

# CHARACTERIZATION OF THE BIAS BETWEEN OXYGEN SATURATION MEASURED BY PULSE OXIMETRY AND CALCULATED BY AN ARTERIAL BLOOD GAS ANALYZER IN CRITICALLY ILL NEONATES

Thomas E. Bachman<sup>1</sup>, Christopher J. L. Newth<sup>2</sup>, Patrick A. Ross<sup>2</sup>, Narayan P. Iyer<sup>2</sup>, Robinder G. Khemani<sup>2</sup>

<sup>1</sup>Czech Technical University in Prague, Kladno, Czech Republic

<sup>2</sup>Children's Hospital Los Angeles, Los Angeles, USA

## Abstract

Continuous monitoring of oxygenation with pulse oximetry is the standard of care for critically ill neonates. A better understanding of its measurement bias compared to arterial oxygen saturation could be helpful both for the clinician and researcher. Towards that end, we examined the electronic database from a large neonatal ICU. From a 24-month period we identified 25,032 paired SpO<sub>2</sub>-SaO<sub>2</sub> measurements from 1,007 infants who were receiving supplemental oxygen during mechanical ventilation. We found that SpO<sub>2</sub> was consistently higher than SaO<sub>2</sub>. The size of the bias was fairly constant when SpO<sub>2</sub> was between 75–93%, above which it dropped steadily. The median size of this bias was 1% SpO<sub>2</sub> during hyperoxemia (SpO<sub>2</sub> 97–100%) with a median variation of 1.3% above and below. During periods of hypoxemia (SpO<sub>2</sub> 75–85%) and normoxemia (SpO<sub>2</sub> 89–93%) the bias was approximately 5% SpO<sub>2</sub>, with a median variation of 5% above and below.

## Keywords

Pulse oximetry, neonatal oxygenation, oxygen saturation

## Background

Use of the noninvasive pulse oximeter (SpO<sub>2</sub>) in medicine is ubiquitous. Continuous monitoring of SpO<sub>2</sub> to manage oxygenation is the standard of care in neonatal critical care. [1, 2] These continuous readings are complemented with periodic arterial blood gas analysis, as well as transcutaneous O<sub>2</sub>-CO<sub>2</sub> measurements, and more recently near infrared tissue oxygen assessment. Nevertheless continuous SpO<sub>2</sub> is the primary measure used to manage oxygenation in the neonatal intensive care unit (NICU).

The reliability of pulse oximetry has improved since it was introduced into the NICU 30 years ago. The accuracy of newer generation devices introduced more than a decade ago has been studied showing reasonable agreement with arterial blood near desirable levels, but marked bias outside. [3, 4] A better understanding of

the limitations could perhaps enhance clinical decision-making and also help those who are modeling and developing oxygenation control systems.

We accessed a large clinical database with the goal of providing more detailed current information on the bias between calculated arterial saturation (SaO<sub>2</sub>) and SpO<sub>2</sub> as measured with leading oximetry technology.

## Methods

The data is from the 58-bed NICU at Children's Hospital Los Angeles. [5] It is a University affiliated tertiary care center in the United States. The Investigational Review Board has waived the need for informed consent for aggregate data analysis studies and specifically approved this project.

The oximeter in the patient monitoring system uses Masimo SET technology (Masimo Corporation, Irvine, California). Infants were included in this analysis if they were receiving support from a mechanical ventilator and supplemental oxygen at the time of the measurement. All complete arterial blood gas analyses that included pH, PaCO<sub>2</sub>, PaO<sub>2</sub>, SaO<sub>2</sub> and FiO<sub>2</sub> were identified. SpO<sub>2</sub> values less than 75% were excluded. The SaO<sub>2</sub> was calculated using standard algorithms incorporated into the blood gas system (epoc®, Alere, Watham Massachusetts). These records were linked to another file with 30-second SpO<sub>2</sub> readings (10-second averaging) from the patient monitor. The SpO<sub>2</sub> value used in the analysis was the mean of 4 SpO<sub>2</sub> readings in the minute before and after the arterial blood collection.

