■ LABORATORY REPORT

A Laboratory Comparison of Three Pulmonary Artery Oximetry Catheters

Phillip E. Scuderi, M.D.,* Drew A. MacGregor, M.D.,† David L. Bowton, M.D.,‡ Robert L. James, M.S.§

Background: Measurement of mixed venous hemoglobin oxygen saturation *via* catheters employing reflectance spectrophotometry has been available for more than 10 yr. Despite numerous clinical reports that have presented data showing the poor accuracy of these devices when used clinically, they are still widely used in clinical care. The reason for lack of agreement with measurements made using bench spectrophotometry is unclear. The purpose of this study is to define the performance limitations of three hemoglobin oxygen saturation catheters (Oximetrix 3, SAT-2, and HEMOPRO$_2$) in a controlled laboratory environment using a blood flow loop primed with fresh whole human blood as a model. Our hypothesis is that the performance limitations of these devices represent inherent limitations in the technology, not error introduced by patient anatomy and physiology.

Methods: Blood was equilibrated in a flow loop to four analytic gas mixtures designed to achieve oxygen saturation of approximately 50%, 60%, 70%, and 80%, respectively, with carbon dioxide tension, *p*H, and temperature held constant. Saturation readings from the catheters were collected on-line by microcomputer. Periodic blood samples were withdrawn from the flow loop for analysis on a bench spectrophotometer and subsequent comparison with catheter-derived values.

Results: By all measures, performances of the Oximetrix 3 and SAT-2 systems were comparable (all data are presented as percent saturation unless otherwise noted); bias ± precision was 3.20 ± 2.47 and −1.25 ± 3.36, respectively, *versus* −9.97 ± 7.05 for the HEMOPRO$_2$. The 95% confidence limits based on intracatheter variability were ±3.49, ±2.90, and ±9.13 for the Oximetrix 3, SAT-2, and HEMOPRO$_2$, respectively. The 95% confidence limits based on total variability, although similar

for Oximetrix 3 (±4.83) and SAT-2 (±6.59), were larger for the HEMOPRO$_2$ (±13.82). The 95% confidence intervals for agreement between catheter brands were −2.14, 11.04 (Oximetrix 3 − SAT-2); −0.18, 26.52 (Oximetrix 3 − HEMOPRO$_2$) and −5.24, 22.68 (SAT-2 − HEMOPRO$_2$).

Conclusions: While the Oximetrix 3 and SAT-2 may be acceptable as continuous monitors used to detect changes or trends, none of the three systems is equivalent to conventional bench oximetry for the measurement of hemoglobin oxygen saturation. (Key words: Measurement techniques: oximeters; pulmonary arterial; spectrophotometry. Oxygen: monitoring.)

INTRAVASCULAR determination of hemoglobin oxygen saturation using the principles of reflectance spectrophotometry has been available clinically for more than 10 yr.[1,2] The most frequent use of this technology has been to determine mixed venous oxygen saturation (Sv$_{O_2}$) using pulmonary artery (PA) catheters equipped with fiberoptics for the transmission and reception of appropriate wavelengths of red and infrared light. These devices are said to yield test results similar to bench spectrophotometers such as the IL-282 and IL-482 CO-Oximeters (Instrumentation Laboratory, Lexington, MA) or the Radiometer OSM3 Hemoximeter (Orlando, FL). The bench spectrophotometers make use of the principles of the Beer-Lambert law to determine the relative proportions of the constituent hemoglobin species present in a blood sample. This technique has replaced analytical chemistry methods and has become the *de facto* criterion standard for determining hemoglobin oxygen saturation. Oximetry PA catheters, however, rely on empiric applications of the optical laws of reflection and refraction. Limitations of this technology have been the focus of clinical studies comparing the performance of PA oximetry catheters to bench oximeters.[3-8] Most studies support the premise that PA oximetry catheters are capable of providing clinically useful trending data. It has been reported, however, that the clinical performance of these devices does not agree with the manufacturers' published specifications.[9] Reasons for this lack of agreement remain unclear. PA oximetry relies on complex optical phenomena involving absorption, reflection, and re-

* Associate Professor, Department of Anesthesia.

† Assistant Professor, Departments of Anesthesia and Medicine.

‡ Associate Professor, Departments of Anesthesia and Medicine.

§ Biostatistician, Department of Anesthesia.

fraction of light within a suspension of erythrocytes in plasma. The accuracy of the measurements also is affected by physiologic factors such as changing hemoglobin and hematocrit, erythrocyte velocity and turbulence, pulsatile blood flow, and artifact induced by catheter tip interaction with the blood vessel wall. Clinical studies, although capable of characterizing total measurement error relative to bench oximetry as a criterion standard, cannot distinguish error due to inherent limitations of the technology *versus* the confounding effects of a given patient's anatomy and physiology. The purpose of this study is to define the limits of agreement of PA oximetry catheters relative to bench oximetry under controlled laboratory conditions using a blood flow loop apparatus.

