Four Types of Pulse Oximeters Accurately Detect Hypoxia during Low Perfusion and Motion

Aaron Louie, B.S., John R. Feiner, M.D., Philip E. Bickler, M.D., Ph.D., Laura Rhodes, B.S., Michael Bernstein, B.S., Jennifer Lucero, M.D.

ABSTRACT

Background: Pulse oximeter performance is degraded by motion artifacts and low perfusion. Manufacturers developed algorithms to improve instrument performance during these challenges. There have been no independent comparisons of these devices.

Methods: We evaluated the performance of four pulse oximeters (Masimo Radical-7, USA; Nihon Kohden OxyPal Neo, Japan; Nellcor N-600, USA; and Philips Intellivue MP5, USA) in 10 healthy adult volunteers. Three motions were evaluated: tapping, pseudorandom, and volunteer-generated rubbing, adjusted to produce photoplethsmogram disturbance similar to arterial pulsation amplitude. During motion, inspired gases were adjusted to achieve stable target plateaus of arterial oxygen saturation (Sa_{O_2}) at 75%, 88%, and 100%. Pulse oximeter readings were compared with simultaneous arterial blood samples to calculate bias (oxygen saturation measured by pulse oximetry [Sp_{O_2}] – Sa_{O_2}), mean, SD, 95% limits of agreement, and root mean square error. Receiver operating characteristic curves were determined to detect mild ($Sa_{O_2} < 90\%$) and severe ($Sa_{O_2} < 80\%$) hypoxemia.

Results: Pulse oximeter readings corresponding to 190 blood samples were analyzed. All oximeters detected hypoxia but motion and low perfusion degraded performance. Three of four oximeters (Masimo, Nellcor, and Philips) had root mean square error greater than 3% for Sa_{O_2} 70 to 100% during any motion, compared to a root mean square error of 1.8% for the stationary control. A low perfusion index increased error.

Conclusions: All oximeters detected hypoxemia during motion and low-perfusion conditions, but motion impaired performance at all ranges, with less accuracy at lower Sa_{O_2}. Lower perfusion degraded performance in all but the Nihon Kohden instrument. We conclude that different types of pulse oximeters can be similarly effective in preserving sensitivity to clinically relevant hypoxia. (ANESTHESIOLOGY 2018; 128:520-30)

PULSE oximetry is a noninvasive technology for continuous monitoring of arterial oxygen saturation (Sa_{O_2}), and has been a standard tool used to assess oxygenation and respiratory function in patients.[1–3] Pulse oximeters transmit red and near-infrared light across a tissue bed (*e.g.*, finger, toe, or earlobe) and detect changes in light absorbance to calculate estimated Sa_{O_2} based on photoplethysmography.[2] Healthcare providers rely on readings of oxygen saturation measured by pulse oximetry (Sp_{O_2}) to provide accurate measurements of oxygen saturation, especially in the detection of hypoxemia associated with deterioration of respiratory function.[2]

Patient movement and low perfusion to extremities can generate artifacts that reduce the accuracy of Sp_{O_2} readings.[1] These limitations are particularly relevant in clinical settings like obstetric units (*e.g.*, laboring women or women with neuraxial anesthesia undergoing cesarean delivery), intensive care unit and postanesthesia care unit, where patients experience voluntary and involuntary movement, including tapping, rubbing, shivering, and seizures in adult and pediatric patients, and kicking and crying in neonates.[4] Motion artifacts can result in low signal-to-noise ratios and underestimation of Sa_{O_2} due to venous blood motion. Errors can be exacerbated by low perfusion.[5] Furthermore, these limitations can obscure true signals, triggering false alarms that

What We Already Know about This Topic

- Pulse oximeter performance is degraded by motion artifacts and low perfusion. Manufacturers developed algorithms to improve instrument performance during these challenges.
- There have been no independent comparisons of these devices.

What This Article Tells Us That Is New

- This study determined the performance of four pulse oximeters (Masimo Radical-7, USA; Nihon Kohden OxyPal Neo, Japan; Nellcor N-600, USA; and Philips Intellivue MP5, USA) in 10 healthy adult volunteers.
- All oximeters detected hypoxemia during motion and low-perfusion conditions, but motion impaired performance at all ranges, with less accuracy at lower arterial oxygen saturation. Lower perfusion degraded performance in all but the Nihon Kohden instrument.

may lead to ignored true alarms—all of which can compromise patient safety and increase cost of care.[6,7]

Over the years, manufacturers of pulse oximeters have developed software algorithms to reduce motion artifacts; the devices containing such software are marketed as motion tolerant/resistant.[8] Other groups have studied and compared the performances of various "new generation" pulse oximeters under simulated motion and low perfusion. To date,

these studies have mainly compared Sp_{O_2} readings from volunteer-initiated hand motion to readings of the volunteer's nonmotion hand as the only reference for true Sa_{O_2} value.[8–14] The U.S. Food and Drug Administration (FDA; Silver Spring, Maryland), however, requires that manufacturers validate accuracy specifications of pulse oximeters by comparing each Sp_{O_2} value with a "gold standard" Sa_{O_2} measurement collected by simultaneous co-oximetry of an arterial blood sample.[15] Similar recommendations exist for validating new generation devices claiming resistance to motion and low perfusion artifacts. While the FDA has a specification of root mean square error less than or equal to 3.0% for pulse oximeter accuracy testing across a 70 to 100% Sa_{O_2} range, there is currently no numerical standard for motion and low perfusion conditions.

In this study we evaluated four commercially available motion- and low perfusion–tolerant pulse oximeters during different types of controlled movement and a range of perfusion. Consistent with FDA guidelines, Sp_{O_2} readings from each device were compared with Sa_{O_2} measurements from simultaneous arterial blood sampling.

Materials and Methods

This study was approved by the University of California, San Francisco Committee on Human Research (San Francisco, California). Written informed consent was obtained from all subjects. Ten healthy adult subjects participated in the study: six men and four women, all healthy, who ranged in age between 22 and 44 yr (median age, 27 yr). Skin tone varied to include 10% light, 40% light/medium, 30% medium, 20% medium/dark, and no dark skin tone. Ethnicity varied to include 60% Asian, 30% Caucasian, 10% Latino, and no African American. All subjects were nonsmokers with no evidence of lung disease, obesity, or cardiovascular problems. Sample size was based on FDA guidelines[3] for accuracy testing that required at least 200 data points balanced across each decadal range (70 to 80%, 80 to 90%, and 90 to 100%) of Sa_{O_2} range from 70 to 100%. Subjects were studied using the identical protocol implemented by our laboratory that routinely tests pulse oximeters for FDA 501(k) certification.[16]

Subjects were placed in the semirecumbent position (30° head up) with a nose clip and breathed air–nitrogen–carbon dioxide mixtures through a mouthpiece from a partial rebreathing circuit with a voluntarily increased minute ventilation and 10 to 20 l/min fresh gas inflow, but with carbon dioxide added as needed to maintain normocapnia. The control hand was placed on a stationary arm rest at elbow level, where an indwelling 22-gauge radial artery catheter was placed to sample arterial blood for measurement of Sa_{O_2}. The opposite motion hand rested securely on a motorized table that generated repeatable and continuous vertical hand movements while the elbow remained fixed and the fingertips either tapped or rubbed on a smooth surface.