The intent to group saturation risk categories was prospective. However the three SpO<sub>2</sub> risk groups were selected after reviewing the initial bias data, to insure within group results had consistent bias, if possible. The categories selected were high risk of hypoxemia (75–85% SpO<sub>2</sub>), likely normoxemia (89–93% SpO<sub>2</sub>) and high risk of hyperoxemia (97–100% SpO<sub>2</sub>).

The purpose of this analysis was not to assess the accuracy of SpO<sub>2</sub>, but rather to identify the expected difference between a SpO<sub>2</sub> level and a simultaneous calculated arterial SaO<sub>2</sub>. For that reason we chose not to apply the Bland-Altman convention of using the average of SpO<sub>2</sub> and SaO<sub>2</sub> as the independent variable or a scatter plot with 95% confidence limits. Rather we selected metrics to describe the actual size and variability of the difference between these two measures. Bias size was defined by the mean and median of the difference between SpO<sub>2</sub> and SaO<sub>2</sub>. Bias variability was defined as the median of the absolute value of the differences between the bias size for each respective SpO<sub>2</sub> risk category.

We determined that with a sample size of 100 per category we would be able to detect the difference in bias of  $1 \pm 2$  STD with a power  $>0.8$  and a threshold for significance of  $p < 0.05$ . We chose to use available data from a 24-month period previously collected to insure an adequate sample per category.

Differences between medians were evaluated using Kruskal-Wallis or Mann-Whitney U tests. Statistical tests were conducted with XLSTAT v19.02 (Addinsoft, Paris, France). A two-tailed  $p < 0.05$  was considered statistically significant.

## Results

We evaluated 25,032 SpO<sub>2</sub>-SaO<sub>2</sub> measurements from 1,007 infants. Their median age was 13 days (IQR: 3–34). The median number of SpO<sub>2</sub>-SaO<sub>2</sub> pairs/SpO<sub>2</sub>-bin was 462 (IQR: 384–1006).

Details of the gas exchange of these infants are shown in Table 1. The median values of these para-

meters are all consistent with nominally accepted target values for critically ill neonates. Nevertheless more than 50% of the SaO<sub>2</sub> values were either hypoxemic or hyperoxemic. In addition, other values outside the bounds of the IQR reflect a wide range well beyond normal. The median difference between the SaO<sub>2</sub> and SpO<sub>2</sub> was 2 (IQR 0–6,  $P < 0.001$ ).

Figure 1 shows the bias across the range of individual SpO<sub>2</sub> bins. It includes the mean with its 95% confidence limits, as well as the median. The median bias is consistently lower than the mean bias and generally outside the confidence limits of the mean, confirming marked skew. Whether considering the median or mean, the magnitude of the bias is clearly not constant across the range. Between 75–93%, SpO<sub>2</sub> reads markedly higher than SaO<sub>2</sub>. In contrast, between 94–100% the difference between SpO<sub>2</sub> and SaO<sub>2</sub> is not only much smaller but also tends to drop with increasing SpO<sub>2</sub>.

*Table 1: Gas Exchange Parameters.*

	Median	IQR
SaO <sub>2</sub> (%)	94	83–98
SpO <sub>2</sub> (%)	97	90–100
SpO <sub>2</sub> -SaO <sub>2</sub>	2	0–6
PaCO <sub>2</sub> mmHg	45	40–52
pH	7.380	7.330–7.430
PaO <sub>2</sub> mmHg	74	50–110
FiO <sub>2</sub> (%)	45	35–67
PaO <sub>2</sub> /FiO <sub>2</sub>	168	100–260

*Table 2: Size & Variability of Bias by Oxemia Category.*

Oxemia category	n	Size Median (IQR)	Variability Median (IQR)
Hypoxemic (SpO <sub>2</sub> 75–85%)	4,533	5 (1–11)	5.1 (2.1–8.9)*
Normoxemic (SpO <sub>2</sub> 89–93%)	3,034	6 (1–11)	4.8 (2.2–8.2)*
Hyperoxemic (SpO <sub>2</sub> 97–100%)	12,585	1 (0–3)	1.3 (0.8–2.8)*

\*  $P < 0.001$  between all categories

The size and variability of the bias for the three categories of SpO<sub>2</sub> risk are shown in Table 2. The differences in the median size are predictable, 5, 6 and 1, consistent with the figure. The later (hyperoxemia) was statistically significantly smaller ( $P < 0.001$ ). The median variability in the three SpO<sub>2</sub> risk ranges is relatively large. As a percent of the bias they are hypoxemia (102%), normoxemia (80%) and hyperoxemia (130%). The differences in variability are

statistically different ( $P < 0.001$ ). The lower bound of the IQRs is near zero, and 16% of the biases are negative.

