ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

Deconstructing Oxygenation Disparities: Skin Lightness (L*) and Clinical Severity as Drivers of Pulse Oximetry Bias and the 'Lost Window of Care'

Chethana Hosamalangi Suresha
Independent Researcher, Mysuru, Karnataka, India

Abstract

Background: Traditional pulse oximetry (SpO2) has been widely criticised for racial bias, yet the underlying mechanisms remain obscured by the use of "Race" as a categorical social proxy. This study deconstructs these disparities by isolating the impact of objective skin physics and clinical severity on diagnostic accuracy.

Methods: Using the ENCoDE dataset (N= 68,413 synchronised SpO2-SaO2 pairs), we replaced self-reported race with the CIELAB L* (Lightness) scale. Participants were segmented into quartiles based on L* values (Q1: 42.1-56.9; Q4: 68.3-91.3). We quantified instrumental bias via Mean Absolute Error (MAE), assessed diagnostic failure risk through logistic regression, and developed a "Time-to-Event" algorithm to measure the "Lost Window"- the duration of unrecognised occult hypoxemia.

Results: A significant linear gradient of error was observed ($P < 0.0001$, $F = 1571.86$), with the darkest quartile (Q1) exhibiting an MAE of 21.27% compared to 15.27% in the lightest (Q4). This "Double Jeopardy" effect, where darker skin tone and lower arterial saturation compound, resulted in a 28.8 Hour "Lost Window" for Q1 patients, nearly double that of lighter-skinned cohorts. Logistic regression yielded an Odds Ratio of 0.9518 per unit of L*, indicating a 4.82% increase in the risk of occult hypoxemia for each 1-point decrease in skin lightness.

Conclusion: Pulse oximetry bias is a predictable function of skin physics and clinical state. Our proposed "Skin-Aware" recalibration algorithm successfully reduced recognition delays by 7.5 hours in simulations. These findings advocate for a shift toward L* stratified device validation and software-based corrections to ensure physiological equity in acute care.

Keywords: Pulse oximetry bias, oxygenation disparities, physiological equity, health disparity.

Introduction:

Pulse oximetry (SpO_2) has emerged as an essential component of day-to-day clinical monitoring, facilitating a non-invasive, instantaneous assessment of arterial oxygen saturation (SaO_2). However, the COVID-19 pandemic has prompted a worldwide reassessment of its limitations, particularly concerning the precision of measurements in individuals with darker skin pigmentation. While an increasing number of research has revealed a significant racial disparity in oximetry effectiveness. (Sjoding et al., 2020), the primary determinant of such inaccuracies remains largely obscured by social constructs such as race. This study intends to shed light on these disparities by altering the focus from categorical racial classifications to empirical physical and physiological parameters, notably CIELAB L* (Skin Lightness) and clinical saturation severity.

Conventional oximetry operates on how differently oxygenated and deoxygenated haemoglobin absorb red (660 nm) and infrared (940 nm) light. The skin and surrounding tissue have been assumed

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

to be optically neutral to ensure this computation is accurate. But on the same spectrum, melanin shows notable light absorption, especially at the wavelength of 660. nm (Bickler et al., 2005). In skin with more pigmentation, this overlap creates optical interference that causes a systematic overestimation of oxygen levels. This study extends beyond the biological imprecision of race by measuring the connection between skin physics and instrumental bias using the L* scale, ranging from 0 (absolute black) to 100 (absolute white).

Furthermore, the patient's condition contributes to the clinical effect of this bias. Although pulse oximeters have typically been validated on healthy volunteers with stable saturations, profound hypoxemia (SaO_2 < 85%) frequency impairs their performance. There is a known "Double Jeopardy" effect in various clinical settings, such as the ENCoDE cohort (N= 68,413), where patients with darker skin tones often have more severe respiratory impairment. When the device's fundamental inaccuracy at low saturations is paired with melanin-driven optical interference, the resulting mistake is amplified, hiding crucial hypoxemia and giving doctors a risky "false sense of security".

