



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Smartphone Biosensor With App Meets FDA/ISO Standards for Clinical Pulse Oximetry and Can Be Reliably Used by a Wide Range of Patients



Sara H. Browne, MD; Mike Bernstein, BEng; Samuel C. Pan, MD; Jonathan Gonzalez Garcia, MD; Craig A. Easson, MSc Eng; Chung-Che Huang, MSc Eng; and Florin Vaida, PhD



**BACKGROUND:** Millions of smartphones contain a photoplethysmography (PPG) biosensor (Maxim Integrated) that accurately measures pulse oximetry. No clinical use of these embedded sensors is currently being made, despite the relevance of remote clinical pulse oximetry to the management of chronic cardiopulmonary disease, and the triage, initial management, and remote monitoring of people affected by respiratory viral pandemics, such as severe acute respiratory syndrome coronavirus 2 or influenza. To be used for clinical pulse oximetry the embedded PPG system must be paired with an application (app) and meet US Food and Drug Administration (FDA) and International Organization for Standardization (ISO) requirements.

**RESEARCH QUESTION:** Does this smartphone sensor with app meet FDA/ISO requirements? Are measurements obtained using this system comparable to those of hospital reference devices, across a wide range of people?

**STUDY DESIGN AND METHODS:** We performed laboratory testing addressing ISO and FDA requirements in 10 participants using the smartphone sensor with app. Subsequently, we performed an open-label clinical study on 320 participants with widely varying characteristics, to compare the accuracy and precision of readings obtained by patients with those of hospital reference devices, using rigorous statistical methodology.

**RESULTS:** “Breathe down” testing in the laboratory showed that the total root-mean-square deviation of oxygen saturation ( $SpO_2$ ) measurement was 2.2%, meeting FDA/ISO standards. Clinical comparison of the smartphone sensor with app vs hospital reference devices determined that  $SpO_2$  and heart rate accuracy were 0.48% points (95% CI, 0.38-0.58;  $P < .001$ ) and 0.73 bpm (95% CI, 0.33-1.14;  $P < .001$ ), respectively;  $SpO_2$  and heart rate precision were 1.25 vs reference 0.95% points ( $P < .001$ ) and 5.99 vs reference 3.80 bpm ( $P < .001$ ), respectively. These small differences were similar to the variation found between two FDA-approved reference instruments for  $SpO_2$ : accuracy, 0.52% points (95% CI, 0.41-0.64;  $P < .001$ ) and precision, 1.01 vs 0.86% points ( $P < .001$ ).

**ABBREVIATIONS:** app = application; FDA = US Food and Drug Administration; HR = heart rate; ISO = International Organization for Standardization; LMICs = low- and middle-income countries; PPG = photoplethysmography; RMSD = root-mean-square deviation;  $SpO_2$  =

oxygen saturation; UCSD = University of California, San Diego; UCSF = University of California, San Francisco

**AFFILIATIONS:** From the University of California, San Diego (S. H. Browne, S. C. Pan, J. Gonzalez Garcia, and F. Vaida), La Jolla, CA;

**INTERPRETATION:** Our findings support the application for full FDA/ISO approval of the smartphone sensor with app tested for use in clinical pulse oximetry. Given the immense and immediate practical medical importance of remote intermittent clinical pulse oximetry to both chronic disease management and the global ability to respond to respiratory viral pandemics, the smartphone sensor with app should be prioritized and fast-tracked for FDA/ISO approval to allow clinical use.

**TRIAL REGISTRY:** [ClinicalTrials.gov](https://www.clinicaltrials.gov); No.: NCT04233827; URL: [www.clinicaltrials.gov](https://www.clinicaltrials.gov);

CHEST 2021; 159(2):724-732

**KEY WORDS:** chronic disease management; remote clinical pulse oximetry; respiratory viral pandemics; smartphone sensor with app

FOR EDITORIAL COMMENT, SEE PAGE 477

Millions of smartphones contain photoplethysmography (PPG) biosensors with applications (apps) that accurately measure heart rate (HR) and blood oxygen saturation (SpO<sub>2</sub>).<sup>1,2</sup> High-grade PPG biosensors (Maxim Integrated) are currently integrated into Android smartphones worldwide, totaling more than 300 million smartphones.<sup>3</sup> The PPG sensor measures the distension of arteries and arterioles in the subcutaneous tissue, due to blood flow with each cardiac cycle. As blood flows through the vessels, the pulse pressure is detected by illuminating the skin with the light from two light-emitting diodes. The amount of light transmitted, absorbed, or reflected to a photodiode is measured. A signal-processing app containing an algorithm then detects and interprets the PPG signal to determine SpO<sub>2</sub> and HR values. Finally, additional software within the app then accesses and displays individual biosensor readings on the smartphone screen.

