Use of Equilibrated Blood for Internal Blood-Gas Quality Control

Elizabeth Teng Leary, Collene J. Delaney, and Margaret A. Kenny

We have used equilibrated human blood for blood-gas quality control since 1970. In blood equilibrated 24 h after shedding, gas tensions are stable for 4 to 6 h at 0 to 4 °C; each control specimen is analyzed several times during that period to resolve malfunctions, etc. Three-fourths of all errors in gas-tension measurement detected with equilibrated blood were detected with the highest-tension controls. Equilibrated blood controls signal about one error every 14 d on each instrument. For more complete quality control, we supplement analysis of equilibrated blood with other sorts of controls, comparing results obtained by assaying each patient’s specimen on two instruments being our most effective adjunct. Such comparisons have identified erroneous assays in 3.9% of the specimens tested. The magnitude of interinstrument discrepancies (random errors) have ranged from 9 to 100% of the appropriate determinations. We use control data derived from equilibrated blood analysis for special management purposes (evaluating instruments, quantitating micro- vs. macro-sampling discrepancies, and decreasing instrument-repair costs).

Additional Keyphrases: analytical error • tonometry • blood gases • variation, source of

In 1970, analysis of human blood equilibrated by use of a tonometer became an integral part of blood-gas quality control in our clinical chemistry laboratory. Today about 500 equilibrated specimens are used annually for the internal control program and another 2500 for a community-wide proficiency testing program (1). This paper summarizes some of our observations on tonometry and blood-gas quality control during the past six years.

First, we report experiences with tonometry and tonometered blood per se. Second, we illustrate with control charts and specific examples the feasibility, sensitivity, and effectiveness of using tonometered blood in the overall control scheme. Third, we explain the need for extending to 24 h/day the quality-control procedure of performing duplicate analyses of each specimen, each on a different blood-gas analyzer. Finally, we illustrate less-common applications of tonometered blood-gas control data in solving analytical and management problems.

We hope that this report will help focus attention on the need for and the problems associated with a blood-gas quality control program.

Materials and Methods

Blood. Day-old whole blood collected with ethylenediaminetetraacetate (EDTA) anticoagulant for routine hematology studies is used. The pool ordinarily contains 1- to 3-ml aliquots from at least 20 patients.

Gases. Gas tanks (type G, 220 ft.³; Airco Industrial Gases, Murray Hill, N. J. 07974) are prepared to these approximate specifications (by vol), the remainder being N₂: (a) 2% CO₂ and 7% O₂, (b) 12% CO₂ and 21% O₂, and (c) 7% CO₂ and 35% O₂. The gas compositions are determined to ±0.015% by triplicate micro Scho-lander assay (2). Gases are analyzed in this way for both tonometry and blood-gas analyzer calibration.

Tonometer. The IL 137 tonometer (Instrumentation Laboratory Inc., Lexington, Mass. 02173) that we use has three gas-hydration chambers, three water traps, and three equilibration chambers set in a single 37 °C water bath. Figure 1 illustrates the arrangement of one of the equilibration systems. We substitute standard 250-ml, triple-arm, round-bottom flasks as the equilibration chambers. The 10-ml plastic sampling syringe (D, No. 810L/S; Becton, Dickinson and Co., Rutherford, N. J. 07070) is fitted with a 3-cm long, 13-gauge needle cut to a flat, blunt tip. The needle tip is positioned 1 mm above the glass surface at the bottom of the flask. The final water trap (E) is a 100-ml beaker containing 20 ml of water (less than 2.5 cm deep). Three aliquots of blood are equilibrated in this instrument simultaneously with the gas mixtures a, b, and c listed above.

Tonometry. For simplicity, we describe here the manipulations necessary for equilibrating only one specimen. A gas flow of 600 ml/min is maintained throughout the procedure.
In order, the pH testing sequence on IL instruments is:

1. Calibration with pH 6.84 buffer.
2. Calibration with pH 7.38 buffer.
3. Confirmation of the pH 6.84 reading if pH 7.38 was adjusted.
5. Rinsing with saline.

The blood-gas and pH-testing sequence on the Radiometer ABL-1 is that programmed by the manufacturer.

Duplicate analysis of the same specimen is performed simultaneously in the same sequence on at least one other instrument.

A two-point gas calibration of IL analyzers is done at least once every 8 h.

**Quality Control Program**

**Program.**

1. Three specimens, equilibrated to high, medium, and low tensions of O₂ and CO₂, are analyzed daily to establish relative accuracy in all blood-gas analyzers.
2. Each patient's specimen is analyzed on two instruments. The interinstrument comparison is acceptable if the difference between analyzers is less than 399 Pa (3 mmHg) pCO₂, 8% of the pO₂ observed, and 0.03 pH unit.
3. Bulk buffers (used for two-point pH calibration before each specimen analysis) are compared to precision buffers daily. (Precision buffers are warranted accurate to ±0.005 pH unit and sold in sealed ampules.)
4. Technologists note exceptional specimen quality or instrument performance in the analysis log.
5. Consumables (membranes, etc.) are inventoried so they can be identified by both lot number and shipment date. Electrodes are tested for linear response with equilibrated blood, tagged for identification, and rotated from use to storage.
6. Preventive-maintenance records are kept in detail.
7. Problem-solving logs are kept on each instrument. Symptoms, corrective actions attempted, ultimate solutions and the analyst's name are recorded.
8. Water bath temperature and barometric pressure are checked and recorded at least three times a day.
9. Between-day precision data for equilibrated blood analysis are evaluated and plotted.