Methods and Materials

Blood Flow Loop

A blood flow loop was constructed as shown in figure 1. Four G-cylinders of analytic grade gas mixtures of oxygen, carbon dioxide, and nitrogen were used (Air Products, Winston-Salem, NC). Oxygen concentrations were chosen to yield hemoglobin oxygen saturations of approximately 50%, 60%, 70%, and 80% at a temperature of 37°C. In addition, each cylinder contained 5% CO_2 with the balance of the gas mixture composed of nitrogen. Gas mixtures were brought to 100% saturation with water vapor (partial pressure of 47 mmHg) at 37°C by a heated humidifier before introduction into

the flow loop. The flow loop was primed each day with 500 ml of fresh whole human blood obtained from a different volunteer donor each day in a manner approved by the Clinical Research Practices Committee of the Bowman Gray School of Medicine/North Carolina Baptist Hospital. Written informed consent was obtained from each volunteer.

PA Oximetry Catheter Systems

The most recent versions of the three commercially available, Food and Drug Administration-approved PA oximetry systems (catheters, optical module, and computer) were obtained from their respective manufacturers. Three oximetry PA catheters were studied: (1) Oximetrix 3 P7110-EP-H 7.5 Fr (Abbott Laboratories, North Chicago, IL), (2) SAT-2 93-A-770H 7.5 Fr (Edwards Critical Care, Irvine, CA), and (3) HEMOPRO$_2$ 7.5 Fr (Viggo-Spectramed, Oxnard, CA). The Oximetrix 3 system uses a dual fiberoptic bundle in the PA catheter. Three reference wavelengths of light are generated by the optical module and transmitted down one fiberoptic bundle. The other fiberoptic bundle is used for detection of backscatter of light from the erythrocytes. The SAT-2 system also uses a dual fiberoptic bundle in the PA catheter. Two reference wavelengths are used. The HEMOPRO$_2$ system uses a triple fiberoptic bundle in the PA catheter. One bundle is used for transmission of two reference wavelengths; the other two bundles are used for detection of backscatter.

The performance of each instrument was evaluated by comparing the oxygen saturation data obtained using
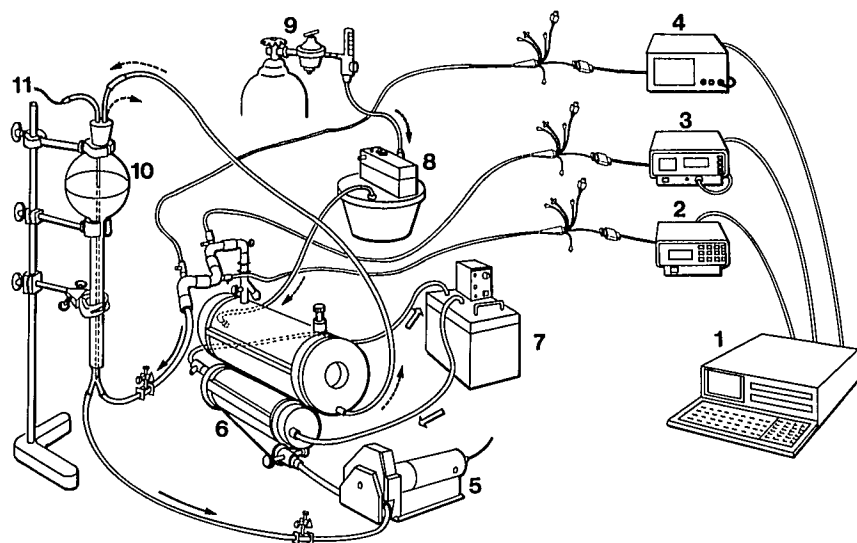

Fig. 1. Flow loop apparatus with data-acquisition system. (1) Compaq Portable II (model 2650, Compaq Computers, Houston, TX), (2) Abbott Oximetrix 3 P7110-EP-H 7.5 Fr, (3) Edwards SAT-2 93-A-770H 7.5 Fr, (4) Viggo-Spectramed HEMOPRO$_2$ 7.5 Fr, (5) Easyload Masterflex pump model 7529-00 and pump drive (model 7592-50, Cole Palmer Instrument, Chicago IL), (6) Sarns Infant Membrane Oxygenator (model 16267, 3M, Minneapolis, MN), (7) Haake water bath (model D1-L, Haake Buchler, Karlsruhe, Germany), (8) Cascade II humidifier (Puritan Bennett, Los Angeles CA), (9) Analytic grade gas mixture with regulator and flow meter, (10) blood reservoir, and (11) all-purpose temperature probe connected to NCC Hi-Lo Temp monitor (model 8200, Mallinckrodt, Argyle, NY). Arrows indicate the direction of blood flow and gas flow within the system.

the PA oximetry system with that from blood samples drawn from the flow loop and analyzed using an IL-282 CO-Oximeter as the criterion standard. This device determines the relative concentrations of oxy-, deoxy-, met-, and carboxyhemoglobin using a four-reference wavelength spectrophotometric technique. The instrument was maintained and calibrated according to the manufacturer's specifications. Yearly maintenance was performed by a factory representative before the start of the study.

