Limiting racial disparities and bias for wearable devices in health science research

Peter J. Colvonen[1,2,3,4,*], **Pamela N. DeYoung**[5], **Naa-Oye A. Bosompra**[5], **Robert L. Owens**[5]

[1]VA San Diego Healthcare System, San Diego, CA,

[2]Department of Psychiatry, University of California, San Diego, San Diego, CA,

[3]VA Center of Excellence for Stress and Mental Health, San Diego, CA,

[4]National Center for PTSD, White River Junction, VT

[5]Department of Medicine: Division of Pulmonary, Critical Care and Sleep Medicine, University of California, San Diego, San Diego, CA

Consumer wearables are devices used for tracking activity, sleep, and other health-related outcomes (e.g. Apple Watch, Fitbit, Samsung, Basis, Mio, PulseOn, Whoop). Wearables promise a myriad of health-related information, including low heart rate alerts, a personal electrocardiogram (ECG) monitor for detecting arrhythmia, sleep tracking (e.g. sleep architecture), and pulse pressure designed to promote healthy living and alert high-risk consumers based on real-time data. Their relative low cost, the collection of longitudinal data, and the ability to display/transmit information suggest a host of benefits if used in clinical practice and to advance remote research. Unfortunately, due to technological limitations of photoplethysmographic (PPG) green light signaling, these health constructs may only be accessible for a population of people with lighter skin tones. We note with deep concern that there is increasing evidence that these devices are not as accurate, and may not work at all, in people with darker skin tones [1, 2]. The reduced accuracy of wearable devices in people with darker skin tones seems to have been known for some time, yet this issue has garnered little attention from the medical community. Apart from limited media coverage [3], anecdotal reports, and minimal research publications [1, 2, 4, 5], the role of dark skin tones on wearable accuracy is a severely underreported phenomenon that requires direct attention and action from industry and the scientific community. Our concern is amplified as these devices are now transitioning from consumer goods into health-related research and their internal algorithms are becoming Food and Drug Administration (FDA) approved. As such, they stand to exacerbate existing structural health disparities of people with darker skin tones and may compound existing structural health disparities for Black Americans [5]. The Black Lives Matter movement is calling for everyone to unveil and dismantle systemic bias throughout society. We challenge the healthcare community to work towards this goal and to ensure that digital health solutions do not reinforce existing

*Corresponding author. Peter J. Colvonen, VA San Diego Healthcare System, 3350 La Jolla Village Drive (116B), San Diego, CA 92161. pcolvonen@health.ucsd.edu.

disparities in care and access as these devices are increasingly used in research and clinical practice.

One of the issues with wearables and skin tone is a technical challenge, in that wearables use a PPG green light signal. Blood, due to its color, readily absorbs green light, where the greater the volume of blood present (i.e. when the heart pumps during systole), the higher the green light absorption and thus heart rate can be measured accordingly. This green light technology is cheap and generally robust, however, skin tone affects the absorption of light differently, which interferes with the algorithm output. Studies have shown that green light lacks precision and accuracy, and may not read at all, when measuring heart rate in darker skin types [2]. There are also several other factors that influence PPG accuracy that may cause an interaction with skin tone, including tattoos, presence of arm hair, sweat, ambient temperature, level of activity, thickness of skin epidermis [6], and body mass [7]. This potentially affects health-related outcomes that rely on PPG including a host of sleep-related indices such as sleep duration, architecture [8], diagnosis of sleep apnea [9], long-term sleep patterns, and other sleep disorders [10]. Despite this, green light technology remains the industry standard in wearables.

A second problem with the current state of wearables research seems to be the lack of diversity in validation studies. The consumer wearable industry *modus operandi* is to bring to market a minimally viable product to the largest population as quickly as possible. Test marketing (there are too few true validation studies published) is often done with a local population that may lack a diversity of skin tones, and objective measurement of skin pigmentation is rarely reported. A lack of accurate information about how wearables work in diverse skin tones may cause unintended consequences by reinforcing existing healthcare disparities for those with darker skin tones. Inaccurate signal detection and validation in those with darker skin might lead to decreased utilization of novel wearable devices, or worse, provide false assurances about the effectiveness of monitoring. Thus, we urge both the industry and researchers to ensure that validation studies are performed in a diverse population.