## Discussion

We believe this is the first large study of  $SpO_2$ - $SaO_2$  bias in critically ill neonates using the current generation of  $SpO_2$  technology. We found that  $SpO_2$  was consistently higher than the calculated  $SaO_2$ . In  $SpO_2$  values associated with neonatal hypoxemia and normoxemia, we found that the average size of bias from  $SaO_2$  was marked but consistent. In contrast at levels of  $SpO_2$  associated with hyperoxemia, the average level of bias was much smaller.

In 2014 Ross et al compared  $SpO_2$  and  $SaO_2$  in 212 infants and children in 5 pediatric ICUs. [3] They analyzed 1,980  $SpO_2$ - $SaO_2$  measurements. The  $SpO_2$  measurements used different pulse oximeters, and  $SaO_2$  was measured directly with co-oximetry. Consistent with our finding they reported smaller bias at higher levels of  $SpO_2$ . They further reported that the bias was larger between 81–85%. This is not particularly evident in our data. The size of these biases was also slightly smaller than we reported. These small differences

could be a result of the lack of accuracy of the calculated  $SaO_2$  as compared to its direct measurement with co-oximetry. Ross also reported significant bias effects associated with infants with cardiovascular anomalies and the miss-use of the oximeter sensor. They did not identify any significant differences between oximeters or infants and children. These former two might also have contributed to the differences in results.

Rosychuk et al reported on a comparison of 1,032 paired  $SpO_2$ - $SaO_2$  measurements with  $SpO_2$  levels between 85–95%. [4] These were very preterm neonates in the first week of life, who had not received a transfusion, and thus have a high level of fetal hemoglobin. The  $SpO_2$  measurements were made with Massimo SET technology. They reported, in contrast to our findings and those of Ross, that the bias was similar above and below a  $SpO_2$  of 90%. However, there were only 101 values between 85–89%, limiting this finding. They also reported a much smaller bias, for  $SpO_2$  levels 91–95%. Some of this difference among studies could be related to the higher fetal hemoglobin [6, 7] in these infants, than in other studies. Like Ross they also measured  $SaO_2$  with co-oximetry, limiting the possibility of this difference being relating to errors between measured and calculated  $SaO_2$ .