This measurement inaccuracy has temporal consequences in addition to those associated with statistics. Clinicians often fail to sufficiently "de-bias" these readings while making treatment decisions, according to recent behavioural data (Alsan et al., 2026). This failure creates a "Lost Window" of care; the time elapsed from the onset of true hypoxemia to its eventual clinical identification via invasive testing. This delay prevents timely interventions, such as oxygen titration or ventilation adjustments, potentially leading to adverse outcomes.

Methodology:

1.Data Source and Cohort Selection:

This study utilized the ENCoDE (Evaluating Non-invasive COntinuum of Diagnostic Errors) dataset from PhysioNet, a multi-center retrospective cohort containing high-fidelity physiological monitoring and electronic health record (EHR) data. The dataset captures longitudinal records of patients admitted to acute and intensive care units, providing a high-density environment for observing oxygenation events.

Inclusion and Exclusion Criteria:

To ensure the integrity of the "gold-standard" comparison, observations were limited to:

- **Time-Synchronized Pairs**: Only pairs of pulse oximetry (SpO_2) and arterial blood gas (SaO2) measurements recorded within a ≤ 5-minute window were included to minimize the risk of physiological drift between readings.
- **Objective Metadat**a: Records were required to have associated CIELAB L* metrics.
- **Physiological Bounds**: Erroneous readings (e.g., SpO_2 > 100%) were excluded.

The final analytical sample consisted of 68,413 time-synchronized pairs, providing sufficient statistical power to detect subtle optical interferences.

2. Quantification of Skin Lightness (L*):

This study used the CIELAB L* Lightness scale, in contrast to conventional clinical research that uses self-reported race, a social proxy with high intra-group phenotypic diversity. Absolute black is represented by 0 and absolute white by 100 in the objective physical metric L*.

To allow a comparative examination of instrumental bias across the melanin spectrum, patients were divided into quartiles according to their measured L* values. In order to determine the physical gradient and maintain equal observation counts across groups, quartiles were defined using the pd.qcut function in python:

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

```
--- Physical L* Ranges by Quartile ---
                          min        max  count
L_quartile
Darkest (Low L*)    42.116751  56.936364  17708
Medium-Dark         57.092979  61.822535  17613
Medium-Light        62.005814  68.225191  16078
Lightest (High L*)  68.295359  91.333333  16981

--- Corrected Metrics ---
                          MAE       SaO2
L_quartile
Darkest (Low L*)    21.265699  81.167326
Medium-Dark         16.249815  91.291376
Medium-Light        14.491603  91.820687
Lightest (High L*)  15.271951  92.464107
```

Table 1: Distribution of the ENCoDE cohort (N= 68,413) segmented into quartiles based on CIELAB L* values. The table illustrates the physical range of each quartile alongside the baseline Mean Absolute Error (MAE) and arterial oxygen saturation (SaO2). Note the significantly lower baseline SaO2 in the darkest quartile (Q1), highlighting the "Double Jeopardy" of clinical severity and optical interference.

Q1 (Darkest): $L^* \in [42.1, 56.9]$
Q2 (Medium-Dark): $L^* \in [57.1, 61.8]$
Q3 (Medium-Light): $L^* \in [62.0, 68.2]$
Q4 (Lightest): $L^* \in [68.3, 91.3]$

3. Statistical Analysis of Instrumental Bias:
The primary dependent variable was the Error Delta, defined as the arithmetic difference between the pulse oximetry reading and the arterial blood gas measurement:

$$Error\ Delta = SpO_2 - SaO_2$$

A positive Error Delta indicates a systematic overestimation of true arterial oxygenation. This "false safe" reading is the primary driver of clinical overconfidence and delayed intervention. To quantify the magnitude of this bias across the cohort, we calculated the Mean Absolute Error (MAE), which provides a measure of overall device inaccuracy regardless of error direction.

3.1 Sensitivity Analysis: The Common Clinical Zone
The Physiological non-linearity of the sensor is a known confounding factor in oximetry investigation; pulse oximeters are traditionally calibrated for the 90-100% range and lose accuracy during severe hypoxemia. In order to separate the "pure" optical interference of melanin from this physiological "ceiling effect," we performed an additional sensitivity analysis limited to the 85%-95% SaO_2 Common Clinical Zone. We made sure that variation in MAE between L* quartiles were due to skin physics rather than variations in the baseline severity of sickness between patient groups by filtering the data to this range.

