

## ARTICLE



# Racial discrepancy in pulse oximeter accuracy in preterm infants

Zachary Vesoulis<sup>1</sup>✉, Anna Tims<sup>2</sup>, Hafsa Lodhi<sup>2</sup>, Natasha Lalos<sup>2</sup> and Halana Whitehead<sup>1</sup>

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

**OBJECTIVE:** Pulse oximetry is commonly used in Neonatology, however recent adult data suggest racial disparity in accuracy, with overestimation of oxygen saturation for Black patients.

**STUDY DESIGN:** Black and White infants <32 weeks gestation underwent simultaneous arterial blood gas and pulse oximetry measurement. Error by race was examined using mean bias,  $A_{rms}$ , Bland–Altman, and linear/non-linear analysis.

**RESULTS:** A total of 294 infants (124 Black, 170 White) were identified with mean GA of  $25.8 \pm 2.1$  weeks and mean BW of  $845 \pm 265$  grams, yielding 4387  $SaO_2$ – $SpO_2$  datapoints.  $SpO_2$  overestimation, measured by mean bias, was 2.4-fold greater for Black infants and resulted in greater occult hypoxemia ( $SpO_2 > 90\%$  when  $SaO_2 < 85\%$ ; 9.2% vs. 7.7% of samples). Sensitivity and specificity for detection of true hypoxemia were similar between groups (39 vs. 38%; 81 vs. 78%).

**CONCLUSION:** There is a modest but consistent difference in  $SpO_2$  error between Black and White infants, with increased incidence of occult hypoxemia in Black infants.

*Journal of Perinatology* (2022) 42:79–85; <https://doi.org/10.1038/s41372-021-01230-3>

## INTRODUCTION

Continuous pulse oximetry has been a valuable tool for Neonatologists since the early 1980s [1, 2]. Titration of supplemental oxygen to maintain a narrow window of oxygen saturation is essential to reduce the risk of retinopathy of prematurity [3, 4], bronchopulmonary dysplasia [5, 6], and death [7, 8]. Continuous pulse oximetry ( $SpO_2$ ) is superior to clinical observation alone; without it, desaturation can only be detected once arterial saturation ( $SaO_2$ ) has dropped below 80% and cyanosis develops [9, 10]. Pulse oximetry also avoids frequent phlebotomy for blood gas analysis, which is painful and causes iatrogenic anemia.

Oxyhemoglobin ( $HbO_2$ ) and deoxyhemoglobin (Hb) absorb red (660 nm) and infrared (940 nm) light differently—saturated blood permits increased transmission of red light but decreased transmission of infrared light. Accordingly, a pulse oximeter consists of red and infrared light emitters and a photoreceiver positioned on opposite sides of an arterial bed and measures the quantity of light transmitted through tissue. As oxygen saturation is the ratio of oxyhemoglobin to total hemoglobin, arterial saturation ( $SaO_2$ ) can be derived from variation in light absorption [11–13].

Melanin is a secondary absorber of near-infrared light and may impact pulse oximeter accuracy. A recent comparison of oxygen saturation determined by pulse oximeter and arterial blood gas (ABG) in adults, demonstrated a notable overestimation of oxygen saturation in patients self-identified as Black [14] and an increased incidence of occult hypoxemia. Another recent study of adults admitted to the ICU with COVID-19 [15] identified suboptimal pulse oximeter accuracy. As this study used a cohort with nearly

70% of individuals identified as Black, Asian or Minority Ethnic, the authors speculated that greater melanin concentration may have contributed to increased inaccuracy [16]. Given the link between hypoxemia and adverse outcomes in preterm infants, overestimation of oxygenation may be problematic.

For this study, we identified a cohort of preterm infants born before 32 weeks gestation who had a simultaneous collection of timed ABG samples and pulse oximetry. We hypothesized that differences in secondary light absorption between Black and White infants will lead to systematic error in pulse oximeter-based determination of arterial oxygen saturation.

## METHODS

### Cohort development

All preterm infants born between 2012 and 2019 with a gestational age of less than 32 weeks, birth weight less than 1500 g, and admitted to the St. Louis Children's Hospital Neonatal Intensive Care Unit (SLCH NICU) were eligible for inclusion.

Infants admitted to the SLCH NICU undergo continuous vital sign monitoring including pulse oximetry. Patient monitors were either Philips IntelliVue MP70 or MX800 (Philips Medical, Andover, MA), but both use a common pulse oximeter, the Nellcor  $SpO_2$  Module (Medtronic, Minneapolis, MN) with the Neonatal-Adult MAX-N adhesive  $SpO_2$  sensor (Covidien, Mansfield, MA). Vital sign data are automatically captured in an electronic database (BedMasterEX, Excel Medical, Jupiter, FL) sampled once per second (1 Hz). During the study period, all infants had standardized oxygen saturation targets to maintain  $SpO_2$  between 90 and 95% with alarm limits set between 88 and 96%. After infants reached 35 weeks post-menstrual age (PMA), the target range was changed to 90–100% with the desaturation alarm set to 88%.

<sup>1</sup>Department of Pediatrics, Division of Newborn Medicine, Washington University, St. Louis, MO, USA. <sup>2</sup>Department of Pediatrics, Washington University, St. Louis, MO, USA. ✉email: vesoulis\_z@wustl.edu

Infants were included in the study if they met gestational age and birth weight criteria, had valid vital sign data, and at least one ABG performed during hospitalization. Standard clinical variables were collected including gestational age, birth weight, sex, antenatal steroid exposure, method of delivery, and Apgar scores. Infants were classified as Black or White based on parental identification on birth certificates. Infants of Hispanic, Asian, or unspecified descent make up a small proportion of admissions to the SLCH NICU and were excluded as there would be an insufficient number for a representative sample. The study was reviewed and approved by the IRB at Washington University under waiver of consent.