Each subject was monitored with four pulse oximeters on the test hand: Masimo Radical-7 with SET software

V7.6.2.1 (USA), Nellcor OxyMax N-600 with software version 1.5.2.2 (USA), Nihon Kohden OxyPal Neo with software Ver79-06, 98-16 (Japan), and Philips Intellivue MP5 with software L.10.75 (USA). Disposable adhesive sensors were used to prevent sensor displacement and were randomly assigned to digits 2 to 5 on every subject's test hand. A reference oximeter (Masimo Radical-7) was mounted randomly to a digit on the control hand using a reusable sensor, and baseline perfusion index (PI) was recorded on that hand.

Three motion modalities were tested: tapping, random, and rubbing. The motion table was programmed to perform (1) machine-generated tapping at a fixed amplitude of ±3 cm and frequency of 2 Hz; (2) machine-generated aperiodic random motion with maximal amplitude and frequency at ±3 cm and 2 Hz, respectively; or (3) volunteer-generated 2-Hz sideways rubbing with the aid of a metronome. Each of these movements produced significant photoplethysmogram disturbance artifacts that were adjusted to be 100 ± 25% of the amplitude of arterial pulsations. We made no efforts to control the perfusion state of the subjects; however, all subjects were well hydrated and the room temperature was about 28°C.

A series of three stable target Sa_{O_2} plateaus between 70 and 100% (approximately 73%, 88%, and 98%) were targeted by the operator, who adjusted the inspired air–nitrogen–carbon dioxide mixture breath-by-breath to achieve the desired saturation. This was done by using end-tidal carbon dioxide and oxygen analysis (Applied Electrochemistry oxygen and carbon dioxide analyzers, USA) and a computer algorithm (LabVIEW 2013, National Instruments, USA) that involves a model oxyhemoglobin dissociation curve and inputted values for hemoglobin P50, arterial-alveolar gas gradients, and base excess.[17]

At each level, arterial blood was sampled after a stable plateau of 30 to 60 s had been achieved, followed by a second sample at the same plateau 30 s later. Parallel Sp_{O_2} readings from all tested oximeters were recorded by hand throughout the protocol. Functional arterial Sa_{O_2} ($Hb_{O_2}/[Hb + Hb_{O_2}]$) was determined by multiwavelength oximetry using a Radiometer ABL-90 (Denmark), which was calibrated according to manufacturer recommendations.

Statistical Analysis

Specific power calculations for a study involving a mixed-effects model are complex and were not undertaken. However, based on our previous published pulse oximeter performance studies using a repeated-measures design and a 10-subject pool, we have found statistically significant differences in pulse oximeter performance that are smaller than clinically relevant effects using same subject size.[18] Additionally, a study of 10 subjects conforms to the FDA's guidelines for study design related to claims for pulse oximetry motion performance.[3] Further, an unpublished pilot study in our laboratory was performed with 10 subjects using an identical motion protocol, and results revealed power to

discriminate differences of 5% in missed readings. The study was not powered to examine differences in sex, skin tone, and ethnicity, but used a subject pool balanced in these factors according to FDA requirements.

Bias was calculated as SpO_2 minus SaO_2 from each oximeter's value and the corresponding arterial blood value. Bias is summarized as mean ± SD, where the SD is considered the precision. Root mean square error was calculated as the square root of the mean difference between $SpO_2 - SaO_2$, squared. The 95% confidence limits of the root mean square error were determined using bootstrapping (random resampling with replacement) with 50,000 repetitions. The 95% limits of agreement (LOA) were calculated as $1.96 \cdot SD$ according to Bland and Altman with adjustments for multiple measurements for each individual according to the "Method Where the True Value Varies."[19] The 95% confidence limits for the LOA were determined using bootstrapping as above.

Bias and the absolute value of the bias under different motions and different ranges of SaO_2 were compared using repeated-measures ANOVA with Tukey-Kramer honestly significant difference for multiple comparisons. Levene's test was used to compare variances between the different motions and SaO_2 ranges.

Receiver operating character (ROC) curves were constructed for each oximeter's SpO_2 determination of hypoxia (defined as $SaO_2 < 90\%$) and severe hypoxia (defined as $SaO_2 < 80\%$). The area under the curve (AUC) and 95% CIs were calculated. A Fisher exact test was used to compare the incidence of low perfusion (PI < 2) between male and female subjects. A P value less than 0.05 was considered statistically significant. Statistical analysis was performed with JMP 11.0 (SAS Institute, USA) and Stata 14 (Statacorp, USA).

Results

Demographics
We obtained data from 10 healthy adult subjects who participated in the study: six men and four women, all healthy, who ranged in age between 22 and 44 yr (median age, 27 yr). Each of the ten subjects had four motion readings from four pulse oximeters subjected to motion and one stationary pulse oximeter recorded at the time of acquiring 190 blood samples.

Comparison of Bias ($SpO_2 - SaO_2$) in Different Motion Modalities
Table 1 represents data comparing the performance of four pulse oximeters during different modalities of motion (tapping, rubbing, and random). For tapping motion, all pulse oximeters had significantly higher absolute mean bias than the non–motion control oximeter. For rubbing motion, all oximeters had significantly higher absolute biases except Nihon Kohden. In random motion, only Masimo displayed significantly higher absolute biases than the reference.

At any SaO_2 level, all pulse oximeters except Nellcor demonstrated higher root mean square errors than the corresponding nonmotion reference results during each of the three motion tests. Nellcor exhibited lower root mean square errors than the reference during random movement. When all motions were analyzed as a whole without concentrating on any particular motion type, all pulse oximeters had higher root mean square errors than the stationary control (root mean square errors of 1.8%; 95% CI, 1.55 to 2.01), with Nihon Kohden's root mean square errors of 2.2% being the lowest among the motion-tested machines.

When comparing performance between the three tested motion modalities, no significant differences in mean absolute bias or precision were found for Masimo and Nihon Kohden, whereas Nellcor and Philips demonstrated significant differences for both these performance parameters. Nihon Kohden missed one reading during rubbing, and Philips missed a number of readings, in all types of motion tests.