**Staff.** Five medical technologists randomly rotate every two months to the tonometer section for one to 30 days of tonometering. From nine to 18 specimens are tonometered each day.

Twenty-eight registered medical technologists of the Clinical Chemistry Division perform blood-gas analyses in our "stat laboratory" on a rotating basis. All actively participate in quality evaluation, preventive maintenance, and problem solving. A supervisor is present 24 h/day. Electronic repairs beyond the scope of laboratory staff are made by the hospital's biomedical electronics technicians.
Calculations. All blood-gas tension data are corrected to standard pressure (101.8 kPa, or 760 mmHg).\(^5\) Between-day data are normalized so that tonometry gas-tank composition changes do not affect the data.

Sources of data. All data in this paper are from the University Hospital's clinical chemistry laboratory, except certain specifically identified data from another laboratory that illustrate certain aspects of quality control. That laboratory does not use interinstrument comparison, strict record-keeping, or complete staff involvement. However, a set of three equilibrated blood specimens prepared at the University Hospital is assayed daily, and a commercial control serum is analyzed daily or whenever difficulty is suspected.

Results

Quality of Blood-Gas Analysis

Representative blood-gas quality control data given in Table 1 show the average precision and accuracy of the four analyzers tested with equilibrated blood on 89 consecutive days. All data, except for 80 mmHg \(p_{CO_2}\), were derived during the same interval. The 80 mmHg \(p_{CO_2}\) control is now used in place of the 30 mmHg \(p_{CO_2}\) control.

Validation of Control Material

We validated the analysis of equilibrated blood pools for monitoring blood-gas analytical quality by determining whether variability in the equilibrated specimen or in tonometry was reflected in the control data.

First, we searched for aberrant control results not caused by blood-gas analyzer performance. We assumed that invalid control material, not simultaneous malfunction, would cause inappropriate readings simultaneously obtained on two or more instruments. "Inappropriate" was defined as inaccuracies greater than three SD (based on Table 1) for both \(p_{O_2}\) and \(p_{CO_2}\) in a single specimen. Results for the three test specimens on 60 consecutive days were examined; no defective specimens were identified.

In a second study, we compared statistics of relative analytical performance when two instruments of the same model (IL 313) were used to assay aliquots of equilibrated control blood or unequilibrated arterial blood (patients' specimens). In Table 2 relative performance is given as the ratio and as the difference in results obtained. The patients' specimen comparisons were made at the same time of day as quality-control assays. Instrument A or B was randomly used to remove the first aliquot of blood. The results show that there was no significant difference in relative instrument performance when evaluated with either equilibrated or unequilibrated blood.

Errors in Control Specimen Preparation

In the past, our equilibrated blood specimens were occasionally prepared with less precision than Table 1 indicates. The improvement was made by eliminating two major sources of error: tiny bubbles and operator bias. The bubble-associated error occurred when a nearly invisible foam of calibration gas formed within the sample after it was drawn into the syringe. Although some of this might have been visible to the tonometer operator, it was impossible to expel it completely. Larger, readily visible bubbles were routinely expelled. Table 3 illustrates typical results. The error was greatest with high \(p_{O_2}\)-containing specimens but absent at all \(p_{CO_2}\) tensions. This type of random false-positive error was eliminated by using sampling needles larger than 16 gauge and exerting as little negative pressure as possible during the sample-withdrawal steps.

The second source of imprecision was directly related to individual technique. Table 4 gives data on the apparent precision of the blood-gas analyzer, derived from

---

5 In the remainder of this presentation the more familiar non-SI unit, mmHg, is used. 1 mmHg = 133 Pa = 133 N/m².

---

<table>
<thead>
<tr>
<th>Theoretical tension mmHg</th>
<th>(\bar{x})</th>
<th>1 SD</th>
<th>CV %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{CO_2}) 15</td>
<td>16</td>
<td>1.0</td>
<td>6.2</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>1.0</td>
<td>3.2</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>1.5</td>
<td>2.9</td>
</tr>
<tr>
<td>80</td>
<td>77</td>
<td>2.4</td>
<td>3.1</td>
</tr>
<tr>
<td>(p_{O_2}) 50</td>
<td>52</td>
<td>1.8</td>
<td>3.5</td>
</tr>
<tr>
<td>150</td>
<td>143</td>
<td>5.0</td>
<td>1.2</td>
</tr>
<tr>
<td>250</td>
<td>224</td>
<td>8.0</td>
<td>3.6</td>
</tr>
</tbody>
</table>

\(n = 89\) days; performance averaged from individual data on two IL 313, one IL 213, and one Radiometer ABL-1 blood-gas analyzers.

<table>
<thead>
<tr>
<th>Ratio (A/B)</th>
<th>A − B</th>
<th>Observed range mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{CO_2})</td>
<td>1.012</td>
<td>0.029</td>
</tr>
<tr>
<td>pH</td>
<td>1.000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

\(n = 30\) days; instruments coded A and B.

\(t\) Theoretical tension of equilibrated blood.
Table 3. Effect of “Microbubble” Generation on Value for Gas Tension

<table>
<thead>
<tr>
<th>Tension, mmHg</th>
<th>Low</th>
<th>Med.</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{PCO}_2$</td>
<td>15</td>
<td>52</td>
<td>146</td>
</tr>
<tr>
<td>$\text{PO}_2$</td>
<td>29</td>
<td>146</td>
<td>52</td>
</tr>
</tbody>
</table>

**Theoretical**

**Observed:**
- without microbubbles: 18 53 30 149 50 242
- with microbubbles: 18 53 31 153 50 290

Tonometer was IL 137; analyzer was IL 313. This same phenomenon was observed with three different analyzers.

