

Maxim Integrated Smartphone Sensor with App Meets FDA/ISO Standards for Clinical Pulse Oximetry and can be Reliably Utilized by a Wide Range of Patients.

Sara H. Browne, Mike Bernstein, Samuel C. Pan, Jonathan Gonzalez Garcia, Craig A. Easson, Chung-Che Huang, Florin Vaida



PII: S0012-3692(20)34436-6

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

Reference: CHEST 3601

To appear in: *CHEST*

Received Date: 18 March 2020

Revised Date: 21 August 2020

Accepted Date: 26 August 2020

Please cite this article as: Browne SH, Bernstein M, Pan SC, Garcia JG, Easson CA, Huang CC, Vaida F, Maxim Integrated Smartphone Sensor with App Meets FDA/ISO Standards for Clinical Pulse Oximetry and can be Reliably Utilized by a Wide Range of Patients. *CHEST* (2020), doi: <https://doi.org/10.1016/j.chest.2020.08.2104>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Copyright © 2020 Published by Elsevier Inc under license from the American College of Chest Physicians.

**Maxim Integrated Smartphone Sensor with App Meets FDA/ISO Standards for Clinical Pulse Oximetry and can be Reliably Utilized by a Wide Range of Patients.**

**(Smartphone sensor with App for clinical pulse oximetry)**

Sara H. Browne<sup>1,2,\*</sup>, Mike Bernstein<sup>3</sup>, Samuel C. Pan<sup>1</sup> Jonathan Gonzalez Garcia<sup>1</sup>, Craig A. Easson<sup>4</sup>, Chung-Che Huang<sup>4</sup>, Florin Vaida<sup>1</sup>

1University of California San Diego, La Jolla, CA, USA, 2Specialists in Global Health, Encinitas, CA, USA, 3 Physio Monitor, LLC, San Ramon, CA, USA, 4 Maxim Integrated, San Jose, CA, USA

\* Corresponding Author (shbrowne@ucsd.edu, University of California San Diego, Altman Clinical Translational Research Institute, University of California, San Diego, 9452 Medical Center Dr. La Jolla, CA 92037. USA)

**Competing Interests**

C-C.H and C. E were employees of Maxim Integrated during the course of this study. The remaining authors declare no competing interests.

**Funding**

The work was funded by a NIH grant supplement R01MH110057-04S to SHB, a grant from Maxim Integrated to cover SP, and a Technology Award from Non-Profit Specialists in Global Health (<https://sigh.global/>) to JGG.

## Abstract

**Background:** Millions of smartphones contain a photoplethysmography (PPG) bio-sensor (Maxim Integrated, San Jose CA) 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 persons effected by respiratory viral pandemics, such as SARS-CoV-2 or Influenza. To be used for clinical pulse oximetry the embedded PPG system must be paired with an App and meet FDA and ISO requirements.

**Research Question:** We evaluated whether this smartphone sensor with App met FDA/ISO requirements and how measurements obtained using this system compared to hospital reference devices across a wide range of persons.

**Study Design and Methods:** We performed laboratory testing addressing ISO and FDA requirements in ten participants using the smartphone sensor with App. Subsequently, we performed an open label clinical study on 320 participants with widely varying characteristics, to compare accuracy and precision of readings obtained by patients, to hospital reference devices using a rigorous statistical methodology.

**Results:** ‘Breathe Down’ testing in the laboratory showed that the total Root Mean Square Deviation (RSMD) of SpO<sub>2</sub> measurement was 2.02%, meeting FDA/ISO standards. Clinical comparison of the smartphone sensor with App versus hospital reference devices determined SpO<sub>2</sub> and heart rate (HR) accuracy was 0.48 % points (CI 0.38 to 0.58;  $p<0.001$ ) and 0.73 bpm

(CI 0.33 to 1.14;  $p < 0.001$ ) respectively; with SpO<sub>2</sub> and HR precision 1.25 versus reference 0.95 points ( $p < 0.001$ ) and 5.99 versus reference 3.80 bpm ( $p < 0.001$ ), respectively. These small differences were similar to the variation found between two FDA approved reference instruments for SpO<sub>2</sub>: accuracy 0.52 points (CI 0.41 to 0.64;  $p < 0.001$ ) and precision 1.01 versus 0.86 ( $p < 0.001$ ).

**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 Registration**

ClinicalTrials.gov NCT04233827

### **Introduction**

Millions of smartphones contain photoplethysmography (PPG) bio-sensors with 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, San Jose, CA), are currently integrated into 15% of android smartphones worldwide, totaling over 300 million smartphones.<sup>3</sup> The PPG sensor measures the distension of the 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 either 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 bio-sensor 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 utility of smartphone 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 (CHF), chronic obstructive pulmonary disease (COPD), or acute disorders such as pneumonia, and post-operative recovery.<sup>4,5,6</sup> Smartphone sensors with Apps for clinical pulse oximetry would allow patients to gather and track their own data to inform out-patient clinic or telemedicine visits, and increase communication with healthcare providers, which may lead to earlier out-patient interventions potentially reducing 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 healthcare infrastructure enabling management support of infections prevalent in these settings such as AIDS defining pneumonias like pneumocystis or cryptococcosis, tuberculosis, as well as chronic cardiopulmonary diseases.<sup>7,8,9,10,11,12,13</sup>

Further, during events of pandemic respiratory viral spread, such as the current SARS CoV-2 or Influenza, remote clinical pulse oximetry may support not only triage, but also the initial management of symptomatic adults.<sup>14,15</sup> SpO<sub>2</sub> correlates with lung injury<sup>15</sup> and in conjunction with clinical exam may be used to differentiate those that 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 follow their symptoms and include

objective report of  $\text{SpO}_2$  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 bed numbers needed, and could 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 low- and middle-income countries (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. Clinical pulse oximetry requirements by FDA /ISO for PPG sensors is  $< 3.5\%$  root mean square deviation (RMSD) of  $\text{SpO}_2$  value. Herein we present data obtained using a smartphone sensor with App (Maxim Integrated, San Jose, CA)<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 (SOC) 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 provides data on how accurately it can be used by the general population.

