We made an assessment of five pulse oximeters in regard to their ability to replace the HP ear oximeter as a noninvasive measurement of SaO2. Trials were performed during isocapnic progressive hypoxia (SaO2 range, 99 to 70 percent) in 22 white and six black subjects. Comparisons between values of SaO2 by oximetry were analyzed by comparing the difference of values by the two methods against their mean. The difference between pulse oximeters and the HP was 2.6 ± 10.3 percent in all subjects. Results in whites were 1.9 ± 10.2 percent and in blacks were 5.1 ± 9.2 percent. The distribution of differences between pulse oximeters and the HP were larger below 80 percent than above 85 percent. We conclude that pulse oximeters give higher values than the HP, a tendency which is more pronounced in black than in white subjects. While the limits of agreement are better at saturations above 85 percent, the 95 percent confidence limits of agreement between pulse oximeters and the HP are rather large (± 10 percent) and unacceptable for assuming that pulse oximeters will provide the same values as found in clinical studies using the HP. (Chest 1990; 97:914-19)

The HP ear oximeter has been used extensively in the management of respiratory disorders for the past decade. It continuously measures oxygen saturation using eight wavelengths of light which are transmitted by fiberoptics to a heated sensor positioned on the pinna of the ear. The accuracy of the HP has been well documented;1-6 unfortunately, it is no longer manufactured. Recently, pulse oximetry was introduced to monitor SaO2. With pulse oximetry, the light source is provided by red and infrared LEDs. The light is transmitted through a vascular bed (finger; toe; nose) to a photodetector positioned on the other side. The pulsatile flow through the tissue bed modulates the light passing through the detector.7,8 Since the thickness and pigmentation of the skin are nonpulsatile, they should not affect the arterial oxygen calculation. Pulsatile characteristics of venous blood flow are assumed to be negligible; however, since only two wavelengths of light are used, as opposed to the eight used by the HP, some errors of measurement can occur in the presence of dyshemoglobins, or events which reduce the amplitude of arterial pulsations below a certain amplitude such as hypotension or ischemia in the sampling site.9

In separate experiments by different investigators, values from the Biop 3, Ohmeda, and Nellcor N100 pulse oximeters have been shown to correlate with an accuracy of ± 2 to ± 5 percent to the HP or from arterial blood gas samples analyzed for oxygen saturation.7,8,11,13 We have compared values from two early pulse oximeters to those from the HP14 and found discrepancies of ± 1 to ± 7.5 percent in the mean difference of values between instruments. We suggested that the degree of accuracy in reference to discrete arterial blood gas measurements does not suggest that continuous measurements of oxygen saturation by the three instruments will be the same. In this later study,14 arterial blood gas measurements were not performed.

The purpose of the present study was to simultaneously compare values from five recently developed pulse oximeters to the HP (as a reasonable "gold standard") to establish their accuracy and response characteristics during progressive isocapnic hypoxemia and their ability to replace the HP as a noninvasive measurement of SaO2.

MATERIALS AND METHODS

Subjects

Studies were performed in healthy male and female subjects, 28 whites and six blacks aged 18 to 40 years. Subjects were recruited by word of mouth and newspaper advertisement. All subjects gave informed consent, and all studies were approved by the Institutional Review Board for Human Experimentation of University Hospitals.

Methods of Recording Oximetric Data

We used recently available pulse oximeters (Criticare Model 501+, Nellcor N100, Nellcor N200, Ohmeda 3700, and the
Physiocare LifeStat 1600, all equipped with a transmittance finger sensor. All values from the pulse oximeters were compared to the HP ear oximeter. The Ohmeda 3700 and Nellcor N200 were used in their "fast"-response modes.

All oximeters were recorded by an eight-channel recorder (Western Graphite Linearecorder) via their respective external connections. We used the 50 to 100 percent saturation scale, since this allowed for greater resolution of the data.