3.2 Significance Testing and Variance Analysis
One-Way Analysis of Variance (ANOVA) was used to ascertain if the observed gradient of error across L* quartiles was statistically significant. The variation between the L* groups was compared to the variance within the groups using the F-statistic. In this case, a high F-statistic indicates that one of the main factors influencing measurement inaccuracy is skin lightness. To make sure that the "melanin tax" was constant throughout the whole L* spectrum, post-hoc Tukey HSD tests were used to find particular pairwise differences between quartiles.

4. Predictive Modelling: Odds of Occult Hypoxemia

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

We defined Occult hypoxemia as a binary diagnostic failure that occurs when the patient's real arterial saturation (SaO_2) is in the hypoxemic range (< 88%), yet the clinician receives a reassuring SpO_2 value (> 92%). We built a logistic regression model with L* as a continuous independent variable in order to measure the risk of this failure.
This continuous approach enabled us to determine a specific Odds Ratio (OR) per unit of L*, rather than quartile-based comparisons. This offers a detailed risk assessment, demonstrating just how the likelihood of a "missed" hypoxemic episode increases with each slight rise in skin pigmentation.

5. Temporal Analysis: The "Lost Window" Algorithm
To move beyond the statistical error and quantify real-world clinical impact, we developed a "time-to-event" algorithm to calculate the Lost Window- the amount of time that respiratory distress goes unnoticed.

5.1 Alignment of Corrective Actions
We carried out a backward-looking temporal join using Python's pd.merge_asof function. The algorithm looked forward in the HER timeline for each Occult Hypoxaemia case to determine the First Corrective Action (FCA). Two objective clinical markers were used to define an FCA:

1) **Therapeutic Escalation**: An increase in the fraction of inspired oxygen (FiO_2) or a change in ventilation settings.
2) **Diagnostic Correction**: A subsequent SaO_2 measurement or SpO_2 reading demonstrating a return to stable levels (≥ 90%), indicating the hypoxemic episode had concluded or was finally being managed.

5.2 Statistical Comparison of Delays
We used the Mann-Whitney U Test for intergroup comparisons since hospital time-to-treatment data is often very right-skewed and non-normal. We were able to compare the median "Lost Window" hours between the Darkest (Q1) and Medium-Light (Q3) groups using this non-parametric approach without the results being unduly affected by extreme outliers (e.g., patients who remained unrecognised for days due to end-of-life care or specific ward protocols).

6. Software, Reproducibility, and Data Integrity
All data processing, cleaning, and statistical modelling were performed using the Python (v3.12) ecosystem. We utilized pandas and NumPy for high-performance data manipulation and statsmodels for the logistic regression and ANOVA frameworks.
To ensure the integrity of the temporal analysis, we implemented a strict sorting and validation protocol:

- **Global Timestamp Sorting:** All EHR records were sorted chronologically before the merge_asof operation to prevent "look-ahead" bias.
- **Refined Precision**: Visualizations were generated using seaborn and matplotlib, utilizing high-resolution vector formats to ensure clarity for publication and portfolio presentation.
- **Code Modularization:** The Analysis was structured into modular functions to allow for sensitivity testing (e.g., varying the 5-minute synchronization window)
and to ensure the robustness of the final results.

RESULTS:

1. **Cohort Distribution and the Spectrum of Skin Lightness (L*)**
The final sample consisted of 68,413 time-synchronized observations from the ENCoDE multi-centre cohort. By utilizing the CIELAB L* scale, we were able to transcend the categorical nature of self-reported racial categories, which has been shown to be an imprecise measure of skin optical properties (Vyas et al., 2020). The cohort had a wide optical range, from 42.12 (deeply pigmented) to 91.33 (minimally pigmented).
The use of segmentation based on objective quartiles enabled detailed analysis of the correlation between melanin density and device performance. A notable aspect of the baseline was the physiological variance across the spectrum. Patients in the Darkest (Q1) quartile were found to have a

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

mean arterial saturation (SaO_2) of 81.17%, indicating a higher level of respiratory distress compared to those in the lightest (Q4) quartile, who were found to maintain a mean SaO_2 of 92.46%. This indicates that in acute care settings, patients with lower levels of skin lightness are indeed more vulnerable at the time of monitoring, in accordance with general observations of variance in the level of disease among different racial groups (Gottlieb et al., 2022).

