

Effects of Skin Pigmentation on Pulse Oximeter Accuracy at Low Saturation

Philip E. Bickler, M.D., Ph.D.,* John R. Feiner, M.D.,† John W. Severinghaus, M.D.‡

Background: It is uncertain whether skin pigmentation affects pulse oximeter accuracy at low HbO₂ saturation.

Methods: The accuracy of finger pulse oximeters during stable, plateau levels of arterial oxygen saturation (Sao₂) between 60 and 100% were evaluated in 11 subjects with darkly pigmented skin and in 10 with light skin pigmentation. Oximeters tested were the Nellcor N-595 with the OxiMax-A probe (Nellcor Inc., Pleasanton, CA), the Novamatrix 513 (Novamatrix Inc., Wallingford, CT), and the Nonin Onyx (Nonin Inc., Plymouth, MN). Semisupine subjects breathed air–nitrogen–carbon dioxide mixtures through a mouthpiece. A computer used end-tidal oxygen and carbon dioxide concentrations determined by mass spectrometry to estimate breath-by-breath Sao₂, from which an operator adjusted inspired gas to rapidly achieve 2- to 3-min stable plateaus of desaturation. Comparisons of oxygen saturation measured by pulse oximetry (SpO₂) with Sao₂ (by Radiometer OSM3) were used in a multivariate model to determine the interrelation between saturation, skin pigmentation, and oximeter bias (SpO₂ - Sao₂).

Results: At 60–70% Sao₂, SpO₂ (mean of three oximeters) overestimated Sao₂ (bias ± SD) by 3.56 ± 2.45% (n = 29) in darkly pigmented subjects, compared with 0.37 ± 3.20% (n = 58) in lightly pigmented subjects (P < 0.0001). The SD of bias was not greater with dark than light skin. The dark–light skin differences at 60–70% Sao₂ were 2.35% (Nonin), 3.38% (Novamatrix), and 4.30% (Nellcor). Skin pigment–related differences were significant with Nonin below 70% Sao₂, with Novamatrix below 90%, and with Nellcor at all ranges. Pigment-related bias increased approximately in proportion to desaturation.

Conclusions: The three tested pulse oximeters overestimated arterial oxygen saturation during hypoxia in dark-skinned individuals.

PULSE oximetry theoretically can compute arterial hemoglobin oxygen saturation from the ratio of the pulsatile to the total transmitted red light divided by the same ratio for infrared light transilluminating a finger, ear, or other tissue. The derived saturation should be independent of skin pigmentation, and many other variables, such as hemoglobin concentration, nail polish, dirt, and jaundice. Several large controlled studies comparing black and white patients (380 subjects)^{1,2} reported no

significant pigment-related errors in pulse oximeters at normal saturation.

However, Severinghaus and Kelleher³ reviewed data from several investigators who had reported anecdotal errors (+3 to +5%) in black patients.^{4–7} Model simulations of errors due to various pigments were reviewed by Ralston *et al.*⁸ Cote *et al.*⁹ reported that nail polish and ink on skin surface can cause errors, a finding confirmed anecdotally by others from fingerprinting ink,¹⁰ henna,¹¹ and meconium.¹² Intravenously injected dyes cause transient errors.¹³ Lee *et al.*¹⁴ found overestimation of saturation, especially at low saturation in pigmented patients (Indian, Malay *vs.* Chinese). The Technology Subcommittee of the Working Group on Critical Care, Ontario Ministry of Health,¹⁵ reported unacceptable errors in pulse oximetry at low saturation in pigmented subjects. Zeballos and Weisman¹⁶ compared the accuracy of the Hewlett-Packard (Sunnyvale, CA) ear oximeter and the Biox II pulse oximeter (Ohmeda, Andover, MA) in 33 young black men exercising at three different simulated altitudes. At an altitude of 4,000 m, where arterial oxygen saturation (Sao₂) ranged from 75 to 84%, the Hewlett-Packard underestimated Sao₂ by 4.8 ± 1.6%, whereas the Biox overestimated Sao₂ by 9.8 ± 1.8% (n = 22). It was stated that these errors, previously reported in whites, were both exaggerated in blacks.