## Methods

### *Laboratory testing of smartphone embedded PPG biosensors against FDA/ ISO standards:*

A “Breathe Down” test, a technique introduced by John Severinghaus MD of the University of California, San Francisco (UCSF) and refined by Dr. Phillip 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 consisting of: a) manufactured part MAX 86100A and b) an App containing a signal processing determination algorithm that detects and interprets PPG signal (MaximFAST) (Maxim Integrated, San Jose, CA; were evaluated as implemented in the ZUK phone (Lenovo, Quarry Bay, Hong Kong). The test set consisted of 10 volunteers comprising African (3), Caucasian (3) and Asian (4) 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 to 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 seconds, before proceeding to the next step. Only the last 30 seconds of each plateau were used in the data analysis. Data was collected at SpO<sub>2</sub> levels near 100% to around 70%. SpO<sub>2</sub> values were recorded continuously, an FDA/ISO approved Massimo Radical reference device (Masimo, Irvine, CA) recorded SpO<sub>2</sub> simultaneously.

HR measurements data was taken using a ProSim8 Fluke Instruments (Fluke, Everett, WA).

*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 following methods established by Bland & Altman 2007<sup>20</sup> providing RMSD and a Bland Altman Plot with upper and lower per limits of agreement per Bland Altman as described in ISO80601-2-61:2017.<sup>16</sup> Second order polynomial fit of the data was calculated.

*Clinical Study Conduct:* A open label clinical protocol was developed and approved by University of California, San Diego (UCSD) Human Research Protections Program, #161573. We obtained measures of HR and SpO<sub>2</sub> from ‘test’ units, android smart phone model Zuk Z2 X (Zuk) (Lenovo, Quarry Bay, Hong Kong) with embedded PPG bio-sensor and App (MAX 86100A and signal processing algorithm MaximFAST) (Maxim Integrated, San Jose, CA). The phones were enclosed within plastic cases (Figure 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 Sign (WASVS) units (model nos. (91)42NTB and (93)2268) (Welch Allyn, New York, USA). Participants were adults over 18 years of age recruited from UCSD Healthcare or Clinical Trials Units (CTU) and provided written informed consent. Initially participants were recruited from out-patient settings (n=250). For each participant, four pairs of measurements (8 data points total) were taken in succession, with no break between sets, in the order indicated in Table 1A. 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 and comparison of test and reference measurement systems, while correcting for other possible sources of variation, including within participant vital sign change over 4 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 (3) two reference units.

Subsequently, participants were recruited from in-patient (IP) and Emergency Room (ER) settings (n=70) and followed a simplified protocol (See Table 1B) and no distinction was made between the Welch Allyn reference units, which were used interchangeably. Each measurement period was 2 minutes in length and respiratory rate (RR) was collected. Participant demographic data, including age, height, weight, gender, ethnicity, hand dominance was 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.



Figure 1: Smartphone with embedded PPG sensor suite and plastic case with braille guide for digit placement used as the Test Instrument

**Table 1.** The order and device combination used to take simultaneous heart rate and SpO<sub>2</sub> measurements in: A) out-patient participants (n=250) and B) in-patient/ER participants (n=70)

|  | Step | Left Index Finger | Right Index Finger |
|--|------|-------------------|--------------------|
|--|------|-------------------|--------------------|

|                                   |   |                  |                  |
|-----------------------------------|---|------------------|------------------|
| A: Out-patient Setting<br>(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      |
| B: In-patient Setting<br>(n=70)   | 1 | Test unit 1      | Reference unit 1 |
|                                   | 2 | Reference unit 1 | Test unit 1      |

*Clinical Study Statistical Methods:* Participants 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 standard deviation (SD) of the instrument:

$$RMSD^2 = Bias^2 + SD^2, \quad (1)$$

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 *out-patient 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

$$Y_{ij} = \beta_1 + b_i + \beta_2 \cdot S_i + \beta_3 \cdot T_i + \beta_4 \cdot R_i + \beta_5 \cdot Side_{ij} + \beta_6 \cdot Step_{ij} + \varepsilon_{ij}. \quad (2)$$

$Y_{ij}$  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,  $R_i=-1/2$  unit 1,  $R_i=1/2$  for unit 2 ( $R_i=0$  for test units);  $Side_{ij}=-0.5$  for left-hand and  $0.5$  for right-hand measurements;  $Step_{ij}$  indicates the step of measurement, as in Table 1A, but centered at 0;  $b_i$  is subject-specific random effect, common to all measurements of participant  $i$ ; and  $\varepsilon_{ij}$  is within-participant error, assumed to have a mean 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 “best guess” of outcome based on 8 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 RMSD = SD);  $\beta_3$  measures bias of test units relative to each other;  $\beta_4$  measures the bias of reference units relative to each other. The precision aspects are captured by the standard deviation 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 versus 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% confidence intervals used the Wald test for the mean (accuracy) coefficients, and the likelihood ratio test for the standard deviation (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 standard deviation. These large outliers are likely due to the incorrect use of the system (e.g., finger placement), and the sensitivity analysis results reflect the accuracy of the two systems under appropriate use conditions.

For the *In-patient 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 out-patient setting but without terms for units,  $T_i$  and  $R_i$ . Due to the small number of outliers for this setting a similar sensitivity analysis is not reported. Statistical significance is considered at  $p < 0.05$  level. All analyses were conducted using the R statistical language<sup>25</sup> and used the *nlme* package<sup>26</sup>.

