PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Accuracy of Pulse Oximetry in Children

Patrick A. Ross, Christopher J.L. Newth and Robinder G. Khemani

Pediatrics 2014;133;22; originally published online December 16, 2013;
DOI: 10.1542/peds.2013-1760

Accuracy of Pulse Oximetry in Children

WHAT'S KNOWN ON THIS SUBJECT: Saturations from pulse oximetry (SpO_2) may overestimate arterial oxygen saturations measured by CO-oximetry (SaO_2). The overestimation can be affected by location of measurement, perfusion, and skin color. Previous studies are limited by small numbers of observations in a hypoxemic range.

WHAT THIS STUDY ADDS: This large sample of hypoxemic patients identified that SpO_2 typically overestimates SaO_2. Bias and precision varied throughout the SpO_2 range. The SpO_2 range of 81% to 85% had the greatest bias: median SpO_2 6% higher than SaO_2 measured by CO-oximetry.

abstract

OBJECTIVE: For children with cyanotic congenital heart disease or acute hypoxemic respiratory failure, providers frequently make decisions based on pulse oximetry, in the absence of an arterial blood gas. The study objective was to measure the accuracy of pulse oximetry in the saturations from pulse oximetry (SpO_2) range of 65% to 97%.

METHODS: This institutional review board–approved prospective, multicenter observational study in 5 PICUs included 225 mechanically ventilated children with an arterial catheter. With each arterial blood gas sample, SpO_2 from pulse oximetry and arterial oxygen saturations from CO-oximetry (SaO_2) were simultaneously obtained if the SpO_2 was ≤97%.

RESULTS: The lowest SpO_2 obtained in the study was 65%. In the range of SpO_2 65% to 97%, 1980 simultaneous values for SpO_2 and SaO_2 were obtained. The bias ($SpO_2 - SaO_2$) varied through the range of SpO_2 values. The bias was greatest in the SpO_2 range 81% to 85% (336 samples, median 6%, mean 6.6%, accuracy root mean squared 9.1%). SpO_2 measurements were close to SaO_2 in the SpO_2 range 91% to 97% (901 samples, median 1%, mean 1.5%, accuracy root mean squared 4.2%).

CONCLUSIONS: Previous studies on pulse oximeter accuracy in children present a single number for bias. This study identified that the accuracy of pulse oximetry varies significantly as a function of the SpO_2 range. Saturations measured by pulse oximetry on average overestimate SaO_2 from CO-oximetry in the SpO_2 range of 76% to 90%. Better pulse oximetry algorithms are needed for accurate assessment of children with saturations in the hypoxemic range. *Pediatrics* 2014;133:22–29

AUTHORS: Patrick A. Ross, MD, Christopher J.L. Newth, MD, FRCPC, and Robinder G. Khemani, MD, MsCl

Children's Hospital Los Angeles and University of Southern California Keck School of Medicine, Department of Anesthesiology Critical Care Medicine, Los Angeles, California

KEY WORDS

oximetry, hypoxia, heart defects, congenital heart disease/defects, pediatric, mechanical ventilation

ABBREVIATIONS

ABG—arterial blood gas
ABG/SpO_2 pairs—simultaneous measurement of ABG and SpO_2
AHRF—acute hypoxemic respiratory failure
A_{rms}—accuracy root mean squared
CCHD—cyanotic congenital heart disease
FDA—US Food and Drug Administration
SaO_2—arterial oxygen saturations from CO-oximetry
SpO_2—saturations from pulse oximetry

Dr Ross participated in data collection locally, performed data analysis, and drafted and revised the manuscript for important intellectual content; Dr Newth conceptualized and designed the study and reviewed and revised the manuscript for important intellectual content; Dr Khemani conceptualized and designed the study, coordinated the data collection at 5 sites, participated in data collection locally, performed data analysis, and drafted and revised the manuscript for important intellectual content; and all authors approved the final manuscript as submitted.

www.pediatrics.org/cgi/doi/10.1542/peds.2013-1760

doi:10.1542/peds.2013-1760

Accepted for publication Oct 3, 2013

Address correspondence to Patrick A. Ross, MD, Department of Anesthesiology Critical Care Medicine, 4650 Sunset Blvd Mailstop 12, Children's Hospital Los Angeles, Los Angeles, CA 90027. E-mail: pross@chla.usc.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: No external funding.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

Practitioners frequently make decisions for children based on oxygen saturations obtained from pulse oximetry. Interventions such as supplemental oxygen, diuretics, or transfer to a higher level of care frequently occur without an arterial blood gas (ABG). Pulse oximeter performance has an impact on clinical decisions.