Comparison of Bias ($SpO_2 - SaO_2$) in Different Ranges of SaO_2
Table 2 compares pulse oximetry performance during any motion across "decadal" SaO_2 ranges of 70 to 80%, 80 to 90%, and 90 to 100%, but with the four SaO_2 plateaus from 68.5% reported with the 70 to 80% range. The introduction of motion caused all four oximeters to produce significantly higher mean absolute bias at the severely hypoxic range (68.5 to 80% SaO_2), whereas moderate hypoxia (80 to 90% SaO_2) combined with motion increased bias for all oximeters except Nihon Kohden. In the normoxic range (90 to 100% SaO_2), Masimo and Nellcor both reported significantly higher mean absolute bias when motion was present. The SD of the bias (precision) increased for all instruments, except Nihon Kohden at 90 to 100%. For every oximeter tested in motion, root mean square errors were higher than the nonmotion reference for all instruments at every SaO_2 range.

Successively lower SaO_2 ranges significantly degraded performance (increased mean absolute bias and root mean square errors and decreased precision) for the nonmotion reference and three of four motion-tested oximeters. Different oxygen saturation ranges did not significantly affect the mean absolute bias or precision for Nellcor, and its measured root mean error values were higher at higher SaO_2 ranges.

For each pulse oximeter tested, the biases from each motion test were plotted against SaO_2 using a modified version of the Bland-Altman method[20] (fig. 1). During motion, bias values deviated further from zero as SaO_2 declined. The three motion modalities of tapping, rubbing, and random impacted the distribution of bias values differently for each pulse oximeter, as reflected by the corresponding upper and lower LOA.

Comparison of Bias ($SpO_2 - SaO_2$) during Low- and Well-perfused States
Table 3 compares the performance of the tested oximeters in different states of blood perfusion. We used a cutoff of

Table 1. Comparison of Bias (SpO$_2$ – SaO$_2$) Information under Different Motion Conditions for All Pulse Oximeters

Oximeter	Tapping	Rubbing	Random	P Value
Masimo Radical-7				
n	61	60	59	—
Dropped	0	0	0	—
Mean bias (%)	1.5†	0.8	–0.6†	0.03
Mean |bias| (%)	2.5	3.0	4.0	0.16
Precision (%)	3.7	5.6	7.4	0.08
A$_{rms}$ (95% CI, %)	4.0 (2.8 to 5.1)	5.6 (3.2 to 8.0)	7.3 (4.7 to 10.0)	—
Lower LOA (95% CI, %)	–5.8 (–7.7 to –3.9)	–10.5 (–14.7 to –6.2)	–15.6 (–22.4 to –8.7)	—
Upper LOA (95% CI, %)	8.9 (6.3 to 11.5)	12.0 (6.5 to 17.5)	14.3 (10.2 to 18.3)	—
P value vs. reference	0.0004	0.003	0.001	—
Nihon Kohden OxyPal Neo				
n	61	59	59	—
Dropped	0	1	0	—
Mean bias (%)	0.4	0.3†	1.3†	0.03
Mean |bias| (%)	1.6	1.6	1.6	0.94
Precision (%)	2.0	2.7	1.7	0.65
A$_{rms}$ (95% CI, %)	2.1 (1.7 to 2.5)	2.7 (1.8 to 3.6)	2.1 (1.7 to 2.6)	—
Lower LOA (95% CI, %)	–3.6 (–4.9 to –2.3)	–5.0 (–7.2 to –2.7)	–2.2 (–2.9 to –1.4)	—
Upper LOA (95% CI, %)	4.5 (3.7 to 5.3)	5.6 (3.8 to 7.4)	4.7 (3.7 to 5.7)	—
P value vs. reference	0.03	0.08	0.90	—
Nellcor N-600				
n	61	60	59	—
Dropped	0	0	0	—
Mean bias (%)	–2.2*	0.2	1.0	< 0.0001
Mean |bias| (%)	3.2	2.9	1.4*	0.006
Precision (%)	5.5	4.5	1.7	< 0.0001
A$_{rms}$ (95% CI, %)	5.9 (3.8 to 8.1)	4.4 (3.2 to 5.7)	2.0 (1.5 to 2.4)	—
Lower LOA (95% CI, %)	–13.2 (–18.2 to –8.3)	–8.6 (–11.5 to –5.7)	–2.3 (–3.0 to –1.7)	—
Upper LOA (95% CI, %)	8.7 (6.1 to 11.4)	9.1 (6.3 to 11.9)	4.4 (3.4 to 5.3)	—
P value vs. reference	0.002	0.0001	0.21	—
Philips Intellivue MP5				
n	57	56	53	—
Dropped	4	4	6	—
Mean bias (%)	1.4	-0.5	1.1	0.12
Mean |bias| (%)	5.7*	2.1	1.4	< 0.0001
Precision (%)	9.5	4.2	2.0	< 0.0001
A$_{rms}$ (95% CI, %)	9.5 (6.9 to 12.1)	4.2 (1.5 to 7.0)	2.3 (1.2 to 3.3)	—
Lower LOA (95% CI, %)	–17.7 (–23.1 to –12.3)	–8.9 (–15.1 to –2.6)	–2.8 (–4.3 to –1.3)	—
Upper LOA (95% CI, %)	20.4 (14.3 to 26.6)	7.8 (3.4 to 12.2)	5.0 (2.7 to 7.3)	—
P value vs. reference	< 0.0001	0.049	0.72	—
Masimo Radical-7 reference				
n	61	60	59	—
Dropped	0	0	0	—
Mean bias (%)	0.6	0.4†	1.1†	0.02
Mean |bias| (%)	1.1	1.1	1.5	0.11
Precision (%)	1.5	1.5	1.9	0.06
A$_{rms}$ (95% CI, %)	1.6 (1.3 to 2.0)	1.5 (1.2 to 1.8)	2.2 (1.7 to 2.7)	—
Lower LOA (95% CI, %)	–2.5 (–3.2 to –1.8)	–2.5 (–3.1 to –1.8)	–2.7 (–3.2 to –2.1)	—
Upper LOA (95% CI, %)	3.6 (2.8 to 4.5)	3.3 (2.5 to 4.0)	5.0 (3.9 to 6.0)	—

Data are mean ± SD, or result (95% CI). Mean bias compared by repeated-measures ANOVA; precision compared by the Levene test; P value vs. reference compares absolute bias vs. the reference pulse oximeter by paired t test.

*Different from other values. †Different from each other.

A$_{rms}$ = root mean square error (with 95% CI by bootstrapping); bias = pulse oximeter measured oxygen saturation (SpO$_2$) – arterial blood oxygen saturation (SaO$_2$) measured with a Radiometer ABL 90; |Bias| = absolute value of the bias; LOA = 95% limits of agreement (with 95% CI by bootstrapping), mean bias ± 1.96 SD (adjusted for repeated measures); precision = SD of the bias.