## Results

### *Laboratory Testing of Biosensor plus App*

Test data was collected from 10 participants during ‘breathe down’ testing (see Methods), comparing in-phone sensors and Massimo 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. Below. In this study 8 plateaus were rejected for lack of stability within 2 SpO<sub>2</sub> Counts, rejecting

the corresponding 16 data points. All 10 subjects completed the study and there were no adverse events.

The root mean square deviation (RMSD) total average for data from all 10 volunteer participants tested was 2.02%. Heart rate error based on analysis of Fluke simulation testing showed that the heart rate error was < 2 bpm RMSD.

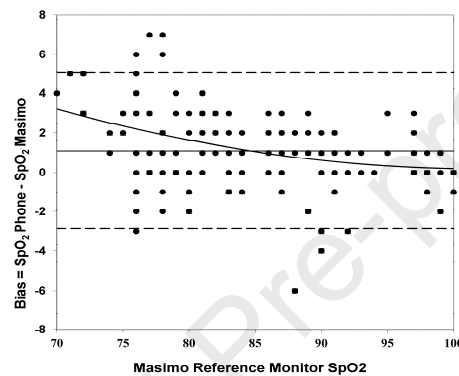


Figure 2: Plot of data collected from 10 volunteers during a ‘breathe down test’ (See Methods). Each point on the plot represents one pair of a Phone reading and Masimo Radical 7 reading. The RMSD of bias between the two instruments in this test was 2.2 SpO<sub>2</sub> counts. The upper and lower limits of agreement per Bland Altman 2007<sup>15</sup> are shown as dashed lines. A 2<sup>nd</sup> order polynomial fit of the data is shown as a solid line.

#### *Clinical Study of Test -vs- Reference measurements*

Participants demographics confirmed a broad age range see Table 2.

**Table 2:** Demographic table participants in both Out-patient and Hospital Settings.

| Demographic Characteristic |                             |                                |                            |
|----------------------------|-----------------------------|--------------------------------|----------------------------|
|                            | All Participants<br>(N=320) | Out-patient Setting<br>(N=250) | Hospital Setting<br>(N=70) |
| Age, years      Mean (SD)  |                             | 46.1 (13.1)                    | 38.5 (16.5)                |

| Range                     | 18-89 | 19-81      | 18-89     |
|---------------------------|-------|------------|-----------|
| Sex, N (%)                |       |            |           |
| Male                      | 129   | 189 (75.6) | 40 (57.1) |
| Female                    | 89    | 59 (23.6)  | 30 (42.9) |
| Other                     | 2     | 2 (0.8)    | 0 (0)     |
| Race and Ethnicity, N (%) |       |            |           |
| Caucasian                 | 129   | 99 (39.6)  | 30 (42.9) |
| Hispanic                  | 104   | 80 (32.0)  | 24 (34.3) |
| African/<br>American      | 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, N (%)     |       |            |           |
| Right                     | 560   | 204 (81.6) | 56 (80)   |
| Left                      | 42    | 35 (14.0)  | 7 (10)    |
| Ambidextrous              | 18    | 11 (4.4)   | 7 (10)    |

In the Out-patient setting Test units gave 960/1000 (96.0%) valid HR and 955/1000 (95.5%) valid SpO<sub>2</sub> readings. Reference units gave 997/1000 (99.7%) valid HR and 997/1000 (99.7%) valid SpO<sub>2</sub> readings. The range of valid HR readings was 41-131 bpm for test, and 38-126 bpm for reference units. The range of valid SpO<sub>2</sub> readings was 90-99% for test, and 92-100% for reference units. In the In-patient/ER 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 per minute, with median 10, standard deviation 4.1, and range 7-30 breaths per minute.

*Bias (Accuracy) of Measurement Systems:* In the Out-patient setting for HR, a significant bias was found for the test relative to reference measurement systems, bias  $\beta_2 = 0.73$  bpm, 95% CI (0.33, 1.14),  $p < 0.001$  (Table 3, Figure 3). A significant bias was found between the two reference units ( $\beta_4 = -0.33$  bpm, 95% CI (-0.66, 0.00),  $p$ -value = 0.049), but not between the two test units ( $\beta_3 = -0.36$  bpm, 95% CI (-0.99, 0.27),  $p$ -value = 0.27), see Table 4.

For SpO<sub>2</sub>, the test units had a significant positive bias compared to reference units ( $\beta_2 = 0.48$  % points, 95% CI (0.38, 0.58) points,  $p < 0.001$ ), see Table 3. No significant bias was found between the two test units, ( $\beta_3 = -0.06\%$ , 95% CI (-0.22, 0.10),  $p = 0.43$ ). However, a significant bias was found between the two reference units ( $\beta_4 = 0.52$  points, 95% CI (0.41, 0.64) points,  $p < 0.001$  (Table 4).

In the In-Patient/ER Setting for HR, no significant bias (accuracy) was found for the test relative to the reference (Table 3, Figure 3A). For SpO<sub>2</sub>, test units had significantly lower readings than reference units, (Table 3, Figure 3B).

*Precision (variation) of Measurement Systems:* In the Out-patient setting for HR, the test measurement system had significantly higher variation (worse precision) than the reference system (standard deviation (SD) = 5.99 bpm, 95% CI (5.61, 6.40) bpm vs. 3.80 bpm, 95% CI (3.56, 4.06) bpm,  $p < 0.001$ , see Table 3 and Figure 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,  $p < 0.001$ . The precision did not differ significantly between the two reference units,  $p = 0.27$  (Table 4). Similar findings were observed in the In-patient/ER setting (See 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, 95% CI (0.89, 1.01) points,  $p < 0.001$  (Table 3 and Figure 3). The precision did not differ between the two test units, 1.26 vs 1.25 points,  $p = 0.90$ , however it differed significantly between the two reference units, SD = 1.01 and 0.86 points,  $p = 0.001$  (Table 4). Similar findings were observed in the In-patient/ER setting (see Table 3, Table 4) and Figure 3A &B).