Direct use of this technology for remote clinical pulse oximetry is highly relevant to the expanding integration of digital technologies into clinical care models. The usefulness of smartphones with embedded sensors and

apps for intermittent clinical pulse oximetry is broad, potentially supporting the effective management of a wide range of chronic cardiopulmonary disease, including congestive heart failure, COPD, or acute disorders such as pneumonia, and postoperative recovery.<sup>4-6</sup> Smartphone sensors with apps for clinical pulse oximetry would allow patients to gather and track their own data to inform outpatient clinic or telemedicine visits, and increase communication with health-care providers, which may lead to earlier outpatient interventions that potentially reduce both morbidity and hospitalization risks. Smartphone sensors with apps for clinical pulse oximetry may be of considerable importance in low- and middle-income countries (LMICs) with less developed health-care infrastructure, enabling management support of infections prevalent in these settings such as AIDS-defining pneumonias like pneumocystis or cryptococcosis, TB, as well as chronic cardiopulmonary diseases.<sup>7-13</sup>

Further, during events of pandemic respiratory viral spread, such as the current severe acute respiratory syndrome coronavirus 2 or influenza, remote clinical pulse oximetry may support not only triage, but also the initial treatment of symptomatic adults.<sup>14,15</sup> SpO<sub>2</sub> correlates with lung injury<sup>15</sup> and in conjunction with clinical examination may be used to differentiate those who require close monitoring and hospital admission from those with milder disease.<sup>6</sup> In circumstances of home quarantine, remote clinical pulse oximetry could allow patients to monitor their symptoms and include objective reports of SpO<sub>2</sub> and HR. Smartphone sensors with apps could allow measurements collected, with patient permissions, to be connected to regional and national data hubs allowing large-scale regional data analysis in near-real time. This could allow estimates of resource distribution, such as hospital beds needed, and could

---

Specialists in Global Health (S. H. Browne), Encinitas, CA; Physio Monitor, LLC (M. Bernstein), San Ramon, CA; and Maxim Integrated (C. A. Easson and C.-C. Huang), San Jose, CA.

**FUNDING/SUPPORT:** This work was funded by a National Institutes of Health grant supplement [R01MH110057-04S] to S. H. B., a grant from Maxim Integrated to cover S. C. P., and a Technology Award from nonprofit Specialists in Global Health (<https://sigh.global/>) to J. G. G.

**CORRESPONDENCE TO:** Sara H. Browne, MD, Altman Clinical and Translational Research Institute, University of California, San Diego, 9452 Medical Center Dr, La Jolla, CA 92037; e-mail: [sbrowne@health.ucsd.edu](mailto:sbrowne@health.ucsd.edu)

Copyright © 2020 The Authors. Published by Elsevier Inc under license from the American College of Chest Physicians. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**DOI:** <https://doi.org/10.1016/j.chest.2020.08.2104>

provide means for medical centers to contact and communicate with patients in home quarantine. The availability of smartphone sensors with apps for clinical pulse oximetry under such circumstances could be critical to LMICs, where medical infrastructure is easily overwhelmed.

Despite the existence of such a large number of smartphones embedded with these sensors with apps, no clinical use is being made of these. The clinical pulse oximetry requirement by the US Food and Drug Administration/International Organization for Standardization (FDA/ISO) for PPG sensors is < 3.5% root-mean-square deviation (RMSD) of the SpO<sub>2</sub>

value. Herein we present data obtained using a smartphone sensor with app (Maxim Integrated)<sup>16,17</sup> and report laboratory testing addressing FDA/ISO requirements.

Subsequently, we report findings from an open-label clinical study across a varied population, using a protocol that obtains the internal measurement error for smartphone sensor with app readings and hospital standard of care reference instruments, and allows comparison of measurement sets to analyze both accuracy and precision. Data obtained using the smartphone sensor with app across a varied population help determine how accurately it can be used by the general population.

## Methods

### Laboratory Testing of Smartphones Embedded With PPG Biosensors Against FDA/ISO Standards

A “breathe down” test, a technique introduced by J. Severinghaus (University of California, San Francisco [UCSF]) and refined by P. Bicker (UCSF),<sup>18,19</sup> was performed in the certified Bickler-Ye Laboratory (Shenzhen University, China) under IRB PN-2019-01 and international human subjects research standards. The smartphone sensor with app consisted of (1) manufactured part MAX 86100A and (2) an app containing a signal-processing determination algorithm that detects and interprets PPG signal (MaximFast) (Maxim Integrated; and was evaluated as implemented in the ZUK phone [Lenovo]). The test set consisted of 10 volunteers comprising black (three), white (three), and Asian (four) racial/ethnic origins. All volunteers provided written informed consent. In each test volunteers placed their index finger over the smartphone sensor system, and a strap was placed over their hand to maintain it in place. The volunteer breathed a mixture of gases with reduced O<sub>2</sub> compared with room air. By manipulating this mixture, the volunteer’s SpO<sub>2</sub> was decreased in steps. At each step, the measurements were allowed to reach a stable plateau, which was held for at least 45 s, before proceeding to the next step. Only the last 30 s of each plateau was used in the data analysis. Data were collected at SpO<sub>2</sub> levels near 100% to about 70%. SpO<sub>2</sub> values were recorded continuously, and an FDA/ISO-approved Masimo Radical reference device (Masimo) recorded SpO<sub>2</sub> simultaneously.

HR measurements data were taken with a ProSim 8 Fluke instrument (Fluke Biomedical).

**Laboratory Testing Data Analysis:** Each plateau was checked for quality of data and was rejected if the plateau was not stable within ± 2% SpO<sub>2</sub>; either the Masimo reference or the test device did not report a value; motion of the volunteer was observed; or perfusion, defined as optical modulation of the transmitted light, was less than 0.5%. Data were analyzed according to methods established by Bland and Altman,<sup>20</sup> providing RMSD and a Bland-Altman plot with upper and lower limits of agreement per Bland Altman as described in ISO 80601-2-61:2017.<sup>16</sup> A second-order polynomial fit of the data was calculated.