Table 2. Comparison of Bias ($SpO_2 - SaO_2$) in Different Ranges of SaO_2

Oximeter	68.5–80	80–90	90–100	P Value
Masimo Radical-7				
n	60	59	71	—
Dropped	0	0	0	—
Mean bias (%)	2.2‡	0.6	−0.8‡	0.0005
Mean |bias| (%)	4.6‡	3.2	1.5‡	0.0001
Precision (%)	6.7	5.8	4.0	0.03
A_{rms} (95% CI, %)	7.0 (4.9 to 9.0)	5.8 (3.0 to 8.6)	4.0 (1.9 to 6.2)	—
Lower LOA (95% CI, %)	−11.2 (−15.2 to −7.3)	−11.0 (−18.5 to −3.5)	−8.7 (−13.6 to −3.7)	—
Upper LOA (95% CI, %)	15.5 (10.9 to 20.1)	12.3 (7.2 to 17.5)	7.1 (3.8 to 10.4)	—
P value *vs.* reference	< 0.0001	0.004	0.04	—
Nihon Kohden OxyPal Neo				
n	59	59	71	—
Dropped	1	0	0	—
Mean bias (%)	0.9	0.8	0.4	0.47
Mean |bias| (%)*	2.4	1.8	0.6	< 0.0001
Precision (%)	3.0	2.2	0.9	< 0.0001
A_{rms} (95% CI, %)	3.1 (2.4 to 3.9)	2.3 (1.9 to 2.8)	1.0 (0.8 to 1.2)	—
Lower LOA	−5.1 (−7.3 to −2.9)	−3.6 (−5.1 to −2.1)	−1.3 (−1.8 to −0.8)	—
Upper LOA	6.9 (5.3 to 8.5)	5.1 (4.2 to 6.1)	2.1 (1.7 to 2.6)	—
P value *vs.* reference	0.04	0.20	0.62	—
Nellcor N-600				
n	60	59	71	—
Dropped	0	0	0	—
Mean bias (%)	1.8†	−0.8	−1.7	< 0.0001
Mean |bias| (%)	2.6	2.6	2.1	0.62
Precision (%)	3.2	4.2	4.7	0.36
A_{rms} (95% CI, %)	3.6 (2.5 to 4.8)	4.2 (2.8 to 5.6)	4.9 (2.8 to 7.1)	—
Lower LOA (95% CI, %)	−4.4 (−6.5 to −2.4)	−9.0 (−12.5 to −5.5)	−10.9 (−15.7 to −6.1)	—
Upper LOA (95% CI, %)	8.0 (5.4 to 10.6)	7.4 (5.3 to 9.4)	7.4 (4.6 to 10.3)	—
P value *vs.* reference	0.02	0.02	0.006	—
Philips Intellivue MP5				
n	58	53	64	—
Dropped	2	6	7	—
Mean bias (%)	1.2	1.3	−0.3	0.29
Mean |bias| (%)	4.6	3.3	1.3†	0.006
Precision (%)	8.6	4.8	3.6	0.001
A_{rms} (95% CI, %)	8.6 (5.8 to 11.4)	4.9 (3.4 to 6.3)	3.5 (0.5 to 6.6)	—
Lower LOA (95% CI, %)	−15.9 (−21.5 to −10.2)	−8.1 (−11.2 to −5.0)	−7.3 (−14.0 to −0.6)	—
Upper LOA (95% CI, %)	18.3 (11.8 to 24.7)	10.6 (7.4 to 13.8)	6.7 (1.6 to 11.8)	—
P value *vs.* reference	0.002	0.0001	0.12	—
Masimo Radical-7 reference				
n	60	59	71	—
Dropped	0	0	0	—
Mean bias (%)	0.9	1.2	0.0†	< 0.0001
Mean |bias| (%)	1.7	1.6	0.6†	< 0.0001
Precision (%)	2.1	1.6	0.9	< 0.0001
A_{rms} (95% CI, %)	2.3 (1.8 to 2.7)	2.0 (1.6 to 2.4)	0.9 (0.7 to 1.1)	—
Lower LOA (95% CI, %)	−3.3 (−4.1 to −2.5)	−2.0 (−2.7 to −1.4)	−1.8 (−2.2 to −1.3)	—
Upper LOA (95% CI, %)	5.1 (4.1 to 6.1)	4.4 (3.6 to 5.3)	1.8 (1.4 to 2.2)	—

Mean bias compared by repeated-measures ANOVA, with multiple comparisons by Tukey-Kramer honestly significant difference; precision compared by the Levene test; P value *vs.* reference compares absolute bias *vs.* the reference pulse oximeter by paired *t* test.

*All different from each other. †Different from other values. ‡Different from each other.