### ABG analysis

Invasive arterial lines are placed for frequent blood sampling and/or arterial blood pressure monitoring. ABG analysis was performed in a consistent manner across all patients—the line was accessed and cleared and a minimum sample volume of 0.5 mL was obtained and immediately brought to the clinical laboratory where it was run on a gas analyzer (ABL800 Flex, Radiometer America, Brea, CA) yielding a measurement of arterial oxygen saturation (SaO<sub>2</sub>). For each patient in the cohort, the measured SaO<sub>2</sub> and the date/time of sample acquisition were recorded.

### SpO<sub>2</sub> processing and bias calculation

Raw recording files were converted to MATLAB format (The MathWorks, Natick, MA) using conversion software (University of Virginia, Charlottesville, VA). A processing script identified the location corresponding to the date/time of the ABG and extracted a time-matched 60-second window of SpO<sub>2</sub> data centered on the ABG timepoint (30 s before and after) which was then averaged. This average of over 60 s was employed to reduce the impact of transient fluctuations in the SpO<sub>2</sub> and to mimic the typical length of time taken to draw an arterial sample for the SaO<sub>2</sub>.

### Accuracy measurements

The accuracy of pulse oximetry to estimate arterial oxygen saturation can be evaluated in several different ways. The primary methods used in FDA clearance are mean bias and accuracy root mean squared, however, we performed an expanded analysis of pulse ox accuracy across six different metrics including:

1. Mean bias (B) – Mean bias is the average difference between SaO<sub>2</sub> and SpO<sub>2</sub> and is calculated using the formula  $B = \frac{\sum_{i=1}^n (SpO_2 - SaO_2)}{n}$ . Positive values resulting from this calculation indicate an overestimate of SaO<sub>2</sub> by the pulse oximeter while negative values indicate an underestimate.
2. Accuracy root mean squared (A<sub>rms</sub>) – A<sub>rms</sub> is a related measure of the average difference between SaO<sub>2</sub> and SpO<sub>2</sub> and is calculated using the formula  $A_{rms} = \sqrt{\frac{\sum_{i=1}^n (SpO_2 - SaO_2)^2}{n}}$  [17]. Given the quadratic nature of this calculation, A<sub>rms</sub> is always a positive value and ranges from 0 (no error at all) and increases as the number of errors increase.
3. Proportional bias – Also called Bland–Altman analysis, this method was developed to quantify differences between measurement methods. Like B and A<sub>rms</sub>, the mean difference between the two methods is first calculated. From these differences, the 95% limits of agreement can be computed as the average difference  $\pm$  1.96 standard deviations. Bland–Altman can be used to identify the relationship of discrepancies between two measurement methods, also called proportional bias. The presence of proportional bias indicates that the degree of disagreement varies over the range of measurements.
4. Prevalence of occult hypoxemia – The occult hypoxemia definition of Sjoding et al. was adapted to the preterm population. In this case, occult hypoxemia was defined as a true SaO<sub>2</sub> < 85% when SpO<sub>2</sub> reads a value in the normal range (SpO<sub>2</sub>  $\geq$  90%).
5. Sensitivity/specificity for detection of occult hypoxemia – The sensitivity of the pulse oximeter to detect true hypoxemia (SpO<sub>2</sub> < 90% when SaO<sub>2</sub> < 85%) was calculated as  $\frac{\text{True positive}}{\text{True positive} + \text{False negative}}$ . The specificity of the pulse oximeter to detect true hypoxemia was calculated as  $\frac{\text{True negative}}{\text{True negative} + \text{False positive}}$ . Sensitivity and specificity calculations were made for the overall cohort and within racial groups.

While B and A<sub>rms</sub> were calculated across the entire range of SpO<sub>2</sub> values, an evaluation of local bias was also conducted on the “clinically relevant” SpO<sub>2</sub> range of 85–100%. An additional limited investigation was performed to study the impact of post-menstrual age (PMA) at the time of sampling on measurement error.

### Statistical approach

Infant characteristics underwent univariate comparison using non-parametric methods including Fisher’s Exact Test for categorical variables and Mann–Whitney *U* test for continuous variables. The proportion of samples where occult hypoxemia occurred was compared between Black and White infants using Fisher’s Exact Test.

The relationship between SpO<sub>2</sub> and SaO<sub>2</sub>, bias, and PMA were modeled using the Pearson correlation coefficient and conventional linear regression. Non-linear regression was performed using a conditional mean function, where the outcome variable (e.g., SaO<sub>2</sub>, bias score) was modeled as a function of SpO<sub>2</sub>. In this approach, a larger dataset is convoluted or broken into smaller subsets, to which low-order polynomials are fit using a least-squares approach, weighting points closer to the subset more than those distant. Smoothed conditional means were calculated using the ggplot2 package for the R statistical package (R version 4.0.3, R Foundation for Statistical Computing, Vienna, Austria) utilizing the locally estimated scatterplot smoothing (LOESS) method.

## RESULTS

### Study cohort description

A total of 639 infants met gestational age and birth weight criteria; 186 infants were excluded as they did not have an ABG performed during their NICU hospitalization, 125 infants were excluded for lack of SpO<sub>2</sub> data at the time of ABG, and 34 infants were excluded for Asian, Hispanic, or not listed background, yielding 294 infants. For the remaining 294 infants, a total of 4,387 SaO<sub>2</sub>–SpO<sub>2</sub> pairs were available for further analysis. The median number of samples per infant was 11 (IQR 4–23) and the median postnatal age at sampling was four days (IQR 2–7), reflecting our typical clinical practice of obtaining an arterial blood gas every 8 h from an arterial catheter in place for an average of four days following birth for VLBW infants. Most of the samples were obtained within the first week of life (75% within seven days, Supplemental Fig. 1).

