Original Research

Accuracy of Smartphone-Based Pulse Oximetry Compared with Hospital-Grade Pulse Oximetry in Healthy Children

Sarah Tomlinson, MD, Sydney Behrmann, BS, James Cranford, PhD, Marisa Louie, MD, and Andrew Hashikawa, MD

Department of Emergency Medicine, Michigan Medicine at University of Michigan, Ann Arbor, Michigan.

Abstract

Background: Pulse oximetry, a ubiquitous, noninvasive method to monitor oxygen saturation (SpO₂), requires larger, nonportable equipment. Smartphone pulse oximeter applications (apps) provide a portable, cost-effective option, but are untested in children. We hypothesize that smartphone pulse oximetry will not be inferior to standard pulse oximetry measured in healthy children.

Materials and Methods: Two main types of pulse oximetry apps, a camera-based app (CBA) that uses a phone camera flash and lens and a probe-based app (PBA) that uses an external plug-in probe, were compared with standard pulse oximetry measured in children ages 2–13 years without a respiratory complaint and a triage SpO₂ ‡ 97% seen in a pediatric Emergency Department. Two investigators obtained heart rate and SpO₂ using each app. Inter-rater reliability was tested using interclass correlations (ICCs), and Bland–Altman method was used to compare app values to triage measurements.

Results: Eighty-one patients were enrolled. ICC for SpO₂ for PBA and CBA were 0.73 and −0.24, respectively. The 95% limits of agreement between the PBA SpO₂ and triage SpO₂ were −2.8 to +2.5 compared with −4.1 to +3.5 for the CBA SpO₂ and triage SpO₂. Mean differences between triage SpO₂ and the PBA SpO₂ (−0.17%) and triage SpO₂ and CBA SpO₂ (−0.33%) were not statistically significant.

Discussion and Conclusions: Smartphone-based pulse oximetry is not inferior to standard pulse oximetry in pediatric patients without hypoxia. Reliability was superior for PBA compared with CBA, with more precise agreement for the PBA compared with the CBA. Future studies should test pulse oximetry apps in a hypoxic pediatric population.

Keywords: telemedicine, home health monitoring, m-health, pediatrics, technology

Introduction

Use of pulse oximetry as a noninvasive method to monitor arterial oxygen saturation (SpO₂) is ubiquitous among surgical, outpatient, and inpatient settings. In developing countries, pulse oximetry often remains unavailable due to cost and lack of resources. The World Health Organization found that nearly 77,000 operating rooms are functioning without pulse oximetry monitoring during anesthesia. Access to pulse oximetry in developing countries could aid in the detection and management of pneumonia in pediatric patients, an illness responsible for over 2.5 million deaths worldwide in children younger than 5 years of age.

In the developed world, pulse oximetry is increasingly used outside of traditional medical settings and has found an increasing role in the home monitoring setting for adult patients with chronic obstructive pulmonary disease (COPD), asthma, and other chronic respiratory conditions. Home pulse oximetry use is increasing in the pediatric population for infants and children with bronchopulmonary dysplasia, acute bronchiolitis, and asthma exacerbations. One pediatric study in high-altitude regions reported that 59% of physicians used continuous home pulse oximetry as part of their management decisions on weaning supplemental oxygen in infants with mild bronchiolitis and hypoxia.

Current pulse oximetry devices in hospital settings and home use require the use of larger, stationary equipment that can be expensive to obtain and cumbersome to transport. Smartphone technology and health-related applications (apps) offers a universally available, relatively low-cost platform for portable pulse oximetry. At least 75% of Americans own a smartphone and have access to a constantly increasing availability of phone applications designed to measure various health parameters. A review of mobile health (m-health) apps published in 2016 showed that 259,000 m-health apps are available on major app stores, with an estimated 3.2 billion downloads by users of m-health apps by the end of 2016. The U.S. Food and Drug Administration does not regulate all of these apps and mobile medical devices, and the accuracies of health-related apps cannot be assumed. Although these untested apps may be marketed for recreational use only, patients and families may be using them to monitor...
and make decisions about their health. Therefore, understanding the accuracy of such apps is imperative.