**Clinical Study Conduct:** An open-label clinical protocol was developed and approved by the University of California, San Diego (UCSD) Human Research Protections Program (#161573). Measures of HR and SpO<sub>2</sub> from “test” units, that is, Android smartphone model ZUK Z2 X (ZUK; Lenovo) with embedded PPG biosensor and app (MAX 86100A

and signal-processing algorithm MaximFast) (Maxim Integrated). The phones were enclosed within plastic cases (Fig 1) containing a circular indent with raised grooves to guide and stabilize finger placement over the PPG sensor. Vital sign measures were obtained simultaneously using two “reference standard of care” Welch Allyn Spot Vital Signs (WASVS) units [models (91)42NTB and (93)2268; Welch Allyn]. Participants were adults over 18 years of age recruited from UCSD Healthcare or Clinical Trials Units and who provided written informed consent. Initially, participants were recruited from outpatient settings (n = 250). For each participant, four pairs of measurements (total, eight data points) were taken in succession, with no break between sets, in the order indicated in Table 1 (top). This is a repeated measure, nested-factorial design, with three factors (measurement system, left/right index finger, experimental step), instrument units nested within measurement system (two units for each system), and repeated measures over participants.<sup>21</sup> This design allows statistical evaluation



Figure 1 – Smartphone with embedded photoplethysmography sensor suite and plastic case with braille guide for digit placement used as the test instrument.

**TABLE 1 ] Order and Device Combination Used to Take Simultaneous Heart Rate and Oxygen Saturation Measurements in Outpatient Participants (n = 250) and Inpatient/ED Participants (n = 70)**

| Setting              | Step | Left Index Finger | Right Index Finger |
|----------------------|------|-------------------|--------------------|
| Outpatient (n = 250) | 1    | Reference unit 1  | Reference unit 2   |
|                      | 2    | Test unit 1       | Test unit 2        |
|                      | 3    | Test unit 2       | Reference unit 1   |
|                      | 4    | Reference unit 2  | Test unit 1        |
| Inpatient (n = 70)   | 1    | Test unit 1       | Reference unit     |
|                      | 2    | Reference unit 1  | Test unit 1        |

and comparison of test and reference measurement systems, while correcting for other possible sources of variation, including within-participant vital sign change over four experimental steps, and left and right index finger differences. The hypotheses tested the accuracy and precision between (1) test and reference units, (2) two test units, and (3) two reference units.

Subsequently, participants were recruited from inpatient and ED settings (n = 70) and followed a simplified protocol (Table 1, bottom) and no distinction was made between the Welch Allyn reference units, which were used interchangeably. Each measurement period was 2 min in length and respiratory rate was determined. Participant demographic data, including age, height, weight, sex, ethnicity, and hand dominance, were collected. Patient participants were verbally screened for the following comorbidities as exclusion criteria: coarctation of the aorta, severe peripheral vascular disease, severe anemia, severe sickle cell disease, and methemoglobinemia. However, none of the patient participants screened reported having these conditions.

**Clinical Study Statistical Methods:** Participant demographics were summarized. Erroneous readings with a recorded value of zero indicating a measurement was not obtained were reported and excluded from further analysis. The measurement error of an instrument is summarized by the root-mean-square deviation of prediction, RMSD, which is the square root of the mean-square deviation (error) between the value reported by the instrument and the true value. The RMSD can be further decomposed into the bias and SD of the instrument:

$$\text{RMSD}^2 = \text{Bias}^2 + \text{SD}^2 \quad (1)$$

where bias refers to the systematic (average) deviation of the instrument from the true value, and SD summarizes the variation of the instrument in repeated measures.<sup>22</sup> Thus, bias measures accuracy and SD measures precision. The design of our clinical studies allows statistical analyses that tease out separately the bias and the SD of the two measurement systems and four units.

For the outpatient setting, accuracy and precision hypotheses were addressed by fitting linear mixed-effects models to HR and SpO<sub>2</sub> outcomes. The model equation is

$$Y_{ij} = \beta_1 + b_i + \beta_2 \times S_i + \beta_3 \times T_i + \beta_4 \times R_i + \beta_5 \times \text{Side}_{ij} + \beta_6 \times \text{Step}_{ij} + \varepsilon_{ij} \quad (2)$$

$Y_{ij}$  is the outcome (HR or SpO<sub>2</sub>) for participant  $i$  at measurement  $j$  ( $j = 1, 2, \dots, 8$ ).  $S_i$  indicates the measurement system ( $S_i = 1$  for test and  $S_i = 0$  for reference units);  $T_i$  indicates which test unit was used,  $T_i = -1/2$  for test unit 1,  $T_i = 1/2$  if test unit 2 was used (for reference unit measurements,  $T_i = 0$ );  $R_i$  indicates the reference units used,  $R_i = -1/2$  for unit 1,  $R_i = 1/2$  for unit 2 ( $R_i = 0$  for test units);  $\text{Side}_{ij} = -0.5$  for left-hand and  $0.5$  for right-hand

measurements;  $\text{Step}_{ij}$  indicates the step of measurement, as in Table 1 (top), but centered at 0;  $b_i$  is a subject-specific random effect, common to all measurements of participant  $i$ ; and  $\varepsilon_{ij}$  is within-participant error, assumed to have a mean of 0. Furthermore, the within-participant errors  $\varepsilon_{ij}$  were allowed to be autocorrelated, with an autoregressive of order 1 (AR1) structure, and to have variances that differ between groups—specifically, test and reference systems or the four measurement units.<sup>23</sup>