Of the 294 included infants, 42% were Black and 58% were White, consistent with the general demographic profile of infants admitted to the SLCH NICU (40% Black, 50% White, 10% Asian, Hispanic, or not listed). The two groups of infants were similar except for slightly lower birth weight (805 g vs. 875 g, *p* = 0.02) and median one-minute Apgar score (2 vs. 3, *p* < 0.01) in the Black infants. All other characteristics were not statistically different. A complete descriptive summary of the cohort can be found in Table 1.

### Occult hypoxemia

The number of data samples was balanced between the two groups, with 2044 samples for Black infants and 2342 samples for White infants. True hypoxemia (defined as SaO<sub>2</sub> < 85%) was noted slightly more often in Black infants, being identified in 312/2044 (15.2%) of samples as compared to 293/2342 (12.5%) of samples for White infants. Occult hypoxemia, (defined as SaO<sub>2</sub> < 85% when SpO<sub>2</sub>  $\geq$  90%) was more common in Black infants, occurring in 188/2044 (9.2%) of samples compared to 181/2343 (7.7%) of samples for White infants, although this difference did not meet statistical significance (*p* = 0.08).

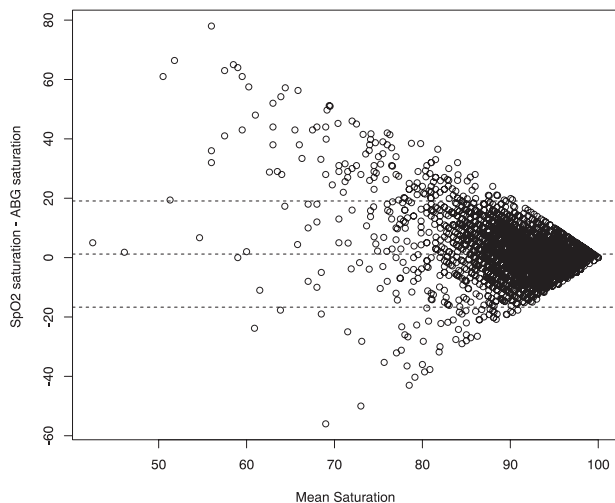
### Sensitivity and specificity for detection of hypoxemia

Of the 4387 SaO<sub>2</sub>–SpO<sub>2</sub> pairs collected in this study, 605/4387 (13.7%) were noted to have true hypoxemia. Overall, the sensitivity of the pulse oximeter for detecting true hypoxemia (defined as SpO<sub>2</sub> < 90% when SaO<sub>2</sub> < 85%) was 38% while the specificity was 89%. In subgroup analysis by race, sensitivity and

**Table 1.** Cohort descriptive statistics.

	Black infants, <i>n</i> = 124	White infants, <i>n</i> = 170	<i>P</i> value
Gestational age, mean (SD), weeks	25.6 (1.9)	25.9 (2.1)	0.09
Birth weight, mean (SD), grams	805 (260)	875 (268)	0.02*
Female sex, <i>n</i> (%)	64 (52)	74 (44)	0.19
Apgar score one-minute, median (range)	2 (0–8)	3 (0–10)	<0.01*
Apgar score five-minutes, median (range)	6 (0–9)	6 (0–9)	0.09
Received antenatal steroids, <i>n</i> (%)	87 (70)	127 (74)	0.49
Vaginal delivery, <i>n</i> (%)	35 (28)	42 (23)	0.68

\* denotes significance at  $p < 0.05$ .



**Fig. 1** Bland–Altman plot demonstrating agreement of ABG and SpO<sub>2</sub>. Note the positive bias overall (indicating overestimation) which worsens as SpO<sub>2</sub> decreases.

specificity were similar for Black infants (39% sensitive, 81% specific) and White infants (38% sensitive, 78% specific).

#### Mean bias and $A_{rms}$

The bias of each individual measurement was calculated as previously noted (SpO<sub>2</sub> saturation–ABG saturation) where a positive value indicates an overestimation of the true arterial saturation, and a negative value indicates an underestimation. The mean bias of the overall sample was 1.19%, indicating an overall overestimation of true oxygen saturation by a little more than 1% for all infants. Local bias was calculated for the “clinically relevant” range of saturations between 85 and 100% and was noted to be somewhat higher at 1.79%.

When comparing mean bias between Black and White infants, overestimation by the pulse oximeter is noted to be 2.4-fold greater for Black infants compared to White infants, (mean bias of 1.73% vs. 0.72%,  $p < 0.01$ ). A similar discrepancy is noted between Black and White infants when looking at local bias for saturations between 85 and 100%, with 1.5-fold greater overestimation for Black infants compared to White infants (2.22% vs. 1.41%,  $p < 0.01$ ).

$A_{rms}$  over the entire range of saturations was notable at 9.2%, exceeding the desired specification of 2–3% [17, 18]. As would be expected,  $A_{rms}$  worsened as saturations decreased with  $A_{rms}$  = 7.9% for SpO<sub>2</sub> 90–100%, 10.8% for SpO<sub>2</sub> 80–89%, 16.9% for SpO<sub>2</sub> 70–79%, and 27.9% for SpO<sub>2</sub> <70%. Consistent with other measures of pulse oximeter accuracy, there was a racial

discrepancy in  $A_{rms}$  values between Black and White infants (9.5% vs. 8.9%).

#### Bland–Altman analysis

In Bland–Altman analysis, an overall bias of 1.19% was confirmed, with a 95% confidence interval noted to be between 0.92–1.46% over the entire range of oxygen saturation. As shown in Fig. 1, the magnitude of the difference quickly goes outside of the control limits ( $\pm 1.96$  SD) once the oxygen saturation drops below 90% with an increasing large overestimation as saturation decreases.