Procedure

Twelve catheters from each of the three manufacturers were evaluated in the flow loop. Three catheters, one from each manufacturer, were tested simultaneously. There are 24 possible ordered sequences of the four levels of oxygen saturation if each oxygen saturation is used only once during the course of a single day's experiment. Twelve of the possible ordered sequences of saturation were chosen to minimize the number of times adjacent saturation levels fell temporally side by side during the experiment. These 12 sequences were randomly assigned to the 12 days necessary to complete the study. This scheme also allowed the three different brands of catheters to be positioned in six ordered sequences within the flow loop. After the flow loop was filled with 500 ml of blood, the flow rate generated by the roller pump was increased gradually to 2 l/min as the air bubbles were purged from the system. Equilibration with the initial gas mixture was begun, temperature in the flow loop was maintained at $37° \pm 1°C$, and the pH of the blood was adjusted to 7.40 ± 0.05 with the addition of 1-mEq aliquots of sodium bicarbonate. The PA oximetry catheter systems were calibrated using the *in vitro* calibration device supplied by the respective manufacturer. Each of the three saturation computers was attached by RS232 interface to a microcomputer for on-line data collection. The three catheters were placed in the flow loop in the sequence previously determined for that day of the experiment. The portion of the flow loop into which the catheters were inserted was shielded from ambient light. The insertion ports were designed to ensure midstream positioning of the tip of the PA oximetry catheter to eliminate the possibility of side wall-induced artifact. Each day's experiment consisted of four 1-h periods at each of the target saturation levels separated by a 5-min transition as the flow loop was brought into equilibration with the next gas mixture in the sequence. Preliminary experiments had deter-

mined gas flows necessary to maintain steady-state within the flow loop as well as gas flows necessary to achieve equilibration to the next target saturation level within 5 min.

Blood samples for comparison were withdrawn from the flow loop at 5-min intervals during the steady-state periods and at 1-min intervals during the transition and equilibration periods. Blood samples were analyzed immediately on an IL-1306 (Instrumentation Laboratory, Lexington, MA) blood gas analyzer and in duplicate on an IL-282 CO-Oximeter. Percent oxyhemoglobin, methemoglobin, and carboxyhemoglobin were determined and recorded for each sample. In addition, total hemoglobin, blood temperature, pH, partial pressure of oxygen, and partial pressure of carbon dioxide were recorded. Sodium bicarbonate was added in 1-mEq increments throughout the experiment as necessary to maintain pH at 7.40 ± 0.05. Computer sampling of the saturation readings from each of the PA catheter oximetry systems was recorded on-line at 15-s intervals throughout the experiment.

Statistical Analysis

Performance characteristics of the PA oximetry catheters relative to CO-Oximetry were evaluated during two experimental situations. Catheters were evaluated during four steady-state conditions, each lasting 1 h, corresponding to the four saturation levels chosen for this experiment, and during three periods of rapid transition in which the flow loop was equilibrated to a new saturation level. Performances of each catheter brand during steady-state and transition were compared.

Blood samples were drawn at the time-indexed intervals noted above using the computer clock as a standard reference. Samples were analyzed in duplicate on the CO-Oximeter. The average of the two values thus determined was used for comparison with the catheter-derived values. As noted above, saturation values from each of the three catheters were recorded by computer at 15-s intervals. The average of two, four, and eight catheter-derived values collected over 15, 45, and 105-s data-acquisition intervals, respectively, were compared with the average CO-Oximeter value. For example, a 15-s data-acquisition interval, consisting of two catheter-derived values, included one value recorded before the blood sample was drawn and another immediately after. A 45-s data-acquisition interval consisted of four data points, two of which were recorded

immediately before drawing the blood sample and two immediately after.

Measurement error was calculated as the difference between the PA oximetry catheter and the CO-Oximeter-derived saturation. Bias (equation 1), defined as the mean difference between catheter and CO-Oximetry determinations of saturation, and precision (equation 2), defined as the standard deviation of the bias, were determined for each brand of catheter.[10] The root mean squared error (RMSE; equation 3), proposed as a measure of overall operational performance or agreement when comparing instrumentation,[11] was calculated as the square root of the average squared bias.

If we let j index the 12 catheters within each group and let i index the n_j comparisons made between each average catheter and the corresponding average CO-Oximeter value, then n_j = 63 total comparisons made for each catheter each day, where n_j = 48 represents comparisons made during steady-state and n_j = 15 represents those made during transitions. Catheter$_{ij}$ − COOx$_{ij}$ denotes the difference between the oximetry PA catheter and CO-Oximetry for the ith measurement in the jth catheter.

$$Bias = \frac{\sum\limits_{j=1}^{N} \sum\limits_{i=1}^{n_j} (Catheter_{ij} - COOx_{ij})}{\sum\limits_{j=1}^{N} n_j}, \quad (1)$$

Precision =

$$\sqrt{\frac{\sum\limits_{j=1}^{N} \sum\limits_{i=1}^{n_j} [(Catheter_{ij} - COOx_{ij}) - Bias]^2}{\sum\limits_{j=1}^{N} (n_j) - 1}}, (2)$$

$$RMSE = \sqrt{\frac{\sum\limits_{j=1}^{N} \sum\limits_{i=1}^{n_j} (Catheter_{ij} - COOx_{ij})^2}{\sum\limits_{j=1}^{N} n_j}}, \quad (3)$$

where n_j represents the number of repeated measurements.