Methods of Producing Hypoxia

We used the rebreathing technique of Rebuck et al. to produce a steady fall in oxygenation by having the subject breathe from a closed system containing 6 L of room air. During the trial, a variable amount of gas was drawn through calcium carbonate crystals to maintain constant isocapnic, end-tidal carbon dioxide levels. During lower oxygen saturations (<85 percent), supplemental air was added to the system to decrease the rate of fall of SaO₂ and to maintain a drop of approximately 1 percent every 17 seconds. Subjects breathed into the rebreathing apparatus until the HP displayed an oxygen saturation of approximately 75 percent; then they were removed from the system and breathed room air until their oxygen saturation returned to baseline. This protocol took approximately six to eight minutes to complete and was repeated two to three times for each subject. Each trial was separated by 10 to 15 minutes of rest. Using this technique of hypoxic induction, arterial desaturation occurred at a rate of 1 percent every 17 seconds. Arterial oxygen samples were drawn during slow stable isocapnic hypoxia only and not during the rapid reoxygenation step, in order to avoid the difference in time response between pulse oximeters and HP (unpublished observations).

Validation of the HP Oximeter

A subgroup of four subjects had radial arterial catheters inserted and serial samples of blood drawn for arterial gas analysis during hypoxic runs, in order to validate the accuracy of the HP in our laboratory. Each subject had three to five samples drawn over two separate hypoxic runs, each separated by 10 to 15 minutes. All arterial samples were placed on ice and analyzed within 30 minutes by a spectrophotometric oximeter (IL 282 CO-oximeter). The accuracy of the IL 282 in our clinical laboratory is ±1.5 percent. Standard criteria for accuracy have a wider range, 2.5 percent. The agreement between the HP and arterial blood gas measurements is portrayed graphically as the difference between the two values ± the average of their measurements (Fig 1). The difference between the HP and arterial blood gas levels was 0.9 ± 4.3 percent, indicating good correlation and 95 percent confidence limits of agreement of about 4 percent.

Statistical Methods

When two different methods of measurement are compared, neither the correlation coefficient nor techniques such as regression analysis are considered to be appropriate. We chose to evaluate our data using the agreement between two methods of clinical measurement. This technique allows a graphic portrayal of differences of two values by two methods against their average. The lack of agreement between methods is assessed by calculating the bias, estimated by the mean difference and the standard deviations of the differences. This distribution of the differences between pulse oximeters and the HP increased as the saturation decreased. To calculate the bias and the limits of agreement when the mean difference is not independent of the size of the measurement, an empirically transformed discrepancy was defined as the n power of the difference between the reading trial oximeter and that of the reference (HP). A scatter of experimental data showing this transformed discrepancy against saturation was observed to have a distribution independent of saturation when n = 4. From this relation the limits of agreement (± Ds) can be estimated over the entire range of saturation. If ± Ds is the limits of agreement at a given saturation level of s (in units of percent), and ± D100 is the corresponding limits at 100 percent saturation, then Ds is estimated by the following relation:

\[ Ds = \left( \frac{D100}{s/100} \right)^\prime \]

D100 was derived from the entire data set via the previous transformation with n = 4. For all oximeters, this value agrees with the value for D100 estimated from measurements between 95 and 100 percent saturation.

Table 1 — Differences between Pulse Oximeters and HP

<table>
<thead>
<tr>
<th>Oximeter and Subjects</th>
<th>Mean ± 2 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nellcor N100</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>6.3 ± 9.5</td>
</tr>
<tr>
<td>White</td>
<td>2.9 ± 10.1</td>
</tr>
<tr>
<td>All</td>
<td>3.6 ± 10.3</td>
</tr>
<tr>
<td>Nellcor N200</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>7.8 ± 8.6</td>
</tr>
<tr>
<td>White</td>
<td>2.3 ± 10.3</td>
</tr>
<tr>
<td>All</td>
<td>3.5 ± 10.9</td>
</tr>
<tr>
<td>Criticare 501+</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>2.2 ± 11.1</td>
</tr>
<tr>
<td>White</td>
<td>1.3 ± 10.3</td>
</tr>
<tr>
<td>All</td>
<td>1.5 ± 10.4</td>
</tr>
<tr>
<td>Ohmeda 3700</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>4.7 ± 7.5</td>
</tr>
<tr>
<td>White</td>
<td>1.6 ± 10.8</td>
</tr>
<tr>
<td>All</td>
<td>2.4 ± 10.4</td>
</tr>
<tr>
<td>Physiocare LifeStat 1600</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>4.3 ± 9.0</td>
</tr>
<tr>
<td>White</td>
<td>1.4 ± 9.0</td>
</tr>
<tr>
<td>All</td>
<td>2.1 ± 9.2</td>
</tr>
<tr>
<td>Mean of all oximeters</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>5.1 ± 9.2</td>
</tr>
<tr>
<td>White</td>
<td>1.9 ± 10.1</td>
</tr>
<tr>
<td>All</td>
<td>2.6 ± 10.1</td>
</tr>
</tbody>
</table>