During our many years of testing pulse oximeter accuracy at oxygen saturations as low as 50%, we have occasionally noted unusually high positive bias, particularly at very low saturation levels, in some but not in other deeply pigmented subjects. This investigation was therefore specifically designed to determine whether errors at low Sao₂ correlate with skin color.

All pulse oximeters marketed in the United States are required by the US Food and Drug Administration to have been tested and to be certified as accurate to less than ±3% root mean square error at Sao₂ values between 70 and 100%. The great majority of calibration and confirmation tests have been conducted in volunteer subjects with light skin pigmentation.

The Food and Drug Administration has recently suggested that studies of pulse oximeter accuracy submitted for Food and Drug Administration device approval include subjects with a range of skin pigmentations, although no quantitative requirement has been distributed. We are aware of no data that support this action.

If there is a significant and reproducible positive bias at low saturation in dark-skinned subjects, inclusion of

* Professor, † Associate Professor, ‡ Professor Emeritus.

Received from the Department of Anesthesia and Perioperative Care, University of California at San Francisco, San Francisco, California. Submitted for publication July 6, 2004. Accepted for publication November 10, 2004. Supported by a UCSF Department of Anesthesia research fund generated from the clinical testing of pulse oximeters. No manufacturers were involved in any part of the concept, design or analysis of this study.

Address reprint requests to Dr. Bickler: Sciences 255, Box 0542, University of California Medical Center, 513 Parnassus Avenue, San Francisco, California 94143-0542. Address electronic mail to: bicklerp@anesthesia.ucsf.edu. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

dark-skinned subjects will increase test group mean root mean square errors, perhaps enough to cause rejection by the Food and Drug Administration. If a reproducible bias is found at low saturation in dark-skinned subjects in all pulse oximeters, warning labels should be provided to users, possibly with suggested correction factors.

Materials and Methods

This study was approved by the University of California at San Francisco Committee on Human Research, and informed consent was obtained from all subjects. Twenty-one healthy, nonsmoking subjects were studied. None of the subjects had lung disease, obesity, or cardiovascular problems. They ranged in age between 24 and 47 yr. Eleven subjects were very darkly pigmented individuals of African-American descent, and the remaining 10 were light-skinned individuals of northern European descent. No Asians or Hispanics were included in the study. The average age of the light-skinned subjects was 28 yr, and that of the dark-skinned subjects was 30 yr.

Subjects were studied semisupine (30° head up) with a nose clip while deliberately hyperventilating air-nitrogen-carbon dioxide mixtures *via* a mouthpiece from a partial rebreathing circuit with 10- to 20-l/min fresh gas inflow. An indwelling 22-gauge radial artery catheter was placed to facilitate arterial blood sampling for measurements of SaO_2 . Five oximeters were mounted on each subject: one Nellcor N-595 with the OxiMax A finger probe (Nellcor Inc., Pleasanton, CA), two model 513s from Novametrix Inc. (Wallingford, CT), and two Onyx models from Nonin Inc. (Plymouth, MN). The Novametrix and Nonin instrument readings were recorded manually. Oxygen saturation measured by pulse oximetry (SpO_2) from the Nellcor oximeters, the end-tidal gases, and estimated SaO_2 were recorded by a computer running LabVIEW 6.0 (National Instruments, Austin, TX). Data from the two Novametrix and two Onyx instruments were averaged separately. Probes were not mounted on the thumbs or little fingers. The Nellcor oximeter was tested in all subjects, and the Novametrix and Nonin instruments were tested in nine light- and seven dark-skinned subjects each.

