Reliability of Six Pulse Oximeters in Chronic Obstructive Pulmonary Disease*

Bernard Hannhart, M.D.;† Hervé Michalski, M.D.;‡ Nicolas Delorme, M.D.;‡ Gilberto Chapparo, M.D.;‡ and Jean-Marie Polu, M.D., F.C.C.P.‡

Six pulse oximeters with finger probes were studied in three groups of 17 hypoxemic patients with COPD aged 50 to 75 years. Transcutaneous arterial oxygen saturation (SpO₂) was measured with the Nellcor N101 (oximeter 1a), the Ohmeda Biox III (oximeter 1b), the Nellcor N200 (oximeter 2a), the Critikon Oxyslutter (oximeter 2b), the Radiometer Oxil100 (oximeter 3a), and the Ohmeda Biox 3700 (oximeter 3b). The SpO₂ was compared with SaO₂ measured in simultaneously withdrawn samples of arterial blood (Radiometer OSM2) at three 20-minute steady-state levels of FIO₂ ranging from 0.21 to 0.40 (SaO₂, 62 to 100 percent). The bias (mean SpO₂-SaO₂ difference) and the error in precision (SD of the differences) were both below 4 percent for instruments 1a and 1b and remained below 1.2 and 3 percent, respectively, for the others. A good agreement between SpO₂ and SaO₂, as reflected by the Bartko intraclass correlation coefficient, was observed in instruments 2a, 3a, and 3b. The individual relationships between SpO₂-SaO₂ differences and SaO₂ appeared to be linear and parallel. With four instruments (1a, 1b, 2a, and 2b), the mean slope of this relationship was negative, showing a systematic instrumental error: the lower the SaO₂, the larger the overestimation of SaO₂. The scattering of the data (precision) principally reflects a subject source of error. In most instruments a technical adjustment could greatly improve instrumental errors and accuracy. The correction of the errors due to between-subject variation would require a system of calibration adjustable by the users to each individual.

\[ \text{SpO}_2 = \text{transcutaneous arterial oxygen saturation} \]

Recently, new microprocessor calculation techniques involved in oximetry have led to the development of the pulse oximeter¹ to measure SpO₂. These instruments combine both principles of spectrophotometry and plethysmography of the pulse amplitude. They monitor the transmission of two wavelengths of light from two emitting diodes (660 and 940 nm) to a receiver cell through vascular beds. The SpO₂ is calculated according to the molecular extinction coefficient of Beer's law by measuring the changes in hemoglobin light absorption from pulse to pulse. The new oximeters have a lightweight finger, ear, or nasal probe, are relatively inexpensive, and are easy to use in clinical practice; however, the studies on the accuracy of SpO₂ readings, comparing transcutaneous oximetric measurements and arterial blood values of SaO₂, are controversial: the good agreement with an accuracy of ±3 to 5 percent reported by some authors²⁻⁵ is not always found by others.⁶⁻¹⁰ Moreover, the results obtained in healthy subjects are not necessarily applicable in hypoxemic patients. The aim of the present study was to compare SpO₂ values measured with six transcutaneous pulse oximeters with SaO₂ measured in an arterial blood sample withdrawn simultaneously (spectrohemoximeter) in patients with COPD who were chronically hypoxemic while breathing room air. Contrary to most of the previous reports in the literature, the instruments were tested not while the patients were made more hypoxemic but while they received an increasing inspiratory level of oxygen. The present comparative study focused on the individual relationships between SpO₂-SaO₂ differences and SaO₂ measured in each patient at three different levels of inspired FIO₂. Thus, the various sources of errors, instrumental or individual, can be differentiated. Finally, some solutions are proposed to improve the reliability in estimation of SaO₂ with pulse oximeters.

**MATERIALS AND METHODS**

This study was performed in 51 white patients with COPD aged 50 to 75 years. They gave their informed consent to participate in the study, which was approved by the local ethics committee. The patients had been treated in an intensive care unit for at least five days for respiratory failure. Smoking had been stopped during that period. The patients were in a stable cardiorespiratory state and received bronchodilator and antibiotic drugs but no vasopressor drug. Six oximeters, partitioned into three groups, were tested in three groups of 17 patients: group 1 was the Nellcor N101 and the Ohmeda Biox III; group 2 was the Nellcor N200 and the Critikon Oxyslutter; and group 3 was the Radiometer Oxil100 and the Ohmeda Biox 3700. Group 1 was the previous generation of the instruments Nellcor N200 and Ohmeda Biox 3700, respectively. In each group the finger probes of the two pulse oximeters simultane-
ously tested during the same session were placed on the same hand and covered with a black towel to avoid any interference from room light. All instruments were operated in their slower response mode. An indwelling cannula was inserted into the radial artery on the same side as the finger probe. A sample of arterial blood was withdrawn during the period of measurement over 10 to 20 seconds, and SaO2 was immediately analyzed with a hemoximeter (Radiometer OSM2). The SpO2 and SaO2 were measured in the steady state, after a 20-minute stabilization period when the pulse signal was appropriate and stable. Three different levels of oxygenation were obtained by changing the FIO2 from 0.21 to 0.40 through a Venturi mask.