#### Linear and non-linear correlation

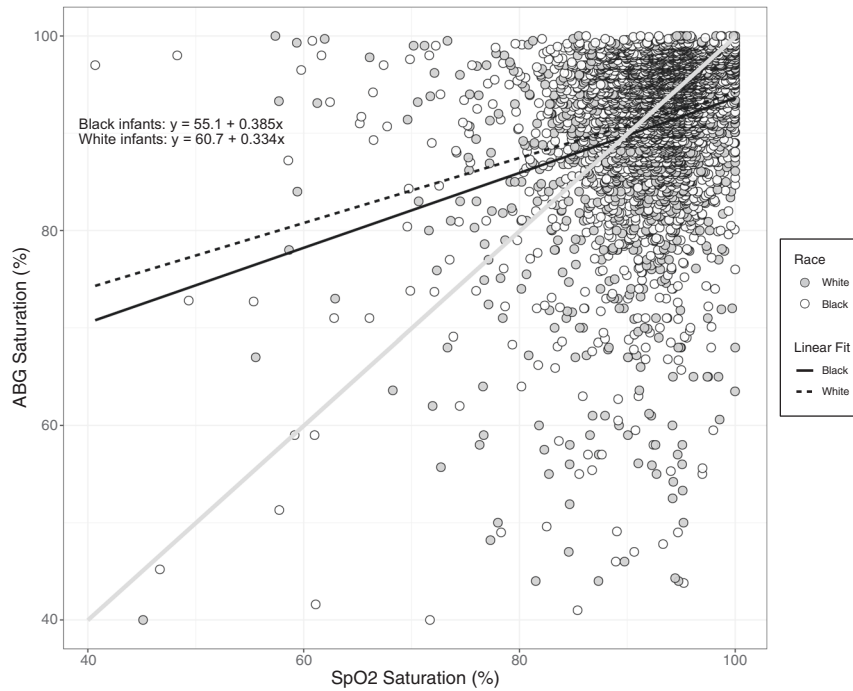
The overall linear correlation between SaO<sub>2</sub> and SpO<sub>2</sub> was moderate, albeit statistically significant ( $R^2 = 0.244$ ,  $p < 0.01$ ). When examined separately, the correlation is slightly stronger for Black infants than White infants ( $R^2 = 0.254$  vs. 0.236). In Fig. 2, a scatterplot of all SaO<sub>2</sub>–SpO<sub>2</sub> values is shown, categorized by race. The slope and intercept of the best fit line demonstrate that not only do Black infants have higher recorded SpO<sub>2</sub> measurements for a given SaO<sub>2</sub>, but this difference also increases as SpO<sub>2</sub> drops.

Non-linear analysis (Fig. 3) demonstrates consistency in bias between Black and White infants only when SpO<sub>2</sub> is between 96 and 100%. When SpO<sub>2</sub> is  $\leq 95\%$ , there is a persistent bias gap between Black and White infants with a higher SpO<sub>2</sub> value for a given SaO<sub>2</sub> in Black infants, although the exact degree varies. While all pulse oximeters follow the same general pattern of overestimation at lower saturations and underestimation at higher saturations, the exact location of this crossover point is different for Black infants, which occurs at an approximate SpO<sub>2</sub> of 90% compared to 92% for White infants. When SpO<sub>2</sub> is less than 96%, there is a widening gap in the degree of error between Black and White infants, with greater underestimation of SpO<sub>2</sub> in the White infants and overestimation in Black infants (Fig. 4).

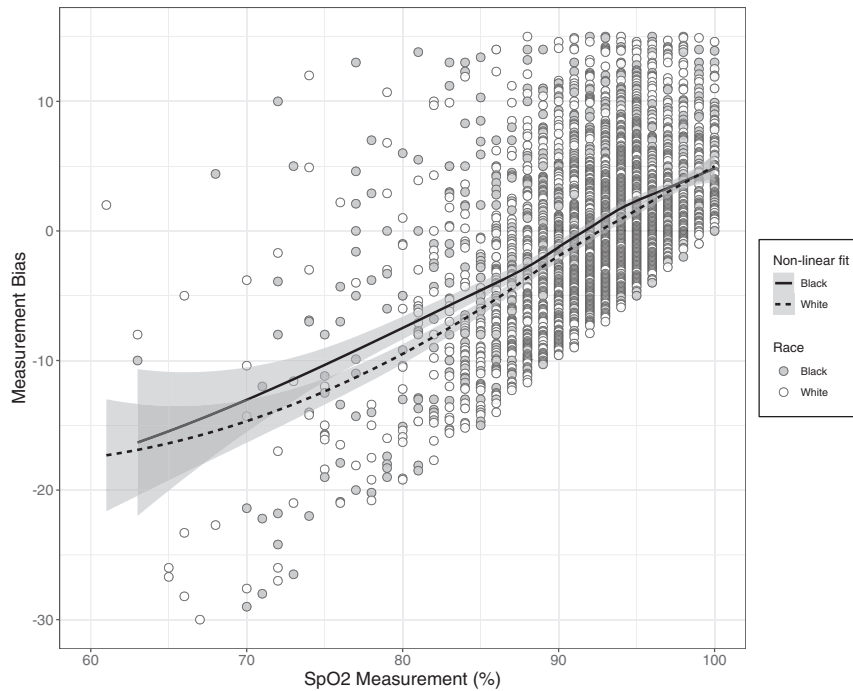
The mean bias by race was evaluated over the range of corrected PMA at the time of measurement. As shown in Supplemental Fig. 2, there is no notable variance by corrected PMA, although there is a significant amount of uncertainty after 35 weeks PMA due to sparse samples.

#### DISCUSSION

These data reveal a complex relationship between pulse oximetry-based estimation of arterial oxygen saturation and race. Over this collection of SaO<sub>2</sub>–SpO<sub>2</sub> pairs, Black infants were noted to have true hypoxemia more often than White infants. Pulse oximetry overestimated the true arterial oxygen saturation of Black VLBW infants by 1% more on average compared to White infants. The direction of error also varies by SpO<sub>2</sub>, with overestimation of the true saturation when SpO<sub>2</sub> is greater than 90% and underestimation at saturations of 90% or below. The greater degree of



**Fig. 2 Scatterplot of oxygen saturation measured by pulse oximeter (x-axis) and ABG (y-axis).** Black infants are shown as gray dots, White infants are shown as white dots. Regression lines are shown for each race group. The line of unity is shown as a diagonal gray line.