Pulse oximeters are developed to perform optimally in a range of oxyhemoglobin saturation from 70% to 100%. The US Food and Drug Administration (FDA) requires documentation of pulse oximeter accuracy defined as accuracy root mean squared (A_{rms}) <3% with an equal number of samples in the decile range of 70% to 100% from adult volunteers through the standard ISO 80601-2-61:2011.[1] As indicated in the FDA's document "Pulse Oximeters – Premarket Notification Submissions [510(k)s]"[2] for devices intended for use with neonates, the FDA recommends performance reports using adult subjects. Additional convenience arterial samples obtained in neonates are recommended when the sensor is new or significantly changed compared with previous devices. It is unclear how well pulse oximeters perform when the majority of observations are in children in a hypoxemic range. Previous work in children has been limited by small samples of observations in a hypoxemic range. However, these earlier manuscripts indicate SpO_2 may systematically overestimate measured arterial oxygen saturation from ABG sampling.[3,4] The SpO_2 overestimation may also be dependent on location of measurement,[4,5] perfusion to the extremity where the pulse oximeter is located, and skin color.[6,7]

The population of children with lower baseline oxygen saturations is growing. There are many children with cyanotic congenital heart disease (CCHD) undergoing and surviving palliative procedures. In addition, practitioners frequently manage children with acute hypoxemic respiratory failure (AHRF) without arterial catheters.[8] In these children, practitioners often increase ventilator support and fraction of inspired oxygen when the oxygen saturation falls out of an acceptable range (<87%). Although some manufacturers have developed pulse oximeters with hypoxemic measurements in mind (Masimo Blue Sensors, Nellcor LoSat Sensors), these sensors are not routine in most hospitals.

The primary objective of this study was to determine the performance of pulse oximetry for children in the range of 65% to 97% compared with arterial oxygen saturation measurements from CO-oximetry (SaO_2). To overcome limitations with a single measure of bias or precision over the entire range of pulse oximetry, the mean bias, local bias, precision, and A_{rms} are reported. The secondary objective was to explore clinical scenarios in which pulse oximetry may be less reliable.

METHODS

Patients

A prospective, observational study in 5 US multidisciplinary PICUs was conducted from August 2009 to October 2010. The participating PICUs were Children's Hospital Los Angeles, Penn State Children's Hospital, University of Virginia Children's Hospital, Monroe Carell Children's Hospital, and Cohen Children's Medical Center of New York. Patients were eligible if they were intubated, mechanically ventilated, had an arterial catheter, had SpO_2 values ≤97%, and were ≥37 weeks' gestational age and <18 years. Patients were excluded if they were receiving extracorporeal membrane oxygenation or were not on a fully supported mode of ventilation. Continuous pulse oximeter recordings were standard of care in all PICUs. Patient demographics were obtained on enrollment and included age, race/ethnicity, gender, weight, and diagnosis. Patients were identified as to whether they had CCHD or AHRF. All participating PICUs had institutional review board approval with waiver of written consent.[9]

Measurements

The decision to obtain an ABG sample with CO-oximetry was left to the primary care team. The bedside provider was trained by a member of the research team before any data collection. Sensor cleanliness and position were verified before obtaining the ABG. The SpO_2 value was prospectively documented at the precise time the ABG was obtained (ABG/SpO_2 pair). ABG/SpO_2 pairs were not recorded if SpO_2 was >97% or there was endotracheal tube suctioning in the 10 minutes before or ventilator changes in the 30 minutes before the ABG. The ABG/SpO_2 pair was not recorded if the pulse oximeter value was not stable before the ABG or if there was concern about the waveform quality. For inclusion in the analysis, SaO_2 was measured using CO-oximetry. The blood gas machines used in the study were the ABL800 (Radiometer Medical Aps, Brønshøj, Denmark), Rapidlab 1265 (Siemens Healthcare, Erlangen, Germany), and Gem 3000 (Instrumentation Laboratory, Lexington, MA).

Additional information recorded with each ABG/SpO_2 pair included temperature, capillary refill in the extremity with the sensor, hemoglobin, pulse oximeter type, end tidal carbon dioxide, and degree of ventilator support. If the patient had CCHD with a physiology that was dependent on flow through a patent ductus arteriosus and there was a difference in pre- and postductal oxygen saturation, the protocol required that the pulse oximeter sensor and the arterial line were on the same side of the ductus. Ventilator information included peak inspiratory pressure, positive end expiratory pressure, pressure support, rate, exhaled tidal volume, mean

airway pressure, fraction of inspired oxygen, and mode. Hemoglobin values were recorded from laboratory samples, and when unavailable, they were obtained from the ABG. Masimo pulse oximeters with the LCNS line of probes (Masimo Corporation, Irvine, CA) were used by 2 PICUs and Nellcor pulse oximeters with the OxiMax line of probes (Covidien-Nellcor, Boulder, CO) were used by 2 PICUs. One PICU used Masimo oximeters with the Nellcor OxiMax line of disposable probes.

Terms

To describe pulse oximeter performance, it is helpful to review a few statistical terms commonly used: bias, precision, accuracy, and A_{rms}. Bias is the $SpO_2 - SaO_2$. SaO_2 as the reference standard from an ABG measured via CO-oximetry. Mean bias is the average $SpO_2 - SaO_2$ using all study observations, over the entire range, and is influenced by where the preponderance of pulse oximetry measurements lies. Local bias is mean $SpO_2 - SaO_2$ over a specific range of SpO_2. Precision is typically 1 SD above and below the mean bias, describing how much random error is in the data. Methods to calculate precision require the bias to be normally distributed, which may not always be true. The limits of agreement are usually taken as 1.96 times the SD unless there is correction for repeated measures. The term "accuracy" is a measure of how far a value is from a reference standard. However, pulse oximeter companies generally present A_{rms}, which is required by agencies that regulate pulse oximeters. The A_{rms} combines the components of bias and precision controlling for the number of samples obtained. A_{rms} of $\leq$3% is required by the FDA.