Data Analysis

In each group of 17 patients, simultaneous values of SpO2 obtained with two pulse oximeters and blood SaO2 were statistically compared using the two-way analysis of variance and Student Newman-Keuls multiple comparison tests. The mean SpO2-SaO2 difference was called "bias," and the error in precision was defined as the SD of the distribution of the SpO2-SaO2 differences. The precision was also represented by the 95 percent confidence intervals of the mean difference: mean difference ± (t × SE), where SE = SD/√n, n = sample size, and t = Student's t with (n-1) degrees of freedom.

Using correlation statistics is not the most appropriate way to define concordance between SpO2 and SaO2, particularly when several measurements are acquired in each subject. Indeed, the coefficient of correlation depends on the range of SaO2 studied, and, moreover, it does not take into account an obvious systematic instrumental or individual error.11 It is therefore a poor index of agreement between measurements.12,13 A better approach consists in the calculation of a single index derived from a repeated-measures analysis of variance: the intraclass correlation coefficient of Bartko.15 It combines an estimation of not only similarity of slopes but also similarity of intercepts of the correlations. The total variance is considered to be partitioned among three sources: the differences among methods; the differences among subjects; and a residual variance. The coefficient, R,, is defined as: R, = (MSs-MSr)/(MSs + MSt + 2MSm), where MSs = mean square due to differences among subjects, MSt = mean square due to differences among methods, and MSr = mean square due to residual variance. The coefficient, R,, is an index of concordance between two measurements. A value of 0.75 or more reflects good agreement.

According to the comparison method formulated by Bland and Altman,14 the differences SpO2-SaO2 were plotted against the corresponding SaO2. Since three measurements at various levels of FIO2 were achieved in each subject, the individual linear relationships between the SpO2-SaO2 differences vs SaO2 could be calculated. The linearity of these relationships was evaluated with the test for trend.14 The parallelism between the individual lines and the between-subject variations was tested using covariance analysis.14 A value of p<0.05 indicated significance for all analyses. Values are reported as the mean ± SD.

RESULTS

The arterial blood gas levels (PaO2, PaCO2, and pH) were similar in the three groups of patients (group 1: 41 ± 7 mm Hg, 50 ± 8 mm Hg, and 7.42 ± 0.03, respectively; group 2: 42 ± 8 mm Hg, 51 ± 9 mm Hg, and 7.43 ± 0.04, respectively; and group 3: 42 ± 7 mm Hg, 51 ± 8 mm Hg, and 7.41 ± 0.04, respectively).

Table 1 summarizes the statistical results for comparison between SpO2 and SaO2. For the two older instruments of group 1, bias and precision ranged from 3.1 to 3.9 percent, and the 95 percent confidence interval of the mean differences was larger than 6 percent. These errors were clearly lower for the four other instruments. The intraclass correlation coefficient of Bartko showed a satisfactory agreement between blood SaO2 values and the pulse oximetric readings with three instruments (Nellcor N 200, Radiometer Oxi100, and Ohmeda Bixo 3700). In each group the two-way analysis of variance showed a significant difference between simultaneous measurements (group 1, p<0.001; group 2, p<0.01; and group 3, p<0.05). The Student Newman-Keuls multiple comparison test showed that the readings of the two pulse oximeters studied together were similar, but they differed significantly from hemoximetric results in groups 1 (p<0.01) and 3 (p<0.05) and with the Critikon Oxyshuttle (p<0.01).