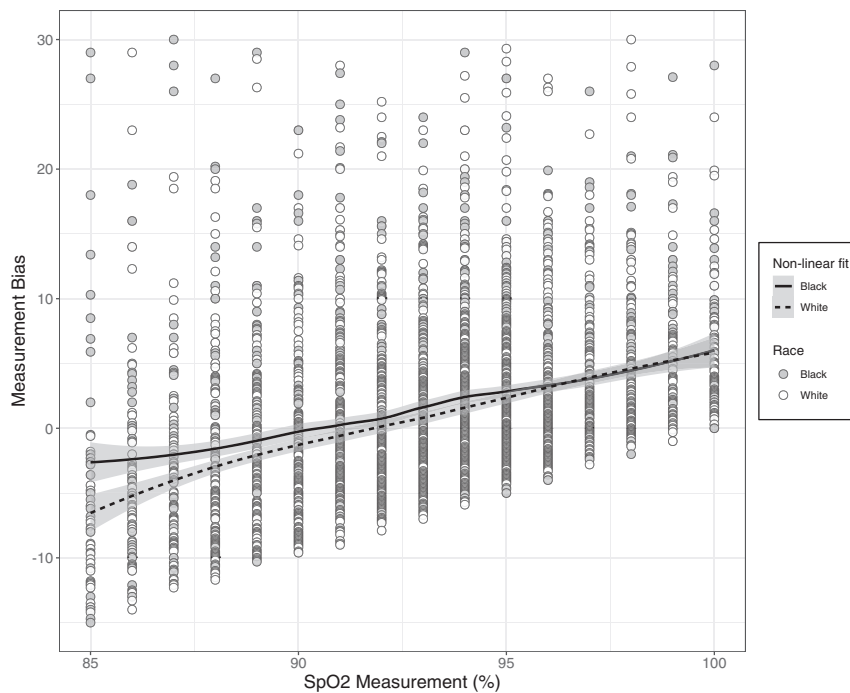
**Fig. 3 Measurement bias (SpO<sub>2</sub> saturation–ABG saturation) is shown, clustered by SpO<sub>2</sub> saturation and race.** Black infants are shown as gray dots, White infants are shown as white dots. Non-linear regression lines are shown for each group (solid for Black infants, dashed for White infants). Shaded areas represent 95% CI.

imprecision for Black infants is well-summarized by the difference in  $A_{rms}$  values (9.5% vs. 8.9%). The aggregate impact of each of these small differences was an increased incidence of occult hypoxemia in Black infants, occurring in 9.2% of samples compared to 7.7% of samples for White infants.

Although the difference in measurement bias between Black and White infants is small, these one-minute samples represent

only a small fraction of each infant’s NICU hospitalization. There are many examples in the literature of the association between hypoxia and severe outcomes including an increased incidence of intraventricular hemorrhage (IVH) [19, 20] and death [7, 8]. In a previous publication, we identified that the difference in time spent with severe hypoxia between infants with severe IVH and those without severe IVH was less than 3% of the total recording





**Fig. 4** Measurement bias is shown, clustered by SpO<sub>2</sub> saturation and race over the clinically relevant SpO<sub>2</sub> range of 85–100%. Black infants are shown as gray dots, White infants are shown as white dots. Non-linear regression lines are shown for each group (solid for Black infants, dashed for White infants). Shaded areas represent 95% CI.

time [21]. Thus, even relatively small increases in hypoxia burden driven by underestimation of true arterial oxygen saturation may have profound consequences. Although the degree of over- or underestimation varies by SpO<sub>2</sub> value, infants spend much of their hospitalization with saturations greater than 85%, thus SpO<sub>2</sub> performance in this range is the primary driving factor and yields a net average overestimation for Black infants.

In the BOOST-II and SUPPORT trials, pulse oximeters with a masking algorithm were used to force separation in the distribution of oxygen saturations while still within what was intended to be the normal range of SpO<sub>2</sub> for preterm infants. The degree of separation generated by the algorithm was small, on the order of 2.5–3% [22], but the difference in the outcome, namely mortality, was much greater (23.1 vs. 15.9%) [8] in the lower saturation group. The race-based systematic error identified in our study (1%) makes up a significant proportion of that separation. Indeed, when the distribution of SpO<sub>2</sub> values for infants in this study are compared, Black infants have a greater distribution of higher SpO<sub>2</sub> values while also a greater number of lower SaO<sub>2</sub> values (Supplemental Fig. 3). In a recent manuscript [23] we described a significant gap in estimated mortality risk between Black and White infants, with greater predicted mortality for Black infants when all other factors were held constant. It is possible that a disparity in extremes of oxygenation is a contributing factor.

Factors altering the absorption of red and near-infrared light should be considered. Even at low gestational ages, infants have mixed hemoglobin subtypes. While the predominant subtype is fetal hemoglobin (HbF), adult hemoglobin (HbA) is also present with estimates ranging between 10 and 20% [24, 25]. In some investigations, the absorption spectra of fetal and adult hemoglobin were found to be identical [26, 27], however, there is some evidence that an admixture of HbF and HbA results in impaired pulse oximetry accuracy (underestimation) by 3–4% [28, 29], although this effect occurs primarily at lower oxygen saturation [30]. Although not measured, the slightly lower gestational age of

Black infants in this cohort may have a higher proportion of HbF contributing to greater SpO<sub>2</sub> error.