Table 4. Effect of Individual Technique on Tonometry Precision

<table>
<thead>
<tr>
<th>$\text{PO}_2$ tension, mmHg</th>
<th>250</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six-week interval</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Specimen equilibrated by</td>
<td>4.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Operator A CV, %, for analysis</td>
<td>3.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Operator B</td>
<td>9.3*</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>4.7</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

* Operator counseled.

Characteristics of the blood pool before or during tonometry were investigated to define factors that might influence equilibrated gas tensions or pH. Results may be summarized:

(a) Specimen dilution or concentration was most often caused by an inappropriate volume of water in the gas-hydration chamber. Specimens were concentrated by rapid gas flow rate.

(b) As much as 10% dilution or concentration of the blood specimen during tonometry did not significantly alter precision of the tonometry.

(c) Although there occasionally was hemolysis during pool preparation or tonometry, final plasma hemoglobin concentrations of 30 to 40 g/liter were not associated with changes in gas tension accuracy or stability.

Daily Control

*Equilibrated blood.* Figure 2 illustrates the application of equilibrated blood in daily quality control, as well as its value when used after daily calibration, maintenance, or repair. On days 10 through 12, results from the IL 313B instrument were at the lower control limit. In response to this, a new electrode was installed on day 13. The control result (x), obtained immediately, was unacceptable, so the electrode was removed and the problem resolved by an unrelated mechanism. On day 16 the defective electrode was inadvertently re-installed during preventive maintenance (y). Again, the analysis of equilibrated blood after maintenance and calibration permitted immediate detection and removal of the defective electrode. The data shown in Figure 2 are also

![Fig. 2. Daily performance of three blood-gas instruments tested with the 50 mmHg $\text{PCO}_2$ control specimen. The three instruments analyzed aliquots of the same equilibrated specimen in the same syringe. X and Y identify significant inaccuracies.](image)
Fig. 3. Analysis of 50 and 80 mmHg pCO₂ equilibrated control specimens performed on a single blood-gas instrument in another laboratory.

Fig. 4. pH determinations made during one month of the 30 mmHg pCO₂/150 mmHg pO₂ equilibrated blood control specimen on three blood-gas instruments.

The control-specimen deterioration could not explain these results, because aliquots of the same equilibrated specimen in the same syringe were analyzed on a second instrument minutes before the data shown were collected. This second instrument’s performance was 76 ± 2 mmHg (x ± 1 SD) and 50 ± 1.5 mmHg pCO₂ during the period shown. These data were within expected performance limits for both gases.

We also evaluated equilibrated pooled blood for pH control. Figure 4 illustrates the usual daily precision achieved with three instruments. There were no specific limits on initial pH of the tonometry blood pool, hemolysis, or on blood concentration or dilution during tonometry when these data were derived. Similar precision was obtained with each control specimen. The useful pH control range extended from approximately 7.54 on the 15 mmHg pCO₂ specimen to about 7.09 on the 80 mmHg pCO₂ specimen. The interinstrument comparison data generated by pH analysis of equilibrated blood are comparable to those obtained by multiple instrument analysis of patient specimens (Table 2).

Interinstrument comparison. Routine preparation of equilibrated blood on evening and midnight shifts was not feasible. Therefore, after establishing relative accuracy with equilibrated blood, we used multiple analysis of patients' specimens on different instruments to extend quality control to 24 h/day. Only 87% of the specimens were actually of sufficient volume to permit comparison. We examined pH, pCO₂, and pO₂ data from 1108 consecutive comparisons made during a three-week period to document the effectiveness of this kind of error detection as an alternative to more frequent use of equilibrated blood. All instruments were used. "Error" was defined as specified in the Methods section.

The existence of an instrument malfunction was confirmed in 30% of the discrepancies by assaying the questioned specimen in a third instrument. Equilibrated blood was used to identify the inaccurate instrument in 10% of the discrepancies. A second blood specimen and a third instrument were used to identify the inaccurate instrument in the remaining cases. In the latter situation, only the initial problem-identifying comparison was counted in this study.

Forty-three errors were identified and confirmed. In two additional incidents, technologists undertook remedial action for results which differed by less than the specified limits. Subsequent attempts at verification failed or are undocumented in both of these instances.
Thus, erroneous results were actually detected in 3.9% of specimens analyzed. Among these, 54% were in $p_{CO_2}$ measurement, 28% in pH, and 18% in $p_{O_2}$. The magnitude of detected errors ranged from 9 to 100% of the correct value, about half exceeding 20% of the assayed value. Errors were positive and negative with equal frequency. Finally, the incidence of errors detected on any instrument was proportional to the number of patients' specimens analyzed on it.

Statistical summaries of interinstrument comparisons are given in Table 2. The data were typical unless a major electrical deterioration occurred in one of the instruments, in which case the errors were greater.

When the comparison was made between two analyzers from different manufacturers, the CV's for the gas tension ratio were similar to those for the comparisons of like instruments in Table 2; the bias was constant and the imprecision was stable and comparable to that of identical instruments. In our laboratory, interinstrument tension ratio CV's of more than 4% resulted from identifiable technical or mechanical difficulty.