A series of 10–12 stable target SaO_2 plateaus between 60 and 100% were sought by an operator who adjusted the inspired air-nitrogen-carbon dioxide mixture breath by breath in response to an analog meter displaying the estimated SaO_2 derived from mass spectrometer end-tidal gas analysis.¹⁷ Input parameters for the computer oxygen dissociation curve included arterial pH estimated from end-tidal partial pressure of carbon dioxide, base excess adjusted if needed after each arterial sample was analyzed, and alveolar-arterial oxygen difference as needed especially at low SaO_2 to attempt to match the predicted with the measured SaO_2 . At each

level, arterial blood was sampled after a plateau of 30–60 s had been achieved, followed by a second sample at the same plateau 30 s later. To insure that each subject had good circulation to the fingers, each hand was wrapped in a warming pad. “Functional” arterial SaO_2 ($HbO_2/(hemoglobin + HbO_2)$) was determined by multiwavelength oximetry (Radiometer OSM-3, Copenhagen, Denmark). Quality control standards were run each day.

Statistical Analysis

Bias is computed as SpO_2 minus SaO_2 from the reading of each oximeter minus the corresponding blood sample value. Bias is reported as mean \pm SD. The relation of bias to SaO_2 was analyzed by linear regression. Bias was also analyzed for decadal subgroups of SaO_2 (60–70, 70–80, 80–90, and 90–100%) using analysis of variance. Differences between oximeters were analyzed by analysis of variance and the Tukey-Kramer multiple comparison technique. Multivariate models were used to analyze the interrelations of bias with skin pigment, oximeter, and SaO_2 . $P < 0.05$ was considered statistically significant. Statistical analysis was performed with JMP 4.0 (SAS Institute, Cary, NC).

Results

Table 1 presents all of the data. The mean of all 1,067 data points at all SaO_2 levels and all three types of oximeters indicated that SpO_2 reads approximately 1% higher in dark- than in light-skinned subjects ($P < 0.0001$). These overall mean bias errors due to pigment were +0.4% with Nonin, +0.6% with Novametrix, and +1.6% with Nellcor.

At lower oxyhemoglobin saturation, greater differences in bias between light- and dark-skinned subjects were apparent. With all three instruments, the effect of skin pigment on bias increased approximately linearly as SaO_2 decreased (fig. 1). In the range of 60–70% SaO_2 , the mean difference in bias between light- and dark-skinned subjects was +1.4% (Nonin), +4.4% (Novametrix), and +4.3% (Nellcor) and 3.2% for all three instruments combined. Dark skin pigment-related bias was statistically significant with Nellcor in all SaO_2 decades and with Novametrix between 60 and 80%, but only in the 70–80% range with Nonin. Multivariate models analyzing bias with respect to skin pigment, SaO_2 (either as a continuous variable or by decadal range), and oximeters showed statistically significant relations for all of the variables.

In figures 2A–C, the data obtained with each instrument are plotted against SaO_2 , using filled circles for dark-skinned subjects. The slopes of the regression lines differ most between dark and light skin with Nellcor and least with Nonin. Compared with light-skinned subjects, the mean effect of desaturation on bias with dark skin is