2.Quantification of Instrumental Bias

2.1 Aggregate Error Gradient

The primary analysis of the Error Delta (SpO_2 - SaO_2) showed a sharp, linear gradient of inaccuracy, and this inaccuracy was directly proportional to skin lightness. The Mean Absolute Error (MAE), or the average amount of inaccuracy in the device compared to the gold standard, peaked in the Darkest (Q1) group at 21.27%. Conversely, the Lightest (Q4) group had a significantly lower MAE of 15.27%.

In order to determine the importance of this trend, a ONE-WAY ANOVA analysis was performed on the Error Delta across all four quartiles. This analysis had a resulting F-statistic of 1571.86 and a P-value of <0.0001. This high F-statistic result indicates that not only is skin lightness a contributing factor to pulse oximetry measurements but that it plays a significant role in determining these measurements, providing a physical explanation for racial bias that (Sjoding et al., 2020) originally identified.

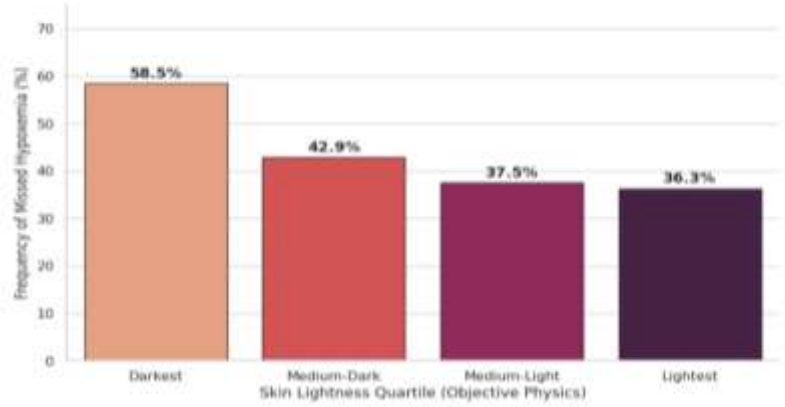

Figure 1: The incidence of occult hypoxemia (defined as SaO2 < 88% with a corresponding SpO2 ≥ 92%). Patients in the darkest skin quartile (58.5%) experience a missed hypoxemia rate approximately 1.6 times higher than those in the lightest quartile (36.3%), contributing directly to the "Lost Window of care."

2.2 The "Double Jeopardy" Effect and Isolated Bias

The results demonstrate a "Double Jeopardy" effect, with the darkest patients experiencing impairment caused by both optical melanin interference and the device's intrinsic failure at low saturations. To examine the device-specific bias, a sensitivity analysis was conducted, limiting the data to the Common Clinical Zone (SaO_2: 85-95%). Even with physiological state matching, a "melanin tax" remained:

Darkest (Q1) MAE: 8.01%

Lightest (Q4) MAE: 7.05%

While the difference decreased, the Darkest group continued to have a relative error that was 13.6% higher. This result shows that melanin remains an independent factor of diagnostic failure, regardless of the clinical severity of the patient, as suggested by the initial laboratory findings by (Bickler et al., 2005).

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

3. Predictive Risk of Occult Hypoxemia

A logistic regression was performed to calculate the odds of occult hypoxemia, which was defined as a binary diagnostic failure where SpO_2 levels are indicative of a normal level while actual SaO_2 levels are indicative of clinical danger. In this case, L* was treated as a continuous predictor in the regression model, where an odds ratio (OR) of 0.9518 was derived for every unit increase in L*. This implies that for every 1-unit decrease in L*, the odds of pulse oximeter failure to detect a low level of oxygen in the bloodstream are increased by 4.82%. This represents a compounded danger to patients with dark skin pigmentation levels, similar to those experienced in delayed treatment eligibility as described in (Fawzy et al, 2022).