The uncertainty in the SpO2 reading on the six pulse oximeters is represented on a scatter plot in Figure 1 by the SpO2-SaO2 differences as a function of SaO2 (three data points in 17 patients for each oximeter). The SpO2-SaO2 differences varied from -6 to +14.5 percent for an SaO2 range of 62 to 100 percent. In most patients the individual relationships between the SpO2-SaO2 differences and the blood SaO2 at three different levels of arterial saturation were fairly linear (test for trend). For each of the instruments in groups 1 and 2, the slopes of the individual least-square lines were similar (covariance analysis). The mean of the slopes, presented for each instrument in Figure 1, was significantly different from zero with groups 1 and 2. The negative mean slope showed the presence of a systematic error. Since

<table>
<thead>
<tr>
<th>Group and Instrument</th>
<th>Bias*</th>
<th>Error of Precision†</th>
<th>95 Percent Confidence Interval †</th>
<th>Bartko's Coefficient‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a Nellcor N101</td>
<td>3.6</td>
<td>3.5</td>
<td>+2.6 . . . +4.6</td>
<td>0.08</td>
</tr>
<tr>
<td>1b Ohmeda Biolox III</td>
<td>3.1</td>
<td>3.9</td>
<td>+2.0 . . . +4.2</td>
<td>0.07</td>
</tr>
<tr>
<td>2a Nellcor N200</td>
<td>0.4</td>
<td>2.2</td>
<td>-0.2 . . . +1.1</td>
<td>0.88</td>
</tr>
<tr>
<td>2b Critikon Oxyshuttle</td>
<td>1.2</td>
<td>3.0</td>
<td>+0.3 . . . +2.0</td>
<td>0.54</td>
</tr>
<tr>
<td>3a Radiometer Oxi100</td>
<td>0.8</td>
<td>2.5</td>
<td>+0.1 . . . +1.5</td>
<td>0.75</td>
</tr>
<tr>
<td>3b Ohmeda Bixo 3700</td>
<td>0.8</td>
<td>2.5</td>
<td>+0.1 . . . +1.5</td>
<td>0.78</td>
</tr>
</tbody>
</table>

*Bias defined as mean difference SpO2-SaO2.
†Error of precision defined as SD of distribution of SpO2-SaO2 differences.
‡95 percent confidence interval of mean difference.
§Intraclass coefficient of Bartko over 0.75 reflects good agreement between SpO2 and SaO2.
the highest values for SaO₂ were, on average, accurately measured with these pulse oximeters, the lower the SaO₂, the larger the overestimation of SpO₂. The parallel individual regression lines presented a dispersion (covariance analysis, p<0.01) reflecting a between-subject variability which did not vary with SaO₂ in these two groups (groups 1 and 2). In group 3, the individual lines were not parallel, and the between-subject scattering increased as SaO₂ decreased.

**Discussion**

Several previous studies concluded with a good agreement between blood SaO₂ and SpO₂ measurements with pulse oximeters in normal man²,³,⁵ and in patients with COPD.⁴,¹⁵ however, the conclusion of most studies was based on the calculation of correlation coefficients between two methods and a test of significance. This method of comparison is commonly used but is inappropriate to represent the accuracy of a technique.¹¹,¹² A high correlation does not mean that the two methods agree and ignores an obvious systematic error. For a good agreement the points have to lie along the line of identity. The residuals scattergram, as presented in Figure 1, which plots differences between pulse oximeter SpO₂ and blood SaO₂ against blood SaO₂ values, is more informative and represents errors over the range of values analyzed. This figure shows that with all instruments, the SpO₂-SaO₂ differences were not normally distributed over the range of measurements; thus, bias and precision were not strictly represented by the means and SD of the SpO₂-SaO₂ differences. The concordance is better estimated with a special index: the intraclass correlation coefficient of Bartko.¹³ Thus, three out of the four newest instruments (groups 2 and 3) presented a good concordance with blood SaO₂.

Because comparative measurements at three different SaO₂ levels were available in each subject, the analysis of the individual relationships, SpO₂-SaO₂ differences vs SaO₂, allowed the distinction to be made between instrumental, within-subject, and between-subject sources of error.

**Instrumental Error**

Mean slopes different from zero and parallelism between the individual lines revealed the existence of
a systematic instrumental error which was independent of the interindividual variation of \( \text{SpO}_2 \). The mean of the slopes obtained in the 17 patients with the same instrument constitutes a better reflection of the systematic error than the global regression line. This error could result from the calibration algorithm built into the oximeter’s software, rather than from random variations. Observing the same discrepancies in another recent study performed in healthy subjects,\(^4\) it is unlikely that this problem is related to an unusual population of patients. The negative slopes observed in groups 1 and 2 implied that an overestimation of \( \text{SaO}_2 \) was found at low \( \text{SaO}_2 \). Although either good accuracy\(^2,3,5\) or, as recently, some underestimation have been demonstrated,\(^6,9\) such overestimation of low \( \text{SaO}_2 \) by oximeters has already been reported.\(^6,7,10,17,18\)