An example of control by interinstrument comparison is shown in Figure 5, a plot of consecutive patient-specimen comparisons made during a 36-h period. For $p_{O_2}$, progressively larger differences between instruments were detected until the comparison on specimen 12 triggered corrective action. The relative errors on instrument B in $p_{O_2}$ analysis of specimens 7, 10, and 12 were -7%, -9%, and -11%, respectively. After correction, $p_{O_2}$ determinations performed during the 24 succeeding hours on 30 consecutive specimens remained within acceptable limits. (The operator ascribed the apparent $p_{O_2}$ discrepancy on specimen 32 to a bubble seen in the sample chamber, i.e., to an invalid comparison.) The $p_{CO_2}$ discrepancy on specimen 28 immediately preceded an unacceptable one-point calibration gas reading. The error was thus identified by interinstrument comparison and by calibration monitoring.

A second example of inter-instrument comparison also illustrates the value of blood per se as the matrix for comparison. Table 6 shows consecutive pH determinations (in the sequence they were obtained) on two instruments in the outside laboratory. Instrument A, a self-calibrating analyzer with built-in trouble-shooting capability, did not signal the existence of a problem during the interval of data collection. Instrument B, a manual analyzer, was used with pH 7.38 and pH 6.84 buffer analysis before patient-specimen analysis. In the sequence shown, there was significant disagreement (0.11 and 0.09 pH units) when blood specimens 1 and 2 were analyzed with the two instruments. The technologist then manually initiated recalibration of instrument A. The resulting calibration data were satisfactory. Because both instruments seemed to be within calibration limits, the technologist next assayed precision buffers and commercial control sera. The resulting data were either unremarkable or inconsistent with the magnitude of the discrepancy in patient-specimen analysis. After assaying patient-specimens 3 and 4, the analyst sent them to the University Hospital laboratory.

### Table 6. Membrane-Associated Error in Blood-pH Determinations Not Detected with Aqueous and Serum Controls

<table>
<thead>
<tr>
<th>Sequence of analysis</th>
<th>Specimen</th>
<th>pH</th>
<th>Instrument A</th>
<th>Instrument B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blood, patient 1</td>
<td>7.49</td>
<td>—</td>
<td>7.60</td>
</tr>
<tr>
<td>2</td>
<td>Blood, patient 2</td>
<td>7.37</td>
<td>—</td>
<td>7.46</td>
</tr>
<tr>
<td>3</td>
<td>Self-calibration solution b</td>
<td>6.85</td>
<td>6.85 ± 0.02</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>Precision buffer</td>
<td>7.39</td>
<td>7.40 ± 0.02</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>Precision buffer</td>
<td>7.39</td>
<td>7.383 ± 0.005</td>
<td>7.38</td>
</tr>
<tr>
<td>6</td>
<td>Versatol, alkalosis c</td>
<td>7.53</td>
<td>7.52 ± 0.02</td>
<td>7.53</td>
</tr>
<tr>
<td>7</td>
<td>Versatol, normal c</td>
<td>7.35</td>
<td>7.38 ± 0.02</td>
<td>7.36</td>
</tr>
<tr>
<td>8</td>
<td>Versatol, acidosis c</td>
<td>7.11</td>
<td>7.14 ± 0.02</td>
<td>7.13</td>
</tr>
<tr>
<td>9</td>
<td>Blood, patient 3</td>
<td>7.43</td>
<td>—</td>
<td>7.49</td>
</tr>
<tr>
<td>10</td>
<td>Blood, patient 4</td>
<td>7.34</td>
<td>—</td>
<td>7.40</td>
</tr>
</tbody>
</table>

* $\bar{x} \pm 2$ SD specified on product label.
* Instrument A automatically calibrates itself with equilibrated buffer.
* Marketed by General Diagnostics Division, Morris Plains, N. J. 07950
where pH determinations within 0.02 pH units of the data from instrument B were made. Investigation showed that the pH electrode membrane of instrument A had been partially occluded with a protein-like film. After the electrode was cleaned and re-installed, pH measurements on instruments A and B agreed to within 0.02 pH units.

In our laboratory, we have detected interinstrument discrepancies of 0.04 to 0.06 pH units during blood analysis without corresponding discrepancies in buffer analysis.

**Effectiveness.** We examined the number of errors prevented by daily quality control, to evaluate the relative effectiveness of assaying equilibrated blood. Table 7 shows the documentable mechanisms and data. Interinstrument comparison data are those cited above. Data for equilibrated blood are derived from the estimate of one error each 3.5 d multiplied by the average workload of 67 specimens per day. The technologist-evaluation data are taken directly from citations in the analysis logs recorded during a one-month interval. The estimated total control effectiveness—59 errors prevented per 1000 specimens received—is representative of average workload conditions without major electronic malfunction on any blood-gas analyzer.

**Analysis time.** Finally, we examined laboratory records to assess whether quality-control practices substantially delayed blood-gas analysis. The average interval from receipt of specimen to conclusion of data reporting was 10.5 min (n = 100). This included inspection of specimen quality, test performance according to the protocol described, with multiple analysis on different instruments, calculations (bicarbonate and percentage O₂ saturations), and correction to patient's temperature (if >37.5 or <36.5 °C). This analysis time is representative of our usual working conditions and analyzing our usual workload of 50% micro-scale specimens, 15% of which are in capillary tubes.

**Special Control Problems**

We used equilibrated blood and control data to study both technical and management problems associated with blood-gas analysis. Three examples illustrate these special applications.