The predicted biological value of outcome (HR or SpO<sub>2</sub>) for generic subject  $i$  is  $\beta_1 + b_i$ . This allows a “best guess” of outcome based on eight measurements, freed of biases of particular units and of other experimental conditions (side, step), and free of within-participant error  $\varepsilon_{ij}$ . It corresponds to the reference system ( $S_i = 0$ ) taken as a gold standard. The parameters  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  address accuracy aspects of hypotheses:  $\beta_2$  measures the bias of test relative to reference units (by definition the reference unit bias is zero and  $\text{RMSD} = \text{SD}$ );  $\beta_3$  measures the bias of test units relative to each other; and  $\beta_4$  measures the bias of reference units relative to each other. The precision aspects are captured by the SD of  $\varepsilon_{ij}$ , which is allowed to differ between test and reference systems (when comparing the two), or between individual units (when comparing the units). The  $R_i = \pm 1/2$  parametrization was chosen so that the gold standard is the average of the two reference units, and so that  $\beta_4$  measures the bias between them. Similarly,  $T_i = \pm 1/2$  so that the bias between systems compares the average of the two test units vs the average of the two reference units, and so that  $\beta_3$  measures the bias between the two test units.

The additional experimental conditions Side and Step are not central to our hypotheses, but they act as potential confounders, and may help in estimating the bias and precision of the instruments. These two factors were excluded if not significant at the 0.20 level in backward model selection.<sup>24</sup> The statistical comparisons and 95% CIs used the Wald test for the mean (accuracy) coefficients, and the likelihood ratio test for the SD (precision) parameters. The choice between AR1 and independence correlation structure for within-participant errors  $\varepsilon_{ij}$  used the likelihood ratio test. A sensitivity analysis was performed, removing outlier observations defined as  $\geq 3$  times their system-specific SD. These large outliers are likely due to the incorrect use of the system (eg, finger placement), and the sensitivity analysis results reflect the accuracy of the two systems under appropriate use conditions.

For the inpatient setting respiratory rates were summarized, with accuracy and precision evaluation simplified to two steps, using a single test and reference unit based on a similar mixed-effects models as for the outpatient setting but without terms for units  $T_i$  and  $R_i$ . Because of the small number of outliers for this setting a similar sensitivity analysis is not reported.

Statistical significance is considered at the  $P < .05$  level. All analyses were conducted with the R statistical language<sup>25</sup> and used the *nlme* package.<sup>26</sup>

## Results

### Laboratory Testing of Biosensor Plus App

Test data were collected from 10 participants during “breathe down” testing (see Methods), comparing in-phone sensors and Masimo Radical pulse oximetry device SpO<sub>2</sub> values. A Bland-Altman plot of the entire data set with second-order polynomial line fit is shown in Figure 2. In this study eight plateaus were rejected for lack of stability within two SpO<sub>2</sub> counts, rejecting the corresponding 16 data points. All 10 subjects completed the study, and there were no adverse events.

The RMSD total average for data from all 10 volunteer participants tested was 2.2%. Heart rate error based on analysis of Fluke simulation testing showed that the heart rate error was < 2 bpm RMSD.

### Clinical Study of Test-vs-Reference Measurements

Participant demographics confirmed a broad range of age and race/ethnicity (Table 2).

In the outpatient setting test units gave 960/1,000 (96.0%) valid HR and 955/1,000 (95.5%) valid SpO<sub>2</sub> readings. Reference units gave 997/1,000 (99.7%) valid HR and 997/1,000 (99.7%) valid SpO<sub>2</sub> readings. The range of valid HR readings was 41 to 131 bpm for test, and 38 to 126 bpm for reference units. The range of valid SpO<sub>2</sub> readings was 90% to 99% for test, and 92% to 100% for reference units. In the inpatient/ED setting test units gave 137/140 (97.9%) valid HR readings (range, 41-131 bpm) and 136/140 (97.1%) valid SpO<sub>2</sub> readings (range, 90%-99%). Reference units gave 138/140 (98.6%) valid HR readings (range, 38-126 bpm) and 138/140 (98.6%) valid SpO<sub>2</sub> readings (range, 92%-100%). The respiratory rate averaged 12 breaths/min (median, 10; SD, 4.1; range, 7-30 breaths/min).

**Bias (Accuracy) of Measurement Systems:** In the outpatient setting for HR, a significant bias was found for the test relative to reference measurement system (bias  $\beta_2 = 0.73$  bpm [95% CI, 0.33-1.14];  $P < .001$ ) (Table 3, Fig 3). Significant bias was found between the two reference units ( $\beta_4 = -0.33$  bpm [95% CI, -0.66 to 0.00];  $P = .049$ ) but not between the two test units ( $\beta_3 = -0.36$  bpm [95% CI, -0.99 to 0.27];  $P = .27$ ) (Table 4).

For SpO<sub>2</sub>, the test units had a significant positive bias compared with reference units ( $\beta_2 = 0.48\%$  points [95% CI, 0.38-0.58% points];  $P < .001$ ) (Table 3). No significant bias was found between the two test units ( $\beta_3 = -0.06\%$  [95% CI, -0.22 to 0.10];  $P = .43$ ). However, a significant bias was found between the two

