Instrumental variability of respiratory blood gases among different blood gas analysers in different laboratories

M.J. Kampelmacher*, R.G. van Kesteren*, E.K.A. Winckers**


ABSTRACT: The aim of this study was to test the hypothesis that differences in oxygen tension (PO2) and carbon dioxide tension (PCO2) values from measurements performed on different blood gas analysers in different laboratories are clinically insignificant.

Samples of fresh whole human tonometered blood (PO2 8.1 kPa (60.8 mmHg); PCO2 5.3 kPa (39.9 mmHg)) were placed in airtight glass syringes and transported in ice-water slush. Blood gas analysis was performed within 3.5 h by 17 analysers (10 different models) in 10 hospitals on one day.

The mean of the differences between the measured and target values was -0.01±0.19 and 0.21±0.13 kPa (-0.06±1.45 and 1.55±1.01 mmHg) for PO2 and PCO2 respectively. The mean of the differences between two samples on one analyser was 0.06±0.06 and 0.04±0.03 kPa (0.47±0.48 and 0.29±0.24 mmHg), respectively. For PO2 and PCO2 the interinstrument standard deviations (sb) were 0.18 and 0.13 kPa (1.38 and 0.99 mmHg), respectively, whereas the intra-instrument standard deviations (s) were 0.06 and 0.03 kPa (0.47 and 0.26 mmHg), respectively. Both for PO2 and PCO2, the ratios of sb2 and s2 were statistically significant (analysis of variance (ANOVA) p<0.001). The standard deviations of a random measurement on a random analyser were 0.19 and 0.14 kPa (1.46 and 1.02 mmHg) for PO2 and PCO2 respectively.

We conclude that the variability in measurement of blood gas values among different blood gas analysers, although negligible, depends much more on inter- than intra-instrument variation, both for oxygen tension and carbon dioxide tension. Technical improvements and adequate quality control programmes, including tonometry, may explain why the variability in blood gas values depends mainly on errors in the pre-analytical phase.


Blood gas analysis is frequently employed in clinical care and research. Arterial oxygen tension (PaO2) and carbon dioxide tension (PaCO2) are frequently used as selection criteria, indices of assessment or endpoints [1–6]. If only one blood gas analyser (BGA) is used, the PaO2 and PaCO2 values reflect the specific performance characteristics of that particular BGA. Although there may be bias (the systematic tendency to over- or underestimate), precision (a measure of reproducibility) is usually excellent [7, 8]. However, if analyses are performed in different BGAs, as often happens in multicentre studies and with patients involved in home care programmes, large differences in mean values and standard deviations, as well as conflicting results, may arise, particularly for oxygen tension (PO2). Manufacturer- and model-specific design and performance characteristics apparently affect the accuracy of the measurement [7–10]. If BGA-specific differences in bias or precision exist, there may be both statistically and clinically significant differences between values obtained by different BGAs. Consequently, research and clinical conclusions may be influenced by the performance characteristics of the BGA models employed.

This study was designed to test the hypothesis that the instrumental variability of respiratory blood gas values is negligible, such that differences in PO2 and carbon dioxide tension (PCO2) measurements performed using different BGAs in different laboratories are clinically insignificant.

Methods

Instruments and materials

Fresh venous blood was used, drawn from one volunteer (MIK) into heparinized collecting tubes. Aliquots of 20 mL blood were equilibrated in a Laué tonometer (Eschweiler and Co., Kiel, Germany) for at least 30 min, with a prehumidified gas mixture containing 8.54% oxygen, 5.61% carbon dioxide and nitrogen to 100% by
volume (from Hoek Loos, Amsterdam, The Netherlands). Both tonometer and humidifier were submerged in a water bath, maintained at 37°C. Tonometered samples were removed anaerobically from the tonometer into two airtight glass syringes (Hamilton, Banaduz AG, Switzerland), which were immediately capped and put into a box with crushed ice. The partial gas pressures of oxygen (P_{O2}, 8.1 kPa (60.79 mmHg)) and carbon dioxide (P_{CO2}, 5.3 kPa (39.94 mmHg)) were calculated by multiplying the barometric pressure minus the pressure of water vapour (6.3 kPa (47 mmHg) at 37°C) by the volume fraction of the respective gas.