The use of relatively low-cost, smartphone apps as portable pulse oximeters has great potential to increase accessibility to vital monitoring equipment. Such devices, if accurate in their measurements, would have significant impact on mobile medical care. In particular, smartphone-based pulse oximetry could improve the evaluation of patients in telemedicine encounters, pharmacy-based health clinics, school-based clinics, summer camps, particularly for those children with disabilities and medical conditions. Finally, these devices could provide more compact, portable equipment options for families with patients traveling with special medical needs.

PURPOSE AND OBJECTIVES

We seek to evaluate two main types of smartphone-based pulse oximeter apps. The first type is a camera-based app (CBA), which utilizes the phone’s own camera lens and flash with no additional device required. The second type is a probe-based app (PBA), which is an app designed to use an external probe that connects directly to the smartphone. The two apps have only been tested in adult populations.\(^{10,11}\) The accuracy of these apps in pediatric populations is currently unknown. Compared with adults, children have different physiological parameters and a narrower range of normal SpO\(_2\) values that could substantially impact accuracy and reliability of pulse oximetry apps.\(^1\) Our study aims to evaluate the accuracy of these two portable, smartphone-based pulse oximetry apps in healthy nonhypoxic children ages 2–13 years. Testing these apps among a population with normal saturations is important before testing can be performed in patients with hypoxia. We hypothesize that smartphone-based pulse oximetry will not be inferior to standard Emergency Department (ED) pulse oximetry measured in healthy children.

Materials and Methods

The study was performed from August 2015 to December 2016, using a convenience sample of patients from Children’s Emergency Services, Michigan Medicine at the University of Michigan, a tertiary care center and level one trauma center with \(~27,000\) visits per year. The study was approved by the University Institutional Review Board. Study subjects included afebrile patients who were 2–13 years of age, English-speaking, and with an SpO\(_2\) \(\geq97\%\) as measured in triage by the stationary pulse oximeter. Subjects were excluded if they were in the ED for a respiratory-related complaint, if they had underlying cardiac, respiratory, hematologic, or metabolic disease, if they were a trauma patient, if capillary refill time in fingers was \(>3\) s, or if they had nail polish on their fingernails. SpO\(_2\) measurement in triage was obtained by nursing as per their standard triage protocol. Patients and their families were approached if identified as study eligible, and written parental or guardian consent was obtained for all patients enrolled in the study. Oral assent was obtained for children 7–9 years of age, and written consent was obtained from children 10–13 years of age. Other baseline demographic data such as patient’s temperature, weight, heart rate (HR), and skin tone were obtained, given previous studies suggesting the possibility of inaccurate pulse oximetry readings in patients with hypo- or hyperthermia, dark skin pigmentation, or tachyarrhythmias.\(^4\)

Our CBA was represented by digiDoc, a Norwegian company specializing in m-health technology that created Pulse Oxi-meter, an app utilizing the iPhone’s onboard technology, including camera lens and flash, to detect HR and SpO\(_2\) levels. No additional device external to the phone is needed and the app is available on Apple platforms for \$4.99\ USD. Our PBA was represented by Masimo, a global medical technology company that produces pulse oximeters for use in clinics and hospitals and has designed a portable, external probe (Masimo iSpO\(_2\)) to detect a patient’s HR and SpO\(_2\) levels when connected to a smartphone with the supporting app (Masimo Personal Health). The small probe attaches to the patient’s finger and comes in both adult and pediatric (<40 kg) sizes. The probe costs \(~\$249\) USD with a free supporting app, and it is available on both Apple and Android platforms.