A_{rms} = root mean square error (with 95% CI by bootstrapping); bias = pulse oximeter measured oxygen saturation (SpO_2) – arterial blood oxygen saturation (SaO_2) measured with a Radiometer ABL 90; |Bias| = absolute value of the bias; LOA = 95% limits of agreement (with 95% CI by bootstrapping), mean bias ± 1.96 SD (adjusted for repeated measures); precision = SD of the bias.
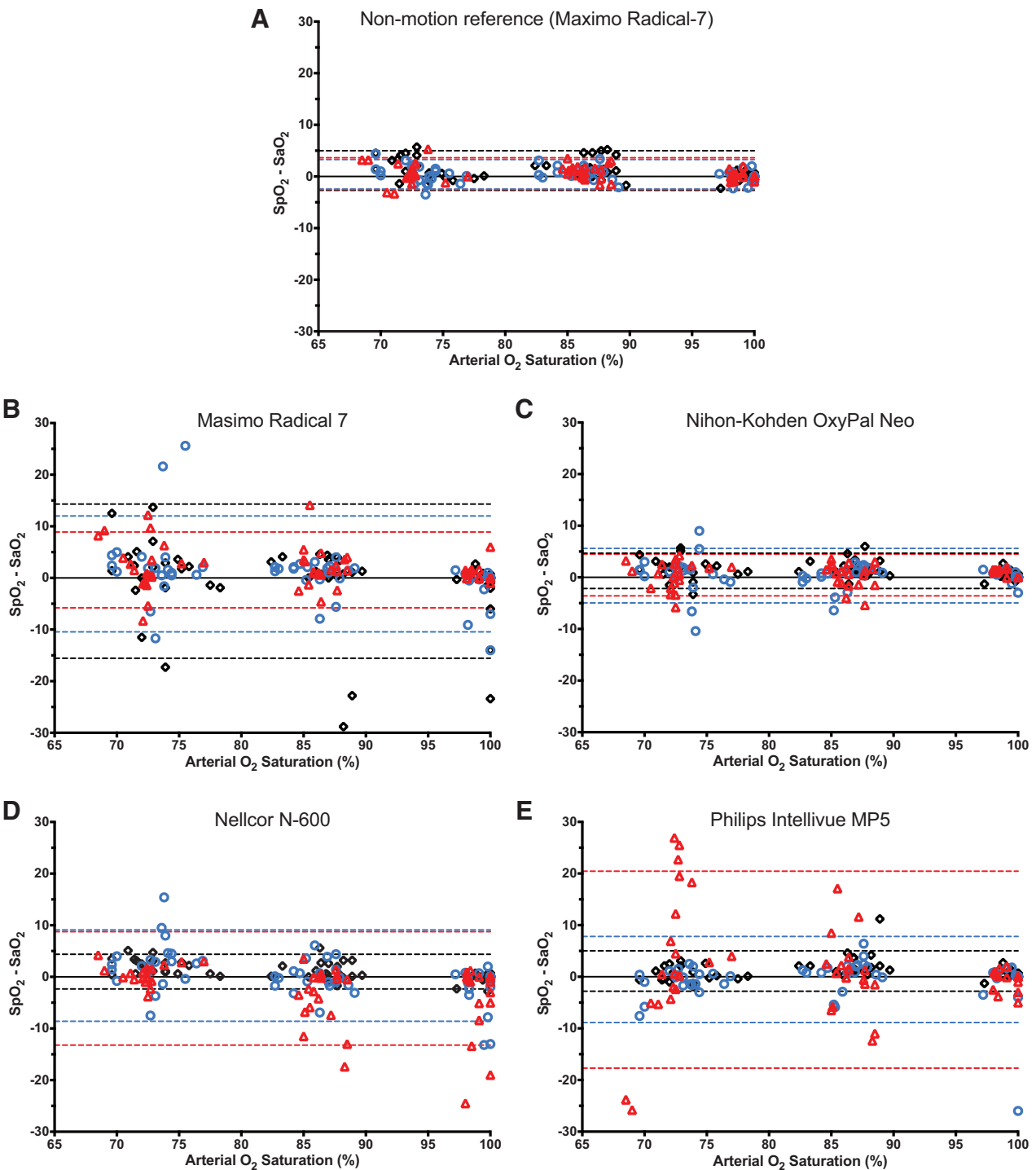
Fig. 1. Bias plot of (Spo₂ – Sao₂) across different ranges of oxyhemoglobin saturation (Sao₂). Bias is calculated as Spo₂ (oxyhemoglobin saturation as measured by the indicated pulse oximeter) minus Sao₂ (oxyhemoglobin saturation as measured by the Radiometer ABL-90 [Denmark] multiwavelength oximetry of arterial blood). Pooled bias values for 10 subjects are plotted for tapping (*red triangles*), rubbing (*blue circles*), and random (*black diamonds*) motions. *Dashed horizontal lines* are the upper and lower limits of agreement with colors corresponding to the symbols of each motion type. (*A*) Nonmotion reference using Maximo Radical-7 (USA), (*B*) Masimo Radical-7, (*C*) Nihon Kohden OxyPal Neo (Japan), (*D*) Nellcor N-600 (USA), and (*E*) Philips Intellivue MP5 (USA).

PI less than 2.0 to represent poor perfusion.[21] Generally, imprecisions are significantly higher with poor perfusion for all the motion-tested pulse oximeters and the nonmotion reference. However, Nihon Kohden during motion did not demonstrate significant differences in precision or mean absolute bias between perfusion index values less than and above 2. Along with the non–motion control, both Masimo and Philips showed significantly different mean absolute bias between the two perfusion categories during motion. Along with one of the male subjects, all four female subjects had significantly reduced poor perfusion (P = 0.048; fig. 2). LOAs were further from zero bias when perfusion index was less than 2 (fig. 3).

Sensitivity and Specificity

The AUC was at least 95% in detecting hypoxemia (Sa_{O_2} < 90%) or severe hypoxemia (Sa_{O_2} < 80%) for every pulse oximeter tested during any motion with all motion types combined (fig. 4). The AUC for the nonmotion reference was 100%. During just the tapping motion only, the AUC ranged from 86.2% (Philips at severe hypoxemia) to 100% (data not shown). During rubbing motion, the AUC ranged from 92.7% (Masimo at severe hypoxemia) to 100%. And during random motion, the AUC ranged from 93.7% (Masimo at severe hypoxemia) to 100%.

During hypoxia, AUC ranged from 93.7% (Nellcor) to 100% at low perfusion (PI < 2), and 98.4% (Philips) to 100% at high perfusion (PI ≥ 2; data not shown). During severe hypoxia, AUC ranged from 90.9% (Masimo) to 100 for poor perfusion, and 97.8% (Nellcor) to 100% with good perfusion.

Discussion

This is the first independent examination of the comparative performance of motion-tolerant and low perfusion–resistant pulse oximeters. Previous studies concerning this type of assessment were industry supported, restricted to a single device, or did not involve a comparison to measured saturation in arterial blood samples. Our findings indicate that four motion-tolerant low perfusion–resistant pulse oximeters behaved similarly during controlled motion and hypoxia. All devices had at least a 95% sensitivity and specificity in detecting hypoxemia (Sa_{O_2} ~ 88%) and severe hypoxemia (Sa_{O_2} ~ 78%) during motion. Although there were some differences in the precision of the four instruments, all instruments detected hypoxia during motion and across a range of perfusion conditions in healthy subjects under controlled laboratory conditions (ROC curves, fig. 4). Analyzing pulse oximeter performance using ROC curves reflects a clinically relevant approach. An important purpose of pulse oximetry is to determine whether a patient is normoxic or hypoxic. Knowing exactly accurate oxygen saturation in a moving or critically ill patient may not be the most important clinical question in many circumstances. The analysis demonstrates excellent sensitivity and specificity of all the devices during

motion. While there were some high Sp_{O_2} readings during hypoxemia (false negatives), which decreased sensitivity, it is also possible that the oximeters would have produced low Sp_{O_2} readings if given longer processing time.

Relevance of Laboratory Motion and Low Perfusion Challenges to Clinical Conditions

This study examined controlled motion in a laboratory setting. In terms of reproducibility and between-instrument comparisons, this study design has advantages compared to studies in spontaneously moving patients. However, the clinical applicability of the findings is less clear than in studies involving spontaneous motion in patients. Our study did involve significant disturbances in the photoplethysmogram, as it produced decreased precision in all oximeters tested. The magnitude of the movement disturbances was adjusted in each subject to produce motion artifacts in the photoplethysmogram signal that were similar in amplitude to the arterial pulse waveforms. Therefore, each oximeter was challenged to reject pulsatile signals that did not correspond to actual arterial pulsations. How this is achieved by the different devices is not known to the user, and are both trade secrets and patent protected. It was interesting that, despite probable differences in the processing software and hardware in the different instruments, all four types of instruments performed similarly well in detecting the changes in saturation between room air and 88% saturation and a change between 88% and 77% saturation. We see no reason why this would be different in clinical situations involving desaturations in patients during movement disturbances.