Analysis

The primary objective of this study was to evaluate the accuracy of simultaneous samples of SpO_2 compared with SaO_2 obtained by CO-oximetry throughout a range of SpO_2 values. Additional objectives were to identify variables that may affect the bias of SpO_2 compared with SaO_2 in the hypoxemic range. Statistical analysis was performed by using Stata version 10 (StataCorp, College Station, TX) and Statistica version 9 (Statsoft, Tulsa, OK).

Overall bias was assessed with a scatter plot of SpO_2 against SaO_2. Local bias was assessed via a box-and-whisker plot examining the difference between SpO_2 and SaO_2 in 7 SpO_2 ranges. To explore inaccuracy, a bar graph was generated of counts of ABG/SpO_2 pairs in which the absolute value of ($SpO_2 - SaO_2$) is $\leq$3% and $>$3% as a function of SpO_2 range. Precision was reported with SD. However, because residuals were not normally distributed, median and interquartile range of ($SpO_2 - SaO_2$) as a whole and for each SpO_2 range were reported. A_{rms} was calculated from the formula

$$A_{rms} = \sqrt{\frac{\sum_{i=1}^{N} (SpO_{2i} - SaO_{2i})^2}{n}}$$

as a whole and for each SpO_2 range.

To evaluate the potential influence of clinical variables on pulse oximetry bias, the ABG/SpO_2 pairs were separated into groups based on absolute value of the bias of $\leq$3% or $>$3%. This separation was used to perform a multivariate logistic regression analysis using a mixed model to examine the effect of other potential confounding variables and to control for repeated measures per patient. The confounding variables included disease category, gender, oximeter and sensor type, capillary refill in the extremity with the pulse oximeter, hemoglobin, temperature, and degree of ventilator support. Confounders were considered for model inclusion if they had a univariate relationship with the outcome ($P < .2$) or if there was strong biologic plausibility to suspect that they may influence the relationship between SpO_2 and SaO_2 (eg, capillary refill).

Finally, to explore whether multiple measurements per subject biased the results, data were filtered to randomly extract a single ABG/SpO_2 pair from each patient that fell within the SpO_2 ranges previously identified. Values for bias and precision (SD) using only 1 sample per patient per range were calculated and compared with the entire data set.

RESULTS

Two hundred twenty-five children were enrolled with 1980 ABG/SpO_2 pairs. The median number of ABG/SpO_2 pairs per patient was 5, interquartile range (2–10), range (1–110). Demographics separated by CCHD or AHRF are shown in Table 1. The source and results of the ABG/SpO_2 pairs including ventilation information are shown in Table 2. The children with CCHD were younger, weighed less, had lower oxygen saturations, had higher hemoglobin values, and required less ventilator support compared with the group with AHRF (all Ps $< .001$). The 122 children with CCHD accounted for 1175 ABG/SpO_2 pairs with a median SpO_2 of 82% and a median SaO_2 of 77%, with 82% of the ABG/SpO_2 pairs having a SpO_2 $\leq$90%. The 103 children with AHRF accounted for 805 ABG/SpO_2 pairs with a median SpO_2 of 95% and a median SaO_2 of 94%, with 14% of the ABG/SpO_2 pairs having a SpO_2 $\leq$90%.

Entire Range of SpO_2

Values for SpO_2 are plotted against SaO_2 for all ABG/SpO_2 pairs in Fig 1. A total of 1304 of the 1980 samples (66%) had a positive bias ($SpO_2 - SaO_2 > 0$). Although the data were not normally distributed, mean and SD are presented to be consistent with other studies. For the entire SpO_2 range 65% to 97% the mean bias was 3.3%, median 2%, the precision (SD) 5.6%, interquartile range (0%–6%), and the A_{rms} 6.5% (Table 3).

TABLE 1 Patient Demographics

	CCHD (n = 122)	AHRF (n = 103)	P
Age, mo, (interquartile range)	1 (1–3)	37 (5–140)	<.001
Weight, kg, (interquartile range)	3.5 (3–4.6)	14.8 (5–36)	<.001
Gender, male, n (%)	74 (61)	63 (62)	.95
Race/ethnicity, n (%)			
African American	8 (7)	12 (12)	.20
Hispanic	59 (48)	35 (34)	
White	47 (39)	47 (45)	
Asian	4 (3)	3 (3)	
Other	4 (3)	6 (6)	

Differences between groups analyzed by Mann-Whitney U Test, $\chi 2$ test or observed/expected $\chi 2$ test.