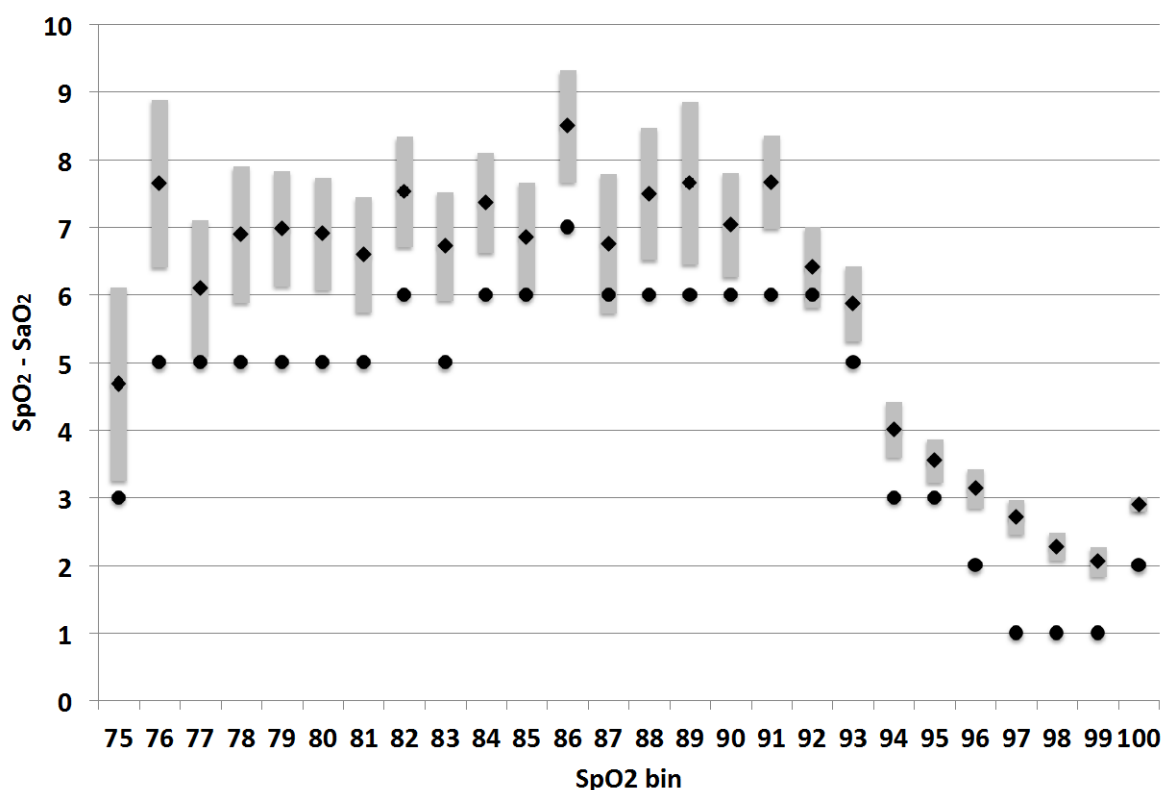


Figure 1: Bias Size median, mean and 95% CL SEM. Median is depicted as a circle and mean as a diamond. The grey bar is the 95% confidence limits of the standard error of the mean.

As noted above, consistent with the finding of Ross, our data reflect markedly difference levels of bias between low and high SpO<sub>2</sub> readings. This suggests may be kinked at 94% SpO<sub>2</sub>. Johnston et al, described a discontinuity in the internal calibration curve of the Massimo SET oximeter technology. [8] They observed that this resulted in a dip in the histogram of SpO<sub>2</sub> between 87–90%. They further confirmed that updated software made available in 2011 to address this issue resulted in little apparent bias with other oximeters. Our data reflects a bias that is very consistent between 87–90% SpO<sub>2</sub>, supporting that the error reported by Johnson was indeed not present. However it suggests the possibility that a similar distortion could now be present between 93–96% SpO<sub>2</sub>, if the transition between the two sections of the calibration curves was not smooth. Our bias levels at SpO<sub>2</sub> values of 75%, 86% and 100% also suggest potential discontinuities. We can not rule out that these effects are related to the use of calculated SaO<sub>2</sub>. However the calculation equation is a continuous function, not subject to discontinuities but still possibly skewed.

We found the variability of the bias was quite large. This finding is consistent with the reports of Ross and Rosychuk. However, the metrics of variability differ among the studies so exact comparisons are difficult. Rosychuk reported that the span of the 95% limits of uncertainty was higher for SpO<sub>2</sub> values between 85–89% than those between 91–95% (18.2%, and 10.4% SpO<sub>2</sub> respectively). This difference is most possibility a result of the limited number of samples in the lower range. Ross reported a IQR of 5–10% in different SpO<sub>2</sub> strata. These were slightly smaller than our comparable findings. Ross reported that the variability was smaller at high levels of SpO<sub>2</sub>, consistent with our finding and those of Rosychuk.

Gerstmann et al, conducted a study similar to ours based on a database from the 1991–1997. [9] Their evaluation of this earlier oximeter technology found “optimal” agreement in a narrow clinical band, and significant bias outside 92–97% SaO<sub>2</sub>. They also reported an increase in bias with decreasing SaO<sub>2</sub>, and SpO<sub>2</sub> values smaller than SaO<sub>2</sub> above 95%. Neither of these is consistent with our results or those of Ross or Rosychuk. It is reasonable to assume the differences are a reflection of the newer oximeter technology.