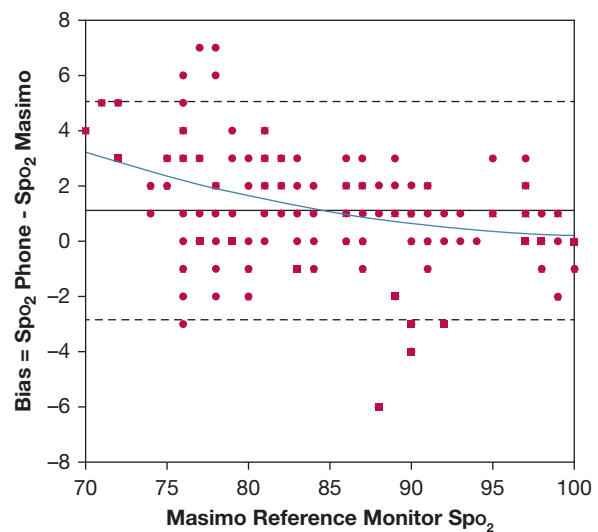


Figure 2 – Plot of data collected from 10 volunteers during a “breathe down test” (see Methods). Each point on the plot represents one pair of a phone reading and Masimo Radical 7 reading. The root-mean-square deviation of bias between the two instruments in this test was 2.2 oxygen saturation counts. The upper and lower limits of agreement per Bland and Altman<sup>20</sup> are shown as dashed lines. A second-order polynomial fit of the data is shown as a solid line. SpO<sub>2</sub> = oxygen saturation.

reference units ( $\beta_4 = 0.52$  points [95% CI, 0.41-0.64 points];  $P < .001$ ) (Table 4).

In the inpatient/ED setting for HR, no significant bias (accuracy) was found for the test relative to the reference (Table 3, Fig 3A). For SpO<sub>2</sub>, test units had significantly lower readings than reference units (Table 3, Fig 3B).

**Precision (Variation) of Measurement Systems:** In the outpatient setting for HR, the test measurement system had significantly higher variation (worse precision) than the reference system (SD = 5.99 bpm [95% CI, 5.61-6.40 bpm] vs 3.80 bpm [95% CI, 3.56-4.06 bpm];  $P < .001$ ) (Table 3, Fig 3). The precision differed significantly between the two test units: SD = 6.92 bpm for test unit 1 and 4.95 bpm for test unit 2 ( $P < .001$ ). The precision did not differ significantly between the two reference units ( $P = .27$ ) (Table 4). Similar findings were observed in the inpatient/ED setting (Table 4).

For SpO<sub>2</sub>, the test measurement system had significantly higher variation (worse precision) than the reference system (SD = 1.25 points [95% CI, 1.17-1.33 points] vs 0.95 points [95% CI, 0.89-1.01 points];  $P < .001$ ) (Table 3, Fig 3). The precision did not differ between the two test units (1.26 vs 1.25 points;  $P = .90$ ); however, it differed significantly between the two reference units (SD = 1.01 and 0.86 points;  $P = .001$ ) (Table 4). Similar findings were observed in the inpatient/ED setting (Tables 3 and 4, Figs 3A and 3B).

**TABLE 2 ] Demographics of Participants in Both Outpatient and Hospital Settings**

| Demographic Characteristic  | All Participants<br>(N = 320) | Participants in Outpatient Setting<br>(n = 250) | Participants in Hospital Setting<br>(n = 70) |
|-----------------------------|-------------------------------|---|--|
| Age, mean (SD), y           | 44.4 (14.3)                   | 46.1 (13.1)                                     | 38.5 (16.5)                                  |
| Range                       | 18-89                         | 19-81   | 18-89  |
| Sex, No. (%)                |                               |   |  |
| Male                        | 229                           | 189 (75.6)                                      | 40 (57.1)                                    |
| Female                      | 89                            | 59 (23.6)                                       | 30 (42.9)                                    |
| Other                       | 2                             | 2 (0.8)   | 0 (0)  |
| Race and ethnicity, No. (%) |                               |   |  |
| White                       | 129                           | 99 (39.6)                                       | 30 (42.9)                                    |
| Hispanic                    | 104                           | 80 (32.0)                                       | 24 (34.3)                                    |
| Black                       | 41                            | 37 (14.8)                                       | 4 (5.7)                                      |
| Asian                       | 24                            | 13 (5.2)  | 11 (15.7)                                    |
| Other                       | 22                            | 21 (8.4)  | 1 (1.4)                                      |
| Hand dominance, No. (%)     |                               |   |  |
| Right                       | 260                           | 204 (81.6)                                      | 56 (80)                                      |
| Left                        | 42                            | 35 (14.0)                                       | 7 (10)                                       |
| Ambidextrous                | 18                            | 11 (4.4)  | 7 (10)                                       |

### Sensitivity Analysis:

1. Test vs reference systems had the same accuracy in HR measurement, with higher precision for the reference, whereas a small significant positive bias in accuracy of SpO<sub>2</sub> measurement persisted, but with similar precision.
2. Reference units had a small significant bias observed in both HR and SpO<sub>2</sub> measures relative to each other, with a small significant difference in precision found in both HR and SpO<sub>2</sub> between units.
3. Test units had no difference in accuracy or precision of HR, but they had a small significant difference in precision for SpO<sub>2</sub> (e-Tables 1, 2).

### Discussion

Results from certified laboratory testing indicate the smartphone sensor with app tested met FDA/ISO standards for pulse oximetry for reflective sensors<sup>16,17</sup> compared with FDA/ISO-approved reference. The testing facility calibrates and tests professional medical pulse oximeters before being cleared for sale by the US FDA and other, worldwide regulatory bodies. Full FDA/ISO approval standards also prescribe the testing we report in addition to at least 200 data points referenced to blood sample analysis.