TABLE 2 Source and Results of ABG/SpO_2 Pairs Including Ventilation

	CCHD (n = 1175)	AHRF (n = 805)	P
Oximeter-Sensor			
Masimo-Masimo	982 (83.6%)	578 (71.8%)	<.001
Nellcor-Nellcor	146 (12.4%)	180 (22.4%)	
Masimo-Nellcor	47 (4%)	47 (5.8%)	
SpO_2%	82 (78–88)	95 (92–96)	<.001
SaO_2%	77 (72–83)	94 (90–96)	<.001
pH	7.39 (7.34–7.44)	7.40 (7.33–7.45)	.24
PaO_2, mm Hg	42 (38–49)	69 (60–83)	<.001
$PaCO_2$, mm Hg	45 (40–52)	55 (45–69)	<.001
Base excess	0 (0–0)	0 (0–0)	1
FiO_2	0.35 (0.25–0.55)	0.6 (0.42–0.67)	<.001
Capillary refill (s)	3 (3–3)	3 (1–3)	<.001
Hemoglobin (g/dL)	15 (14–16)	11 (10–13)	<.001
Temperature °C	37 (37–37)	37 (37–37)	.76
Conventional ventilation	n = 1168	n = 611	
End tidal carbon dioxide (mm Hg)	32 (25–37)	37 (27–45)	<.001
Peak inspiratory pressure (cm H_20)	24 (20–26)	30 (24–36)	<.001
Positive end-expiratory pressure (cm H_20)	5 (4–5)	8 (5–10)	<.001
Mean airway pressure (cm H_20)	9 (8–11)	15 (11–20)	<.001
Tidal volume (mL/kg)	10 (8–10)	8 (7–10)	<.001
Pressure support (cm H_20)	10 (10–10)	10 (5–11)	<.001
Ventilator rate (breaths/min)	20 (15–25)	20 (16–25)	.89
High frequency oscillatory ventilation	n = 7	n = 194	
Mean airway pressure (cm H_20)	28 (21–31)	30 (25–33)	.20
Amplitude	51 (48–54)	60 (50–65)	.07
Frequency (Hz)	9 (9–9)	8 (6–9)	.04

Number and percentage of oximeter-sensor combinations are calculated by column. Data (not normally distributed) are presented as median (interquartile range). Differences between groups analyzed by Mann-Whitney U test or observed/expected $\chi 2$ test. FiO_2, fraction of inspired oxygen.

Smaller Ranges of SpO_2

There were significant differences in bias, precision (SD), and accuracy based on SpO_2 range. The local bias ($SpO_2 - SaO_2$) was lowest in the SpO_2 ranges of 65% to 70% and 96% to 97% with median values of 0% and greatest in the SpO_2 range of 81% to 85% with a median value of 6% (Fig 2). Precision (SD) and A_{rms} were worst in the SpO_2 range of 81% to 85%, with poor accuracy in all ranges <91% (Table 3).

The data were filtered to randomly extract a single ABG/SpO_2 pair from each patient that fell within the SpO_2 ranges. Five hundred twenty-six ABG/SpO_2 pairs remained with a mean bias of 2.8%, median 2%, precision (SD) 5.6%, interquartile range (–1% to 6%), and A_{rms} was 6.3% (nearly identical to values including repeated measurements). Findings were also similar within each of the SpO_2 ranges (analysis not shown).

Multivariate Modeling

The ABG/SpO_2 pairs were separated into 2 groups based on whether the absolute value of the bias was ≤3% or >3% (Table 4). For ABG/SpO_2 pairs in which the bias was >3%, children with CCHD accounted for 78% of the samples (719 of 922) compared with children with AHRF (P < .001). The proportion of instances when the bias is >3% is highest in the midrange of SpO_2 (75%–90%; Fig 3).

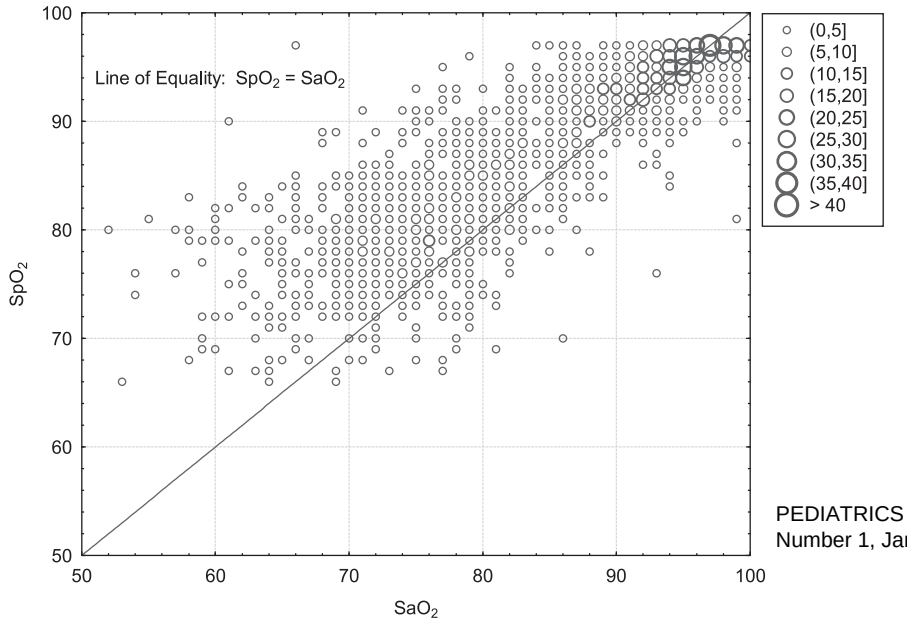
From multivariate modeling, variables associated with higher likelihood of bias >3% include CCHD, prolonged capillary refill, and having a SpO_2 between 81% to 85%, 86% to 90%, or 91% to 95% (compared with a SpO_2 of 96% to 97%; Table 5). Variables associated with a lower likelihood of bias were African American race/ethnicity, male gender, and the combination of Masimo oximeter with a Nellcor sensor. After controlling for these variables, mean airway pressure, hemoglobin, PICU site, temperature, fraction of inspired oxygen, age <2 months, other races/ethnicity, and other oximeter combinations did not contribute to the model.