The steady-state, transition, and combined values for bias, precision, and RMSE were calculated for each brand of catheter using each of the three data-acquisition intervals described above. There were 756 individual comparisons for each data-acquisition interval (n_j = 63 values for each catheter × n = 12 catheters

of each brand used during the course of the study). Of these, 576 comparisons were made during steady-state periods and 180 were made during transitions.

Ninety-five-percent confidence intervals of the observed biases were determined overall, for each brand of catheter, and individually, for each catheter tested within the brand. These confidence intervals were calculated as ±1.96σ, using for σ either the square root of the overall variance of the catheter brand or the square root of the intracatheter variance as an estimate of σ.

The overall and intracatheter variances were determined using SAS 6.08 (Proc Glm, SAS Institute, Cary, NC). For each catheter brand, an analysis of variance was performed on the 45-s-interval bias with individual catheter as a random factor. The intracatheter variance was estimated by the analysis of variance error mean square. Intercatheter variability was estimated by the variance components of the analysis of variance. Overall variability was estimated by the square of the overall precision. These variances were used to calculate 95% confidence limits as described above.

The agreement between any two catheter brands was summarized by pairing their saturation measurements (45-s interval means) for each sample. The mean differences between these catheter measurements and their 95% confidence intervals for individual measurement differences were determined.

The adherence of catheter performance to the manufacturers' published specifications was tested using binomial tests. Each manufacturer's catheter failure rate was determined with its 95% confidence intervals. A manufacturer's catheter failure rate was defined as the percentage of its catheters that failed to meet specifications. The 95% confidence intervals include all the possible true population failure rates that may by random chance give the observed failure rate with a probability of 0.05 or greater. This range of possible true population failure rates was determined by recursive testing of possible failure rates until the 95% confidence interval bounds were found.

Results

The performance characteristics for each of the systems are summarized in table 1. Bias, precision, and RMSE for each brand of catheter are shown for 15-s, 45-s, and 105-s data-acquisition intervals. All values are in percent saturation ($HbO_2/(HbO_2 + Hb) \cdot 100\%$) unless otherwise noted. Three data-acquisition intervals

Table 1. Performance Characteristics of the Three Oximetry Systems

Catheter Brand	State	15 s Bias ± Precision	RMSE	45 s Bias ± Precision	RMSE	105 s Bias ± Precision	RMSE
Abbott (Oximetrix 3)	Steady	3.31 ± 2.50	4.15	3.30 ± 2.46	4.12	3.30 ± 2.45	4.11
	Transition	2.83 ± 2.53	3.79	2.86 ± 2.47	3.77	2.85 ± 2.43	3.74
	Combined	3.20 ± 2.52	4.07	3.20 ± 2.47	4.04	3.20 ± 2.45	4.03
Baxter (SAT2)	Steady	−1.12 ± 3.30	3.49	−1.12 ± 3.29	3.48	−1.12 ± 3.29	3.47
	Transition	−1.72 ± 3.56	3.94	−1.71 ± 3.54	3.92	−1.71 ± 3.51	3.89
	Combined	−1.26 ± 3.37	3.60	−1.25 ± 3.36	3.58	−1.25 ± 3.35	3.57
Viggo-Spect (HEMOPRO$_2$)	Steady	−9.82 ± 7.19	12.17	−9.79 ± 7.18	12.14	−9.78 ± 7.12	12.10
	Transition	−10.60 ± 6.59	12.47	−10.57 ± 6.58	12.44	−10.63 ± 6.52	12.46
	Combined	−10.00 ± 7.06	12.24	−9.97 ± 7.05	12.21	−9.97 ± 7.00	12.18

RMSE = root mean squared error.

were chosen for analysis to minimize the effect of occasional spontaneous fluctuations in the saturation readings from each of the catheter systems. Performance also was compared during steady-states and transitions to determine whether constant or rapidly changing conditions affected performance. The data-acquisition interval did not affect level of agreement (*i.e.*, bias and precision) between any of the catheter systems and CO-Oximetery, and unless otherwise noted, all subsequent comparisons refer to data collected at 45-s data-acquisition intervals. The differences in performance noted between steady-states and transitions were also minimal.

The Oximetrix 3 and SAT-2 had comparable performance as measured by bias and precision, whereas HEMOPRO$_2$ showed much poorer performance. The SAT 2 and HEMOPRO$_2$ systems tended to underestimate saturation compared to CO-Oximetry, and the Oximetrix 3 system tended to overestimate saturation compared to CO-Oximetry. The precision of the three instruments based on the combined data for 45-s data-acquisition intervals was ±2.47, ±3.36, and ±7.05 for the Oximetrix 3, SAT-2, and HEMOPRO$_2$, respectively.

RMSE, proposed as a measure of overall performance, was similar for the Oximetrix 3 and SAT-2 groups. The HEMOPRO$_2$ group again demonstrated poorer performance by this measure.

The 95% confidence limits based both on intracatheter and total variability using a 45-s data-acquisition interval were calculated and are shown in table 2. The intracatheter 95% confidence limits for the combined (steady-state plus transition) data are similar for the Oximetrix 3 and SAT-2 (±3.49 and ±2.90, respectively), with the HEMOPRO$_2$ showing more variability (±9.13). The 95% confidence limits based on total

variability for the Oximetrix 3, SAT 2, and HEMOPRO$_2$ are more disparate (±4.84, ±6.59, and ±13.82, respectively).