**Micro- vs. macro-sampling bias.** When equilibrated blood was sampled by the micro technique recommended for the IL 313, the pCO₂ determinations were lower than when the same specimen was sampled by the macro technique. The values for micro pO₂ determinations were higher than those for macro-sampled determinations except for those with tensions >200 mmHg. The micro determinations were less accurate when compared to the theoretical tensions of equilibrated controls. We ascribed the error to both contamination of the micro specimen with air and electrode membrane hysteresis after air sampling as prescribed by the protocol. To decrease the error, we changed the micro-sampling protocol from the recommended sequence, which required air aspiration before micro-sampling was manually initiated, to a sequence in which a modified flush solution was aspirated instead of air. The flush solution was modified by continuously bubbling calibration gas (12% O₂, 5% CO₂, and 83% N₂) through it at 37 °C to generate an O₂ tension of about 110 mmHg and CO₂ tension of about 40 mmHg. On 17 patients' specimens with pO₂ values ranging from 68 to 190 mmHg by microanalysis, the unmodified microanalysis results averaged 7 mmHg greater than macroanalysis results. The maximum observed error was +14 mmHg at 113 mmHg pO₂, with 12.4% relative error. The same specimens analyzed by the modified protocol had a mean error of +2 mmHg pO₂ and a maximum observed error of +8 mmHg, with 7.7% relative error. On patients' specimens ranging from 19 to 61 mmHg pCO₂ by macroanalysis, the unmodified microanalysis results averaged 4 mmHg less than the macroanalysis results. The maximum observed error was −8 mmHg at 62 mmHg pCO₂ with −12.9% relative error. The modified protocol had a mean error for pCO₂ of 0 mmHg and a maximum observed error of 2 mmHg, with +5.6% relative error.

**Interinstrument bias.** Pre-purchase evaluation of the IL 213 analyzer with equilibrated blood showed that its pO₂ determinations were less than the theoretical values and less than results obtained with the IL 313 analyzers. Control data for equilibrated blood collected during 19 consecutive weeks showed that the pO₂ determinations with the IL 213 were 3, 8, and 7 mmHg less than the results for 50, 150, and 250 mmHg specimens, respectively, with the IL 313. For 37 patients' specimens with tensions between 52 and 244 mmHg (on the IL 313) the IL 213 results averaged 8 mmHg less than the IL 313. We decreased the bias by substituting between specimens a gas with a greater proportion of O₂ (21% instead of 12%) for IL 213 calibration. After this modification, equilibrated blood control data collected for 20 consecutive weeks showed that, with the IL 213, pO₂ determinations were 1, 4, and 5 mmHg less than the results for the 50, 150, and 250 mmHg specimens, respectively, with the IL 313. On 37 patients' specimens with tensions between 49 and 150 mmHg (as measured with the IL 313 analyzer) the average IL 213 bias was only −2 mmHg pO₂.

**Cost containment.** Some preventive maintenance and all major acute electronic repair of the blood-gas instruments were performed by University Hospital biomedical engineering technicians. As a cost-con-
tainment endeavor and where equivalent quality could be maintained, these instrument specialists routinely substituted components available through local electronics distributors for those supplied by the instrument manufacturers. The validity of these substitutions was judged by instrument performance as monitored by equilibrated blood quality control data and by preventive maintenance records. Table 8 compares local supplier charges and blood-gas-instrument manufacturer charges for several items. During the four years of this substitution and self-repair program, the internal control data have shown no change in instrument performance.

Discussion

The analysis of equilibrated blood serves four unique functions in our quality-control program. First, it identifies inaccuracies definitively. Second, it monitors instrument response (voltage output stability, blood/gas factor magnitude, etc.) with control material that has a composition that essentially matches that of patients' specimens. Third, it establishes whether the blood-gas instruments have suitable linearity over gas tension ranges that encompass 99% of the values for patients' specimens. Fourth, it generates long-term performance statistics for administrative, research, and management purposes. Lacking the resources to independently document the quality of equilibrated blood specimens, we infer that these functions are well served from (a) the reproducibility of control data derived over several years, (b) the low incidence of documentably unsuitable equilibrated specimens, (c) the similarity of analyzer control data derived with equilibrated and unequilibrated blood, and (d) the positive comparison of our control data for equilibrated blood with performance reports by other investigators.

Our precision limits given in Table 1 compare favorably with those of Noonan and Burnett (3). Using equilibrated bicarbonate/sodium chloride solutions during six months, they obtained values of 110 ± 5 mmHg $p_{CO_2}$ and 170 ± 6 mmHg $p_O_2$ (± 1 SD) with two Model 165 analyzers (Corning Scientific Instruments, Medfield, Mass. 02052). A commercial buffer control (Blood G.A.S. Control; General Diagnostics, Morris Plains, N. J. 07950), is described on the package insert (set 4K008) as having limits (± 1 SD) of 17 ± 1, 40 ± 1.5, and 53 ± 2.5 mmHg for $p_{CO_2}$, and 55 ± 2.5, 99 ± 2.5, and 145 ± 5 mmHg for $p_O_2$.