**Protocol**

Blood samples from the glass syringes were measured successively in 10 different hospitals. Three hospitals used more than one BGA. This procedure was performed twice on the same day. Five hospitals were visited during the first round and four hospitals during the second. The first and final measurements of each round were performed in the clinical laboratory of our hospital. The other hospitals, which were located within a distance of 20 km, were then visited. Each round took less than 3.5 h. In each BGA, two measurements were performed. All samples were injected into the instruments by experienced laboratory technicians, who were told that the samples were part of a research study. Barometric pressure was taken from the latest calibration printout. Each BGA was deemed to be operating properly in accordance with the laboratory’s quality-control procedures, before each pair of samples was introduced. As this study was designed explicitly to assess the performance of BGAs under normal operating conditions, each pair of samples was interspersed by clinical specimens, and among regular scheduled calibration and/or quality control procedures. No other procedures of this kind were performed and no measurement was preceded by such a procedure.

**Statistical analysis**

The data are expressed as mean±standard deviation (SD). "Deltas" (δP_{O2} and δP_{CO2}) are the differences between the measured values of the samples and the target values (t) for P_{O2} and P_{CO2}. Bias was defined as the mean difference between the BGA-determined P_{O2} and the tP_{O2} value, or between the BGA-determined P_{CO2} and the tP_{CO2} value. Precision was defined as the differences between two measurements in one BGA. For both P_{O2} and P_{CO2}, the ratios of s_b^2 and s^2 were statistically significant (p<0.001). The overall so (s_b) values of random measurements in a random BGA, were 0.19 and 0.14 kPa (1.46 and 1.02 mmHg) for P_{O2} and P_{CO2}, respectively. Because s was much smaller than s_b, (s_b=√(s^2+s_b^2)) depended mainly on s_b.

**Discussion**

The purpose of this study was to test the reproducibility of a single blood gas measurement in a random BGA in any hospital in the Netherlands under routine clinical conditions. We were not interested in the bias of each BGA, but in the variability (standard deviation) between the BGAs tested. Our results show that this variability is clinically insignificant, both for P_{O2} and P_{CO2}. This study was part of a multicentre study on long-term oxygen therapy, in which P_{A,O2}<8.0 kPa (60 mmHg)

<table>
<thead>
<tr>
<th>Hospital No.</th>
<th>Instrument</th>
<th>tP_{O2}</th>
<th>tP_{CO2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABL3 (A)</td>
<td>-0.16</td>
<td>-0.03</td>
</tr>
<tr>
<td>1</td>
<td>ABL3 (B)</td>
<td>-0.02</td>
<td>+0.37</td>
</tr>
<tr>
<td>1</td>
<td>ABL3 (C)</td>
<td>-0.16</td>
<td>+0.07</td>
</tr>
<tr>
<td>1</td>
<td>ABL3 (D)</td>
<td>+0.05</td>
<td>+0.27</td>
</tr>
<tr>
<td>1</td>
<td>ABL-330</td>
<td>-0.21</td>
<td>+0.23</td>
</tr>
<tr>
<td>1</td>
<td>Corning 288</td>
<td>-0.18</td>
<td>+0.17</td>
</tr>
<tr>
<td>2</td>
<td>Corning 178</td>
<td>-0.20</td>
<td>+0.01</td>
</tr>
<tr>
<td>3</td>
<td>AVL947</td>
<td>-0.30</td>
<td>+0.33</td>
</tr>
<tr>
<td>4</td>
<td>Corning 178</td>
<td>-0.03</td>
<td>+0.24</td>
</tr>
<tr>
<td>5</td>
<td>Corning 178</td>
<td>+0.12</td>
<td>+0.15</td>
</tr>
<tr>
<td>6</td>
<td>IL1312</td>
<td>+0.08</td>
<td>+0.30</td>
</tr>
<tr>
<td>7</td>
<td>IL1306 (E)</td>
<td>+0.16</td>
<td>+0.21</td>
</tr>
<tr>
<td>8</td>
<td>IL1306 (F)</td>
<td>+0.16</td>
<td>+0.25</td>
</tr>
<tr>
<td>9</td>
<td>IL1301</td>
<td>-0.20</td>
<td>+0.03</td>
</tr>
<tr>
<td>10</td>
<td>ABL30</td>
<td>+0.23</td>
<td>+0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.29</td>
<td>+0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.22</td>
<td>+0.12</td>
</tr>
</tbody>
</table>

δPt_{O2} and δPt_{CO2} are the differences between the BGAs tested. Our results show that this variability is clinically insignificant, both for P_{O2} and P_{CO2}. This study was part of a multicentre study on long-term oxygen therapy, in which P_{A,O2}<8.0 kPa (60 mmHg).
is used as a selection criterion. If significant variability between the blood gas values of the different BGAs could be demonstrated, we might have been forced to demand that the $P_a\text{O}_2$ of all patients was measured in our hospital before a patient could enter the study. However, on the basis of our results there was no reason for such a requirement.