Effect of Different Motion Types on Performance

We analyzed three different motion types for each of the four devices and found that motion increased mean bias and root mean square errors for all pulse oximeters. Some small differences in the performance of the oximeters during different types of motion were observed, but an obvious pattern was not apparent. This was surprising because we assumed that the instruments use different types of algorithms to determine saturation. The root mean square error was more than 3% for many of the motions in all devices with the exception of Nihon Kohden. Although Philips's root mean square error was under 3.0% for the pseudorandom motion, there were many incidents of failure to display a saturation value ("Dropped," table 1). Our motion protocol involved two types of repetitive hand motion that were employed in previous studies of pulse oximeter performance under motion and low perfusion conditions.[9,14]

Perfusion Index

We did not specifically design the study to examine the effects of low perfusion on pulse oximeter performance. However, variations in perfusion index in our subjects enabled us to examine the role of this variable in the precision of the instruments. Several previous studies have

Table 3. Comparison of Bias ($SpO_2 - SaO_2$) Information at PI Less than 2 and for 2 and Above

Oximeter	PI < 2	PI ≥ 2	P Value
Masimo Radical-7			
n	95	95	—
Dropped	0	0	—
Mean bias (%)	0.0	1.2	0.54
Mean |bias| (%)	4.4	1.7	< 0.0001
Precision (%)	7.4	2.9	< 0.0001
A_{rms} (95% CI, %)	7.4 (5.4 to 9.3)	3.1 (2.0 to 4.2)	—
Lower LOA (95% CI, %)	−15.2 (−20.0 to −10.4)	−4.5 (−6.8 to −2.2)	—
Upper LOA (95% CI, %)	15.2 (11.4 to 19.1)	7.0 (4.6 to 9.4)	—
P value *vs.* reference	< 0.0001	0.002	—
Nihon Kohden OxyPal Neo			
n	94	95	—
Dropped	1	0	—
Mean bias (%)	0.9	0.5	0.47
Mean |bias| (%)	1.6	1.5	0.35
Precision (%)	2.1	2.2	0.55
A_{rms} (95% CI, %)	2.3 (1.8 to 2.7)	2.2 (1.7 to 2.8)	—
Lower LOA (95% CI, %)	−3.3 (−4.3 to −2.3)	−3.9 (−5.6 to −2.2)	—
Upper LOA (95% CI, %)	5.1 (4.0 to 6.2)	4.9 (3.7 to 6.0)	—
P value *vs.* reference	0.90	0.0004	—
Nellcor N-600			
n	95	95	—
Dropped	0	0	—
Mean bias (%)	−0.7	0.1	0.62
Mean |bias| (%)	2.7	2.1	0.99
Precision (%)	5.1	3.4	< 0.0001
A_{rms} (95% CI, %)	5.1 (3.5 to 6.8)	3.4 (2.5 to 4.3)	—
Lower LOA (95% CI, %)	−10.7 (−14.7 to −6.7)	−6.6 (−8.4 to −4.8)	—
Upper LOA (95% CI, %)	9.2 (6.9 to 11.6)	6.8 (4.7 to 8.8)	—
P value *vs.* reference	0.02	< 0.0001	—
Philips Intellivue MP5			
n	80	95	—
Dropped	15	0	—
Mean bias (%)	0.8	0.6	0.048
Mean |bias| (%)	4.1	2.1	0.006
Precision (%)	7.8	4.0	< 0.0001
A_{rms} (95% CI, %)	7.8 (5.4 to 10.2)	4.0 (2.4 to 5.6)	—
Lower LOA (95% CI, %)	−14.9 (−20.4 to −9.4)	−7.3 (−9.8 to −4.8)	—
Upper LOA (95% CI, %)	16.4 (11.2 to 21.7)	8.5 (4.8 to 12.2)	—
P value *vs.* reference	0.0006	0.0009	—
Masimo Radical-7 reference			
n	95	95	—
Dropped	0	0	—
Mean bias (%)	1.4	−0.1	< 0.0001
Mean |bias| (%)	1.6	0.8	0.0005
Precision (%)	1.7	1.1	< 0.0001
A_{rms} (95% CI, %)	2.2 (1.9 to 2.6)	1.1 (1.0 to 1.3)	—
Lower LOA (95% CI, %)	−2.0 (−2.4 to −1.5)	−2.4 (−2.9 to −1.9)	—
Upper LOA (95% CI, %)	4.8 (4.2 to 5.5)	2.2 (1.8 to 2.6)	—

Data are mean ± SD, or result (95% CI). Mean bias compared by Wilcoxon rank sum; precision compared by the Levene test; P value *versus* reference compares absolute bias *versus* the reference pulse oximeter by paired *t* test.

A_{rms} = root mean square error (with 95% CI by bootstrapping); bias = pulse oximeter measured oxygen saturation (SpO_2) − arterial blood oxygen saturation (SaO_2) measured with a Radiometer ABL 90; |Bias| = absolute value of the bias; LOA = 95% limits of agreement (with 95% CI by bootstrapping), mean bias ± 1.96 SD (adjusted for repeated measures); PI = perfusion index; precision = SD of the bias.
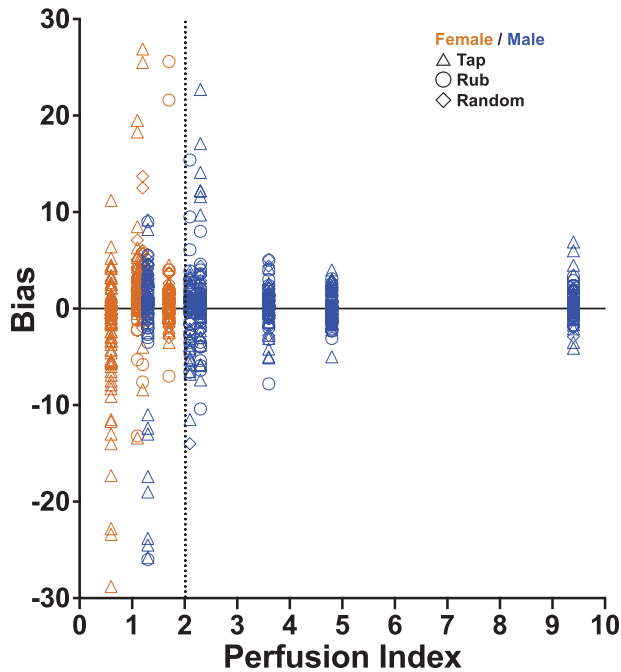
Fig. 2. Bias plot of absolute bias by perfusion index (PI) ranges across different sexes: *orange*, female, and *blue*, male, for each motion type: tapping (*triangle*), rubbing (*circle*), and random (*diamond*). *Dashed line* indicates PI = 2.