We do not consider equilibrated blood a true primary standard against which to evaluate blood-gas analyzer accuracy, because the tonometry process per se necessarily remains an uncertainty. However, it unquestionably establishes external target tensions. Data for five of the six equilibrated tensions suggest that they are established at or close to the theoretical value. Indeed, we evaluate our performance relative to the theoretical value for all tensions except at 250 mmHg $p_O_2$. Several factors contribute to the apparent inaccuracy of this oxygen tension: blood pool composition, gas phase calibration, transducer hysteresis, sampling-transmission pressures, gas loss to the hydraulics system, fixed sensing time, and/or blood temperature adjustment. Delineation of the relative contributions of these factors, inherent in the instruments used, is beyond the scope of this paper. Although these phenomena affect all readings, they have a greater influence on $p_O_2$ analysis than on $p_{CO_2}$ and are more pronounced at higher than at lower tensions for either gas. The term "blood/gas factor," defined by Adams (4) as $100 \times \frac{[p_{CO_2} \text{gas} - p_O_2 \text{gas}]}{p_O_2 \text{gas}}$, is used here and elsewhere in the text to describe the differential transducer response to gas vs. response to whole blood without implying specific cases of the phenomenon. Under our laboratory conditions, the blood/gas factor is held relatively constant (less than 5% relative error, except at 250 mmHg $p_O_2$), and blood determinations are not "corrected" mathematically to adjust for it.

However, the inaccuracy at 250 mmHg $p_O_2$ is consistent with the observations of Veefkind et al. (5), who showed almost identically depressed readings. The accuracy reported here also compares favorably with the report of Noonan and Burnett (3). When equilibrating their buffer saline for quality control, they observed a −10.2% relative error for 170 mmHg $p_O_2$ and a −11.5% relative error for 110 mmHg $p_{CO_2}$. Subsequently, Mueller et al. (6) recalculated these data as −4.8% and −14.3% relative error, respectively. In contrast, Adams and Morgan-Hughes (7) achieved a blood/gas factor of 4.7% when equilibrating blood with 99% $O_2$.

We have noted two sources of error in specimen preparation that have in the past artificially obscured blood-gas analyzer evaluation and several possible sources of error reported by others which have not. The observation of the "microbubble" effect on the accuracy of high $p_O_2$ tensions was independent of reports by Ravin and Briscoe (8). We believe that both the bubble size and the enhanced $O_2$ solubility incurred while the equilibrated blood is stored on ice contribute to the error. The partial pressure of $CO_2$ in solution is less temperature dependent than is the partial pressure of $O_2$. The microbubble effect is illustrated here because it was the most frequent cause of tonometry error identified during the last six years. However, our records show that the staff performing tonometry can consistently eliminate microbubbles and avoid the other reported sources of random error. When they do not, the resulting tonometric imprecision completely invalidates the blood-gas quality control. In addition, we note that

<table>
<thead>
<tr>
<th>Component</th>
<th>From Instrument Manufacturer</th>
<th>From alternative vendor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nixie driver</td>
<td>$44.10</td>
<td>$2.21</td>
</tr>
<tr>
<td>Pump kit</td>
<td>26.50</td>
<td>3.30</td>
</tr>
<tr>
<td>Numeric display</td>
<td>21.70</td>
<td>7.15</td>
</tr>
<tr>
<td>Mercury battery, 1.35 V</td>
<td>17.40</td>
<td>1.65</td>
</tr>
<tr>
<td>Fuse, 0.5 A (5/box)</td>
<td>8.10</td>
<td>0.90</td>
</tr>
</tbody>
</table>
the time after equilibrating that technologists spend examining the high \( p_{O_2} \) specimen and eliminating any microbubbles may contribute to a minor constant \( p_{O_2} \) loss from the high control.

The nature of the blood pool may explain the relative stability of equilibrated gas tensions reported here. Pooling aliquots from so many specimens masks the unique blood cell and enzymatic composition or pH of individual specimens. Twenty-four hours after blood is collected, blood-cell metabolism may be limited by substrate depletion as well as by temperature and the EDTA-caused absence of calcium or magnesium. During preparation and tonometry, the pool warms to 25 to 37°C for about 30 min. This should be insufficient time to deplete dissolved and hemoglobin-bound gases detectably in most equilibrated pools.

The apparent rarity of deterioration caused by bacterial growth in our equilibrated blood, however, should not be interpreted as a confirmation of Swatman’s (9, 10) recent assertion that bacteria, especially \textit{Pseudomonas}, cannot absorb \( O_2 \) from equilibrated blood because tonometry changes the \( pH \) making it different from patient blood. On the contrary, \textit{Pseudomonas} is not fastidious with respect to \( pH \) and is commonly cultured by microbiologists on media between \( pH \) 6.8 (nutrient agar) and 7.4 (Mueller–Hinton medium). Two of our equilibrated controls have \( pH \) values within these limits. We have not had any blood-gas testing problem (including bacterial contamination of the analytical chambers) that was identified by our interinstrument comparison program but that could not be confirmed with equilibrated blood analysis.

The stability of day-old equilibrated blood (for about 4 h) compares favorably with the 2 min reported for a commercial quality-control medium (11) or the 15 min reported for equilibrated bicarbonate/chloride solution (3). Indeed, the relative stability of equilibrated tensions in our pools greatly enhances the practicality of blood-gas quality control. The time of both control-specimen preparation and control analysis can easily be adjusted to the contingencies of emergency laboratory operation. Performance of several instruments can be confidently compared by use of the same control specimen, even when analysis on one instrument is delayed by patient-specimen testing. An instrument can be serially tested during and after problem solving, recalibration, etc. However, contamination of a control specimen by air or KCl through repeated sampling is possible. This is more pronounced when sampling is done by aspiration (IL 313) than by injection (Radiometer ABL-1), as illustrated in Table 5. Because the high and low tensions remain markedly abnormal where interinstrument inconsistencies are most pronounced, the specimens are still useful in arbitrating interinstrument comparisons from 4 to 12 h after preparation.