Our study has some limitations. First the SpO<sub>2</sub> data is an average of the two minutes coincident with the arterial blood gas, and no effort was made to eliminate comparisons where the SpO<sub>2</sub> or the infant were unstable at that time, as others have done. Doing so would have reduced the variability, but probably not the size of the bias. Thus our findings reflect real clinical situations and not idealized laboratory comparisons. Second the SaO<sub>2</sub> values that we reported were calculated by the blood gas analyzer, not directly measured by a co-oximeter. These calculations take into account pH and PaCO<sub>2</sub> but not other factors

affecting the SaO<sub>2</sub>-PaO<sub>2</sub> relationship. [10] Direct measurement of SaO<sub>2</sub> with a co-oximeter is rarely used today in the NICU. It is comforting that our results were consistent with those reported by Ross using a co-oximeter. Finally a single average of 4 values of SpO<sub>2</sub> over 2 minutes is seldom used in the NICU, rather staff are well versed at integrating problems from motion artifact and poor perfusion when accessing the trend, mitigating the impact of the point variability we reported.

## Conclusions

We found that SpO<sub>2</sub> readings are usually higher than calculated SaO<sub>2</sub> across the entire clinical range. The size of the bias is quite similar in SpO<sub>2</sub> ranges associated with hypoxemia and normoxemia, but is decreased significantly in higher ranges associated with hyperoxemia. We hope this data will provide additional insight to clinicians as they assess neonatal oxygenation based on continuous non-invasive pulse oximetry. We also expect this information should be helpful to those designing neonatal oxygen control systems and modeling neonatal oxygenation.

## Acknowledgement

This work was not supported by a grant or outside funding.

## References

- [1] Sweet, D. G., Carnielli, V., Greisen, G, et al.: *European consensus guidelines on the management of neonatal respiratory distress syndrome in preterm infants – 2016 update*. Neonatology 2017; 111(2), 107–125.
- [2] Cummings, J. J., Polin, R. A.: *Committee on Fetus and Newborn. Oxygen targeting in extremely low birth weight infants*. Pediatrics 2016; 138(2), e20161576.
- [3] Ross, P. A., Newth, C. J. L., Khemani, R.: *Accuracy of pulse oximetry in children*. Pediatrics 2014; 133, 22–29.
- [4] Rosychuk, R. J., Hudson-Mason, A., Eklund, D., Lacaze-Masmonteil, T.: *Discrepancies between arterial oxygen saturation and functional oxygen saturation measured with pulse oximetry in very preterm infants*. Neonatology 2012; 101(1), 14–18.
- [5] Bachman, T. E., Newth, C. J. L., Ross, P. A. et al.: *Pulse-oximetry readings and hypoxemia and hyperoxemia: NICU observational study*. 2017 In-Press.
- [6] De Halleux, V., Truttmann, A., Gagnon, Bard, H.: *The effect of blood transfusion on the hemoglobin oxygen dissociation curve of very early preterm infants during the first week of life*. Semin Perinatol. 2002 Dec;26(6):411-5.
- [7] Shiao, S. Y.: *Effects of fetal hemoglobin on accurate measurements of oxygen saturation in neonates*. Perinat Neonatal Nurs. 2005 Oct-Dec;19(4):348-61.

- [8] Johnston, E. D., Boyle, B., Juszczak, E. et al.: *Oxygen targeting in preterm infants using the Masimo SET Radical pulse oximeter*. Arch Dis Child Fetal Neonatal Ed 2011; 96: F429–F433.
- [9] Gerstmann, D., Berg, R., Haskell, R., et al.: *Operational valuation of pulse oximetry in NICU patients with arterial access*. J Perinatol 2003; 23, 378–383.
- [10] Severinghaus, J. W.: *Simple and accurate equations for human blood O<sub>2</sub> dissociation computations*. J. Appl. Physiol., 46, 1979, p. 599–602.

*Thomas E. Bachman MSc.*  
*Department of Biomedical Technology*  
*Faculty of Biomedical Engineering*  
*Czech Technical University in Prague*  
*nám. Sítná 3105, CZ-272 01 Kladno*  
*&*  
*Economedtrx*  
*www.medtrx.org*  
*Lake Arrowhead, CA USA 92352-1269*

*E-mail: [tbachman@me.com](mailto:tbachman@me.com)*  
*Phone: +1 909 337-0828*