Figure 1. Differences between HP and arterial blood gas measurements as performed on IL 282 in four subjects. Difference between HP readings and IL 282 (in percent saturation) is plotted on ordinate, vs average of two methods (in percent saturation) on abscissa, in black (solid squares) and white (open squares) subjects.
RESULTS

Results of trials performed during isocapnic pro-
gressive hypoxia (99 to 70 percent) in 22 white and six
black subjects are presented in Table 1. This table
summarizes the differences between the pulse oxim-
eters (Nellcor N100, Nellcor N200, Criticare 501 +,
Ohmeda 3700, and Physiocontrol Lifestat 1600) and
the HP measurements. The differences for each ox-
imeter are presented as mean ± 2 SD. The difference
between pulse oximeters and the HP was 2.6 ± 10.3
percent in all subjects. This indicates that continuous
measurement by pulse oximeters gives higher values
than the HP and that the 95 percent confidence limits
are wider than when HP values are validated against
blood gas values. The tendency for higher values is
more prominent in black subjects than in whites, with
results in whites being 1.9 ± 10.1 and in blacks
5.1 ± 9.2 percent.

Figure 2 presents the difference between pulse
oximeters and the HP vs the average of each pulse
oximeter and the HP in all subjects. The data is
presented graphically in a separate panel for each
oximeter. This way of displaying the data showed an
increase of the scatter of differences with the decrease
in saturation in all pulse oximeters. Following this are
wider limits of agreement at low saturation and nar-
rower limits of agreement at high saturation.

To analyze our data when the scatter of difference
changes in relation to the size of the measurement
(SaO2), we employed an empirical transformation of
the differences between each pulse oximeter and the HP
that yielded a distribution independent of saturation.

Figure 3 presents the transformed data vs the average
of oximetric readings and the HP in all subjects. The
raw data (the difference between pulse oximetry and
HP readings) were transformed and are plotted for
each individual pulse oximeter against the average of
the two methods in a separate panel. The units of the
transformed data are irrelevant; hence, the values of
the transformed data have little direct clinical signifi-
cance. Graphic analysis of Figure 3 shows the distri-
bution of the transformed data to be independent of
the saturation; hence, we can apply the analysis
described in the methods section to the transformed
data.

The limits of agreement of the transformed data have
to be related to the original scale of measurement.
Estimation of the limits of agreement for each oximeter
at some individual saturations calculated from the
transformed data (as described in the methods section)
are presented in Table 2. This table shows the differ-
ences between the pulse oximeters and the HP to be
smaller at high saturations and to increase steeply with
a decrease in saturation; for example, the limits of
agreement of the Physiocontrol at 100 percent are
±5.8 percent, while at 75 percent of saturation, the
limits for the same oximeter are ±13.8 percent (2.4
times wider).

DISCUSSION

Rationale for Comparison by Agreement

In clinical measurements, comparisons of a new
 technique of measurement with an established one is
often needed to see whether they agree sufficiently
for the new methods to replace the old. Many studies use the correlation coefficient (r) between the two methods as an indicator of agreement. Bland and Altman\(^+(\) suggested that this may be inadequate, since the correlation coefficient measures the strength of a relation between two variables and not the agreement between them. According to their analysis, high correlation does not mean that the two compared methods agree. Plotting the difference between methods against the true value helps assess the differences between methods. The mean of these differences will be the relative bias, and their SD is the estimate of error. If the difference between methods is independent of the size of the measurement, 95 percent of the differences would lie between ±2 SD of the mean. These limits (±2 SD) are defined as the limits of agreement. If the limits of agreement are clinically acceptable, then the new method could replace the old.