Because the majority of observations in the lower ranges of SpO_2 came from children with CCHD, a second multivariate model was performed restricted to only the subgroup of children with CCHD. Overall results were similar with a higher likelihood of bias associated with prolonged capillary refill and SpO_2 ranges of 81% to 85% and 86% to 90%. Furthermore, a lower likelihood of bias was associated with African American race/ethnicity and male gender.

DISCUSSION

There is significant variability in the bias, precision (SD) and accuracy of pulse oximetry as a function of SpO_2. SpO_2 on average overestimates SaO_2 measured with CO-oximetry. This local bias is most positive in the SpO_2 range

Message: Se méfier de la SpO2 equi surestime la SaO2 !



PEDIATRICS Volume 133, Number 1, January 2014

FIGURE 1

SpO_2 plotted against SaO_2 for all ABG/SpO_2 pairs. The line indicates the line of equality between SpO_2 and SaO_2. Bubble size indicates number of values in that range. A total of 1304 of the 1980 samples (66%) lie above the line, indicating a positive bias for SpO_2 measurements.

TABLE 3 Percent Difference of SpO_2 – SaO_2 for Each ABG/SpO_2 Pair

SpO_2 Range %	Count	Median Bias %	IQR	Mean Bias %	SD %	A_{rms} %
Entire range 65–97	1980	2	0 to 6	3.3	5.6	6.5
6–70	41	0	−3 to 4	0.2	6.6	6.5
71–75	120	2	−1 to 6	2.7	6.5	6.9
76–80	319	4	1 to 8	4.8	6.0	7.7
81–85	336	6	3 to 11	6.6	6.3	9.1
86–90	263	4	0 to 7	4.2	5.6	7.0
91–95	540	1	−1 to 4	1.8	4.2	4.5
96–97	361	1	−1 to 3	1.0	3.4	3.6

Presented are count of ABG/SpO_2 pairs, median bias, interquartile range (IQR), mean bias, SD, and A_{rms} presented for SpO_2 65% to 97% and subset ranges.

between 81% to 85%, although it is present in the SpO_2 range of 76% to 90%. The precision (SD) and accuracy are also poor in this range. Although the median bias is low in the SpO_2 range from 65% to 75%, there is poor precision (SD) and accuracy. This variability in bias, precision, and accuracy remains after controlling for capillary refill, diagnosis, pulse oximeter, gender, and race/ethnicity.

These findings may have significant clinical implications. For example, if the SpO_2 is measured at 85%, on average the SaO_2 would be 79% and 50% of the time the SaO_2 would lie between 75% and 83%. However, to achieve 95% certainty, SaO_2 values would lie between 64% and 89%. In contrast, if the SpO_2 is measured at 94%, on average the SaO_2 would be 93% and 50% of the time the SaO_2 would lie between 90% and 95% and 95% of the time the SaO_2 would lie between 89% and 99%. In the lowest SpO_2 range of 65% to 70%, the median bias is close to zero, but the poor precision (SD) and accuracy could affect clinical care. Physicians caring for CCHD patients often make different treatment decisions based on whether the SpO_2 is 65% vs 70%. Improved accuracy in this range could help decision-making.

The increased bias and poor precision in the lower SpO_2 range may add insight into recent findings from neonatology demonstrating potential harm with lower oxygen saturations.[10–12] Three large multicenter randomized trials have targeted ranges of SpO_2 to prevent retinopathy of prematurity. Two of the 3 trials demonstrated increased mortality in the oxygen saturation group from 85% to 89%, compared with 91% to 95%.[10,12] The increased bias when SpO_2 is <90% may result in SaO_2 values that are much lower than anticipated.