**Within-Subject Error**

The linearity of most individual relations, \( \text{SpO}_2-\text{SaO}_2 \) differences vs \( \text{SaO}_2 \), reflected a within-subject homogeneity. Thus, taking the instrumental error into account, a change in \( \text{SaO}_2 \) from a previous value can be accurately estimated in one given patient by pulse oximetry.\(^15,16\) That confirms the interest in noninvasive transcutaneous oximetry in clinical settings to follow the evolution of episodic hypoxemia or to conduct physiologic dynamic tests such as the ventilatory response to hypoxia or exercise.

**Between-Subject Error**

Besides the instrumental error (negative mean slope), the scattering of lines represented a between-subject source of error which was largely independent on the \( \text{SaO}_2 \) level in groups 1 and 2. On the contrary, in group 3, the \( \text{SpO}_2-\text{SaO}_2 \) differences were not related to a systematic error and were essentially due to individual variability.

Whether the error had a within-subject or a between-subject source, the subject source was confirmed by the similarity between the individual behavior observed with the two instruments simultaneously analyzed in the same patients in each group. This variability could not be attributed to skin pigmentation,\(^19\) the presence of elevated bilirubin,\(^20\) or carboxyhemoglobinemia,\(^21,22\) methemoglobinemia,\(^23\) or exogenous dyes,\(^24\) since all patients were white people with normal bilirubinemia, had stopped smoking for more than five days, and had not received any injection. Moreover, the systemic hemodynamic state does not seem to affect the oximetric readings.\(^3\) This uncertainty could be related to the various conditions of acquisition by the finger probe and depend on the local circulation at the site of the probe.

**Possibility of Correction of the \( \text{SpO}_2 \) Readings**

The instrumental origin of a systematic error, increasing while \( \text{SaO}_2 \) falls, could be mathematically corrected. Such corrections could be made by simple revision of the algorithm used by the software. If a mathematical correction is made for each of the six instruments, according to the mean slope and mean intercept of the individual regression lines between \( \text{SpO}_2-\text{SaO}_2 \) differences and \( \text{SaO}_2 \), the bias becomes negligible, and Bartko’s coefficient is clearly ameliorated (0.85, 0.81, 0.97, 0.94, 0.95, and 0.82, respectively). One of the early instruments, the Ohmeda Biox III, probably benefitted from such an improvement to become the Ohmeda Biox 3700.

Nevertheless, the aforementioned mathematical correction brings only a small advantage to the subject source of error, mainly represented by the slight improvement of the precision (3.4, 3.7, 2.0, 2.5, 2.4, and 2.5, respectively). After the technical correction the slope of the individual lines approaches zero, indicating that in one given patient, the absolute value of \( \text{SaO}_2 \) was misread by a constant error. Thus, for repeated measurements of \( \text{SpO}_2 \), an individual correction could be processed from an initial \( \text{SpO}_2-\text{SaO}_2 \) difference. In our study, when in each patient the \( \text{SpO}_2 \) values are diminished by the first \( \text{SpO}_2-\text{SaO}_2 \) difference observed during room air breathing, the error in precision is additionally reduced for five instruments (1.9, 2.6, 1.6, 2.2, 2.5, and 2.3, respectively). Therefore, for very rigorous research purposes, precise \( \text{SaO}_2 \) assessment from oximetric readings requires an individual adjustment by direct comparison against one arterial blood sample value.\(^15,25\) For this purpose the pulse oximeters should possess a calibration system allowing the users to make such an adjustment.

Finally, a residual source of error was due to within-subject variability. This affects precision and is difficult to correct; however, when it is possible, repeating and averaging several measurements in the same conditions could improve the precision.

As indicated by the present study, within the limitations of 5 percent accuracy, new pulse oximeters appear to be a helpful tool to evaluate \( \text{SaO}_2 \) in clinical practice in chronically hypoxicem patients; however, minor technical improvements would procure substantially better reliability for some instruments.

**References**


CHEST / 99 / 4 / APRIL, 1991 845

---

European Association of Cardiothoracic Anesthesiologists

The sixth annual meeting of EACTA will be held in Milan, Italy, June 4-7 at the Hotel Executive. For information contact the Organizing Secretariat: OIC Incentive, Viale Majno 21, 20122 Milan, Italy (39-2) 76100181.90.