The findings of this study are in agreement with two studies that compared BGA accuracy for $P_O2$ or $P_CO2$, with a common source of tonometered blood, and which found small differences between BGA means [9, 11]. The good within-instrument precision for both $P_O2$ and $P_CO2$ is consistent with several previous studies, that have used repetitive and consecutive analysis of blood or quality-control solutions uninterrupted by clinical blood analyses [9, 12]. Nevertheless, our results are remarkable, all the more so because several studies comparing the measured tonometered $P_O2$ values from BGAs of different makes indicate that, despite the fact that imprecision may be relatively small, significant differences in bias exist between analysers [7, 8, 12].

There are analyser-specific factors that can influence accuracy, particularly for the measurement of $P_O2$. The following instrument differences may all contribute to model-specific differences: sample size; sample introduction technique; sample warming; chamber rinsing; analysis time; chamber size; contamination from residual material within the measuring chamber prior to sample introduction (memory effect); inherent drift characteristics of the electrode; electrode signal processing; calibration methods; intra-instrument variation over time; and ease of instrument repair and maintenance [8–11]. Technical improvement of modern analysers and adequate quality-control programmes could explain why the variability among the different types of BGAs in our study was substantially lower than that previously reported in other proficiency programmes. Because the overall variability between BGAs was determined mainly by interinstrument variability, frequent calibration and quality control, including tonometry, are more important than repetitive measurements in the same BGA. In this study, tonometered blood was used because it is the technique for establishing the inaccuracy and imprecision of an individual BGA [8–13]. Tonometry of fresh heparinized whole blood instead of stored whole blood offers a more "physiological" method, since the material and, especially, the $P_O2$, at which the haemoglobin is half saturated with oxygen ($P_S02$) are the same as in actual patient samples [9]. Moreover, by using tonometered blood, we tried to mimic the clinical setting as closely as possible.

Several factors may have decreased the variation between the BGAs tested. Firstly, only one target value was used. Moreover, at $P_O2$ levels <8.0 kPa (60 mmHg) differences between measured and tonometered $P_O2$ values tend to be small [14]. At this low $P_O2$, the dissociation of oxyhaemoglobin will stabilize the level of physically dissolved oxygen, which is the oxygen fraction measured by the electrode [15]. In contrast to the $P_O2$ values, $\delta P_CO2$ is affected much less by the $P_CO2$ level [14]. Secondly, this study was undertaken on one particular day. Furthermore, all BGAs tested were located at virtually the same altitude and the samples were transported in two airtight glass syringes. Finally, imprecision might have been larger if more than two samples had been measured by the same BGA. The samples were, however, measured under routine clinical conditions, which could potentially have increased overall variation [14]. In addition, tonometered samples were kept in ice-water slush for a considerable time. Although $P_CO2$ and pH values hardly change within 4 h, theoretically $P_O2$ values could have changed more than 2% [16–18]. The $P_O2$ and $P_CO2$ values measured by the same BGA at the beginning and at the end of every round, however, did not differ by more than 0.09 and 0.07 kPa (0.7 and 0.5 mmHg), respectively.

Although pre-analytical errors cannot be ruled out completely, in this study, variability was affected predominantly by analytical errors. This is different from the situation in clinical practice where pre-analytical errors play a major role, especially for $P_O2$. In the pre-analytical phase, the blood gas results may be influenced, for example, by: the ventilatory status of the patient before and during blood collection; the blood $P_O2$ level; the technique of specimen collection; the nature of the specimen container; preparation of the container with anticoagulant; sample handling; cooling of the sample; storage and transport of the specimen; and severe leucocytosis or thrombocytosis [16–26]. This study provides indirect support for the idea that standardization and quality control in the preanalytical phase represent the best strategy to optimize reproducibility of measurements of blood gas values in a clinical setting.

We conclude that the instrumental variability of blood gas values among different types and models of blood gas analysers in different laboratories depends much more on inter- than on intra-instrument variation, and is negligible, for both oxygen tension and carbon dioxide tension. High technical standards and adequate quality-control programmes, including tonometry, may explain why the variability in blood gas values is related mainly to the pre-analytical phase.

Acknowledgements: The authors would like to thank Ir.S. van Ooij and the staff of the Department of Clinical Chemistry, University Hospital Utrecht, for technical assistance.

References