reported that low perfusion degrades pulse oximeter performance and results in nondisplayed saturation values. As far as we are aware, the only previous study quantitatively examining the relationship of decreased perfusion index to increased bias was the study by Hummler *et al.*[21] In the Hummler *et al.* study, bias increased when PI was less than 2. We defined a perfusion index of less than 2 as low

perfusion; this cutoff is supported by the large decrease in precision in readings at these lower PI values (fig. 3). As has been our experience with many years of pulse oximetry testing, our female subjects were more likely to have a low perfusion index (less than 2%). Low perfusion was associated with increasing mean absolute bias and decreased precision (table 3). This is an important association because it implies that sex affects pulse oximeter performance through the mechanism of low perfusion. Although it is possible to heat the hands with heating pads or warm air, it may not always be practical to do so in patients.

Limitations

Our study had several limitations. The first was that, of necessity in a controlled laboratory setting, controlled motion was used. The motions may not reflect they types of motion that present challenges to pulse oximeter performance in patients, and therefore the extrapolation of our findings to clinical conditions is limited. However, our motion protocols created a substantial disturbance to the photoplethysmograph waveform, one that was of the same magnitude as the pulse signals. In addition, our motion controller software permitted inclusion of a "pseudorandom motion" protocol, wherein motion is randomly jerky, of varying amplitude and direction within specified extremes. This may be similar to some types of patient motion. Based on our knowledge of how motion-resistant algorithms work, our motion disturbances were expected to be a significant challenge to the instruments. A study in actively moving patients involving a blinded comparison of different types of motion tolerant/resistant pulse oximeters is needed.
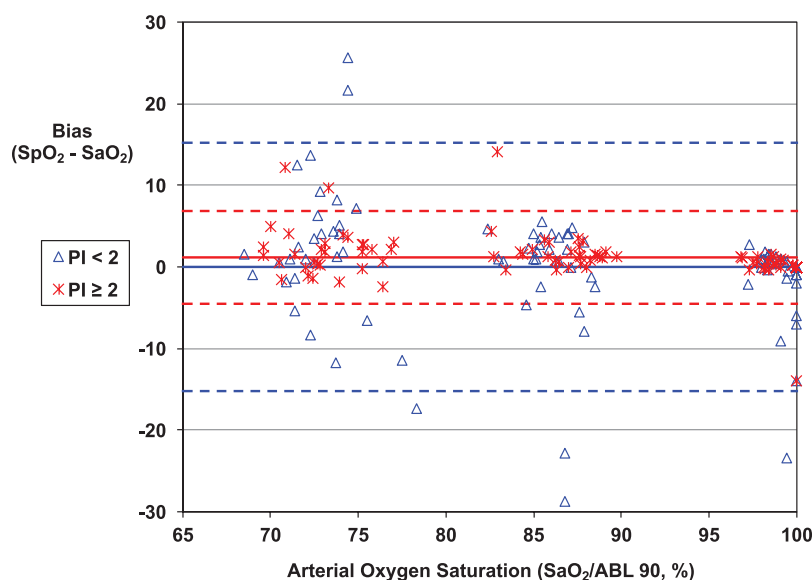

Fig. 3. Bias plot of (Spo$_2$ – Sao$_2$) across different ranges of oxyhemoglobin saturation (Sao$_2$). Bias is calculated as Spo$_2$ (oxyhemoglobin saturation as measured by the indicated pulse oximeter) minus Sao$_2$ (oxyhemoglobin saturation as measured by the Radiometer ABL-90 [Denmark] multiwavelength oximetry of arterial blood). Bias values are separated by perfusion index (PI). PI < 2 (*blue triangle*) and PI ≥ 2 (*red star*). *Dashed horizontal lines* are the upper and lower limits of agreement.
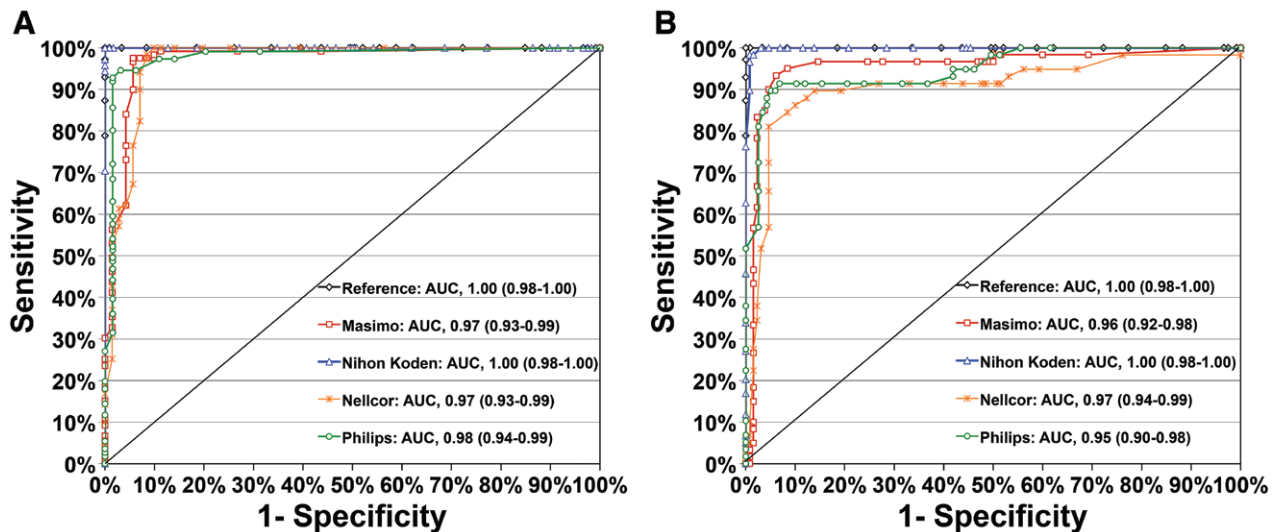
Fig. 4. Receiver operating characteristic (ROC) curves showing the sensitivity and specificity for the oxygen saturation measured by pulse oximetry (Spo$_2$) reading to detect mild hypoxemia, defined as arterial blood oxygen saturation (Sao$_2$) less than 90% (*A*) and severe hypoxemia, defined as Sao$_2$ less than 80% (*B*) during motion (all motion types combined). The area under the curve (AUC) for each pulse oximeter, with 95% confidence limits, is shown on the plots. Reference: nonmotion Masimo Radical-7 (USA; *black line*), Masimo Radical-7 (*red line*), Nihon Kohden OxyPal Neo (Japan; *blue line*), Nellcor N-600 (USA; *orange line*), and Philips Intellivue MP5 (USA; *green line*).