*Sensitivity Analysis:* 1) Test versus 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 the precision was similar; 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)

**Table 3.** Comparison of bias (accuracy), standard deviation (precision), and root mean square deviation of the Test (In phone) and Reference (Welch Allyn Spot Vital Signs) measurement systems, for heart rate and SpO<sub>2</sub>, in out-patient (n=250) and in-patient (n=70) studies. The Reference system by definition has no bias and RMSD = SD. Bias and standard deviation correspond to  $\beta_2$  and SD( $\epsilon_{ij}$ ) in model equation (2) (See Methods).

|                                      | Bias (Accuracy)<br>Test vs Reference System |           |         | Standard Deviation (Precision)<br>Each System |                   |         | Root Mean Square Deviation<br>Each System |                   |
|--------------------------------------|---------------------------------------------|-----------|---------|-----------------------------------------------|-------------------|---------|-------------------------------------------|-------------------|
|                                      | Test                                        | Reference | P-value | Test                                          | Reference         | P-value | Test                                      | Reference         |
| <b>Out-patient Study</b>             |                                             |           |         |                                               |                   |         |                                           |                   |
| Heart Rate, bpm (95% CI)             | 0.73 (0.33, 1.14)                           | 0 (ref)   | <0.001  | 5.99 (5.61, 6.40)                             | 3.80 (3.56, 4.06) | <0.001  | 6.03 (5.47, 6.60)                         | 3.80 (3.56, 4.06) |
| SpO <sub>2</sub> , % points (95% CI) | 0.48 (0.38, 0.58)                           | 0 (ref)   | <0.001  | 1.25 (1.17, 1.33)                             | 0.95 (0.89, 1.01) | <0.001  | 1.34 (1.21, 1.47)                         | 0.95 (0.89, 1.01) |
| <b>In-patient Study</b>              |                                             |           |         |                                               |                   |         |                                           |                   |
| Heart Rate, bpm (95% CI)             | -0.19 (-1.59, 1.22)                         | 0 (ref)   | 0.79    | 6.58 (5.78, 7.49)                             | 4.99 (4.39, 5.69) | 0.004   | 6.58 (4.94, 8.23)                         | 4.99 (4.39, 5.69) |
| SpO <sub>2</sub> , % points (95% CI) | -0.94 (-1.41, -0.47)                        | 0 (ref)   | <0.001  | 2.62 (2.31, 2.97)                             | 0.89 (0.78, 1.01) | <0.001  | 2.78 (2.21, 3.36)                         | 0.89 (0.78, 1.01) |

Ref – reference; bpm – beat per minute; Test – In phone; Reference – Welch Allyn Spot Vital Signs. The bias comparison used the Wald test of mixed effects linear model

**Table 4.** Comparison of bias (accuracy) and standard deviation (precision) for HR and SpO<sub>2</sub> *within* the Test unit (smartphone sensor with App), and *within* the reference unit (Welch Allyn Spot Vital Signs) measurement systems in the out-patient study (n=250). Bias corresponds to

$\beta_3$  (test units) and to  $\beta_4$  (reference units), and standard deviations correspond  $SD(\epsilon_{ij})$  in model equation (2) (See Methods).

|                             | Model           | Bias (Accuracy) Within System        |         | Std. Deviation (Precision) Within System |                   |         |
|-----------------------------|-----------------|--------------------------------------|---------|------------------------------------------|-------------------|---------|
|                             |                 | Unit <sub>1</sub> -Unit <sub>2</sub> | P-value | Unit <sub>1</sub>                        | Unit <sub>2</sub> | P-value |
| Heart Rate (bpm)            | Test Units      | -0.36 (-0.99, 0.27)                  | 0.27    | 6.92 (6.42, 7.46)                        | 4.95 (4.60, 5.34) | <0.001  |
|                             | Reference Units | -0.33 (-0.66, 0.00)                  | 0.049   | 3.86 (3.58, 4.16)                        | 3.65 (3.39, 3.94) | 0.27    |
| SpO <sub>2</sub> (% points) | Test Units      | -0.06 (-0.22, 0.10)                  | 0.43    | 1.26 (1.04, 1.52)                        | 1.25 (1.03, 1.51) | 0.90    |
|                             | Reference Units | 0.52 (0.41, 0.64)                    | <0.001  | 1.01 (0.84, 1.22)                        | 0.86 (0.71, 1.04) | 0.001   |

Ref – reference; bpm – beat per minute; Test Units – Smartphone sensor with App; Reference Units – Welch Allyn Spot Vital Signs. The bias comparison used the Wald test of mixed effects linear model. The standard deviation comparison used the likelihood ratio test of the mixed-effects linear model

Figure 3(a):

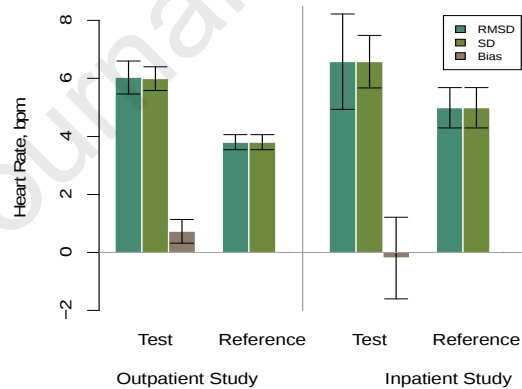


Figure 3(b):

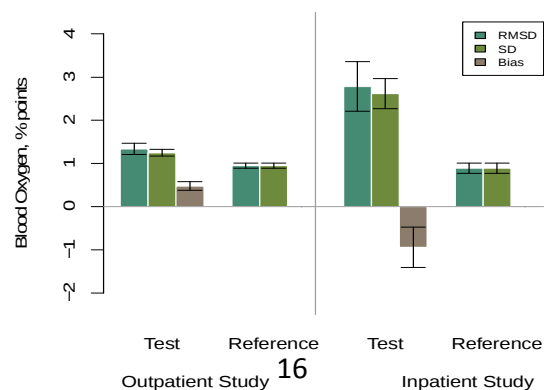


Figure 3: Comparison of root mean square deviation (RMSD), standard deviation (SD), of the Test (In phone) and Reference (Welch Allyn Spot Vital Signs) measurement systems, for a) heart rate and b) SpO<sub>2</sub>, in the out-patient (n=250) and in-patient (n=70) studies. The error bars correspond to 95% confidence intervals. RMSD decomposes into SD and bias (See Methods, formula 1). The Reference system has no bias, by definition, and thus RMSD equals SD.