The difference plots of measurement error (Saturation [Catheter$_{ij}$ − COOx$_{ij}$] *versus* CO-Oximeter), steady-state, and transition are shown in figure 2. Steady-state and transition values show similar distributions. The confidence limits based on total variability and on intracatheter variability are shown and are plotted relative to the overall bias exhibited by each instrument.

Individual catheter performance was tested against manufacturers' specifications. The percentage of measurement errors for each catheter falling outside specifications was calculated. One hundred percent of the HEMOPRO$_2$ catheters, 67% of the Oximetrix 3 catheters, and 42% of the SAT-2 catheters exceeded their manufacturer's specifications. Performance (bias and precision) of each of the individual catheters is summarized in figure 3. As noted above, each of the systems had an overall tendency to either over- or underestimate the CO-Oximetry saturation (*i.e.*, positive or negative bias). Performance by catheter within a given brand showed considerable variability. For example, catheters 5 and 9 of the Oximetrix 3 group had worse precision than the other catheters in that group, whereas catheter 12 in the SAT-2 group had a bias directionally different from the other catheters in its group. Similarly, the bias and precision for some catheters in the HEMOPRO$_2$ group were better than the measures for the group as a whole.

The 95% confidence intervals of the differences for paired oxygen saturation measurements between brands was −2.14, 11.04 (Oximetrix 3 − SAT-2); −0.18, 26.52 (Oximetrix 3 − HEMOPRO$_2$); and −5.24, 22.68 (SAT-2 − HEMOPRO$_2$). The confidence

Table 2. 95% Confidence Intervals due to Intracatheter and Total Variability for Individual 45-s Data Acquisition Intervals

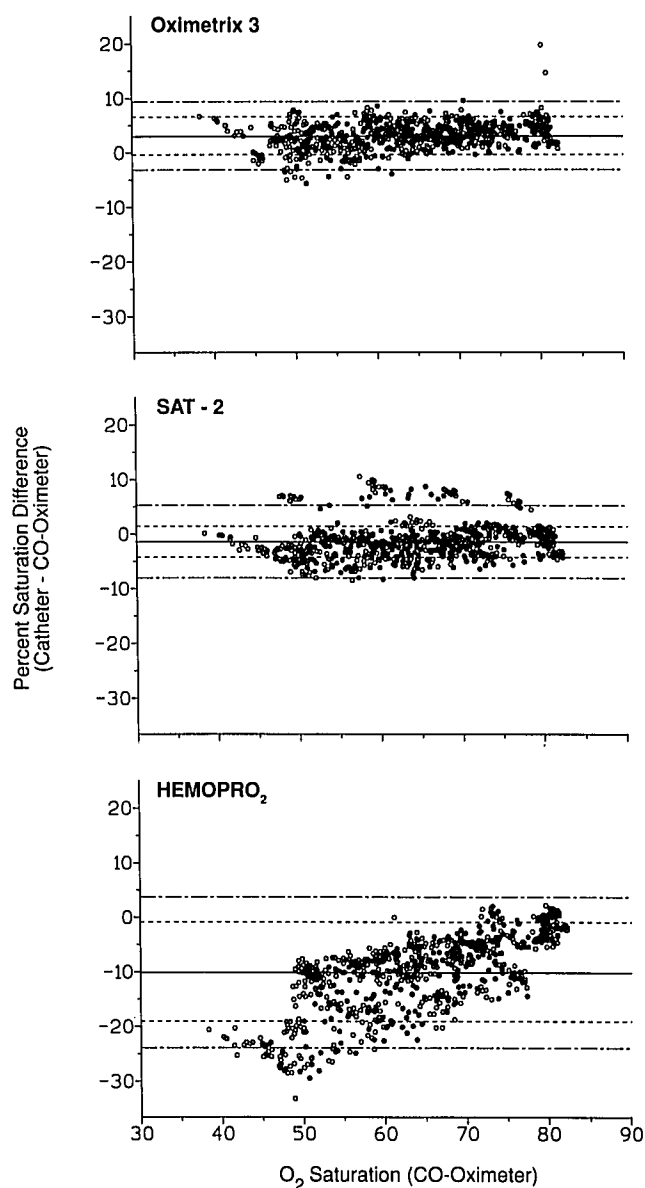|	95% Confidence Intervals (% saturation)		
	Oximetrix 3	SAT2	HEMOPRO$_2$
Steady state			
Intracatheter	±3.43	±2.70	±9.74
Total variability	±4.82	±6.45	±14.07
Transition			
Intracatheter	±3.65	±3.29	±6.35
Total variability	±4.84	±6.94	±12.90
Combined			
Intracatheter	±3.49	±2.90	±9.13
Total variability	±4.84	±6.59	±13.82

intervals were obtained using paired data obtained at 45-s data-acquisition intervals during both transitions and steady-states.