We wished to know if the new pulse oximeter could replace the HP in continuous measurement of \(\text{SaO}_2\). Hence, we plotted the difference between the pulse oximeters and the HP vs their average. The average of the two measurements was chosen as the best approximation to the true value of \(\text{SaO}_2\), since both methods have an error of measurement. If there is an association between the differences and the size of the measurements, a transformation (of the raw data) that would be independent of the measurement should be used to calculate the limits of agreement.\(^+\) In our study the distribution of the differences between the pulse oximeters and the HP increased as saturation decreased (Fig 2). In order to analyze these data in the same way as the difference between oximeters and arterial blood gas measurements that are independent of the size of the measurement (saturation), a transformed discrepancy was defined as the \(n^{th}\) power of the difference between the reading trial oximeter and that of the reference (the HP). A scatter of the experimental data showing this transformed discrepancy vs the mean (saturation) was observed to have a distribution independent of saturation when \(n = 4\) (Fig 3). From this relation the clinically meaningful values of the limits of agreement were estimated over the entire range.

Our intent was to estimate the accuracy of pulse oximeters under conditions of their unique clinical usefulness of continuous monitoring of oxygen saturation. For this purpose, HP was chosen as the historic standard to which the other oximeters were compared. The difference between the individual pulse oximeters and the HP was \(2.6 ± 10.3\) percent in all subjects. The

---

**Table 2—Estimated Limits of Agreement at Various Saturations**

<table>
<thead>
<tr>
<th>Oximeter</th>
<th>100</th>
<th>95</th>
<th>95</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiocontrol LifeStat 1600</td>
<td>5.8</td>
<td>6.8</td>
<td>9.4</td>
<td>13.8</td>
</tr>
<tr>
<td>Criticare 501+</td>
<td>7.4</td>
<td>8.6</td>
<td>12.0</td>
<td>17.6</td>
</tr>
<tr>
<td>Nellcor N100</td>
<td>6.1</td>
<td>7.0</td>
<td>9.8</td>
<td>14.3</td>
</tr>
<tr>
<td>Nellcor N200</td>
<td>5.8</td>
<td>6.8</td>
<td>9.4</td>
<td>13.8</td>
</tr>
<tr>
<td>Ohmeda 3700</td>
<td>6.2</td>
<td>7.2</td>
<td>10.0</td>
<td>14.6</td>
</tr>
<tr>
<td>Mean of all oximeters</td>
<td>6.3</td>
<td>7.3</td>
<td>10.1</td>
<td>14.8</td>
</tr>
</tbody>
</table>

---

*CHEST / 97 / 4 / APRIL, 1990*
data indicate that pulse oximaters give higher values than the HP, and there is a greater probability of differences between the individual pulse oximeters and the HP as oxygen saturation falls below 85 percent.

Furthermore, the mean difference between oximetric readings including the HP and arterial blood gas levels were different in whites (1.9 ± 10.2 percent) and in blacks (5.1 ± 9.2 percent). These results indicate that the pigmentation of the skin affects pulse oximetry, with mean results being systematically higher in blacks than in white subjects. These results are contrary to previous studies comparing the HP oximeter to arterial blood gas levels, studies which suggested that the pigmentation of the skin does not affect the accuracy of the oximeter; however, they do fit results of a study published by the Hewlett-Packard Company showing an effect related to the skin's pigmentation in calibrating their instrument for black subjects and whites. This example of the difference between pulse oximeters and the HP may be due in this case to a difference in adjustments to baseline intensity necessary for pulse oximetric readings; however, note that the limits of agreement are similar in white and black subjects, so that the effect of the skin's color is to increase the bias without affecting the limits of agreement. Since this effect of the skin's pigmentation appears to result in a consistent bias, it would be possible to adjust for it in clinical practice or research.

An additional finding was that as saturation fell, the differences between pulse oximetric and HP values increased. Similar findings were found when oximetric measurements (pulse or HP) are compared to arterial blood gas saturation calculated from arterial blood gas levels. Since the scatter of differences of continuous measurements of oxygen saturation by pulse oximetry vs. the HP increased as arterial saturation decreased, the limits of agreement are wider at low saturations and narrower at high saturations. The limits of agreement at each saturation, which are presented in Table 2, indicate acceptable limits of agreement at high oxygen saturation but unacceptable levels lower than 85 percent saturation.