## Discussion

Results from certified laboratory testing indicate the smartphone sensor with App tested met ISO and FDA standards for pulse oximetry for reflective sensors<sup>16,17</sup> compared to FDA/ISO approved reference. The testing facility calibrates and tests professional medical pulse oximeters before being cleared for sale by 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 approximately 300 million smart phones worldwide currently.<sup>3</sup> To evaluate whether this system could be utilized 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 to 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 out-patient and in-patient settings, the error of measurement (RMSD) was driven by measurement precision, with bias having a minor role. In the out-patient 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 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/ER 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 consequence of small but significant test unit differences detected are of no clinical importance, and tolerable within high level medical settings. The reference instruments use FDA approved devices from two different manufacturers, Massimo (Irvine, CA) and Nelcor (Minneapolis, MN), that noninvasively measure the 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 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 seconds. Our data indicate the precision (measurement variation) of smartphone pulse oximetry was worse when participants had to hold their finger on the sensor for two minutes (in-patient/ER setting), likely explained by difficulty associated with maintaining a digit in place for two minutes. In our experience PPG signal should be reliably obtained within 30-60 seconds of finger placement. The smartphone sensor with App tested in this study had no refinements for assessing PPG quality or exclude 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 embedded biosensor plus App as a spot checker, it is designed for the user to hold their finger in place for approximately 30 seconds and get a single reading.

Smartphone sensors with Apps enabling clinical pulse oximetry over multiple geographical locations on demand is of immense practical medical importance globally, particularly in LMICs. During respiratory viral pandemics smartphones 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 shows 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/ ISO organizations.

### **Data Availability**

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

### **Acknowledgements**

Specialists in Global Health (SiGH) (<https://sigh.global>) initiated this collaboration and brought the team that produced this research together, having identified the potential contribution to global health this work could make. The work was funded by a NIH grant supplement R01MH110057-04S to SHBrowne, a grant from Maxim Integrated Inc to SPan and a Technology Award from Specialists in Global Health (SiGH).

**Author Contributions** SHB takes responsibility for the content of the manuscript, including the data and analysis. Overall study concept design and acquisition of sponsorship: SHB, FV and MB. Data acquisition, analysis, and/or interpretation of data: SHB, FV, MB, SCP, JGG, and C-CH. Drafting of the work: SHB, FV, and MB. CAE and C-CH contributed expertise and edits to the contents of this manuscript.

**Competing Interests** C-CH and CAE were employees of Maxim Integrated during the course of this study. The remaining authors declare no competing interests.

### **References**

1. Alafeef, M. & Fraiwan, M. Smartphone-based respiratory rate estimation using photoplethysmographic imaging and discrete wavelet transform. *Journal of Ambient Intelligence and Humanized Computing* **41**, 387–395 (2017).

2. Ansermino, J. M. Universal Access to Essential Vital Signs Monitoring. *Anesthesia & Analgesia* **117**, 883–890 (2013).
3. Kooistra, J. Newzoo's 2018 Global Mobile Market Report: Insights into the World's 3 Billion Smartphone Users | Newzoo. *Newzoo* (2018). Available at: <https://newzoo.com/insights/articles/newzoos-2018-global-mobile-market-report-insights-into-the-worlds-3-billion-smartphone-users/>. (Accessed: 20th December 2019)
4. Hervás, R. *et al.* Mobile Monitoring and Reasoning Methods to Prevent Cardiovascular Diseases. *Sensors* **13**, 6524–6541 (2013).
5. 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. *Medical & Biological Engineering & Computing* **56**, 547–569 (2018).
6. Kakria, P., Tripathi, N. K. & Kitipawang, P. A Real-Time Health Monitoring System for Remote Cardiac Patients Using Smartphone and Wearable Sensors. *International Journal of Telemedicine and Applications* **2015**, 1–11 (2015).
7. Foran, M. *et al.* Prevalence of undiagnosed hypoxemia in adults and children in an under-resourced district hospital in Zambia. *International Journal of Emergency Medicine* **3**, 351–356 (2010).
8. Maartens, G. *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. *Southern African Journal of HIV Medicine* **19**, (2018).
9. Chouaid, C. *et al.* Cost Effectiveness of Noninvasive Oxygen Saturation Measurement During Exercise for the Diagnosis of *Pneumocystis carinae* Pneumonia. *American Review of Respiratory Disease* **147**, 1360–1363 (1993)