Table 1. Bias, Sp_o₂ minus Sa_o₂, for Three Oximeters

Oximeter	Skin	60–70%	70–80%	80–90%	90–100%	All (50–100%)
Nonin Onyx	All	2.32 ± 1.84 (27)	0.67 ± 1.13 (76)	−0.59 ± 0.93 (112)*	−0.71 ± 1.08 (103)†	−0.08 ± 1.45 (319)
	Light	1.91 ± 1.60 (19)	0.41 ± 1.15 (49)	−0.70 ± 0.88 (65)	−0.83 ± 1.20 (66)	−0.21 ± 1.42 (200)
	Dark	3.29 ± 2.12 (8)	1.14 ± 0.96 (27)	−0.44 ± 0.99 (47)	−0.50 ± 0.79 (37)	0.15 ± 1.48 (119)
	P value	0.07	0.005	NS	NS	0.03
Novametrix 513	All	2.22 ± 3.47 (27)	2.66 ± 1.80 (76)†	1.37 ± 1.35 (112)*	0.38 ± 1.02 (103)	1.40 ± 1.94 (319)†
	Light	0.93 ± 3.20 (19)	2.13 ± 1.64 (49)	1.33 ± 1.16 (65)	0.43 ± 0.99 (66)	1.15 ± 1.79 (200)
	Dark	5.29 ± 1.74 (8)	3.63 ± 1.69 (27)	1.42 ± 1.58 (47)	0.28 ± 1.09 (37)	1.83 ± 2.12 (119)
	P value	0.001	<0.001	NS	NS	0.002
Nellcor N-595	All	0.07 ± 3.74 (33)†	0.70 ± 2.64 (113)	0.86 ± 1.52 (142)*	0.15 ± 1.28 (140)	0.49 ± 2.18 (429)
	Light	−1.62 ± 3.42 (20)	−0.34 ± 2.26 (54)	0.24 ± 1.25 (73)	−0.16 ± 0.98 (72)	−0.27 ± 2.05 (220)
	Dark	2.68 ± 2.58 (13)	1.66 ± 2.62 (59)	1.52 ± 1.51 (69)	0.48 ± 1.48 (68)	1.29 ± 2.03 (209)
	P value	0.0005	<0.0001	<0.0001	0.003	<0.0001
Combined	All	1.44 ± 3.32 (87)	1.26 ± 2.24 (265)	0.57 ± 1.53 (366)	−0.04 ± 1.23 (346)	0.59 ± 2.0 (1067)
	Light	0.37 ± 3.20 (58)	0.70 ± 2.03 (152)	0.29 ± 1.38 (203)	−0.18 ± 1.17 (204)	0.21 ± 1.89 (620)
	Dark	3.56 ± 2.45 (29)	2.01 ± 2.30 (113)	0.93 ± 1.64 (163)	0.17 ± 1.29 (142)	1.13 ± 2.03 (447)
	P value	<0.0001	<0.0001	<0.0001	<0.01	<0.0001

Data are presented as mean ± SD (n). P values are for differences between light- and dark-skinned individuals in specified range of arterial oxygen saturation (Sa_o₂).

* All oximeters are significantly different, P < 0.05. † Indicated oximeter is statistically different from other two instruments, P < 0.05.

NS = not significant; Sp_o₂ = oxygen saturation measured by pulse oximetry.

$\Delta\text{Sp}_{o_2} = \Omega(100 - \text{Sa}_{o_2})$, where Ω is 0.064 with Nonin, 0.123 with Novametrix, and 0.125 with Nellcor. Ω could be considered as a correction factor for clinical use to correct (reduce) low Sp_o₂ values in dark-skinned patients. However, the linear equation would overcorrect data at intermediate Sa_o₂ with Nonin.

The scatter of data increases at low Sa_o₂ as shown in table 1. The SD averages approximately 1.2 in all instruments at high Sa_o₂ but increases in the 60–70% range to 2.5 (Nonin), 3.0 (Novametrix), and 3.7 (Nellcor). However, the SD is not greater with dark than light skin at low Sa_o₂.

We found small but statistically significant differences between oximeters. When data from light- and dark-skinned subjects were combined, the mean bias of Novametrix (1.40 ± 1.94%) was statistically higher than that of Nellcor (0.49 ± 2.18) and Nonin (0.08 ± 1.45%). Small differences were also found within decadal Sa_o₂ intervals.

Discussion

Pulse oximetry theory predicts that the ratio of the pulsatile to the total transmitted red light divided by the same ratio for infrared light should be dependent only on arterial saturation. In practice, there are many minor deviations from this ideal. Major known variability is caused by anemia, light scattering, venous and tissue pulsation by mechanical force from nearby arteries, pulsatile variations in tissue thickness in the light path other than in the arteries, nail polish, and skin pigment.³ Therefore, pulse oximeter design is often based on empirically determined correction factors obtained by *in vivo* comparison of oximeter readings with arterial oxyhemoglobin saturation (Sa_o₂) of volunteer subjects. In

our 18 yr of testing pulse oximeter accuracy, and probably in other testing laboratories, the majority of subjects have been light skinned. Most pulse oximeters have probably been calibrated using light-skinned individuals, with the assumption that skin pigment does not matter. The current data show that skin pigment introduces a consistent positive bias at low Sa_o₂ in the Nonin, Novametrix, and Nellcor instruments. We infer that most pulse oximeters show similar pigment-related bias at low saturation. A previous study of pulse oximeter accuracy in dark-skinned individuals only measured bias at room air saturation levels and therefore concluded that oximeter accuracy is independent of skin color.²