Discussion

Techniques for measuring oxygen saturation using spectrophotometric techniques were first described in 1962.[1,2] Advancements in fiberoptic and light-emitting diode technology led to the development of PA catheters capable of monitoring Sv_{O_2} continuously, using the principles of the optical phenomena of reflection. Measurement of blood oxygen saturation is possible because of the differential reflection and absorption of various wavelengths of red and infrared light by oxyhemoglobin and reduced hemoglobin. In theory, a solution containing a mixture of oxyhemoglobin and deoxyhemoglobin has absorption or reflection characteristics that can be quantified to yield the relative percentage of each species present. PA Oximetry systems use light-emitting diodes to generate appropriate wavelengths of light, which are then delivered to the PA by a fiberoptic bundle contained in the catheter. The light backscattered from the hemoglobin present in the erythrocytes is collected by a detecting fiberoptic bundle also contained in the PA catheter. Analysis of the reflected light and the application of appropriate algorithms permit the calculation of the relative proportions of oxy- and deoxyhemoglobin.

Previous research evaluated the clinical performance of PA oximetry catheters in various settings, including animal models[12] and human studies.[3-7] These investigations, which did not employ periodic recalibration

of the PA oximetry catheters,[3,4,6-8] showed that the Oximetrix 3 system produced closer agreement to bench oximetry than did the Edwards device; however, these studies all evaluated the older SAT-1 version of the Ed-

Fig. 2. Measurement error of the three groups across the range of values tested. Shown as the difference between catheter- and CO-Oximeter-derived oxygen saturation. Dashed lines = 95% confidence limits based on total variability; dotted lines = 95% confidence limits based on intracatheter variability; solid line = bias. Comparisons were made using a 45-s data-acquisition interval. Steady-state = open circles; transition = filled circles.
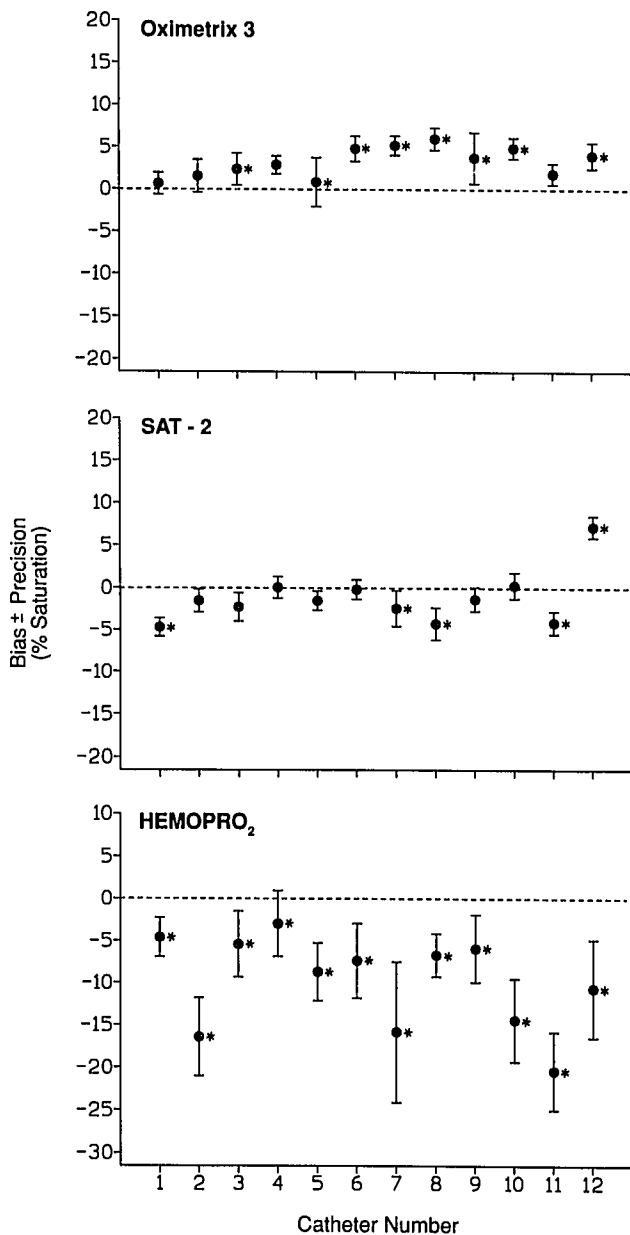
Fig. 3. Performance by catheter for the 5-h study. Each plot represents all data collected for each catheter using a 45-s data-acquisition interval and combined states (steady-state plus transition). *Catheters whose performance exceeded manufacturer's specifications.

the previously reported differences in performance were attributable to inherent differences in two- *versus* three-wavelength catheter systems.[12,13] However, a later clinical study showed comparable performance of the Oximetrix 3 and SAT-2 devices.[9]

Many factors are capable of influencing the performance of *in vivo* oximetry, including hemoglobin concentration, changes in hemoglobin concentration, artifact as a result of backscatter from the blood vessel wall, blood flow velocity, and changes in blood pH.[6,7] Each manufacturer has developed proprietary technology and computer algorithms in an attempt to deal with the potential error introduced by these and other factors. The purpose of this study was to compare the performance of the three currently available PA oximetry systems (Oximetrix 3, SAT-2, and HEMOPRO$_2$) under controlled laboratory conditions, because it was unclear why these instruments demonstrated the lack of agreement with bench spectrophotometry reported in clinical studies. Two possibilities exist. First, the technology employed in these devices is inherently sound, and under controlled laboratory conditions, performance would be comparable to bench oximetry. However, when the devices are used *in vivo*, the complex effects and interactions of a patient's physiology and anatomy lead to a decrement in performance. Second, the technology is not yet capable of providing closer agreement with bench spectrophotometry, and the results of clinical trials represent limitations in the technology itself. Because the catheter systems were tested simultaneously, under controlled laboratory conditions, the differences in performance relative to CO-Oximetry represent technologic limitations. The similarity of our findings to those reported in our 1992 clinical study[9] suggest that the performance characteristics demonstrated in clinical comparisons originate in inherent inaccuracy of this technology, as opposed to an individual patient's anatomy or physiology.