The multivariate model shows that pulse oximetry performs less well in children with CCHD and prolonged capillary refill. Children with CCHD were the largest proportion in the SpO_2 range from 81% to 85%, although this range remained the most likely for bias even when restricting the analysis only to children with CCHD. Therefore, it appears the inaccuracy is most related to the range of SpO_2 and not the presence of CCHD. Prolonged capillary refill makes physiologic sense because greater bias is anticipated with poorer

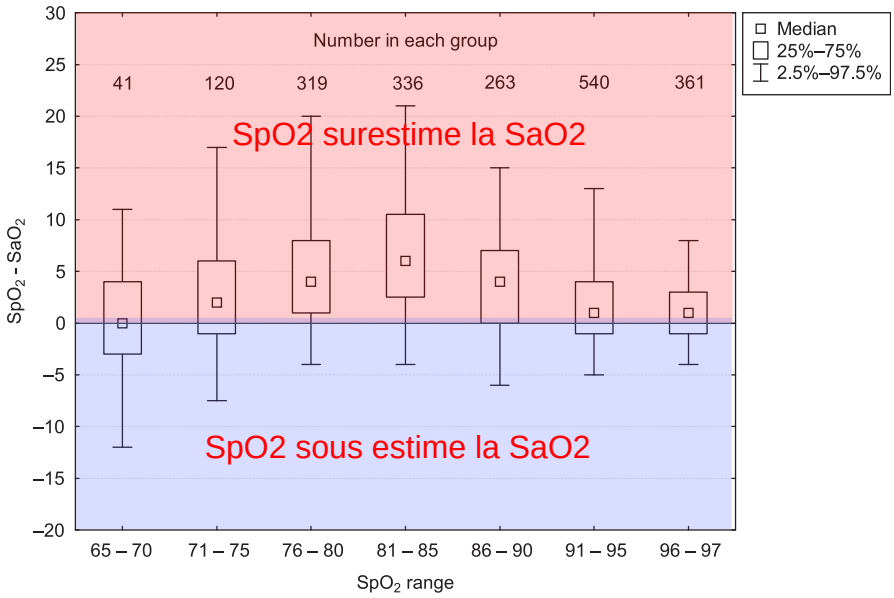


FIGURE 2
Box plot of ($SpO_2 - SaO_2$) by SpO_2 range increments for all patients. The number of ABG/SpO_2 pairs in each increment is above the whisker. Bias varies throughout the SpO_2 range with the largest bias in 81% to 85% SpO_2 range. The bias is small for $SpO_2 \geq 91\%$.

perfusion.[3] Age <2 months showed a univariate effect, but this was not supported by the regression model. Age was meant as a surrogate for fetal hemoglobin, which may also be partially captured by other elements of the multivariate model. African American race/ethnicity and male gender were associated with lower likelihood of bias of the pulse oximeter. It is unclear what conclusions can be drawn from this. Race/ethnicity but not degree of skin color was recorded. There is no clear biological mechanism in which gender should have an impact on pulse oximetry performance. It is possible these are surrogates for an unmeasured confounding variable.

The findings of this study are consistent with other, smaller studies. Das et al[5] in 2010 studied pulse oximeter bias based on sensor location. Unfortunately, only 8 children had oxygen saturations <90%. However, in their small sample size, SpO_2 was always greater than SaO_2. Sedaghat-Yazdi et al[4] in 2008 studied pulse oximeter bias based on sensor location. However, only 24 samples had SaO_2 <90%. In 2004, Torres et al[3] showed poor pulse oximeter accuracy with 77 SaO_2 samples <90%. Our study echoes these findings and has the largest sample size to date by an order of magnitude.

Although pulse oximeters are marketed as meeting standards of accuracy throughout a range of oxygen saturations, this study highlights the need for manufacturers to present more data than the A_{rms} for the entire range. Because mean bias and A_{rms} will be influenced by large number of samples with SpO_2 values >91%, where pulse oximeters perform best, a single number for accuracy does not tell the entire picture. Furthermore, accuracy of pulse oximetry algorithms are often demonstrated using adult healthy volunteers breathing a hypoxic gas mixture, which is not the clinical environment in which they are used.

The local bias, precision (SD), and accuracy values presented are much larger than anticipated. The data for this study were acquired in a clinical setting, in the context of a reasonably controlled research study. Accuracy of pulse oximetry algorithms are generally demonstrated in a laboratory setting using adult healthy volunteers breathing a hypoxic gas mixture. This study was unable to meet the rigorous conditions of a research laboratory setting,[13] but the findings highlight the real-world clinical application of this scenario. Although it is not reasonable to generate hypoxia in children to produce an improved algorithm for pulse oximetry, there are pediatric cardiac surgery centers with large numbers of patients with CCHD. Perhaps future development could occur with a population such as this to generate algorithms that are more accurate for all children.