10. McCollum, E. D. *et al.* Pulse oximetry for children with pneumonia treated as out-patients in rural Malawi. *Bulletin of the World Health Organization* **94**, 893–902 (2016).
11. Lazzerini, M., Sonogo, M. & Pellegrin, M. C. 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* **10**, e0136166 (2015).
12. Aggarwal, D., Gupta, A., Janmeja, A. K. & Bhardwaj, M. Evaluation of tuberculosis-associated chronic obstructive pulmonary disease at a tertiary care hospital: A case-control study. *Lung India: Official Organ of Indian Chest Society* **34**, 415–419 (2017).
13. Chin, A. T. *et al.* Chronic lung disease in adult recurrent tuberculosis survivors in Zimbabwe: a cohort study. *The International Journal of Tuberculosis and Lung Disease* **23**, 203–211 (2019).
14. Sala, H. *et al.* Initial Clinical Management of Symptomatic Adult Patients during Influenza A (H1N1) Epidemics. *The Journal of Emergency Medicine* **41**, 435–440 (2011).
15. Verhoeven, D., Teijaro, J. R. & Farber, D. L. Pulse-oximetry accurately predicts lung pathology and the immune response during influenza infection. *Virology* **390**, 151–156 (2009).
16. ISO - International Organization for Standardization. ISO 80601-2-61:2017. ISO (2017). Available at: <https://www.iso.org/standard/67963.html>. (Accessed: 10th January 2020)
17. Center for Devices and Radiological Health. Pulse Oximeters - Premarket Notification Submissions [510(k)s]. *U.S. Food and Drug Administration* (2013). Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug>. (Accessed: 12th December 2019).

18. Severinghaus, J. W., Naifeh, K. H. & Koh, S. O. Errors in 14 pulse oximeters during profound hypoxia. *Journal of Clinical Monitoring* **5**, 72–81 (1989).
19. Severinghaus, J. W. Gadgeteering for Health Care. *Anesthesiology* **110**, 721–728 (2009).
20. Bland, J. M. & Altman, D. G. Agreement Between Methods of Measurement with Multiple Observations Per Individual. *Journal of Biopharmaceutical Statistics* **17**, 571–582 (2007).
21. Montgomery, D. C. *Design and analysis of experiments*. **10**, 530 (Wiley, 2020).
22. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning, second edition: data mining, inference, and prediction*. 24 (Springer, 2009).
23. Pinheiro, J. C. & Bates, D. M. *Mixed-effects models in S and S-PLUS*. 201 (Springer, 2000).
24. Maldonado, G. & Greenland, S. Simulation Study of Confounder-Selection Strategies. *American Journal of Epidemiology* **138**, 923–936 (1993).
25. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at <https://www.R-project.org/>.
26. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-148, <https://CRAN.R-project.org/package=nlme>.

**Supplementary e-Table 1:** Sensitivity analysis: Out-patient Study (n=250) - Comparison of bias (accuracy), standard deviation (precision), and root mean square deviation of the Test (Smartphone biosensor with APP) and Reference (Welch Allyn Spot Vital Signs) measurement systems, for heart rate and SpO<sub>2</sub>. This sensitivity analysis excluded outlier measurements (observations  $\geq 3$  standard deviations in absolute value) as follows: HR: n=24 observations from 15 participants excluded, (Test n=18, Reference, n=6); SpO<sub>2</sub>: n=18 observations from 18 participants excluded (Test n=14, Reference n=4). Bias and standard deviation correspond to  $\beta_2$  and SD( $\epsilon_{ij}$ ) in model equation (2). The bias comparison used the Wald test of mixed effects linear model. The standard deviation comparison used the likelihood ratio test of the mixed-effects linear model.

| Outliers Removed                     | Bias (Accuracy)<br>Test vs Reference System |           |         | Standard Deviation (Precision)<br>Each System |                      |         | Root Mean Square<br>Deviation<br>Each System |                      |
|--------------------------------------|---------------------------------------------|-----------|---------|-----------------------------------------------|----------------------|---------|----------------------------------------------|----------------------|
|                                      | Test                                        | Reference | P-value | Test                                          | Reference            | P-value | Test                                         | Reference            |
| <b>Out-patient Study</b>             |                                             |           |         |                                               |                      |         |                                              |                      |
| Heart Rate, bpm (95% CI)             | 0.18<br>(-0.10, 0.46)                       | 0 (ref)   | 0.21    | 4.04<br>(3.78, 4.32)                          | 3.32<br>(3.11, 3.55) | <0.001  | 4.04<br>(3.66, 4.43)                         | 3.32<br>(3.11, 3.55) |
| SpO <sub>2</sub> , % points (95% CI) | 0.54<br>(0.45, 0.63)                        | 0 (ref)   | <0.001  | 0.94<br>(0.90, 0.99)                          | 1.04<br>(0.99, 1.09) | 0.058   | 1.08<br>(0.98, 1.18)                         | 1.04<br>(0.99, 1.09) |

Ref – reference; bpm – beat per minute; Test – Smartphone sensor with App; Reference – Welch Allyn Spot Vital Signs.

**Supplementary e-Table 2:** Sensitivity analysis: Out-patient Study (n=250) - Comparison of bias (accuracy) and standard deviation (precision) for heart rate and SpO<sub>2</sub> *within* the Test (Smartphone sensor with App) and the Reference (Welch Allyn Spot Vital Signs) measurement systems. This sensitivity analysis excluded n=24 outliers for heart rate, and n=18 outliers for

SpO<sub>2</sub>. Between-units bias corresponds to  $\beta_3$  (test units) and to  $\beta_4$  (reference units), and standard deviations correspond SD( $\epsilon_{ij}$ ) in model equation (2) (See Methods). The bias comparison used the Wald test of mixed effects linear model. The standard deviation comparison used the likelihood ratio test of the mixed-effects linear model.

| Outliers Removed            | Model           | Bias (Accuracy) Within System        |         | Std. Deviation (Precision) Within System |                   |         |
|-----------------------------|-----------------|--------------------------------------|---------|------------------------------------------|-------------------|---------|
|                             |                 | Unit <sub>1</sub> -Unit <sub>2</sub> | P-value | Unit <sub>1</sub>                        | Unit <sub>2</sub> | P-value |
| Heart Rate (bpm)            | Test Units      | 0.10 (-0.32, 0.52)                   | 0.63    | 4.16 (3.84, 4.49)                        | 4.04 (3.74, 4.37) | 0.53    |
|                             | Reference Units | -0.43 (-0.71, -0.16)                 | 0.002   | 3.16 (2.92, 3.41)                        | 3.53 (3.27, 3.82) | 0.009   |
| SpO <sub>2</sub> (% points) | Test Units      | -0.02 (-0.15, 0.11)                  | 0.76    | 1.10 (1.05, 1.15)                        | 0.99 (0.94, 1.04) | 0.034   |
|                             | Reference Units | 0.52 (0.41, 0.63)                    | <0.001  | 1.00 (0.95, 1.05)                        | 0.87 (0.83, 0.91) | 0.005   |

Ref – reference; bpm – beat per minute; Test Units – In phone; Reference Units – Welch Allyn Spot Vital Signs.