Values for both SpO\(_2\) and HR were recorded from available triage vital signs. Next, research team members used a dedicated research smartphone to access the two pulse oximetry apps to obtain each study subject’s values for SpO\(_2\) and HR during the patient’s stay in the ED. Measurements using the two apps were obtained as per the manufacturer’s guidelines (Appendix). For the PBA, the probe was placed on the patient’s third or fourth digit of the nondominant hand, with the production of a continuous waveform on the phone app. Once the waveform had stabilized for 5 s, the HR and SpO\(_2\) values were recorded. If handedness was not known, the left hand was used for measurement. For the CBA, the patient’s second digit of the right hand was used. The CBA produced an isolated value for SpO\(_2\) and HR rather than continuous monitoring, so those values were recorded. If there was a technical issue obtaining data using either app, a second measurement was obtained and that was the measurement included in the overall analysis. Data were not recorded if vital signs could not be obtained after two separate attempts, and these records were removed from the analysis. The two apps were paired with either an iPhone 5 or an iPhone 6 for data collection. Data were recorded and maintained using Research Electronic Data Capture (REDCap).
We estimated that we would need a sample size of \( \sim 80 \) subjects. This was based on simulation work showing that this sample size provides a reasonable balance between maximizing precision of parameter estimates, and minimizing resources required for adequate assessment of feasibility.\(^{12} \) A small subset of patient data was obtained by two members of the study team, with the goal of measuring reliability between two independent raters. Reliability between these two raters was calculated using the intraclass correlation (ICC). The remainder of the data analysis was performed using Bland–Altman analysis, a method comparison study that compares a new measurement technique (CBA, PBA) with an established one (triage vital signs).\(^{13} \) The mean difference was calculated between the CBA and triage vital signs for both HR and SpO\(_2\) values, and the same was calculated for the PBA. Next, the 95% limits of agreement (LOAs) were calculated, which were defined as "the range within which most differences between measurements by the two methods will lie."\(^{14} \) If the range of differences defined by the LOAs was not considered clinically important, the two methods can be used interchangeably. Precision of the estimated LOAs was assessed by calculating the exact 95% confidence intervals (95% CIs) using procedures and tables published by Carkeet.\(^{15} \) Spearman’s correlation was used to determine if there was any linear association between mean SpO\(_2\) or HR and bias, to determine that bias did not change with any change in mean SpO\(_2\) or HR values.

### Results

A total of 101 patients were approached for study enrollment, and 20 patients were excluded. Eighty-one subjects were included in the final analysis. Study patients were excluded for the following reasons: receiving patient care \((n = 3)\), discharged before able to obtain study data \((n = 1)\), parent not present in ED \((n = 1)\), patient ultimately did not meet inclusion criteria \((n = 2)\), parent refusal to participate \((n = 10)\), child refusal to participate \((n = 2)\), or other \((n = 1)\). This last patient was ultimately excluded because the patient’s fingers were too small to use the apps, and no measurements could be recorded as a result.

Participants were on average 8.1 years old (standard deviation [SD] = 3.4), weighed 32.2 kg (SD = 17.2), and most (81.5%) were right-handed. In terms of gender and race, 54.3% were male, and 76.5% were white. Table 1 outlines the demographics of the study population. Average triage SpO\(_2\) was 98.9 (SD = 1.1), and average triage HR was 94.1 (SD = 18.5). To assess ICC between two independent team members, values for pulse oximetry and HR were obtained with both devices by 2 team members for 18 participants. Results showed that, for pulse oximetry, the ICCs for the PBA and CBA were 0.73 and −0.24,