The frequency of error detection by equilibrated controls shown here implies that either \( p_{O_2} \) or \( p_{CO_2} \) testing inaccuracy occurs on each instrument once every 14 d. We believe this is typical performance reflecting our best efforts at quality control. It does not reflect major electronic malfunction of the blood-gas analyzers, which causes recurrent and progressively more frequent inaccuracies. Our records indicate that each instrument has three or more malfunctions of this type each year. Errors of the frequency and magnitude illustrated in Figure 3 are prevented by withdrawing symptomatic analyzers from service and having them repaired. The causes of inaccuracies detected by equilibrated blood vary somewhat from one instrument to another, most being directly related to instrument design. However, some problems are common to all—inappropriate calibration, temperature deviations, etc.

The sensitivity of monitoring high gas tensions (80 mmHg \( CO_2 \) and 250 mmHg \( O_2 \)) for quality control is noteworthy. Of all the problems identified in the interval examined here, 76% involved inaccuracies at high tensions. This confirms our observation in proficiency studies (1) that nonlinear response is the most common type of systematic error observed. We further note that the magnitude of errors observed is greater at high tensions, even though relatively minor deviations from theoretical on the 15 mmHg \( p_{CO_2} \) control would cause comparatively large relative errors. In our laboratory, few inaccuracies occur only at lower tensions, and when they do, they rarely exceed 4 mmHg \( p_{CO_2} \).

Until we began using the 80 mmHg \( p_{CO_2} \) control in 1975, equilibrated blood control data indicated that \( p_{CO_2} \) analysis had fewer inaccuracies than did \( p_{O_2} \) analysis. Current data show inaccuracies occur with equal frequency for both gases. We believe both analyses are and have been equally reliable. In retrospect, it seems probable that interinstrument comparisons often triggered correction of deteriorating \( p_{CO_2} \) analysis before the 50 mmHg equilibrated controls registered any problem performance.

The implications of Figure 5 that whole blood is a more sensitive control matrix than gas is consistent with reports by others on blood/gas factor fluctuations (4, 12, 13). The progressive drift illustrated was not signaled or prevented by analysis of calibration gas (34 mmHg \( p_{CO_2} \) and 89 mmHg \( p_{O_2} \)) before and after each patient specimen as required by the analytical protocol. Progressive protein build-up on electrode membranes shown in this example is a very common cause of such error in our laboratory. However, we do not use tonometry specifically to define changes of the blood/gas factor with time because (a) interinstrument comparisons with patient’s blood effectively monitor changing electrode response to blood, (b) technologist intervention in reaction to interinstrument comparisons limit the magnitude of fluctuations generated, and (c) preventive maintenance now minimizes long-term blood/gas response changes signaled by equilibrated blood.

The data in Table 6 are entirely analogous to those of Figure 5 except that the calibration material analyzed between bloods is aqueous buffer. Durst (14) reported that differential blood/buffer response may be caused by the significantly lower pH/temperature coefficient of phosphate calibration buffers relative to blood.
ter instrument cible.” Accurate instrument for half
ported assays in addition, insensitive (3).
Interinstrument detection frequency is 8.1%.
However, performing test phenomenon with
are 6.2%.
It describes deteriorating pH changes. We suspect
that, for example, one ample and one volume-limited
specimen from the umbilical artery, taken consecutively
from a premature infant, could differ significantly for
analytical reasons. Before we introduced the new pro-
tocol, such artifacts would occur.

Augmenting the equilibrated blood control mecha-
nism with interinstrument comparison retains blood as
the test matrix, diminishes labor costs, and allows high
frequency control monitoring 24 h/d, and permits early
detection of developing problems. It is our principal
mechanism for random error control. However, because
relatively few patients have very high or very low
blood-gas tensions, this mechanism is comparatively
insensitive to problems developing at extreme levels,
which are better defined with equilibrated blood. In
addition, the possibility of air contamination during
duplicate sampling requires that technologists use care
in performing the analyses and caution in interpreting
them.

The between-day imprecision of interinstrument
comparison (as ratios) described here is smaller than
that observed by Noonan and Burnett (3). They re-
ported 8.1 and 6.2% CV’s for pCO₂ and pO₂,
respectively, on Model 165 instruments (n = 30 d, Corning Scientific
Instruments). Table 2 shows 4.1 and 3.5% CV’s for pCO₂
and pO₂ ratios on IL 313 (Instrumentation Lab., Inc.)
instruments. Noonan and Burnett also reported that
half of the discrepancies detected were “nonrepro-
cducible” and that the remainder accurately signaled need
for corrective action. The control effectiveness of in-
terinstrument comparisons reported here is 95.3% if the
two inappropriate technologist responses to satisfactory
comparisons are included; it is 100% if they are not. The
contrast in these reports may be related to blood-gas
instrument design per se, or to analytical technique.

The high (50%) incidence of pCO₂ errors detected by
interinstrument comparisons implies that CO₂ tension
assays are subject to more random error than are O₂ or
pH determinations. However, a similar study in 1971
(3) showed equal error incidence for pCO₂, pO₂, and pH
analysis. That study also showed a total error-detection
rate of 4.7%. It is possible that the difference between
the two studies reflects changes in preventive mainte-
nance or deteriorating quality of CO₂ electrode mem-
branes currently in use. We suspect the latter and cannot
otherwise explain the discrepancies.