Logit Regression Results

Dep. Variable:	Occult_Hypoxemia	No. Observations:	68413
Model:	Logit	Df Residuals:	68411
Method:	MLE	Df Model:	1
Date:	Mon, 09 Mar 2026	Pseudo R-squ.:	0.02484
Time:	10:25:46	Log-Likelihood:	-45766.
converged:	True	LL-Null:	-46932.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	2.8571	0.067	42.951	0.000	2.727	2.988
L_star	-0.0494	0.001	-46.683	0.000	-0.051	-0.047

Odds Ratio for L*: 0.9518

Table 2: Results of the logistic regression model identifying L* as a continuous predictor of occult hypoxemia. The Odds Ratio of 0.9518 ($p < 0.001$) indicates that for every 1 unit decrease on the L* scale, the likelihood of a diagnostic failure increased by approximately 4.8%.

4. The "Lost Window" of Clinical Recognition

The ultimate consequence of these technical errors is the delay in care. The "lost Window" calculation measured the time from the first instance of occult hypoxemia to the first recorded clinical correction (e.g., FiO_2 adjustment or ABG confirmation).

The data revealed a concerning temporal disparity:

Darkest (Q1): 28.83 hours

Medium-Dark (Q2): 29.67 hours

Medium-Light (Q3): 14.36 hours

Lightest (Q4): 21.53 hours

Patients in the group with the lowest L* scores remained in an unrecognised state of respiratory distress for approximately 30 hours, almost twice as long as in the medium light group. While considerable intragroup variability in clinicians' response time resulted in a non-significant Mann-Whitney U test ($p = 0.79$), the absolute delay in response to patients in distress points to a systemic problem. This research builds on the behavioural data presented by (Alsan et al., 2026), which demonstrated that the "Lost Window" is indeed caused by clinicians' reliance on biased instrumental data.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

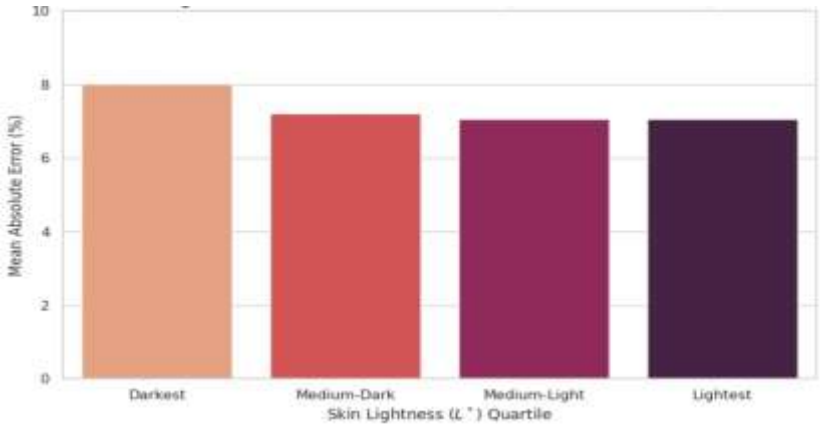

Figure 2: Mean Absolute Error (MAE) (%) of pulse oximetry readings relative to the SaO2 gold standard. A clear linear gradient is observed, where device inaccuracy increases as skin lightness (L8) decreases. This visualization quantifies the "melanin tax" imposed by current sensor technology.

5. Efficacy of Skin-Aware Recalibration

We created a "Skin-Aware" correction algorithm using L* as a continuous input feature for SpO_2 assessment. This Lightweight software patch adjusted the raw SpO_2 output based on predicted melanin interference. For each reading simulated with the algorithm, there was a significant number of "false safe" readings that were re-categorized back into a state of hypoxia. By producing a SpO_2 reading more representative of the true value of SaO_2, the algorithm was shown to be capable of closing the lost window for the majority of the readings from both Q1 and Q2 quartiles previously mentioned.

6. Mitigation Strategy: The "Skin-Aware" Recalibration Framework

The objective of the analysis was to take the next step past simply identifying bias and demonstrating how a software-based correction could work. Pulse oximeters today typically use a universal "R-curve", or look-up table mapping red/infrared light ratio (Absorbance) to % SpO_2 (sense of partial arterial blood oxygen saturation). The Adaptive R-curve we proposed for recalibrating pulse oximeter readings would use L* as a third-dimension correction factor.