Skin pigmentation may also play a role in pulse oximeter accuracy. Melanin is a pigment produced by melanocytes located in the basal epidermis after exposure to ultraviolet radiation (UV, 280–315 nm) [31]. Although melanogenesis accelerates during periods of increased UV-B exposure (suntan), different amounts of basal melanin production result in varying skin tones [32]. Although the peak absorption frequency of melanin is between 400 and 600 nm (ultraviolet band), it also absorbs light across visible and infrared frequencies and is a particularly strong absorber in the near-infrared range [33]. As light absorption in the near-infrared spectrum is used by the pulse oximeter to identify the quantity of oxy- and deoxyhemoglobin, additional absorption by a secondary chromophore (melanin) disrupts the expected relationship. This potential problem has been borne out in experimental data demonstrating that increasing amounts of melanin lead to increasing error in spectroscopic measurements of hemoglobin species [33].

Melanocytes can be identified in the embryonic epidermis in the first trimester [34]. There has been limited examination of the developmental trajectory of melanin by melanocytes at different gestational ages, although a small study suggests that significant differences in skin reflectance (indicative of sufficient melanin to alter light absorption) between White and Black patients do not occur until 32 weeks corrected gestational age [35].

Complicating this analysis is the frequency with which these infants receive phototherapy for jaundice. Phototherapy lamps are engineered to generate light in the same frequency range that unconjugated bilirubin maximally absorbs light (340–540 nm) [34]. This spectral range overlaps with the band which activates melanocytes including UV-A and B (290–400 nm) [36]. There are a number of reports of an increase in melanin production following typical clinical phototherapy treatment in Black and Asian infants [37]. Increased melanin production in response to phototherapy may accelerate the change in skin reflectance beyond what might

be expected for gestational age, introducing the possibility of pulse oximetry error at younger than expected gestational ages.

The results of this study raise important questions about the utility and reliability of pulse oximetry in critical care. The modest correlation between SaO<sub>2</sub> and SpO<sub>2</sub> has been noted in other studies of infants and children [38–40], especially at lower saturation, and is likely the result of using healthy volunteers for calibration which may not be representative of the sick infant [11]. The impact of skin pigmentation has been investigated in several previous studies of adults and older children, although not in preterm infants. Studies of adults have reported mixed findings; some have identified overestimation of oxygen saturation by 2–10% in participants with greater skin pigmentation [41, 42]. Other studies [43, 44] have failed to replicate this difference, although both studies noted lower SpO<sub>2</sub> signal quality in adults with greater skin pigmentation. There has been only a single investigation of racial differences in pulse oximeter performance for infants [45] and it did not demonstrate a difference between infants with light and dark pigment. However, this study had a small sample size and strict inclusion criteria (term infants, no anemia or hypotension, stable SpO<sub>2</sub> for 2 minutes prior to sample) which are not typical for premature infants.

Although the validation data for the pulse oximeter used in this study is not publicly available, there are several other published neonatal validation studies using Nellcor-based pulse oximeters. Unfortunately, the race of the study population is either not provided in these reports [46–52] or Black infants are under-represented in the study cohort [29]. Without intentional oversampling of Black infants or stratified analysis of infants by race, disparities in device performance are not apparent.

Infants in this study received a race classification based on self-report in birth certificate documentation. This binary approach does not capture the continuum of skin pigmentation, which likely has a variable impact on pulse oximeter accuracy. Several different systems for quantifying skin color have been suggested including the Fitzpatrick Skin Phototype [53], which uses a combination of visual appearance and response to ultraviolet light (burning vs. tanning). An alternative approach is the individual typology angle (ITA) which utilizes digital photography of a skin sample under highly controlled lighting conditions to quantify the amount of red-green, yellow-blue, and lightness-darkness (L\*a\*b\* color) using software [54]. Although considerably more complex, melanin levels can also be quantified spectroscopically. Indeed, there are research near-infrared spectroscopy devices, tissue oxygen monitors which operate on similar physics principles to pulse oximetry, which utilize algorithms to quantify [55] and remove the influence of melanin in measurements [56, 57]. Notably, none of these approaches have been evaluated in the neonate.

There are several limitations of this study. First, blood gas samples can be contaminated by the accidental introduction of air into the syringe. This is a known problem in all blood gas analyses and samples are carefully examined (and potentially discarded) before the assay is run, minimizing this risk [58].

Second, the SpO<sub>2</sub> sensor placement is rotated every twelve hours to prevent skin injury. In positions other than the right upper extremity, the probe is in a post-ductal position. For preterm infants with a patent ductus arteriosus, there is the possibility of a mismatch between pre- and post-ductal measurements. Sensor placement is not routinely charted in the medical record and PDA screening is not universal, thus it is impossible to reconstruct when or how often this occurred.

Finally, infants in this study were assigned a binary race classification based on birth certificate data. The retrospective nature of this study prevents qualitative or quantitative assessment of skin tone for further interrogation. Future prospective studies should quantify melanin content, ideally spectroscopically, for a more granular understanding of the relationship between

skin tone and device performance disparity. It is essential that preterm neonates are not excluded from correction algorithms investigation.

In conclusion, we find a small but consistent racial disparity in oxygen saturation measurement by pulse oximetry and an increased incidence of occult hypoxemia in Black preterm infants. There is increasing awareness of racial disparities in outcomes of preterm infants, particularly the risk of mortality [59–61]. Knowledge of the potential for occult hypoxemia may lead to changes in oxygen saturation targeting, with more attention paid to the avoidance of low-normal saturations to reduce the risk of adverse outcomes.