We chose to study extremes of skin pigment, dark and light, to most easily detect an effect with a small number of subjects. As shown in figures 2A–C, we found a bias ranging from −4 to +8% in dark-skinned subjects and a slightly smaller range in light-skinned subjects. However, as noted above, the SD was not found to be greater in dark-skinned subjects at any Sa_o₂ range. This implies that the bias error does not arise simply from inadequate signal intensity and confirms the problem as a real pigment-related difference in pulse oximeter optical factors, which deserves attention and possible provision of correction factors, tables, or even built-in user-optional adjustments. Furthermore, the larger bias in oximeter readings at low saturation did not arise because of data from a few subjects. Scatter plots of data from separate subjects indicate some intersubject variability but no obvious outliers.

Because we tested only three types of oximeters, our results may not apply to pulse oximeters made by other manufacturers. However, all three instruments tested showed increasing positive bias in oximeter readings at low Sa_o₂ in subjects with dark skin pigmentation. This may be a property common to a wide variety of modern pulse

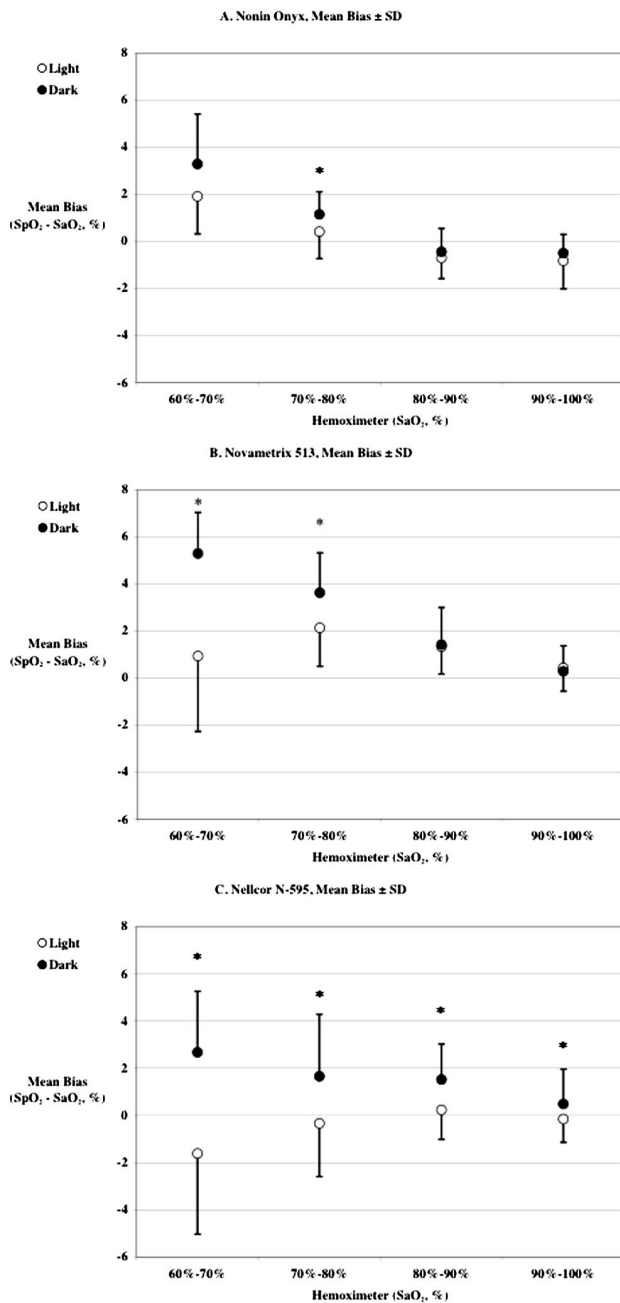


Fig. 1. Mean bias \pm SD for the three oximeters in different ranges of oxyhemoglobin saturation. Bias is calculated as SpO₂ (oximeter measured value of oxyhemoglobin saturation) minus SaO₂ (oxyhemoglobin saturation measured by a Radiometer OSM-3 multiwavelength oximeter). Light-skinned subjects are indicated by open circles, dark-skinned subjects by closed circles. (A) Nonin