Another limitation was that the study did not encompass a full range of perfusion values in all subject types; the subjects with low perfusion were predominately female. Therefore, the generalizability of our conclusion that low perfusion performance was similar among all devices is limited. Our protocol also maintained plateaus for defined periods of time. It is possible that some of the pulse oximeters might have determined more accurate and precise readings if provided a longer period of processing time. We did not include a nonmotion resistant pulse oximeter on the motion hand as control. This was not practical because we did not want an instrument on the thumb. As mentioned, even the motion-resistant oximeters examined showed decreased performance in our study, so we do not believe that this was a significant limitation. Our study was not powered to specifically examine differences in skin tone, ethnicity, and sex, as these are not repeated-measures tests and would require a significantly larger sample size. The subject diversity was designed only to follow FDA guidelines regarding subject competition. The study was also not specifically designed to compare one oximeter to the other, and differences could have been influenced by other factors.

Another study limitation is that the study may not have been powered to identify differences between oximeter performance during motion, even though the study was designed with a sample size sufficient to detect statistical, but perhaps not clinically relevant, differences in oximeter performance in detecting changes in oxygenation.[18] It seems unlikely to us that the small differences in performances observed in the study have any clinical significance.

Conclusions

Motion and low perfusion degraded the performance of four types of pulse oximeters that are marketed as motion-resistant devices, but all four types tested detected hypoxia with greater than 95% specificity. Low perfusion was associated with less precision. We conclude that different types of algorithms to read through motion and low perfusion are similarly capable of detecting significant changes in oxygenation under controlled laboratory conditions.

Research Support

Competing Interests

Drs. Feiner, Bickler, and Lucero have previously funded independent university-based research with funds collected from companies who contracted with our lab to perform evaluation of instrument performance. In the last five years the following companies contracted with the Hypoxia lab: Masimo (Irvine, California), Nonin (Plymouth, Minnesota), Bluepoint Medical (Centreville, Virginia), Xhale (Wales, England) Nihon Kohden (Tokyo, Japan), Motara (Milwaukee, Wisconsin), Medico (Tokyo, Japan), Philips (Amsterdam, The Netherlands), Renu (Everett, Washington), Scanadu (Sunnyvale, California), Sensepc (Rostock, Germany), Sentec (Therwil, Switzerland), Sharp (Osaka, Japan) Sleep Med (Peabody, Massachusetts), Solaris (San Francisco, California), Springer (Berlin, Germany), True Wearables (Rancho Santa Margarita, California), Unimed (Shenzhen, China), Verifood (Vancouver, British Colum-

Correspondence

Address correspondence to Dr. Lucero: University of California, San Francisco, 533 Parnassus Avenue, San Francisco, California 94117. jennifer.lucero@ucsf.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Jubran A: Pulse oximetry. Crit Care 2015; 19:272

2. Nitzan M, Romem A, Koppel R: Pulse oximetry: Fundamentals and technology update. Med Devices (Auckl) 2014; 7:231–9

3. US Department of Health and Human Services FaDA: Pulse Oximeters-Premarket Notification Submissions [510(k)s] Guidance for Industry and Food and Drug Administration Staff. Bethesda, Maryland, March 4, 2013

4. Tobin RM, Pologe JA, Batchelder PB: A characterization of motion affecting pulse oximetry in 350 patients. Anesth Analg 2002; 94(1 Suppl):S54–61

5. Cannesson M, Talke P: Recent advances in pulse oximetry. F1000 Med Rep 2009; 1:pii: 66. doi: 10.3410/M1-66

6. Petterson MT, Begnoche VL, Graybeal JM: The effect of motion on pulse oximetry and its clinical significance. Anesth Analg 2007; 105(6 Suppl):S78–84

7. Rheineck-Leyssius AT, Kalkman CJ: Influence of pulse oximeter lower alarm limit on the incidence of hypoxaemia in the recovery room. Br J Anaesth 1997; 79:460–4

8. Giuliano KK, Higgins TL: New-generation pulse oximetry in the care of critically ill patients. Am J Crit Care 2005; 14: 26–37; quiz 38–9

9. Shah N, Ragaswamy HB, Govindugari K, Estanol L: Performance of three new-generation pulse oximeters during motion and low perfusion in volunteers. J Clin Anesth 2012; 24:385–91

10. Barker SJ, Shah NK: The effects of motion on the performance of pulse oximeters in volunteers (revised publication). ANESTHESIOLOGY 1997; 86:101–8

11. Barker SJ: "Motion-resistant" pulse oximetry: A comparison of new and old models. Anesth Analg 2002; 95:967–72

12. Townshend J, Taylor BJ, Galland B, Williams S: Comparison of new generation motion-resistant pulse oximeters. J Paediatr Child Health 2006; 42:359–65

13. Jopling MW, Mannheimer PD, Bebout DE: Issues in the laboratory evaluation of pulse oximeter performance. Anesth Analg 2002; 94(1 Suppl):S62–8

14. Gehring H, Hornberger C, Matz H, Konecny E, Schmucker P: The effects of motion artifact and low perfusion on the performance of a new generation of pulse oximeters in volunteers undergoing hypoxemia. Respir Care 2002; 47:48–60

15. Batchelder PB, Raley DM: Maximizing the laboratory setting for testing devices and understanding statistical output in pulse oximetry. Anesth Analg 2007; 105(6 Suppl):S85–94

16. Bickler PE, Feiner JR, Severinghaus JW: Effects of skin pigmentation on pulse oximeter accuracy at low saturation. ANESTHESIOLOGY 2005; 102:715–9

17. Severinghaus JW: Simple, accurate equations for human blood O2 dissociation computations. J Appl Physiol Respir Environ Exerc Physiol 1979; 46:599–602

18. Bickler PE, Feiner JR, Lipnick MS, Batchelder P, MacLeod DB, Severinghaus JW: Effects of acute, profound hypoxia on healthy humans: Implications for safety of tests evaluating pulse oximetry or tissue oximetry performance. Anesth Analg 2017; 124:146–53

19. Bland JM, Altman DG: Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007; 17:571–82

20. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307–10

21. Hummler HD, Engelmann A, Pohlandt F, Högel J, Franz AR: Decreased accuracy of pulse oximetry measurements during low perfusion caused by sepsis: Is the perfusion index of any value? Intensive Care Med 2006; 32:1428–31