<table>
<thead>
<tr>
<th>Table 1. Patient Characteristics ((n=81))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>Unknown/not reported</td>
</tr>
<tr>
<td><strong>Race</strong></td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>More than one race</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Black/African American</td>
</tr>
<tr>
<td>Native Hawaiian/other Pacific Islander</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
</tr>
<tr>
<td>Unknown/not reported</td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>2–2.9 years old</td>
</tr>
<tr>
<td>3–3.9 years old</td>
</tr>
<tr>
<td>4–4.9 years old</td>
</tr>
<tr>
<td>5–5.9 years old</td>
</tr>
<tr>
<td>6–6.9 years old</td>
</tr>
<tr>
<td>7–7.9 years old</td>
</tr>
<tr>
<td>8–8.9 years old</td>
</tr>
<tr>
<td>9–9.9 years old</td>
</tr>
<tr>
<td>10–10.9 years old</td>
</tr>
<tr>
<td>11–11.9 years old</td>
</tr>
<tr>
<td>12–12.9 years old</td>
</tr>
<tr>
<td>13–13.9 years old</td>
</tr>
<tr>
<td><strong>Skin type</strong></td>
</tr>
<tr>
<td>Type 1</td>
</tr>
<tr>
<td>Type 2</td>
</tr>
<tr>
<td>Type 3</td>
</tr>
<tr>
<td>Type 4</td>
</tr>
<tr>
<td>Type 5</td>
</tr>
<tr>
<td>Type 6</td>
</tr>
</tbody>
</table>

*Fitzpatrick scale: type 1–6, where type 1 is the lightest and 6 is the darkest.
respectively. For HR, the ICCs for the PBA and CBA were 0.99 and 0.62, respectively.

**BLAND–ALTMAN PLOTS FOR PBA AND TRIAGE SPO₂**
Using the Bland–Altman approach, results showed that the mean difference between the PBA SpO₂ and the triage SpO₂ was -0.17% (standard error [SE] = 0.16, 95% CI = -0.49 to 0.14), indicating that measures of SpO₂ from the PBA were, on average, slightly lower than those from triage. Spearman's correlation between the mean SpO₂ and the difference scores was 0.02. As shown in Figure 1, the Bland–Altman plot indicated that the 95% LOAs were -2.8 to 2.5. The 95% CI for the lower LOA was -2.3 to -3.4, and the 95% CI for the upper LOA was 1.9 to 3.0.

**BLAND–ALTMAN PLOTS FOR CBA AND TRIAGE SPO₂**
Results showed that the mean difference between the CBA SpO₂ and the triage SpO₂ was -0.33% (SE = 0.22, 95% CI = -0.76 to 0.10), indicating that the measures of SpO₂ from the CBA were also, on average, slightly lower than triage measures. Spearman's correlation between the mean SpO₂ and the difference scores was 0.29. As shown in Figure 2, the Bland–Altman plot indicated that the 95% LOAs were -4.1 to +3.5. The 95% CI for the lower LOA was -3.5 to -5.0, and the 95% CI for the upper LOA was 2.8 to 4.3.

**BLAND–ALTMAN PLOTS FOR PBA AND TRIAGE HR**
For HR data, results from the Bland–Altman analysis showed that the mean difference between the PBA HR and the triage HR was -3.0 bpm (SE = 1.2, 95% CI = -5.5 to -0.5), indicating that measures of HR from the PBA were, on average, lower than those from triage. Spearman's correlation between the mean HR and the difference scores was -0.15. As shown in Figure 3, the Bland–Altman plot indicated that the 95% LOAs were -25.3 to +19.3. The 95% CI for the lower LOA was -21.5 to -30.1, and the 95% CI for the upper LOA was 15.6 to 24.2.

**BLAND–ALTMAN PLOTS FOR CBA AND TRIAGE HR**
Results also showed that the mean difference between the CBA HR and the triage HR was -12.0 bpm (SE = 2.5, 95% CI = -16.9 to -7.1), indicating that measures of HR from the CBA were, on average, lower than those from triage. Spearman's correlation between the mean HR and the difference scores was -0.06. As shown in Figure 4, the Bland–Altman plot indicated that the 95% LOAs were -55.7 to +31.7. The 95% CI for the lower LOA was -48.3 to -65.2, and the 95% CI for the upper LOA was 21.3 to 41.2.