**Table 1.** The order and device combination used to take simultaneous heart rate and SpO<sub>2</sub> measurements in: A: out-patient participants (n=250) and B in-patient/ER participants (n=70)

|                                   | Step | Left Index Finger | Right Index Finger |
|-----------------------------------|------|-------------------|--------------------|
| A: Out-patient Setting<br>(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        |
| B: In-patient Setting<br>(n=70)   | 1    | Test unit 1       | Reference unit 1   |
|                                   | 2    | Reference unit 1  | Test unit 1        |

**Table 2:** Demographic table participants in both Out-patient and Hospital Settings.

| Demographic Characteristic |                                                                 |                                  |                                                                 |                                                               |
|----------------------------|-----------------------------------------------------------------|----------------------------------|-----------------------------------------------------------------|---------------------------------------------------------------|
|                            |                                                                 | All Participants<br>(N=320)      | Out-patient Setting<br>(N=250)                                  | Hospital Setting<br>(N=70)                                    |
| Age, years                 | Mean (SD)<br>Range                                              | 18-89                            | 46.1 (13.1)<br>19-81                                            | 38.5 (16.5)<br>18-89                                          |
| Sex, N (%)                 | Male<br>Female<br>Other                                         | 129<br>89<br>2                   | 189 (75.6)<br>59 (23.6)<br>2 (0.8)                              | 40 (57.1)<br>30 (42.9)<br>0 (0)                               |
| Race and Ethnicity, N (%)  | Caucasian<br>Hispanic<br>African/<br>American<br>Asian<br>Other | 129<br>104<br>41<br><br>24<br>22 | 99 (39.6)<br>80 (32.0)<br>37 (14.8)<br><br>13 (5.2)<br>21 (8.4) | 30 (42.9)<br>24 (34.3)<br>4 (5.7)<br><br>11 (15.7)<br>1 (1.4) |
| Hand Dominance, N (%)      | Right<br>Left<br>Ambidextrous                                   | 560<br>42<br>18                  | 204 (81.6)<br>35 (14.0)<br>11 (4.4)                             | 56 (80)<br>7 (10)<br>7 (10)                                   |

**Table 3.** Comparison of bias (accuracy), standard deviation (precision), and root mean square deviation of the Test (In phone) and Reference (Welch Allyn Spot Vital Signs) measurement systems, for heart rate and SpO<sub>2</sub>, in out-patient (n=250) and in-patient (n=70) studies. The Reference system by definition has no bias and RMSD = SD. Bias and standard deviation correspond to  $\beta_2$  and SD( $\epsilon_{ij}$ ) in model equation (2) (See Methods).

|                                      | Bias (Accuracy)<br>Test vs Reference System |           |         | Standard Deviation (Precision)<br>Each System |                      |         | Root Mean Square Deviation<br>Each System |                      |
|--------------------------------------|---------------------------------------------|-----------|---------|-----------------------------------------------|----------------------|---------|-------------------------------------------|----------------------|
|                                      | Test                                        | Reference | P-value | Test                                          | Reference            | P-value | Test                                      | Reference            |
| <b>Out-patient Study</b>             |                                             |           |         |                                               |                      |         |                                           |                      |
| Heart Rate, bpm (95% CI)             | 0.73<br>(0.33, 1.14)                        | 0 (ref)   | <0.001  | 5.99<br>(5.61, 6.40)                          | 3.80<br>(3.56, 4.06) | <0.001  | 6.03<br>(5.47, 6.60)                      | 3.80<br>(3.56, 4.06) |
| SpO <sub>2</sub> , % points (95% CI) | 0.48<br>(0.38, 0.58)                        | 0 (ref)   | <0.001  | 1.25<br>(1.17, 1.33)                          | 0.95<br>(0.89, 1.01) | <0.001  | 1.34<br>(1.21, 1.47)                      | 0.95<br>(0.89, 1.01) |
| <b>In-patient Study</b>              |                                             |           |         |                                               |                      |         |                                           |                      |
| Heart Rate, bpm (95% CI)             | -0.19<br>(-1.59, 1.22)                      | 0 (ref)   | 0.79    | 6.58<br>(5.78, 7.49)                          | 4.99<br>(4.39, 5.69) | 0.004   | 6.58<br>(4.94, 8.23)                      | 4.99<br>(4.39, 5.69) |
| SpO <sub>2</sub> , % points (95% CI) | -0.94<br>(-1.41, -0.47)                     | 0 (ref)   | <0.001  | 2.62<br>(2.31, 2.97)                          | 0.89<br>(0.78, 1.01) | <0.001  | 2.78<br>(2.21, 3.36)                      | 0.89<br>(0.78, 1.01) |