The smartphone sensor with app tested is embedded within millions of smartphones worldwide currently.<sup>3</sup> To evaluate whether this system could be used by a wide

range of people to provide reliable, robust clinical measurements, we used a low-risk, inexpensive, but rigorous protocol and statistical analysis to compare measurement accuracy and precision with that of hospital medical reference devices. Our methodology also allowed determination of accuracy and precision of measurements within each system. The model used a predicted biological value based on eight sets of measurements within the same individual to produce a “true biological reading” free of bias associated with any one instrument.

In both outpatient and inpatient settings, the error of measurement (RMSD) was driven by measurement precision, with bias having a minor role. In the outpatient setting we found a small positive bias in the accuracy of HR and SpO<sub>2</sub> measurement by the smartphone sensor with app, which disappeared in the sensitivity analysis of HR, but persisted in SpO<sub>2</sub> measurement. Precision of the smartphone sensor with app was slightly less than that of reference units for HR and SpO<sub>2</sub>, a difference that persisted only for SpO<sub>2</sub> following sensitivity analysis. In the hospital/ED setting the accuracy for HR was the same for both “test” and “reference” systems. SpO<sub>2</sub> measurement, however, showed a small but consistent bias. However, our analysis revealed similar significant differences in accuracy and precision between the two Welch Allyn reference devices. This provides clarification that the

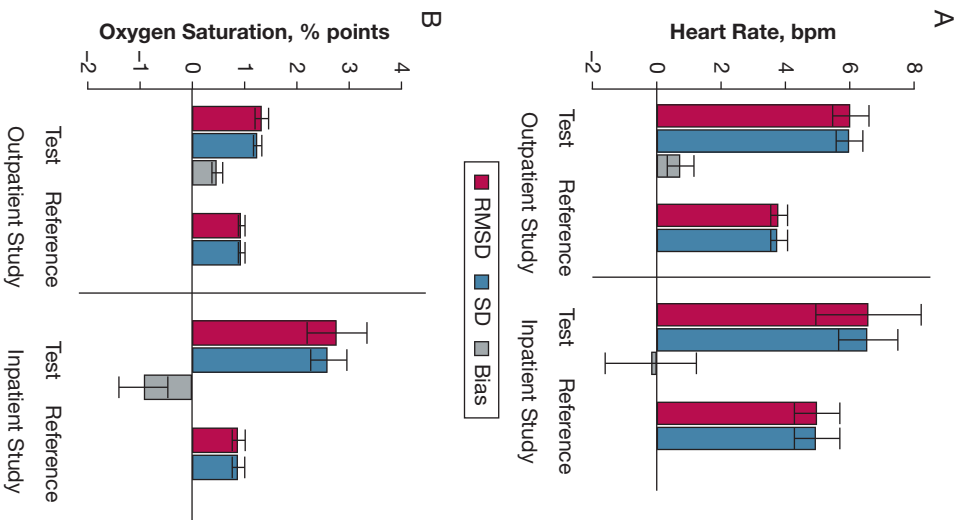
**TABLE 3 ]** Comparison of Bias (Accuracy), SD (Precision), and Root-Mean-Square Deviation of Test and Reference Measurement Systems, for Heart Rate and Oxygen Saturation, in Outpatient (n = 250) and Inpatient (n = 70) Studies<sup>a</sup>

| Study Setting                           | Bias (Accuracy)<br>Test vs Reference System |                        |         | SD (Precision)<br>Each System |                     |         | Root-Mean-Square Deviation<br>Each System |                     |
|---|---|------------------------|---------|-------------------------------|---------------------|---------|---|---------------------|
|   | Test <sup>b</sup>                           | Reference <sup>b</sup> | P Value | Test                          | Reference           | P Value | Test                                      | Reference           |
| <b>Outpatient study</b>                 |   |                        |         |                               |                     |         |   |                     |
| Heart rate, bpm<br>(95% CI)             | 0.73 (0.33 to 1.14)                         | 0 (ref)                | < .001  | 5.99 (5.61 to 6.40)           | 3.80 (3.56 to 4.06) | < .001  | 6.03 (5.47 to 6.60)                       | 3.80 (3.56 to 4.06) |
| SpO <sub>2</sub> , % points<br>(95% CI) | 0.48 (0.38 to 0.58)                         | 0 (ref)                | < .001  | 1.25 (1.17 to 1.33)           | 0.95 (0.89 to 1.01) | < .001  | 1.34 (1.21 to 1.47)                       | 0.95 (0.89 to 1.01) |
| <b>Inpatient study</b>                  |   |                        |         |                               |                     |         |   |                     |
| Heart rate, bpm<br>(95% CI)             | -0.19 (-1.59 to 1.22)                       | 0 (ref)                | .79     | 6.58 (5.78 to 7.49)           | 4.99 (4.39 to 5.69) | .004    | 6.58 (4.94 to 8.23)                       | 4.99 (4.39 to 5.69) |
| SpO <sub>2</sub> , % points<br>(95% CI) | -0.94 (-1.41 to -0.47)                      | 0 (ref)                | < .001  | 2.62 (2.31 to 2.97)           | 0.89 (0.78 to 1.01) | < .001  | 2.78 (2.21 to 3.36)                       | 0.89 (0.78 to 1.01) |

The bias comparison used the Wald test of mixed-effects linear model. ref = reference; SpO<sub>2</sub> = oxygen saturation.

<sup>a</sup>The reference system by definition has no bias and RMSD = SD. Bias and SD correspond to  $\beta_2$  and SD ( $\epsilon_{ij}$ ), respectively, in model Equation 2 (see Methods).

<sup>b</sup>Test = in-phone measurement system; Reference = Spot Vital Signs measurement system (Welch Allyn).



