "Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan," the FDA anticipates that medical devices and software will require a specialized review process as existing processes were not designed for these technologies[2]. The FDA calls for a process similar to its most stringent premarket pathway (PMA), which every manufacturer must complete to scientifically demonstrate device safety and effectiveness[2]. We recommend an extension to the PMA application regarding the subsection on clinical investigations[2]. Although the clinical investigation part must adhere to specific requirements that are promulgated by the FDA, none of these requirements includes consideration that disease may affect people differently. Although the FDA recommends that sponsors include clinically relevant populations no later than the end of phase two, it is not mandating it[2]. Instead, the FDA emphasizes that the IEC, the committee of outside experts tasked with ensuring subject safety and rights, has several members that reflect the sociocultural diversity of the communities from which research participants are drawn. The clinical investigation section of the PMA can potentially be improved in terms of its study protocols, patient information and study design (Table 1).

Efforts to improve representation in clinical testing are not unprecedented; in 2020, the FDA offered nonbinding recommendations to the pharmaceutical industry to increase clinical trial diversity concerning vaccinations in the wake of disproportionately large numbers of people of colour being most severely affected by COVID-19 (ref. [5]). In their COVID-19 vaccine trials, Pfizer and Moderna sought to represent the demographic spread of the US population, in part by partnering with historically Black colleges and universities[5]. Moderna also delayed its enrollment period to ensure thresholds of representation were met for various minority groups[5]. Ultimately, these efforts led Pfizer and Moderna to enroll a heterogeneous mixture of clinical trial participants that mimicked the racial composition of the US, within a margin of error between 2–3% for racial minorities[5]. These efforts are laudable, but since the approval of the COVID-19 vaccines, the FDA's recommendation has largely been met with words instead of actions, with many organizations solely acknowledging the importance of diversity[2]. Also, a complex interplay of factors, including a lack of perception of the issue among investigators and history-rooted mistrust among patients, limits minority inclusion in trials[5]. For instance, physicians have been shown to present patients of colour with opportunities for clinical trials less frequently than white patients[5]. Enrollment efforts often suffer from a lack of cultural appropriateness or an inability to tackle language and health literacy disparities[5]. Finally, trials are averse to selecting from hospitals more likely to care for people of colour, as they are less likely to have health insurance[5].

To facilitate representative samples of racially diverse populations, as seen in the COVID-19 Pfizer and Moderna trials, the FDA should foster collaborations among pharmaceutical companies and minority-serving institutions to enroll additional patients and broaden manufacturing cycles to prioritize recruitment of underrepresented patients. While the roll-out of medical devices and software offers society an array of new technologies, the pharmaceutical industry and the FDA must acknowledge that barriers exist for the most vulnerable. To realize the equity in health outcomes all patients deserve, the FDA must look to examine and test medical devices and software across racially diverse populations.

Alessandro Hammond [1,2], **Bhav Jain** [3], **Leo Anthony Celi** [3,4,5] & **Fatima Cody Stanford** [6] ✉

[1]Harvard University, Cambridge, MA, USA. [2]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Boston, MA, USA. [3]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [6]Massachusetts General Hospital, MGH Weight Center, Department of Medicine-Division of Endocrinology-Neuroendocrine, Department of Pediatrics-Division of Endocrinology, Nutrition Obesity Research Center at Harvard (NORCH), Harvard Medical School, Boston, MA, USA.

✉e-mail: fstanford@mgh.harvard.edu

Published online: 7 February 2023

References

1. Jewett, C. Doctors Warn F.D.A. of Health Risks Posed by Flawed Oxygen Devices. *The New York Times* (November 2022); https://www.nytimes.com/2022/11/02/health/pulse-oximeters-black-patients.html
2. US Food & Drug Administration. Artificial intelligence and machine learning in software as a medical device. *FDA* https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (2021).
3. Gichoya, J. W. et al. *Lancet Digit. Health* **4**, e406–414 (2022).
4. Advisory Board. These medical devices produce racially biased results. Here's why that matters. *Daily Briefing* https://www.advisory.com/Daily-Briefing/2022/09/28/medical-devices (2022).
5. Artiga, S., Kates, J., Michaud, J. & Hill, L. Racial diversity within COVID-19 vaccine clinical trials: key questions and answers. *KFF* https://www.kff.org/racial-equity-and-health-policy/issue-brief/racial-diversity-within-covid-19-vaccine-clinical-trials-key-questions-and-answers/ (2021).
6. Abràmoff, M. D. et al. *npj Digit. Med.* **5**, 72 (2022).
7. Moore, J. X. et al. *JNCI Cancer Spectr.* **4**, pkaa010 (2010).

Acknowledgements

L.A.C. acknowledges funding from National Institute of Health NIBIB R01 EB017205, and F.C.S. acknowledges funding from National Institutes of Health NIDDK P30 DK040561 and U24 DK132733.

Competing interests

The authors declare that they have no competing interests.