## REFERENCES

- Dawson JA, Davis PG, O'Donnell CPF, Kamlin COF, Morley CJ. Pulse oximetry for monitoring infants in the delivery room: a review. *Arch Dis Child Fetal Neonatal Ed.* 2007;92:F4–7.
- Salyer JW. Neonatal and pediatric pulse oximetry. *Respir Care.* 2003;48:386–96. discussion 397–398
- Sears JE, Pietz J, Sonnie C, Dolcini D, Hoppe G. A change in oxygen supplementation can decrease the incidence of retinopathy of prematurity. *Ophthalmology.* 2009;116:513–518.
- Supplemental therapeutic oxygen for prethreshold retinopathy of prematurity (STOP-ROP), a randomized, controlled trial. I: primary outcomes. *Pediatrics* 2000;105:295–310.
- Warner BB, Stuart LA, Papes RA, Wispé JR. Functional and pathological effects of prolonged hyperoxia in neonatal mice. *Am J Physiol.* 1998;275:L110–117.
- Bhandari V, Elias JA. Cytokines in tolerance to hyperoxia-induced injury in the developing and adult lung. *Free Radic Biol Med.* 2006;41:4–18.
- SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network, Carlo WA, Finer NN, Walsh MC, Rich W, Gantz MG, et al. Target ranges of oxygen saturation in extremely preterm infants. *N Engl J Med.* 2010;362:1959–69.
- BOOST II United Kingdom Collaborative Group, BOOST II Australia Collaborative Group, BOOST II New Zealand Collaborative Group, Stenson BJ, Tarnow-Mordi WO, Darlow BA, et al. Oxygen saturation and outcomes in preterm infants. *N Engl J Med.* 2013;368:2094–104.
- Changizi M, Rio K. Harnessing color vision for visual oximetry in central cyanosis. *Med Hypotheses.* 2010;74:87–91.
- O'Donnell CPF, Kamlin COF, Davis PG, Carlin JB, Morley CJ. Clinical assessment of infant colour at delivery. *Arch Dis Child Fetal Neonatal Ed.* 2007;92:F465–467.
- Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. *MDER.* 2014;7:231–9.
- Chan ED, Chan MM, Chan MM. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respiratory Med.* 2013;107:789–99.
- Sahni R. Continuous noninvasive monitoring in the neonatal ICU. *Curr Opin Pediatr.* 2017;29:141–148.
- Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *N Engl J Med.* 2020;383:2477–2478.
- Philip KEJ, Bennett B, Fuller S, Lonergan B, McFadyen C, Burns J, et al. Working accuracy of pulse oximetry in COVID-19 patients stepping down from intensive care: a clinical evaluation. *BMJ Open Res.* 2020;7:e000778.
- Philip KEJ, Tidswell R, McFadyen C. Racial bias in pulse oximetry: more statistical detail may help tackle the problem. *BMJ* 2021;372:n298.
- Batchelder PB, Raley DM. Maximizing the laboratory setting for testing devices and understanding statistical output in pulse oximetry. *Anesthesia Analgesia.* 2007;105:S85–S94.
- Luks AM, Swenson ER. Pulse oximetry for monitoring patients with COVID-19 at home. potential pitfalls and practical guidance. *Ann ATS.* 2020;17:1040–1046.
- Ng IHX, da Costa CS, Zeiler FA, Wong FY, Smielewski P, Czosnyka M et al. Burden of hypoxia and intraventricular haemorrhage in extremely preterm infants. *Arch Dis Child Fetal Neonatal Ed* 2019. <https://doi.org/10.1136/archdischild-2019-316883>.
- Ballabh P. Intraventricular hemorrhage in premature infants: mechanism of disease. *Pediatr Res.* 2010;67:1–8.
- Vesoulis ZA, Bank RL, Lake D, Wallman-Stokes A, Sahni R, Moorman JR, et al. Early hypoxemia burden is strongly associated with severe intracranial hemorrhage in preterm infants. *J Perinatol.* 2019;39:48–53.
- Whyte RK, Nelson H, Roberts RS, Schmidt B. Benefits of oxygen saturation targeting trials: oximeter calibration software revision and infant saturations. *J Pediatrics.* 2017;182:382–384.
- Whitehead HV, McPherson CC, Vesoulis ZA, Cohlan BA, Rao R, Warner BB, et al. The challenge of risk stratification of infants born preterm in the setting of