**Figure 3 – A and B.** Comparison of root-mean-square deviation (RMSD) and SD of the test (in-phone) and reference (Spot Vital Signs unit; Welch Allyn) measurement systems, for heart rate (A) and SpO<sub>2</sub> (B), in the outpatient study (n = 250) and inpatient study (n = 70). The error bars correspond to 95% CIs. RMSD decomposes into SD and bias (see Methods, Equation 1). The reference system has no bias, by definition, and thus RMSD equals SD. See Figure 2 legend for expansion of abbreviation.

consequence of small but significant test unit differences detected is of no clinical importance, and tolerable within high-level medical settings. The reference instruments include FDA-approved devices from two different manufacturers, Masimo and Nelcor, that noninvasively measure oxygen saturation by a red and infrared light source, photo detectors, and a probe to transmit light through a translucent, pulsating arterial bed in the digit; this may explain observed measurement differences.

Clinical investigations used an inexpensive plastic cell phone case to allow guidance of the digit over the smartphone sensor with app. For adequate measures of



**TABLE 4 ]** Comparison of Bias (Accuracy) and SD (Precision) for Heart Rate and Oxygen Saturation Within Test Unit (Smartphone Sensor With App) and Within Reference Unit (Welch Allyn Spot Vital Signs) Measurement Systems in Outpatient Study (n = 250)<sup>a,b</sup>

| Vital Sign                  | Model           | Bias (Accuracy) Within System |         | SD (Precision) Within System |                     |         |
|-----------------------------|-----------------|-------------------------------|---------|------------------------------|---------------------|---------|
|                             |                 | Unit 1-Unit 2                 | P Value | Unit 1                       | Unit 2              | P Value |
| Heart rate, bpm             | Test units      | -0.36 (-0.99 to 0.27)         | .27     | 6.92 (6.42 to 7.46)          | 4.95 (4.60 to 5.34) | < .001  |
|                             | Reference units | -0.33 (-0.66 to 0.00)         | .049    | 3.86 (3.58 to 4.16)          | 3.65 (3.39 to 3.94) | .27     |
| SpO <sub>2</sub> , % points | Test units      | -0.06 (-0.22 to 0.10)         | .43     | 1.26 (1.04 to 1.52)          | 1.25 (1.03 to 1.51) | .90     |
|                             | Reference units | 0.52 (0.41 to 0.64)           | < .001  | 1.01 (0.84 to 1.22)          | 0.86 (0.71 to 1.04) | .001    |

The bias comparison used the Wald test of mixed-effects linear model. The SD comparison used the likelihood ratio test of the mixed-effects linear model. See Table 3 legend for expansion of abbreviation.

<sup>a</sup>Bias corresponds to  $\beta_3$  (test units) and to  $\beta_4$  (reference units), and SD corresponds to  $\epsilon_{ij}$  in model Equation 2 (see Methods).

<sup>b</sup>Test units, smartphone sensor with app; reference units, Spot Vital Signs (Welch Allyn).

SpO<sub>2</sub> exposure of the sensor to maximal areas of capillary circulation in the finger pulp is essential and the digit needs to remain immobile for approximately 30 s. Our data indicate that the precision (measurement variation) of smartphone pulse oximetry was worse when participants had to hold their finger on the sensor for 2 min (inpatient/ED setting), likely explained by the difficulty associated with maintaining a digit in place for 2 min. In our experience, PPG signal should be reliably obtained within 30 to 60 s of finger placement. The smartphone sensor with app tested in this study had no refinements for assessing PPG quality or excluding meaningless readings. Improved app software, performing basic sensitivity analyses allowing evaluation of finger contact or outlier values necessitating repeated measurements, could ensure even greater ease of use and reading reliability. In the real world, people will use the smartphone with embedded biosensor plus app as a spot checker; it is designed for the user to hold their finger in place for approximately 30 s and obtain a single reading.

Smartphone sensors with apps enabling clinical pulse oximetry over multiple geographical locations on demand are of immense practical medical importance globally, particularly in LMICs. During respiratory viral pandemics, smartphone sensors with apps allowing clinical pulse oximetry may support medical practitioners in triage and initial clinical management.<sup>14,15</sup> Moreover, smartphone sensors with apps could easily be connected to regional hospitals and

national data hubs, allowing large-scale regional data analysis in near-real time to estimate resource requirements, and may empower patients by storing intermittent historical remote data to present to hospital triage or medical practitioners, and could support management of chronic cardiopulmonary diseases and postoperative clinical follow-up care, including lung transplantation.

## Conclusion

Our data show that the smartphone sensor with app tested met laboratory FDA/ISO standards and could be used to obtain highly accurate and repeatable measurements across a varied population. Full FDA/ISO approval would require additional laboratory testing to incorporate at least 200 data points referenced to blood sample analysis, which could be completed in a few weeks. Given the immediate practical medical importance of remote intermittent clinical pulse oximetry, industry should be encouraged and supported to pursue full FDA/ISO approval of this smartphone sensor with app. This approval should be prioritized and fast-tracked by the FDA and ISO.

## Data Availability

The data files are held by UCSD in a data repository. For access, please e-mail the Antiviral Research Center (AVRC) Regulatory Group: [avrcregulatory@ucsd.edu](mailto:avrcregulatory@ucsd.edu).