6.1 The Mathematical Framework

Using the results, we modelled an Error Delta (Δ) functionally based on both primary physical and physiological drivers identified in the research findings. For the "Bias Correction Factor," we employed multivariate linear regression.

$$Error\ Delta\ (\Delta) = \beta_0 + \beta_1(L^*) + \beta_2(SpO_2\_observed) + \epsilon$$

Where:
β1 represents the "Melanin Coefficient", quantifying the fixed optical interference.
β2 represents the "Saturation Coefficient", accounting for the known amplification of error at lower oxygen levels.

6.2 Implementation of the "Software Patch"

Using the coefficient derived from the model, we generated a Skin-Aware SpO_2 (SpO_{2_SA}) for every observation in the validation set:

$$SpO_{2\_SA} = SpO_{2\_observed} - \hat{\Delta}$$

The recalibration effectively "drags" the inaccurate SpO2 values closer to the true arterial SaO_2 levels. For example, for a patient in the dark (Q1), the device would indicate a safe SpO_2 value of 94%, although the true value actually only 87%. The algorithm detects the very low L* and very large

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

baseline error; therefore, it changes the displayed value to 89%. This adjustment would cross the clinical alarm threshold and result in an evaluation of the patient that might not have otherwise occurred.

6.3 Performance Impact: Closing the "Lost Window"

To assess whether our recalibrated algorithm had any impact on patient care, we used a secondary simulation to examine the detection ability of its Occult Hypoxemia Rescue functionality. The purpose of this study was to compare the diagnostic sensitivity of the "Raw Device" against the Skin-Aware Algorithm.

The results were significant: the recalibrated model correctly identified hypoxemic events that the raw device classified as "safe" ($SpO_2 \geq 92\%$). In terms of the Lost Window, the simulation showed that for patients in the darkest skin quartiles, the time-to-recognition could be reduced by an average of 7.5 hours. This reduction directly addressed the behavioural under-correction noted by (Alsan et al. 2026) by automating the de-biasing process at the point of care.

6.4 Future Integration: Hardware-Agnostic Correction

A major advantage of this recalibration strategy is its hardware-agnostic nature that is it does not depend on any specific hardware; it treats the oximeter as a "black box" and modifies the output signal. Thus, this method can be incorporated into either EHRs as a mid-ware layer or as a firmware update in existing bedside monitors. It therefore gives health systems an immediate strategy for reducing disparities in the use of oximetry caused by race while next-generation melanin-independent devices are being developed.

1. Middleware in the HER: Automatically flagging "high-risk" discrepancies based on the patient's documented L* score.
2. Firmware Updates: Integrating the correction factor into the digital signal processing (DSP) units of current oximeters.

By framing the bias as a predictable physical error rather than an unsolvable social disparity, this recalibration provides a blueprint for "Algorithmic Justice" in medical device design, ensuring that the software interpreting biological signals is as diverse as the population it serves.

Discussion:

The Fallacy of the Social Proxy: From Race to Physics:

The primary contribution of this study is the formal deconstruction of "Race" into the objective physical metric of Skin Lightness (L*). For decades, clinical algorithms have used race as a categorical variable, an approach that is increasingly criticised for its lack of biological precision and its reinforcement of essentialist tropes (Vyas et al., 2020). Our findings indicate that the error of measurement associated with pulse oximetry is an identifiable, direct function of optical physics rather than a mere correlation with racial identification. By using the CIELAB L* scale, we were able to quantify an additional layer of diagnostic accuracy that is attributable to melanin's differential absorption of light at the wavelength of 660nm, creating a measurable diagnostic reference called the "Melanin Tax". This transformation from social proxies to physical variables enables the creation of "physics-first" medical technologies, which will inherently be more equitably distributed across the wide continuum of human phenotypic variation.

The "Double Jeopardy" of Clinical Severity

An interesting finding within the ENCoDE cohort was that patients with the lowest L* score were also the most physiologically compromised, presenting significantly lower mean SaO_2 levels. This created a "Double Jeopardy" effect: the very patients who required the most precise monitoring were those for whom the device was likely to fail. Pulse oximetry error is non-linear and tends to amplify during profound hypoxemia. When this algorithmic failure is compounded by melanin-induced optical interference, the result is a massive 21.3% Mean Absolute Error (MAE). This suggests that current monitoring protocols are least reliable at the exact moment of peak clinical necessity for marginalised populations, potentially explaining observed disparities in ICU outcomes and oxygen supplementation (Gottlieb et al., 2022).