**OTHER FACTORS AFFECTING RESULTS**
There were no statistically significant differences in bias, based on which model of smartphone was used (iPhone 5 vs. iPhone 6), all p’s > 0.05. In addition, there were no statistically significant differences in bias among patients with darker skin tones (Fitzpatrick type 4–6) compared with lighter skin tones (Fitzpatrick type 1–3), all p’s > 0.05. Correlational results were calculated using Spearman’s correlation, and indicated that for both the PBA and CBA measures, there were no systematic associations between the magnitude of bias and the mean levels of SpO₂ or HR.
Discussion

To our knowledge, this is the first study that explores the accuracy of smartphone-based portable pulse oximeters in a population of healthy children. Accuracy of these apps is critical for both monitoring children in home settings and also, if this technology would be used, in hospitals as well as clinics in the developing world. Unlike standard pulse oximetry, which is operated and interpreted by healthcare professionals, these apps could also be used by individuals without substantial medical knowledge to provide context for data interpretation, and therefore, the need for reliability across individual measures is great. The ICC calculations show that reliability is good for the PBA but low for the CBA, even when the two investigators tested the same patient within 1 to 2 min of each other. This raises concern of the accuracy and reliability of the CBA measurements. We hypothesized that this discordance between the two investigators was likely because the CBA requires precise placement of the patient’s finger over the camera lens and flash, which may be difficult to do in small children due to the size of their fingers or the need to remain motionless to obtain an accurate reading. The PBA uses technology to reduce motion artifact, which may have contributed to its increased accuracy and reliability, especially in a pediatric population.

Analysis of the accuracy of both apps compared to triage vital signs showed measures of SpO₂ and HR from both PBA and CBA were, on average, lower than the triage measures of SpO₂ and HR. However, the degree of bias
in measures of SpO₂ and HR was smaller for the PBA compared with the CBA, and the estimated lower and upper LOAs for measures of SpO₂ and HR were overall more precise. We hypothesize that the improved accuracy of the PBA is due to the stability provided from the actual probe compared to using the camera lens and flash, as previously stated; the probe more closely resembles the technology used in standard triage pulse oximetry, which may also have contributed to its improved accuracy. Skin tone was also assessed as a potential confounder of accuracy, as some studies have noted that this is a possibility, although at lower SpO₂ values than those studied here.⁴ This was, however, not found to an effect on the accuracy of the two apps.

When looking to clinically apply the difference in SpO₂ measurements between the instruments and the triage vital signs, we find that the difference in the PBA is acceptable (±2 points from the triage values tested), whereas the difference in the CBA is not (±4 points from the triage values tested). This would mean that SpO₂ of 97–100% in triage could be measured as low as 95% on the PBA, whereas it could be measured as low as 93% on the CBA. Given that children have a narrower range of normal SpO₂, that value of 93% would be considered abnormal in an otherwise healthy child,¹ and may prompt unnecessary medical workup. A value as low as 95% is within acceptable range for a healthy child, and thus the PBA demonstrates reliability in a healthy pediatric population.

We feel that ensuring the accuracy of smartphone-based pulse oximetry apps in pediatric population has significant implications for m-health technology and patient care, particularly with the substantial rise in the availability of medical services to individuals in nontraditional forums. Budgetary cuts to schools have led to the decrease in the availability of school nurses nationally, with more telemedicine options being explored.¹⁶ The field of telemedicine would benefit from a low-cost, portable pulse oximeter option to provide additional critical information about a patient’s respiratory status. This technology could easily be used by medical providers in school-based clinics or summer camps, particularly in the care of children with chronic medical conditions, such as asthma. Finally, this technology could serve as a cheaper, portable option compared with bulky equipment currently used for children at home requiring home health monitoring, supplemental oxygen, or continuous pulse oximetry.