Onyx: Bias values for light-skinned and dark-skinned subjects were significantly different only in the oxyhemoglobin saturation range of 70–80% (* $P < 0.05$). (B) Novamatrix 513: Bias values for light-skinned and dark-skinned subjects were significantly different in the oxyhemoglobin saturation ranges of 60–70% and 70–80% (* $P < 0.01$). (C) Nellcor N-595: Bias values for light-skinned and dark-skinned subjects were significantly different in all oxyhemoglobin saturation ranges (* $P < 0.01$).

oximeters, given that the basic technology is similar. In addition, the mean differences in readings between light- and dark-skinned individuals were larger than the differ-

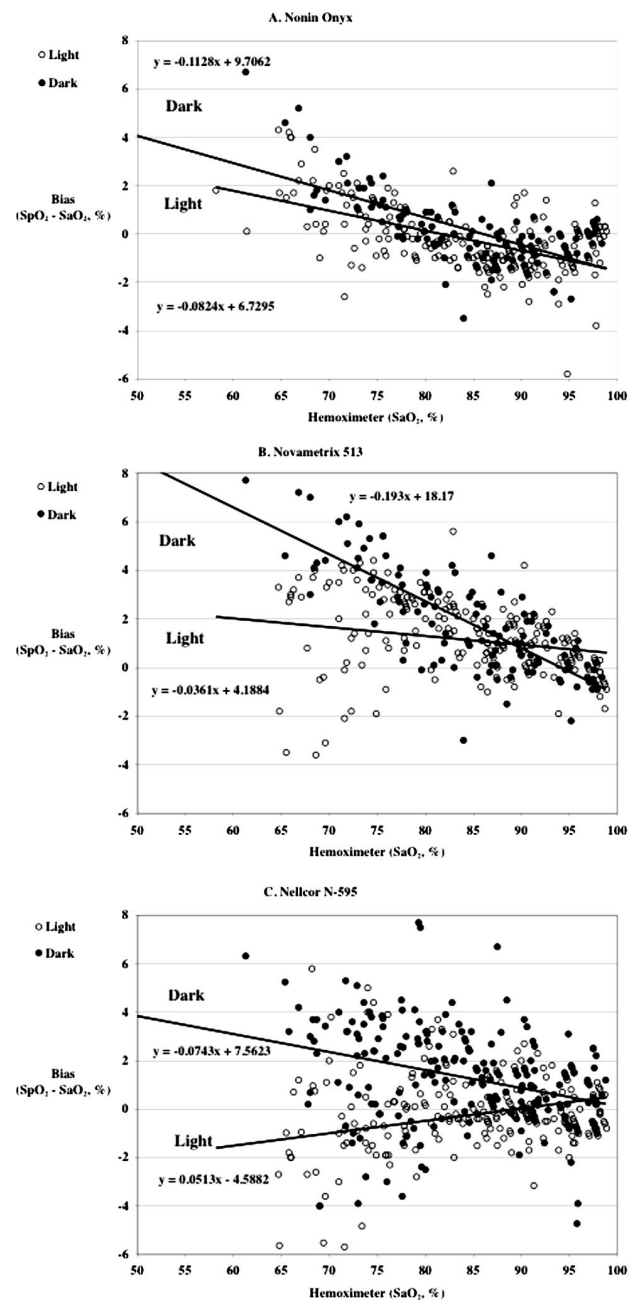


Fig. 2. Bias for the three oximeters is plotted versus oxyhemoglobin saturation, SaO₂, measured by a Radiometer OSM-3 multiwavelength oximeter (hemoximeter). Bias is calculated as SpO₂ (oximeter measured value of oxyhemoglobin saturation) minus SaO₂. Light-skinned subjects are indicated by open circles, dark-skinned subjects by closed circles. (A) Nonin Onyx, (B) Novamatrix 513, (C) Nellcor N-595. Linear regressions were performed separately for light-skinned and dark-skinned subjects and are statistically significant ($P < 0.001$). Multivariate analysis showed statistically significant effects for both skin pigment and oxyhemoglobin saturation ($P < 0.01$).

ences between the different oximeters. The results of this study call for further investigation of the effect of skin color on oximeter accuracy at low SaO₂ in other oximeter types. It is also possible that fingers have a different bias than ears or other monitoring sites.