## Acknowledgments

**Author contributions:** S. H. B. takes responsibility for the content of the manuscript, including the data and analysis. Overall study concept design and acquisition of sponsorship: S. H. B., F. V., and M. B. Data acquisition, analysis, and/or interpretation of data: S. H. B., F. V., M. B., S. C. P., J. G. G., and C.-C. H. Drafting of the work: S. H. B., F. V., and M. B. C. A. E. and C.-C. H. contributed expertise and edits to the content of this manuscript.

**Financial/nonfinancial disclosures:** The authors have reported to *CHEST* the following: C.-C. H. and C. A. E. were employees of Maxim Integrated during the course of this study. None declared (S. H. B., M. B., S. C. P., J. G. G., F. V.).

**Role of sponsors:** The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

**Other contributions:** Specialists in Global Health (SiGH) (<https://sigh.global>) initiated this collaboration and brought together the team that produced this research, having identified the potential contribution to global health this work could make.

**Additional information:** The e-Tables can be found in the Supplemental Materials section of the online article.

## References

- Alafeef M, Fraiwan M. Smartphone-based respiratory rate estimation using photoplethysmographic imaging and discrete wavelet transform. *J Ambient Intell Human Comput*. 2020;11:693-703.
- Ansermino JM. Universal access to essential vital signs monitoring. *Anesth Analg*. 2013;117(4):883-890.
- Kooistra J. Newzoo's 2018 global mobile market report: insights into the world's 3 billion smartphone users. September 11, 2018. <https://newzoo.com/insights/articles/newzoos-2018-global-mobile-market-report-insights-into-the-worlds-3-billion-smartphone-users/>. Accessed September 25, 2020.
- Hervás R, Fontecha J, Ausín D, Castanedo F, Bravo J, López-de-Ipiña D. Mobile monitoring and reasoning methods to prevent cardiovascular diseases. *Sensors (Basel)*. 2013;13(5):6524-6541.
- Tomasic I, Tomasic N, Trobec R, Krpan M, Kelava T. Continuous remote monitoring of COPD patients: justification and explanation of the requirements and a survey of the available technologies. *Med Biol Eng Comput*. 2018;56(4):547-569.
- Kakria P, Tripathi NK, Kitipawang P. A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *Int J Telemed Appl*. 2015;2015:373474.
- Foran M, Ahn R, Novik J, et al. Prevalence of undiagnosed hypoxemia in adults and children in an under-resourced district hospital in Zambia. *Int J Emerg Med*. 2010;3(4):351-356.
- Maartens G, Stewart A, Griesel R, et al. Development of a clinical prediction rule to diagnose *Pneumocystis jirovecii* pneumonia in the World Health Organization's algorithm for seriously ill HIV-infected patients. *South Afr J HIV Med*. 2018;19(1):851.
- Chouaid C, Maillard D, Housset B, Febvre M, Zaoui D, Lebeau B. Cost Effectiveness of noninvasive oxygen saturation measurement during exercise for the diagnosis of *Pneumocystis carinii* pneumonia. *Am J Respir Dis*. 1993;147(6 Pt 1):1360-1363.
- McCollum ED, King C, Deula R, et al. Pulse oximetry for children with pneumonia treated as outpatients in rural Malawi. *Bull World Health Organ*. 2016;94(12):893-902.
- Lizzerini M, Sonogo M, Pellegrin MC. Hypoxaemia as a mortality risk factor in acute lower respiratory infections in children in low and middle-income countries: systematic review and meta-analysis. *PLoS One*. 2015;10(9):e0136166.
- Aggarwal D, Gupta A, Janmeja AK, Bhardwaj M. Evaluation of tuberculosis-associated chronic obstructive pulmonary disease at a tertiary care hospital: a case-control study. *Lung India*. 2017;34(5):415-419.
- Chin AT, Rylance J, Makumbirofa S, et al. Chronic lung disease in adult recurrent tuberculosis survivors in Zimbabwe: a cohort study. *Int J Tuberc Lung Dis*. 2019;23(2):203-211.
- Sala H, Segovia Roca J, Zerbo C, et al. Initial clinical management of symptomatic adult patients during influenza A (H1N1) epidemics. *J Emerg Med*. 2011;41(4):435-440.
- Verhoeven D, Teijaro JR, Farber DL. Pulse-oximetry accurately predicts lung pathology and the immune response during influenza infection. *Virology*. 2009;390(2):151-156.
- International Organization for Standardization. ISO 80601-2-61:2017: Medical electrical equipment—Part 2-61: particular requirements for basic safety and essential performance of pulse oximeter equipment. 2017. <https://www.iso.org/standard/67963.html>. Accessed September 25, 2020.
- Center for Devices and Radiological Health, US Food and Drug Administration. Pulse oximeters: premarket notification submissions [510(k)s]: guidance for industry and Food and Drug Administration staff. March 2013. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug>. Accessed September 25, 2020.
- Severinghaus JW, Naifeh KH, Koh SO. Errors in 14 pulse oximeters during profound hypoxia. *J Clin Monit*. 1989;5(2):72-81.
- Severinghaus JW. Gadgets for health care: the John W. Severinghaus lecture on translational science. *Anesthesiology*. 2009;110(4):721-728.
- Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17(4):571-582.
- Montgomery DC. *Design and Analysis of Experiments*. 10th ed. New York: John Wiley & Sons; 2019:530.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer; 2009:24.
- Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. New York: Springer; 2000:201.
- Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol*. 1993;138(11):923-936.
- R Core Team R. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. <https://www.R-project.org/>. Accessed September 25, 2020.
- Pinheiro J, Bates D, DebRoy S, Sarkar D; R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-148. 2020. <https://CRAN.R-project.org/package=nlme>. Accessed September 25, 2020.