Method comparison studies, when possible, will use an accepted standard for comparison. For the purposes of this study, we have considered multiwavelength bench spectrophotometry (*e.g.*, CO-Oximetry) as a criterion standard for the determination of hemoglobin oxygen saturation. The statistical methods we used to assess agreement between PA catheter oximetry and CO-Oximetry are based on this assumption.

Instrument reliability often is assessed using the manufacturer's published specifications. The high rate of failure to meet specifications in our study reveals the importance of performing independent evaluations.

wards system. One of these studies[4] undertook a comparison of three oximetry systems and showed that the Oximetrix 3 system produced closer agreement to bench oximetry than did the Edwards device or the Viggo-Spectramed device. It has been hypothesized that

Our statistical methodology illustrates the use of simple binomial distributions to test for adherence to manufacturer's specifications. The P values and confidence limits were derived directly from the binomial distribution, and thus the values are exact. Alternatively, it should be noted that, with reasonable sample sizes, as in our study, the normal approximation for the binomial distribution usually will give reasonably accurate P values and confidence limits.

Despite the controlled nature of this performance comparison, important differences between the three PA oximetry systems and CO-Oximetry were demonstrated. The summary performance data presented in table 1 show that each of these devices exhibits a systematic tendency to either over- or underestimate CO-Oximetry. This nonzero bias occurs whether the comparisons are made during periods of steady-state oxygen saturation or rapid change in saturation. The Oximetrix 3 system consistently overestimated CO-Oximetry, whereas the SAT-2 and HEMOPRO$_2$ systems underestimated CO-Oximetry. The bias of the HEMOPRO$_2$ is considerably greater in magnitude (*i.e.,* approximately -10% saturation) than either the Oximetrix 3 (approximately 3% saturation) or the SAT-2 (approximately -1.5% saturation). The HEMOPRO$_2$ also exhibits a nonconstant bias across the range of values tested (fig. 2). The bias demonstrated by this device appears proportional to the value tested, ranging from -25% saturation at saturation values of approximately 50% to near 0 saturation at saturation values of approximately 80%. It is important to emphasize that a large positive or negative bias in a measurement is of relatively little importance provided it is consistent, because it is a simple matter to correct the observed result by the amount of the previously observed bias. Lack of reproducibility of a measurement, however, represents a fundamental flaw in the measurement process and cannot be corrected by simple mathematical manipulation. The precision estimates presented here quantitate the variability of the bias of each of the three instruments tested. The larger the value of the precision estimate, the more variable the catheter system measurement is relative to CO-Oximetry.

Variability of catheter measurements also can be quantitated by calculating 95% confidence limits relative to CO-Oximetry. These intervals can be determined for intracatheter as well as total group variability. The square root of the average intracatheter variance is a measure of reproducibility within single catheters. As shown in table 2, the intracatheter reproducibility

of saturation relative to CO-Oximetry (*i.e.,* 95% confidence limits) is better than would be expected if one looked only at these limits for the entire group. The comparison of 95% confidence limits of the three systems shows better intracatheter reproducibility than group performance. To put this another way, each individual catheter from a given manufacturer functions with a greater level of reproducibility than would be predicted from the pooled group performance. The individual catheter itself becomes a covariate of saturation. For instance, if the therapeutic goal is to ensure that a patient's actual Sv$_{O_2}$ is 70%, it would be necessary for an Oximetrix 3 catheter to read 78% (*i.e.,* 70% (target value) + 3.2% (bias) + 4.84% (95% confidence limit) = 78%) to ensure with 95% confidence that the patient's actual Sv$_{O_2}$ was at least 70%. In the case of the HEMOPRO$_2$, which has a negative bias, it would be necessary for the catheter to read approximately 74% to ensure the same level of certainty regarding the target Sv$_{O_2}$ of 70%. In addition, because of the large negative bias and wide 95% confidence limits demonstrated by the HEMOPRO$_2$, a catheter reading as low as 56% could correspond to a 70% target Sv$_{O_2}$. Similar analogies apply in cases in which the decision to treat may be triggered if certain low Sv$_{O_2}$ values are reached.

Conversely, if the oximetry PA catheters are used as trending devices, a smaller change can be used to predict an actual trend in the patient's Sv$_{O_2}$. For instance, the Oximetrix 3 and SAT-2 systems will allow the detection, with 95% confidence, of a directional trend with a change in readings of approximately $\pm 3\%$. Thus, a decline in the oximeter system reading from 70% to 67% likely represents an actual change in the patient's condition. It does not, however, predict the actual value of the patient's Sv$_{O_2}$.