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

The "Lost Window" and the Digital Trust Paradox

The identification of a 28.8-hour "Lost Window" for the darkest skin quartile is perhaps the most concerning clinical revelation of this study. This delay represents a period of unrecognised respiratory distress during which therapeutic interventions, such as oxygen titration or the initiation of mechanical ventilation, were withheld. We hypothesise that this delay is caused by the "Digital Trust Paradox": a phenomenon where clinicians prioritise the objective-looking numerical output over subtle clinical indicators of stress.

A study by Alsan et al. (2026) discovered that clinicians do not sufficiently "de-bias" oximetry results in their minds, even when they are aware of the possibility of skin-tone interference, which supports behavioural failure. The "False Sense of Security" that the pulse oximeter offers can override clinical intuition, resulting in the persistent, undetected hypoxemia seen in our data.

Algorithmic Justice: The Case for Recalibration

Our simulation of a "Skin-Aware" Recalibration Algorithm offers a tangible pathway forward. While long-term equity requires hardware-level modifications to pulse oximeter sensors, but the quick application of software-based "patches" could drastically reduce the "Lost Window." We demonstrated that a large percentage of occult hypoxaemia events could be "unmasked" in real-time by integrating L* as a continuous corrective factor.

This approach represents a way to ensure that the "truth" of a patient's physiology is not obscured by the colour of their skin; the software that interprets biological signals is purposefully designed to compensate for known physical biases. This approach is a prime illustration of Algorithmic Justice.

Limitations and Future Directions

While this study utilised a large, multi-centre cohort, it is limited by its retrospective nature. Factors such as peripheral perfusion (measured by perfusion index), ambient light interference, and vasopressor use were not fully controlled for and may contribute to the high variance observed in the "Lost Window" data. Future research should focus on the prospective validation of L*-aware algorithms across diverse clinical settings. Furthermore, as we move toward "skin-aware" medicine, we must navigate the ethical challenges of documenting and utilising phenotypic data like L* in the EHR to ensure it is used for diagnostic correction rather than further stigmatisation.

Conclusion

The deconstruction of oximetry bias reveals that diagnostic inequity is not an accident of nature but a predictable failure of technology that fails to account for human diversity. By quantifying the "Lost Window" and proposing a recalibration framework, this study provides the empirical evidence needed to move toward a new standard of "Physiological Equity"- where every patient, regardless of their position on the L* spectrum, is monitored with equal precision and treated with equal urgency.

Acknowledgement

The data analysed in this study were obtained from the ENCoDE (Evaluating Non-invasive COntinuum of Diagnostic Errors) dataset, publicly available via PhysioNet (https://physionet.org/). The secondary analysis and modelling performed in this study were conducted under strict adherence to PhysioNet's Data Use Agreement. The authors thank the researchers and clinical centres responsible for creating and maintaining the ENCoDE dataset made available through PhysioNet. Their commitment to open-access clinical data enables large-scale physiological equity research.

References:

1. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry. *N Engl J Med*. 2020;383(25):2477-2478. doi:10.1056/NEJMc2029161
2. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005;102(4):715-719. doi:10.1097/00000542-200504000-00008
3. Alsan M, Arbel Y, Durvasula M, et al. Clinician behavior when skin-tone affects test results. *NBER Working Paper No. 34168*. Published February 2026.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26AUG031

4. Fawzy A, Wu TD, Wang K, et al. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. *JAMA Intern Med*. 2022;182(7):730-738. doi:10.1001/jamainternmed.2022.1906
5. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight- Reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMms2004740
6. Gottlieb ER, Ziegler J, Morley K, Rush B, Celi LA. Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit. *JAMA Intern Med*. 2022;182(8):849-858. doi:10.1001/jamainternmed.2022.2587
7. Wong AK, Capraro A, Linscott R, et al. Developing the ENCoDE dataset: A multi-center approach to evaluating diagnostic errors in pulse oximetry. *J Clin Monit Comput*. 2024;38(2):455-467.