There are limitations to this study. First, the multicenter nature of the study increases generalizability, but the results may be confounded by institutional variation. For example, in 1 hospital, a sensor was used that was different from the manufacturer of the oximeter machine, making it difficult to determine the source of inaccuracy. However, this is a real-world situation, and results from that hospital were not large enough to skew the overall trend. Subgroup analysis by pulse oximeter

TABLE 4 Absolute Value of (SpO_2 – SaO_2) Difference as ≤3% or >3% by Condition

Condition	Absolute Value of (SpO_2 – SaO_2) Difference		*P*
	≤3% (*n* = 1058)	>3% (*n* = 922)	
CCHD	456 (43.1%)	719 (78.0%)	<.001
AHRF	602 (56.9%)	203 (22.0%)	
Age			
<2 mo	453 (42.8%)	513 (55.6%)	<.001
>2 mo	605 (57.2%)	409 (44.4%)	
Gender, male	802 (75.8%)	584 (63.3%)	<.001
Race/Ethnicity			
African American	201 (19.0%)	86 (9.3%)	<.001
Hispanic	292 (27.6%)	282 (30.6%)	
White	484 (45.7%)	463 (50.2%)	
Asian	39 (3.7%)	50 (5.4%)	
Other	40 (3.8%)	41 (4.4%)	
Unknown	2 (0.2%)	0	
Capillary refill (s)			
1	396 (37.4%)	221 (24.0%)	<.001
3	525 (49.6%)	540 (58.6%)	
4	115 (10.9%)	127 (13.8%)	
5	19 (1.8%)	22 (2.1%)	
6	3 (0.3%)	12 (1.3%)	
PICU			
1	513 (48.5%)	538 (58.3%)	<.001
2	74 (7.0%)	20 (2.2%)	
3	273 (25.8%)	236 (25.6%)	
4	79 (7.5%)	62 (6.7%)	
5	119 (11.2%)	66 (7.2%)	
Oximeter-sensor combination			
Masimo-Masimo	786 (74.3%)	774 (83.9%)	<.001
Nellcor-Nellcor	198 (18.7%)	128 (13.9%)	
Masimo-Nellcor	74 (7.0%)	20 (2.2%)	
SpO_2 range			
65–70	19 (1.8%)	22 (2.4%)	<.001
71–75	52 (4.9%)	68 (7.4%)	
76–80	141 (13.3%)	178 (19.3%)	
81–85	90 (8.5%)	246 (26.7%)	
86–90	110 (10.4%)	143 (15.5%)	
91–95	356 (33.6%)	184 (20.0%)	
96–97	290 (27.4%)	71 (7.7%)	

Data are count of ABG/SpO_2 pairs (percent total). Differences between groups analyzed by $\chi 2$ test or observed/expected $\chi 2$ test.

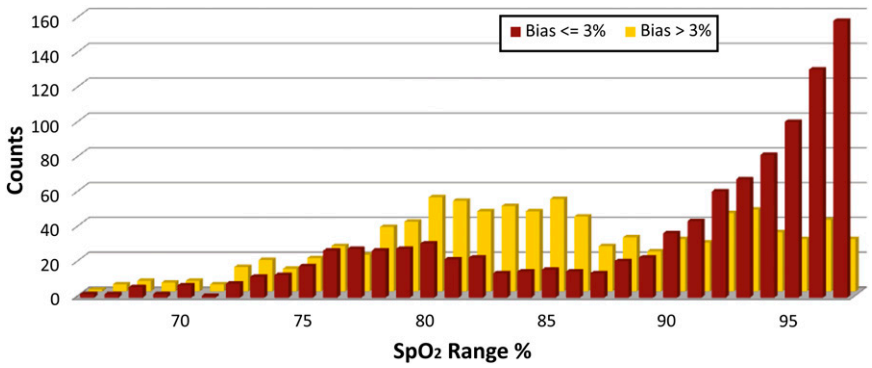


FIGURE 3
Count of ABG/SpO_2 pairs by absolute value of the bias ≤3% or >3% over SpO_2 range. The proportion of observation varies over the range with an increased number at higher SpO_2. The effect of local bias would be missed if the entire range were presented as a single number.

TABLE 5 Multivariate Model Controlling for Patient Effect

Variable	Coefficient	*P*
CCHD	1.82	<.001
Capillary refill	0.26	<.001
Male gender	−0.30	.023
Race/ethnicity		
White	Baseline	NA
African American	−0.55	.003
Hispanic	−0.15	NS
Asian	−0.26	NS
Other	−0.03	NS
Oximeter-Sensor		
Masimo-Masimo	Baseline	NA
Masimo-Nellcor	−1.45	.011
Nellcor-Nellcor	0.04	NS
SpO_2 range		
96%–97%	Baseline	NA
65%–70%	0.23	NS
71%–75%	0.16	NS
76%–80%	0.05	NS
81%–85%	0.98	.001
85%–90%	1.08	<.001
91%–95%	0.75	<.001

Entire group of patients. Positive values for the coefficient imply higher likelihood of outcome (bias >3%). Variables that were considered but not included are mean airway pressure, hemoglobin, PICU site, temperature, and age <2 mo. NA, not applicable

brand was not performed because the majority of samples were measured with 1 oximeter brand. However, there was similar SpO_2 overestimation in the second manufacturer, and in that subgroup, this sample size is as large as any previously reported. Moreover, pulse oximetry type (except the combination of Masimo oximeter and Nellcor sensor) was not significant in the multivariate model.