A confounding factor in accurately understanding the limitations of wearables on skin tone is the current standard of measuring skin tone: the Fitzpatrick Skin Type Scale (FST) [11]. Developed in 1975 by individuals with white skin for individuals with white skin [12], the FST is a subjective scale that classifies six skin type categories according to skin pigmentation and skin's reaction to sun exposure. There is a substantive literature examining the racial biases and limitations of the FST [12–14]. The phototype designation with six categories has been shown to have only a weak correlation with skin color with large within-group variance of skin tone [13, 15, 16]. Further, the FST and other subjective measures (e.g. Taylor Pigmentation Scale) have been shown to be inaccurate and biased based on the administrator [17]. As such, the use of the FST may give false assurances about skin tone and wearable effectiveness. For example, a small study ($N = 53$) examined the accuracy of several wearables using the FST and found, "no statistically significant difference in accuracy across skin tones." [4] We fear that these conclusions are misleading because too few people with the darkest skin tones were included ($n = 9$), as assessed by the FST rather than objective measurement. The most objective solution is the use of reflectance

spectrometry which accurately identifies skin color/tone using multiple color wavelengths for classification [18], and should be the gold standard for all validation studies examining skin tone.

There is a wearable revolution in healthcare coming, with the FDA and National Institutes of Health (NIH) calling for increased remote clinical outcomes to patient-generated measures that do not require clinician oversight; the Covid-19 crisis is accelerating the appeal for remote monitoring in clinical trials [19]. The application of wearables in clinical and research arenas offers enormous potential to increase health information and access to interventions; however, we caution that incomplete metrics about these devices may cause unintended consequences by reinforcing existing healthcare disparities for those with darker skin tones. Action is needed as increasing numbers of studies are undertaken without information about skin tone and wearable performance. For example, Fitbit alone has approximately 476 published studies and 449 studies registered on ClinicalTrials.gov [4] and the Apple Watch Heart Study recruited 419 927 people to track irregular heart rhythms [20]. There are further opportunities for inequalities as large US-based companies (e.g. IBM and Target) are starting to offer trackers as part of wellness programs. The historical standards of research and reporting on wearables must be improved upon.

Consumer wearable companies seem to be aware their devices lack accuracy in people with dark skin tones and a few companies are actively working with varying approaches to manage this deficiency. For example, Fitbit reported that it has increased the power of their green light transmitter, and newer versions of the Apple Watch have an additional infrared light sensor (which may be less affected by skin tone and more accurate than green light). Both Garmin [21] and Everion [22] are the most forthcoming by directly letting consumers know their validation studies show PPG data may not be accurate in individuals with dark skin. At the current time, however, there is minimal data publically available to understand the scope and magnitude of the potential problem, and whether the solutions above actually work as the pace of technology advances faster than researchers can keep up [3].

As scientists, we hold colleagues and ourselves to a high level of integrity, accountability, and ethical standards, when conducting human research and reporting our findings. With regard to wearables, our challenge to the scientific community is to consider how the inaccuracies of PPG technology for individuals with dark skin contribute to health disparities. These disparities in access to health care exacerbate the harm that social structures and policies cause to the health of Black Americans [5]. We believe that educating ourselves, and the research community at large, is critical. This includes: (1) directly working with wearables companies to improve upon their effectiveness and consumer reach to support people of color; (2) decreasing use of the Fitzpatrick scale and increasing reporting of more objective, non-offensive, standards of skin tone; (3) urging companies to advance their technology (e.g. using multiple wavelengths for varying skin tone [23], improved fit, or using hospital-grade technology); 4) holding the research community accountable for addressing and reporting bias; and 5) making sure that people of varying skin tones are included in validation and effectiveness research. Technological advancements are muted if their inherent biases continue historical structural health disparities [24]. It is

only through direct action to define and mitigate these biases that we will all benefit equally from the coming revolution in healthcare.