We tried to ascertain whether staff response to in-
accurate results on equilibrated blood differed from
their response to inappropriate results detected with
interinstrument comparisons. Although there appears
to be a more rapid response to the former, it is not
documentable. Figure 5, however, illustrates the phe-
nomenon. The presence of an error was first noted on
specimen 10 but not corrected until after analysis of
specimen 12. Results for specimens 7, 10, and 12 ob-
tained on instrument B were not reported to physicians.
Similar data and the technologists’ own notes indicate
that they tend to correct problems soon but not always
immediately after identifying them when using inter-
instrument comparisons for control. There is no anal-
ogous response delay when an inaccuracy is detected
with equilibrated blood. From this observation, we infer
that target tensions are more helpful for quality control
than comparable tensions because they elicit a more
rapid, confident response.

Technologist response is, in fact, vital to all aspects
of quality control described here. Table 4 shows one
form of direct staff impact. The technique of our to-
nometer operator completely obliterated the usefulness
of equilibrated blood in monitoring blood-gas analysis
variability and in establishing long-term performance
statistics. In a proficiency-testing study, we illustrated
an analogous technologist impact on gas analysis per se
(1). Figure 2 illustrates how easily systematic errors can
be inadvertently introduced by technologists in a busy
service laboratory. It also illustrates the promptness
with which attentive personnel use control information
to identify even uncommon, unpredictable sources of
error. Table 6 shows that technologists directly consti-
tute an effective control mechanism when they evaluate
specimen quality and identity. In our experience, ade-
quate quality cannot be achieved without a well-trained,
thoughtful, and motivated staff. We concur with
Mohler’s observation, based on proficiency testing ex-
perience, that “quality control begins with well-trained
analysts” (15).

We could list many technical or management appli-
cations of blood-gas performance statistics derived from
equilibrated blood control data. We cited the macro- vs.
the micro-sampling differences because we know of no
laboratories that control micro analysis per se. In our
laboratory, the analytical precision of macro- and
micro-sampling seems to be the same, whether evalu-
ated by equilibrated blood or interinstrument com-
parisons. (The blood container does affect precision,
however, so that micro-sampling from capillary tubes
is less precise than from tuberculin syringes.) The ac-
curacy also varies with sample size. The discrepancy
between macro- and modified micro-sampling data
could be reduced further by using fluid completely
(rather than partly) equilibrated with calibration gas
at 37 °C, but this is impractical. By diminishing the
micro-sampling inaccuracy, we reduce the possibility
that, for example, one ample and one volume-limited
specimen from the umbilical artery, taken consecutively
from a premature infant, could differ significantly for
analytical reasons. Before we introduced the new pro-
tocol, such artifacts would occur.
Enhancing the $p_{O_2}$ test agreement between the IL 213 and 313 analyzers ensures consistency. It also facilitates interinstrument comparisons, because the almost negligible bias is easier to evaluate. Moreover, with instrumental bias minimized, removing one analyzer for maintenance or repair has no impact on the results being generated.

We cite the financial savings resulting from local purchase of instrument components to illustrate that our blood-gas quality-control expenses are not an unmitigated financial deficit. We do not recommend such cost-recovery activity for all laboratories because it requires staff experience and adequate test information. Without these, such endeavors could engender more patient risk than financially-limited use of quality control. Nevertheless, decreased down-time, better control of consumables (membranes, tubing, buffers, etc.), and recovery of costs for documentally defective consumables are all financial benefits derived from our equilibrated blood control program.

Finally, we draw attention to the relative effectiveness of documentable quality-control mechanisms (Table 7). Although the equilibrated blood mechanism eliminates few errors directly in our setting, its absence could result in many errors. The relative value of equilibrated blood is directly related to the number of other blood-gas quality-control measures used in a laboratory. From our proficiency testing experience, we believe that the error incidence could rise to 100% unless an external accuracy standard, preferably blood, is used. Certainly, interinstrument comparisons are worthless alone because the comparison discrepancies cannot be arbitrated.

If considered in medical context, the data in Table 7 help justify frequent quality control of blood-gas determinations. First, the error rate reported here is 60% higher than Grannis et al. (16) found for such other clinical laboratory tests as sodium, potassium, or calcium assay. Unlike other laboratory tests, which are effectively controlled in batches, blood-gas analysis would seem to have a high incidence of random error. Unlike other laboratory tests, a high proportion of the results will not fall within the predicted normal ranges. In fact, we found no specimens with simultaneously normal pH, $p_{CO_2}$, and $p_{O_2}$ among 100 adult specimens evaluated with age- and sex-adjusted normal values. By contrast, 85% of the emergency amylase determinations and 74% of the emergency sodium determinations were within normal limits. The high incidence of abnormals is also associated with a high probability that some form of medical intervention, either passive or active, will rapidly result from the clinician’s assessment of the data. Thus, blood-gas testing is comparatively unique and its very uniqueness warrants additional consideration of testing quality. Unfortunately, although few hospitalists would perform serum sodium or aspartate analyses without simultaneously assaying high and low control standards and/or sera, many will generally perform blood-gas analyses with few or no controls. We question this practice because of our experience with the state of the art of blood-gas testing and our observations of the medical context in which our data are used.

In summary, we have found equilibrated blood to be a sensitive, inexpensive, reproducibly prepared, and reasonably stable control matrix. By supplementing its use with interinstrument analysis of patient blood, we retain the benefits of blood as a control medium and decrease the incidence of random errors that might otherwise be reported. Overall, we have found that extensive control of blood-gas analysis is necessary and achievable.

This work was supported in part by NIGMS Training Grant in Clinical Chemistry ST01-GM00776 and by NIGMS Research Fellowship 4F22-GM01572.

References


---