This effect of low saturations could be explained by a reduction in the signal-to-noise ratio in pulse oximetry. As saturation decreased, less red light is able to pass through the fingertip, and the AC signal becomes weaker. To compensate, the LED-driving current and the photodetector gain are increased to maintain the AC signal in a usable range. As the oximetric gain increases, incidental physiologic and electrical noise also increases, thereby decreasing the oximeter's stability and accuracy.

Another possible explanation for the wider limits of agreement could involve the time response of the various oximeters during progressive hypoxia. During progressive hypoxia the rate of change of oxygen saturation rises as saturation falls, and this could account for the increased scatter at low saturation; however, we suspect that slow progressive hypoxia, as performed by us, is a steady-state condition and assume that there is no time lag between pulse oximeters and the HP under these conditions. If these conditions are not steady-state, then agreement would be less favorable.

The rate of change in oxygen saturation in our study is similar to that used in physiologic and clinical studies of exercise and of hypoxic responsiveness. The present study documents the need for instrument specificity in determining abnormal patterns, even under these rather steady-state conditions. More dynamic conditions, such as in sleep apnea, may increase the difference further. In addition, normative values derived with the HP are not directly comparable to normative values with the new pulse oximeters, and factors like the skin's pigmentation and the range of hypoxemia may be a source of systematic error in protocols involving the newer pulse oximeters.

In conclusion, while both HP and pulse oximetry may show a good agreement with discrete measurement of SaO2, continuous measurement by pulse oximetry gives consistently higher values and a greater probability of differences with the HP. In comparison to the HP, pulse oximetric values can be higher in blacks than in white subjects. In addition, as saturation decreases, there is increasingly worse agreement between the HP and pulse oximeters.

ACKNOWLEDGMENTS: We thank Nellcor Corp. of Hayward, Calif, for the use of the Nellcor N200 pulse oximeter. We also thank David P. Brown, M.S.B.E. and Peter W. Cheung, Ph.D., from the Department of Electrical Engineering, University of Washington, Seattle, for their advice and support and Jean L. Arnold, B.S.E.E. from Case Western Reserve University, Cleveland, for her analysis of the data and support.

REFERENCES
1 Anderson JV, Fall D, Hall FK. Continuous non-invasive monitoring of hemoglobin saturation in newborn infants. Clin Res 1984; 32:905A
7 Mihm FC, Halperin BD. Non-invasive monitoring of respiratory failure with pulse oximetry and capnography. Anesthesiology 1983; 59:A136
8 Pologie JA. Pulse oximetry: technical aspects of machine design.
National Center for Analysis of Home Mechanical Ventilation

The National Center for Analysis of Home Mechanical Ventilation is a nonprofit research organization which is investigating suspected ventilation system malfunctions including problems associated not only with positive pressure ventilators, but also patient circuits, exhalation valves, oxygen, tracheostomy tubes and medical condition of patients receiving mechanical ventilation in the home. The goal of the Center is to improve the application of home mechanical ventilation in a safe, effective manner. Analysis of suspected malfunctions or problems in the home will result in educational efforts to improve the expertise and training of the home care providers as well as respiratory therapists and other health professionals. Results of national data collection of home ventilation will be reported in journal articles and at national meetings.

Physicians, therapists, or nurses who have experienced problems in patients receiving mechanical ventilation in the home are encouraged to contact the Center to record these problems in the national database. Data are being collected through homecare vendors who become members of the Center.

Current members of the Center include: Aequitron Medical, Inc.; Bear Medical Systems; Glaskoe Health Care; LIFECARE; LINCARE; Puritan-Bennett Corporation; Baxter Pharmascal Division; Cryogenic Associates; Hudson Oxygen; Instrumentation Industries, Inc.; Schering Laboratories; Shiley, Inc.; Siemens Life Support.

The following professional organizations are also participating in the Center: American Academy of Physical Medicine and Rehabilitation; American Association for Respiratory Care; American College of Chest Physicians; Gazette International Networking Institute.

To report ventilation system malfunctions or obtain more information on membership in the Center, contact Barry Make, M.D., Director, or Karen Glenn, R.R.T., Data Coordinator at (303) 398-1949. The Center is housed at the National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver, CO 80206.

References