- competing and disparate healthcare outcomes. *The. J Pediatrics*. 2020;223:194–196.
24. Phillips HM, Holland BM, Jones JG, Abdel-Moiz AL, Turner TL, Wardrop CA. Definitive estimate of rate of hemoglobin switching: measurement of percent hemoglobin F in neonatal reticulocytes. *Pediatr Res*. 1988;23:595–597.
  25. Bard H. Postnatal fetal and adult hemoglobin synthesis in early preterm newborn infants. *J Clin Invest*. 1973;52:1789–1795.
  26. Zijlstra WG, Buursma A, Meeuwse-van der Roest WP. Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin. *Clin Chem*. 1991;37:1633–1638.
  27. Harris AP, Sendak MJ, Donham RT, Thomas M, Duncan D. Absorption characteristics of human fetal hemoglobin at wavelengths used in pulse oximetry. *J Clin Monit*. 1988;4:175–177.
  28. Whyte RK, Kirpalani H, Asztalos EV, Andersen C, Blajchman M, Heddle N, et al. Neurodevelopmental outcome of extremely low birth weight infants randomly assigned to restrictive or liberal hemoglobin thresholds for blood transfusion. *Pediatrics*. 2009;123:207–213.
  29. Shiao S-YPK, Ou C-N. Validation of oxygen saturation monitoring in neonates. *Am J Crit Care*. 2007;16:168–178.
  30. Nijland R, Jongsma HW, Nijhuis JG, Oeseburg B, Zijlstra WG. Notes on the apparent discordance of pulse oximetry and multi-wavelength haemoglobin photometry. *Acta Anaesthesiol Scand Suppl*. 1995;107:49–52.
  31. Ou-Yang H, Stamatias G, Kollias N. Spectral responses of melanin to ultraviolet A irradiation. *J Investigative Dermatol*. 2004;122:492–496.
  32. Montagna W, Prota G, Kenney JA. *Jr Black Skin: Structure and Function*. Elsevier Science: Saint Louis, 2014. <http://qut.eblib.com.au/patron/FullRecord.aspx?p=1190974> (accessed 5 Feb 2021).
  33. Matas A, Sowa MG, Taylor G, Mantsch HH. Melanin as a confounding factor in near infrared spectroscopy of skin. *Vibrational Spectrosc*. 2002;28:45–52.
  34. Martin RJ, Fanaroff AA, Walsh MC. Fanaroff and Martin's neonatal-perinatal medicine: diseases of the fetus and infant. Saunders/Elsevier: Philadelphia, 2011.
  35. Post PW, Krauss AN, Waldman S, Auld PAM. Skin reflectance of newborn infants from 25 to 44 weeks gestational age. *Hum Biol*. 1976;48:541–557.
  36. Rosen CF, Seki Y, Farinelli W, Stern RS, Fitzpatrick TB, Pathak MA, et al. A comparison of the melanocyte response to narrow band UVA and UVB exposure in vivo. *J Invest Dermatol*. 1987;88:774–779.
  37. Lyngsnes Randeberg L, Bruzell Roll E, Norvang Nilsen LT, Christensen T, Svaasand LO. In vivo spectroscopy of jaundiced newborn skin reveals more than a bilirubin index: Reflectance spectroscopy and jaundice. *Acta Paediatrica*. 2007;94:65–71.
  38. Perkins GD, McAuley DF, Giles S, Routledge H, Gao F. Do changes in pulse oximeter oxygen saturation predict equivalent changes in arterial oxygen saturation? *Crit Care*. 2003;7:R67.
  39. Rosychuk RJ, 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;101:14–19.
  40. Ross PA, Newth CJL, Khemani RG. Accuracy of pulse oximetry in children. *Pediatrics*. 2014;133:22–29.
  41. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005;102:715–719.
  42. 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;105:518–23.
  43. Adler JN, Hughes LA, Vtivecchia R, Camargo CA, Jr. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Academic Emerg Med*. 1998;5:965–970.
  44. Ries AL, Prewitt LM, Johnson JJ. Skin color and ear oximetry. *Chest*. 1989;96:287–290.
  45. Foglia EE, Whyte RK, Chaudhary A, Mott A, Chen J, Propert KJ, et al. The effect of skin pigmentation on the accuracy of pulse oximetry in infants with hypoxemia. *J Pediatrics*. 2017;182:375–377.e2.
  46. Ramanathan R, Durand M, Larrazabal C. Pulse oximetry in very low birth weight infants with acute and chronic lung disease. *Pediatrics*. 1987;79:612–617.
  47. Gerstmann D, Berg R, Haskell R, Brower C, Wood K, Yoder B, et al. Operational evaluation of pulse oximetry in nicu patients with arterial access. *J Perinatol*. 2003;23:378–383.
  48. Bohnhorst B, Peter CS, Poets CF. Detection of hyperoxaemia in neonates: data from three new pulse oximeters. *Arch Dis Child Fetal Neonatal Ed*. 2002;87:F217–219.
  49. Hay WW, Rodden DJ, Collins SM, Melara DL, Hale KA, Fashaw LM. Reliability of conventional and new pulse oximetry in neonatal patients. *J Perinatol*. 2002;22:360–366.
  50. Hay WW, Brockway JM, Eyzaguirre M. Neonatal pulse oximetry: accuracy and reliability. *Pediatrics*. 1989;83:717–722.
  51. Bucher HU, Fanconi S, Baeckert P, Duc G. Hyperoxemia in newborn infants: detection by pulse oximetry. *Pediatrics*. 1989;84:226–230.
  52. Poets CF, Wilken M, Seidenberg J, Southall DP, von der Hardt H. Reliability of a pulse oximeter in the detection of hyperoxemia. *J Pediatrics*. 1993;122:87–90.
  53. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol*. 1988;124:869.
  54. Del Bino S, Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br J Dermatol*. 2013;169:33–40.
  55. Zonios G, Bykowski J, Kollias N. Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy. *J Investigative Dermatol*. 2001;117:1452–1457.
  56. Soyemi OO, Landry MR, Yang Y, Idwasi PO, Soller BR. Skin color correction for tissue spectroscopy: demonstration of a novel approach with tissue-mimicking phantoms. *Appl Spectrosc*. 2005;59:237–244.
  57. Yang Y, Landry MR, Soyemi OO, Shear MA, Anunciacion DS, Soller BR. Simultaneous correction of the influence of skin color and fat on tissue spectroscopy by use of a two-distance fiber-optic probe and orthogonalization technique. *Opt Lett*. 2005;30:2269.
  58. Biswas CK, Ramos JM, Agroyannis B, Kerr DN. Blood gas analysis: effect of air bubbles in syringe and delay in estimation. *Br Med J (Clin Res Ed)*. 1982;284:923–927.
  59. Alexander GR, Wingate MS, Bader D, Kogan MD. The increasing racial disparity in infant mortality rates: composition and contributors to recent US trends. *Am J Obstet Gynecol*. 2008;198:51.e1–51.e9.
  60. Matoba N, Collins JW. Racial disparity in infant mortality. *Semin Perinatol*. 2017;41:354–359.
  61. Anderson JG, Rogers EE, Baer RJ, Oltman SP, Paynter R, Partridge JC, et al. Racial and ethnic disparities in preterm infant mortality and severe morbidity: a population-based study. *Neonatology*. 2018;113:44–54.

#### AUTHOR CONTRIBUTIONS

ZAV and HWV developed the research idea. AT, HL, and NL performed chart reviews, collected, and organized data. ZAV performed the analysis and wrote the initial draft of the manuscript. HWV, AT, HL, and NL critically reviewed the manuscript. All of the authors have approved of the final manuscript, as submitted.

#### FUNDING

NIH grant: K23 NS111086 [Vesoulis].

#### COMPETING INTERESTS

Dr. Vesoulis provides independent consulting for Medtronic plc in an unrelated capacity. This project was not discussed with Medtronic during any stage of design, analysis, or writing. None of the authors have any other real or perceived conflicts of interest to disclose.

#### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41372-021-01230-3>.

**Correspondence** and requests for materials should be addressed to Zachary Vesoulis.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.