Funding

References

1. Shcherbina A, et al.Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med. 2017;7(2):3.

2. Fallow BA, et al.Influence of skin type and wavelength on light wave reflectance. J Clin Monit Comput. 2013;27(3):313–317. [PubMed: 23397431]

3. Hailu RFitbits and other wearables may not accurately track heart rates in people of color. 2019. https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/.AccessedJuly 20, 2020.

4. Bent B, et al.Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med. 2020;3:18. [PubMed: 32047863]

5. Feiner JR, et al.Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg. 2007;105(6 Suppl):S18–S23, tables of contents. [PubMed: 18048893]

6. Moço AV, et al.Skin inhomogeneity as a source of error in remote PPG-imaging. Biomed Opt Express. 2016;7(11):4718–4733. [PubMed: 27896011]

7. Castaneda D, et al.A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron. 2018;4(4):195–202. [PubMed: 30906922]

8. Fonseca P, et al.Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. Sleep. 2017;40(7). doi:10.1093/sleep/zsx097.

9. Shokoueinejad M, et al.Sleep apnea: a review of diagnostic sensors, algorithms, and therapies. Physiol Meas. 2017;38(9):R204–R252. [PubMed: 28820743]

10. Boe AJ, et al.Automating sleep stage classification using wireless, wearable sensors. NPJ Digit Med. 2019;2:131. [PubMed: 31886412]

11. Fitzpatrick TB. Sun and skin. Journal de Medecine Esthetique. 1975;2:33–34.

12. Pichon LC, Landrine H, Corral I, Hao Y, Mayer JA, Hoerster KD. Measuring skin cancer risk in African Americans: is the Fitzpatrick Skin Type Classification Scale culturally sensitive. Ethn. Dis2010;20(2):174–179. [PubMed: 20503899]

13. Ware OR, et al.Racial limitations of Fitzpatrick skin type. Cutis. 2020;105(2):77–80. [PubMed: 32186531]

14. Galindo GR, et al.Sun sensitivity in 5 US Ethnoracial groups. Cutis. 2007;80(1):25–30. [PubMed: 17725060]

15. Yun IS, et al.Skin color analysis using a spectrophotometer in Asians. Skin Res Technol. 2010;16(3):311–315. [PubMed: 20637000]

16. Xiao K, et al.Characterising the variations in ethnic skin colours: a new calibrated database for human skin. Skin Res Technol. 2017;23(1):21–29. [PubMed: 27273806]

17. Fider NA, et al.Differences in color categorization manifested by males and females: a quantitative World Color Survey study. Palgrave Commun. 2019;5(1): 1–10.

18. Pershing LK, et al.Reflectance spectrophotometer: the dermatologists' sphygmomanometer for skin phototyping?J Invest Dermatol. 2008;128(7):1633–1640. [PubMed: 18200057]

19. Turner JR. New FDA guidance on general clinical trial conduct in the era of COVID-19. Ther Innov Regul Sci. 2020;54:723–724. [PubMed: 32557317]

20. Perez MV, et al.; Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med. 2019;381(20):1909–1917. [PubMed: 31722151]

21. Garmin. The heart rate sensor on my watch is not accurate. 2020. https://support.garmin.com/en-US/?faq=xQwjQjzUew4BF1GYcusE59.AccessedAugust 1, 2020.

22. Everion. What are the limitations & restrictions of the device?2020. https://biovotion.zendesk.com/hc/en-us/articles/213178049-What-are-the-limitations-restrictions-of-the-device- AccessedAugust 1, 2020.

23. Yan L, Hu S, Alzahrani A, Alharbi S, Blanos P. A multi-wavelength opto-electronic patch sensor to effectively detect physiological changes against human skin types. Biosensors. 2017;7(2):22.

24. Morris NThe race problem with Artificial Intelligence: "Machines are learning to be racist Read more."2020. https://metro.co.uk/2020/04/01/race-problem-artificial-intelligence-machines-learning-racist-12478025/?ito=cbshare.AccessedAugust 1, 2020.