The assessment of agreement between clinical measurements also may be influenced by the passage of time.[14] Previous studies[4,5,12] demonstrated that the performance of *in vivo* oximetry systems may deteriorate with the passage of time. During our 5-h study period, there was no significant change in the bias observed in any of the three systems tested. However, this does not preclude the possibility of drift in any of these systems over longer time intervals.

Because therapy may be instituted or withheld on the basis of information obtained from this type of monitor, it is essential to understand the performance characteristics of available systems. The overall performance of Oximetrix 3 and SAT-2 compared to CO-Oximetry

is similar (table 1). However, this similarity may have resulted from the controlled nature of the experiment. Our model maintained stable hemoglobin, pH, partial pressure of carbon dioxide, and temperature. The possibility of vessel wall artifact was also eliminated. This study does not preclude the possibility of differences in performance between two- and three-wavelength systems, as previously reported in some clinical studies.[12,13] There is a directional difference in bias between the Oximetrix 3 group (positive bias) and the SAT-2 group (negative bias). The magnitude of the bias and precision for these two instruments are somewhat different; however, RMSE, which has been proposed as an overall indicator of performance relative to a criterion standard, is quite similar (Oximetrix 3, 4.04; SAT-2, 3.58). The performance achieved by the HEMOPRO$_2$ is worse when compared by this test (RMSE = 12.21).

The disparity in performance between manufacturers is emphasized further by the wide 95% confidence limits for intergroup agreement. These limits define the level of agreement that could be expected if oxygen saturation were measured simultaneously by catheters from different manufacturers. Although, clinically, no patient would have two oxygen saturation catheters in place simultaneously, it is possible that an institution may have more than one brand available for use. The lack of agreement between brands would make interpretation of clinical data across groups of patients difficult.

In conclusion, limitations in the performance of oximetry PA catheters appear to be inherent in the technology. Levels of agreement with CO-Oximetry under controlled laboratory conditions are no better than those reported in clinical comparison studies. Without clinical studies to establish acceptable levels of agreement between PA oximetry catheters and bench oximeters, no statistical conclusions can be drawn regarding the clinical performance of these devices. However, the lack of agreement with manufacturers' published specifications ($\pm2\%$ compared to CO-Oximetry) demonstrated here and the wide 95% confidence limits for total variability indicate that these devices cannot be used to replace bench spectrophotom-

etry as the reference method. Determination of a patient's actual Sv$_{O_2}$ requires the use of a conventional bench oximeter such as a CO-Oximeter. However, if these devices are used primarily as trending monitors, when identification of directional changes in saturation is of paramount importance, both the Oximetrix 3 and SAT-2 appear to function within acceptable clinical limits based on intracatheter 95% confidence limits.

References

1. Enson Y, Briscoe WA, Polanyi ML, Cournand A: *In vivo* studies with an intravascular and intracardiac reflection oximeter. J Appl Physiol 17:552–558, 1962

2. McArthur KT, Clark LC, Lyons C, Edwards S: Continuous recording of blood oxygen saturation in open-heart operations. Surgery 51:121–126, 1962

3. Karis JH, Lumb PD: Clinical evaluation of the Edwards Laboratories and Oximetrix mixed venous oxygen saturation catheters. J Cardiothorac Anesth 2:440–444, 1988

4. Rouby J-J, Poete P, Bodin L, Bourgeois JL, Arthaud M, Viars P: Three mixed venous saturation catheters in patients with circulatory shock and respiratory failure. Chest 98:954–958, 1990

5. Reinhart K, Moser N, Rudolph T, Bredle D, Specht M, Gramm HJ, Goecke J, Eyrich K: Accuracy of two mixed venous saturation catheters during long-term use in critically ill patients. ANESTHESIOLOGY 69:769–773, 1988

6. Martin WE, Cheung PW, Johnson CC, Wong KC: Continuous monitoring of mixed venous oxygen saturation in man. Anesth Analg 52:784–793, 1973

7. Landsman ML, Knop N, Kwant G, Mook GA, Zijlstra WG: A fiberoptic reflection oximeter. Pflugers Arch 373:273–282, 1978

8. Hecker BR, Brown DL, Wilson D: A comparison of two pulmonary artery mixed venous oxygen saturation catheters during the changing conditions of cardiac surgery. J Cardiothorac Anesth 3:269–275, 1989

9. Scuderi PE, Bowton DL, Meredith JW, Harris LC, Brockschmidt-Evans J, Anderson RL: A comparison of three pulmonary artery oximetry catheters in intensive care unit patients. Chest 102:896–905, 1992

10. Altman DG, Bland JM: Measurement in medicine: The analysis of method comparison studies. Statistician 32:307–317, 1983

11. Sheiner LB, Beal SL: Some suggestions for measuring predictive performance. J Pharmacokir.et Biopharm 9:503–12, 1981

12. Gettinger A, DeTraglia MC, Glass DD: *In vivo* comparison of two mixed venous saturation catheters. ANESTHESIOLOGY 66:373–375, 1987

13. Oldham PD: Measurement in Medicine: The Interpretation of Numerical Data. London, English Universities, 1968, pp 148–152

14. Altman DG, Royston JP: The hidden effect of time. Stat Med 7:629–637, 1988