A second limitation is that the data were collected using a waiver of consent. In turn, no patient-identifying information was retained, preventing us from gathering variables not initially included in the study. Such examples include perfusion index, location of pulse oximeter probe, amount of fetal hemoglobin, presence of other hemoglobin species, use of inotropic medications, presence of a patent ductus arteriosus, and pulse pressure. These are potential confounding variables that were not captured. These areas should be investigated in future studies with pulse oximetry.

A third limitation is that fetal hemoglobin was not measured. Pulse oximeter algorithms are generated from adult volunteers with the assumption of 2 types of hemoglobin: oxy and deoxyhemoglobin. They may not perform well with carboxyhemoglobin, methemoglobin, or possibly fetal hemoglobin. There are limited studies showing the performance of pulse oximeters in the presence of fetal hemoglobin,[14,15] with conflicting results. Although it is possible that fetal hemoglobin has an impact on pulse oximetry it is unlikely to explain all of the findings, given that results are similar when restricting to only older patients (Supplemental Information). Furthermore, there was no difference in oxygen dissociation curves based on age ≤2 months or >2 months (Supplemental Information), supporting the idea that the findings are not solely explained by fetal hemoglobin.

CONCLUSIONS

Pulse oximeters appear to be accurate against CO-oximetry in a higher SpO_2 range such as >91%. However, in the lower SpO_2 range (76%–90%), the local bias ($SpO_2 - SaO_2$) is ~5% and the precision (SD) and accuracy are poor when SpO_2 is <90%. The manufacturers of these devices should improve algorithms in this range.

ACKNOWLEDGMENTS

The authors thank those people who organized and participated in the research study: "Comparison of SpO_2 to PaO_2 Based Markers of Lung Disease Severity for Children With Acute Lung Injury," which provided the data for this article: Neal J. Thomas, MD, MsC, Vani Venkatachalam, MD, Jason P. Sciememe, MD, Ty Berutti, MD, MS, James B. Schneider, MD, Douglas F. Willson, MD, Mark W. Hall, MD, Debbie Spear, Jill Raymond, Christine Traul, Jeff Terry, and Paul Vee.

REFERENCES

1. International Organization for Standardization (ISO). Particular requirements for basic safety and essential performance of pulse oximeter equipment (ISO 80601-2-61-2011). Geneva, Switzerland: ISO; 2013
2. Center for Devices and Radiological Health. Pulse oximeters—premarket notification submissions [510(k)s]. (document issued March 4, 2013). Food and Drug Administration; 2013. Available at: www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM081352.pdf. Accessed November 1, 2013
3. Torres A Jr, Skender KM, Wohrley JD, et al. Pulse oximetry in children with congenital heart disease: effects of cardiopulmonary bypass and cyanosis. *J Intensive Care Med*. 2004;19(4):229–234
4. Sedaghat-Yazdi F, Torres A Jr, Fortuna R, Geiss DM. Pulse oximeter accuracy and precision affected by sensor location in cyanotic children. *Pediatr Crit Care Med*. 2008;9(4):393–397
5. Das J, Aggarwal A, Aggarwal NK. Pulse oximeter accuracy and precision at five different sensor locations in infants and children with cyanotic heart disease. *Indian J Anaesth*. 2010;54(6):531–534
6. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005;102(4):715–719
7. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesth Analg*. 2007;105(suppl 6):S18–S23
8. Khemani RG, Markovitz BP, Curley MA. Characteristics of intubated and mechanically ventilated children in 16 pediatric ICUs. *Chest*. 2009;136:765–771
9. Khemani RG, Thomas NJ, Venkatachalam V, et al; Pediatric Acute Lung Injury and Sepsis Network Investigators (PALISI). Comparison of SpO2 to PaO2 based markers of lung disease severity for children with acute lung injury. *Crit Care Med*. 2012;40(4):1309–1316
10. Carlo WA, Finer NN, Walsh MC, et al; SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Target ranges of oxygen saturation in extremely preterm infants. *N Engl J Med*. 2010;362(21):1959–1969
11. Schmidt B, Whyte RK, Asztalos EV, et al; Canadian Oxygen Trial (COT) Group. Effects of targeting higher vs lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. *JAMA*. 2013;309(20):2111–2120
12. The BOOST II United Kingdom, Australia, and New Zealand Collaborative GroupsOxygen saturation and outcomes in preterm infants. *N Engl J Med*. 2013;368(22):2094–2104
13. Batchelder PB, Raley DM. Maximizing the laboratory setting for testing devices and understanding statistical output in pulse oximetry. *Anesth Analg*. 2007;105(suppl 6):S85–S94
14. Shiao SY. Effects of fetal hemoglobin on accurate measurements of oxygen saturation in neonates. *J Perinat Neonatal Nurs*. 2005;19(4):348–361
15. Rajadurai VS, Walker AM, Yu VY, Oates A. Effect of fetal haemoglobin on the accuracy of pulse oximetry in preterm infants. *J Paediatr Child Health*. 1992;28(1):43–46

Accuracy of Pulse Oximetry in Children

Patrick A. Ross, Christopher J.L. Newth and Robinder G. Khemani

Pediatrics 2014;133;22; originally published online December 16, 2013;
DOI: 10.1542/peds.2013-1760