Future studies should focus on testing portable pulse oximeters in the hypoxic pediatric patient where the possible range for SpO₂ values is larger than the range used in this study. Future research may also seek to incorporate this technology as a home oxygen monitor, with a cost savings analysis performed for patients and healthcare systems.

**LIMITATIONS**

There are limitations to this study. First, while testing these instruments, we identified that the time between when triage vitals and study measurements were obtained was not standardized. Some children had study measurements taken within 1 h of triage vitals, and others had several hours later, with a median time between measurements for the CBA of 95 min, and a median time between measurements for the PBA of 99 min. While we would not expect SpO₂ values to change significantly, HR values could have changed due to medical interventions (e.g., pain control). This may account for some
increased variation seen in the HR data sets. We also had one child (2 years old) who was too small to use the two apps, because the fingers were too small to pick up a reading from either app. This indicates that there is a lower limit for size of the patients using these apps to maintain accuracy. Finally, these results are not applicable to patients with hypoxia or respiratory-related complaints. Further studies are needed to determine if accuracy of SpO2 is preserved when the range of values is widened by including a hypoxic patient population.

Of note, there was no statistical difference between the measurements obtained on an iPhone 5 versus iPhone 6, indicating that the apps can be used across different generations of smartphone. Future studies can incorporate use of other smartphone brands, but we hypothesize that the phone’s model will not affect the accuracy of the measurements obtained.

Conclusions
To our knowledge, this is the first study that seeks to evaluate the accuracy of portable, smartphone-based, pulse oximeters in a pediatric population. Reliability was good for the PBA that uses an external probe connected to the smartphone, but poor for the CBA that uses the phone’s camera and flash. While both apps had, on average, lower measures of SpO2 and HR compared to triage vital signs, our data support that the agreement was more precise for PBA compared with the CBA. Future studies should target a hypoxic pediatric patient population to see if the same level of accuracy exists across a wider range of values for SpO2.

Disclosure Statement
No competing financial interests exist.

REFERENCES

Address correspondence to: Sarah Tomlinson, MD
Department of Emergency Medicine
University of Michigan
1500 E Medical Center Drive
TC B1380, SPC 5305
Ann Arbor, MI 48109
E-mail: shamilt@med.umich.edu

Received: June 29, 2017
Revised: August 8, 2017
Accepted: August 9, 2017
Online Publication Date: December 7, 2017
Appendix: Manufacturer Guidelines for Use of the Apps


Using iSpO2

Before you begin, see Safety Information, Warnings, and Cautions on page 3.

1. Insert the iSpO2 connector into the dock connector port of the Apple product.
   
   To avoid electrical shock, do not use the iSpO2 when the Apple product is plugged into an AC power outlet.
   
   To avoid electrical shock when using the iSpO2, do not plug anything else into any port of the Apple product, including a headphone jack or docking port.


3. Place your finger in the iSpO2.

The measurements will display on the Apple product as a waveform measurement and numeric display.

2. Masimo iSpO2. Taken from instruction manual (included with purchase of probe).
Best Practices For Accurate Readings

- Ensure that your finger and the sensor are aligned (i.e. the iSpO2 is parallel to the finger). Ensure that the iSpO2 cable is not twisted, as that can cause incorrect alignment.
- Ensure that your finger is inserted correctly and completely into the iSpO2. The tip of your finger should be touching the soft rubber stop inside the sensor.
- Select your testing finger, in the following priority:
  1. Non-dominant ring finger (non-dominant is the hand used less often).
  2. Non-dominant middle finger
  3. Dominant ring finger (dominant is the hand used more often).
  4. Dominant middle finger
- Do not use on an anatomically incorrect finger (e.g. damaged, clubbed, deviated, etc.).
- If artificial nails or excessive fingernail polish are present, select another site or remove the polish/artificial nails.
- Do not use on a finger with occlusive jewelry, such as a ring.
- Your finger should be cleaned of debris and dry prior to sensor placement.
- Do not expose the iSpO2 to moisture, humid environments or liquids.