The magnitude of the oximeter error in all three oximeters due to dark skin pigmentation is relatively small at saturations greater than 80% and probably of no clinical significance. However, in individuals with darkly pigmented skin, bias of up to 8% was observed at lower saturations, which may be quite significant under some circumstances. For example, in congenital heart disease, many patients have stable low saturation values, and accuracy in an outpatient setting or during surgery may be desired. In high-altitude medicine, oximeters are frequently relied on for accurate readings in both research and clinical settings.

In conclusion, dark skin pigmentation results in overestimation of arterial oxygen saturation, especially at low saturation in the three tested pulse oximeters. A notice warning of this effect may currently be the best available action.

The authors thank Renchell J. Andres, B.S. (Medical Student, University of California at San Francisco, San Francisco, California), for help with recruiting subjects and data collection.

References

1. Adler JN, Hughes LA, Vivilecchia R, Camargo CAJ: Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Acad Emerg Med* 1998; 5:965-70
2. Bothma PA, Joynt GM, Lipman J, Hon H, Mathala B, Scribante J, Kromberg J: Accuracy of pulse oximetry in pigmented patients. *S Afr Med J* 1996; 86:594-6
3. Severinghaus JW, Kelleher JF: Recent developments in pulse oximetry. *ANESTHESIOLOGY* 1992; 76:1018-38
4. Jubran A, Tobin MJ: Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. *Chest* 1990; 97:1420-5
5. Cahan C, Decker MJ, Hoekje PL, Strohl KP: Agreement between noninvasive oximetric values for oxygen saturation. *Chest* 1990; 97:814-9
6. Ries AL, Prewitt LM, Johnson JJ: Skin color and ear oximetry. *Chest* 1989; 96:287-90
7. Emery JR: Skin pigmentation as an influence on the accuracy of pulse oximetry. *J Perinatol* 1987; 7:329-30
8. Ralston AC, Webb RK, Runciman WB: Potential errors in pulse oximetry: III. Effects of interferences, dyes, dyshaemoglobins and other pigments. *Anaesthesia* 1991; 46:291-5
9. Coté CJ, Goldstein EA, Fuchsman WH, Hoaglin DC: The effect of nail polish on pulse oximetry. *Anesth Analg* 1988; 67:683-6
10. Battito MF: The effect of fingerprinting ink on pulse oximetry. *Anesth Analg* 1989; 69:265
11. Goucke R: Hazards of henna. *Anesth Analg* 1989; 69:416-7
12. Johnson N, Johnson VA, Bannister J, McNamara H: The effect of meconium on neonatal and fetal reflectance pulse oximetry. *J Perinat Med* 1990; 18:351-5
13. Scheller MS, Unger RJ, Kelner MJ: Effects of intravenously administered dyes on pulse oximetry readings. *ANESTHESIOLOGY* 1986; 65:550-2
14. Lee KH, Hui KP, Tan WC, Lim TK: Factors influencing pulse oximetry as compared to functional arterial saturation in multi-ethnic Singapore. *Singapore Med J* 1993; 34:385-7
15. Noninvasive blood gas monitoring: a review for use in the adult critical care unit. Technology Subcommittee of the Working Group on Critical Care, Ontario Ministry of Health. *Can Med Assoc J* 1992; 146:703-12
16. Zeballos RJ, Weisman IM: Reliability of noninvasive oximetry in black subjects during exercise and hypoxia. *Am Rev Respir Dis* 1991; 144:1240-4
17. Severinghaus JW, Naifeh KH, Koh SO: Errors in 14 pulse oximeters during profound hypoxia. *J Clin Monit* 1989; 5:72-81