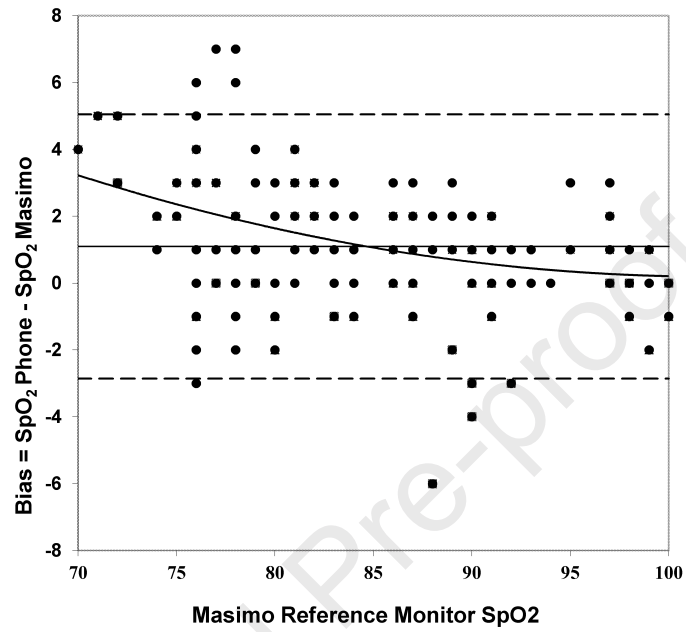
Ref – reference; bpm – beat per minute; Test – In phone; Reference – Welch Allyn Spot Vital Signs. The bias comparison used the Wald test of mixed effects linear model

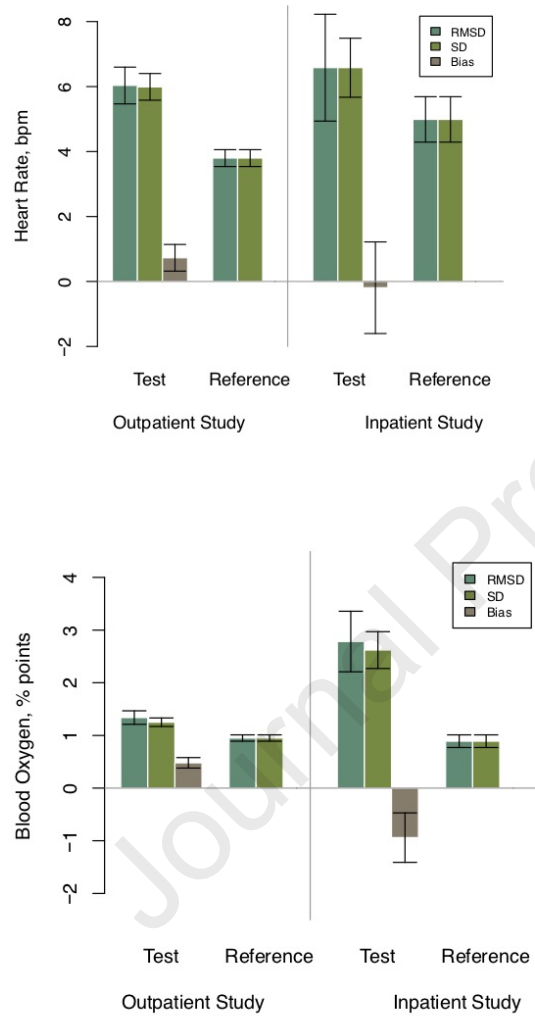
**Table 4.** Comparison of bias (accuracy) and standard deviation (precision) for HR and SpO<sub>2</sub> *within* the Test unit (smartphone sensor with App), and *within* the reference unit (Welch Allyn Spot Vital Signs) measurement systems in the out-patient study (n=250). Bias corresponds to  $\beta_3$  (test units) and to  $\beta_4$  (reference units), and standard deviations correspond SD( $\epsilon_{ij}$ ) in model equation (2) (See Methods).

|                             | Model           | Bias (Accuracy) Within System        |         | Std. Deviation (Precision) Within System |                   |         |
|-----------------------------|-----------------|--------------------------------------|---------|------------------------------------------|-------------------|---------|
|                             |                 | Unit <sub>1</sub> -Unit <sub>2</sub> | P-value | Unit <sub>1</sub>                        | Unit <sub>2</sub> | P-value |
| Heart Rate (bpm)            | Test Units      | -0.36 (-0.99, 0.27)                  | 0.27    | 6.92 (6.42, 7.46)                        | 4.95 (4.60, 5.34) | <0.001  |
|                             | Reference Units | -0.33 (-0.66, 0.00)                  | 0.049   | 3.86 (3.58, 4.16)                        | 3.65 (3.39, 3.94) | 0.27    |
| SpO <sub>2</sub> (% points) | Test Units      | -0.06 (-0.22, 0.10)                  | 0.43    | 1.26 (1.04, 1.52)                        | 1.25 (1.03, 1.51) | 0.90    |
|                             | Reference Units | 0.52 (0.41, 0.64)                    | <0.001  | 1.01 (0.84, 1.22)                        | 0.86 (0.71, 1.04) | 0.001   |

Ref – reference; bpm – beat per minute; Test Units – Smartphone sensor with App; Reference Units – Welch Allyn Spot Vital Signs. The bias comparison used the Wald test of mixed effects linear model. The standard deviation comparison used the likelihood ratio test of the mixed-effects linear model







## Abbreviations

**bpm** (beats per minute); **CHF** (congestive heart failure); **CI** (Confident Interval); **COPD** (chronic obstructive pulmonary disease); **CTU** (Clinical Trials Units ); **ER** (Emergency Room); **FDA** (Food and Drug Administration); **HR** (Heart Rate); **IP** (In-patient); **LMICs** (low and middle income countries ); **ISO** (International Organization for Standards); **PPG** (photoplethysmography); **RR** (Respiratory Rate); **RSMD** (Root Mean Square Deviation); **SARS-CoV-2** (Severe Acute Respiratory Syndrome Coronavirus 2); **SD** (Standard Deviation); **SOC** (standard of care); **SpO<sub>2</sub>** (Oxygen Saturation); **UCSF** (University of California, San Francisco); **UCSD** (University of California, San Diego)