Volume calibration alone may be misleading

G. VAN DEN BOOM*, L. M. VAN DER STAR*, H. FOLGERING‡, C. P. VAN SCHAYCK* AND C. VAN WEEL*

*Department of General Practice and Social Medicine and
†Department of Pulmonology Dekkerswald, University of Nijmegen, The Netherlands

The use of spirometry is becoming more and more widespread in non-laboratory situations such as general practice or occupational medicine. In these non-laboratory situations, volume calibration with a 3000 ml syringe is often the only feasible method to ensure that the spirometer produces valid and reproducible data. Sophisticated equipment to calibrate forced manoeuvres with standard waveforms are not present.

In this study, we assessed whether volumetric calibration is a guarantee for valid and comparable spirometric results.

Two portable spirometers were tested. On 8 consecutive test days, both spirometers were calibrated with a 3000 ml syringe in accordance with the American Thoracic Society (ATS) guidelines. The comparability of the spirometric results (forced expiratory volume in 1 S, FEV₁) was tested in two ways. Firstly, the spirometers were compared to each other using the results from 43 volunteers on the same 8 test days. The spirometers were presented in a randomized order and volunteers were asked to perform a series of reproducible manoeuvres in both spirometers. Paired observations were analysed, using Bland and Altman plots. Secondly, the spirometers were compared to a 'gold standard', a computer-driven syringe (CDS).

Calibration with the 3000 ml syringe showed that both spirometers complied with the ATS criteria for volume calibration for diagnostic spirometry. However, paired FEV₁ data obtained in subjects showed a systematic, volume-dependent difference between the two spirometers (mean difference: 289 ml, P < 0.001, systematic difference: 8.6%, P < 0.0001). This systematic difference was confirmed by the comparisons with the CDS.

Volume calibration may be misleading. The results from volume calibration may meet the ATS criteria, but this is no guarantee that data from forced manoeuvres are accurate. If CDS equipment to simulate standard waveforms is not available, it is recommended that biological calibration is performed regularly and, if possible, that paired data from two (or more) different spirometers are compared.

Introduction

Over the last decades, spirometry has attained a prominent position in the diagnosis and monitoring of obstructive lung diseases. The market introduction of portable and affordable spirometers has made the use of spirometry more and more widespread. As a result, spirometric evaluation has increasingly become a standard procedure in practices outside the clinic. To ensure the quality of the spirometric results, regular calibration of the spirometers is a mandatory aspect (1-3). In the specialized clinics, sophisticated equipment and knowledge with regard to testing the precision, accuracy and linearity of spirometers is available. In non-laboratory situations, however, equipment such as computer-driven syringes (CDS) are often not present.

Then, the only feasible option for out-clinic physicians to ensure the accuracy of a spirometer, is to purchase a spirometer which complies with widely accepted equipment criteria such as the ATS or ERS criteria and to perform volume calibration with a 3000 ml syringe regularly. When the volume calibration procedure (repeatedly) indicates that the spirometer complies with the appropriate criteria, the spirometer is expected to perform well and to produce valid and reproducible results for volumes as well as for flows.

In this study, we investigated whether satisfactory results from volume calibration using a 3000 ml syringe obtained from two portable spirometers are a guarantee for valid and interchangeable spirometric results.

Methods

SPIROMETERS

Two portable spirometers were tested in this study. Both spirometers met the ATS recommended equipment criteria (4) according to the manuals when they left the
factory (equipment recommendations: diagnostic spirometry, page 1111). The first spirometer has a heated Fleisch pneumotach sensor which converts a flow signal into a series of digital signals, interpreted by a microprocessor. The built-in software included an auto-calibration procedure. This procedure was performed at the start of the study: the spirometer did not require adjustment. During the 8 consecutive test days, this procedure was not repeated. After the 8 test days, the auto-calibration procedure was run once more, again resulting in no adjustment. The second spirometer measures flow by counting the number of swirls through a heated sensor. The swirls were generated by static blades inside the mouthpiece. This spirometer was ATS correct and required no auto-calibration according to its manual. Both spirometers have a similar flow range (0-8 l/s, 0-4-8 l respectively) and were able to display and print flow-volume and volume-time spiromgrams. As only one spirometer of each of the two brands has been tested, the brand names are omitted to prevent erroneous conclusions regarding the two brands.

DESIGN

In this study, three sets of experiments were performed:

Volume calibration of the spirometers

Volume calibration of the two spirometers was performed according to the standard procedures in the manual. For the calibration procedures, a 3000 ml syringe was used (Calibringe 1, Vacumed, Ventura, CA, U.S.A.), as recommended in the respective manuals and the ATS criteria for equipment (Diagnostic spirometry, page 1111) (4). This was done on 8 separate days, just before measuring volunteers. The standard procedure was performed with a relatively slow plunger speed (approximately 500 ml s\(^{-1}\)) and repeated with a relatively high plunger speed (approximately 3000 ml s\(^{-1}\)). The subsequent two volumes per spirometer per day were recorded and evaluated using the ATS criteria (within \(\pm 3\%\)). To ensure that volumes were BTPS correct, the Fleisch pneumotach sensor heating was switched off during calibration, as was recommended in the manual. As mentioned, the second spirometer was BTPS correct.

Comparability of spirometric data

A total of 43 volunteers participated in the study. To ensure a wide range of volumes and flows, volunteers of different age groups, sex and smoking status were selected. Thirty-seven were healthy, while the remaining six had a confirmed diagnosis of COPD. Most volunteers had never performed spirometry. After detailed instruction by a thoroughly trained assistant, volunteers were allowed a number of practice manoeuvres to get acquainted with the procedure. Only after the volunteers were able to perform qualitatively correct and reproducible manoeuvres (according to the ATS criteria for Maneuvre Performance and Acceptability and Reproducibility, pages 1119-1122), they were asked to perform two forced expiratory manoeuvres from maximal inspiration in each spirometer. The order in which the spirometers were presented was randomized. The four resulting manoeuvres were recorded and analysed pairwise. As the paired observations were collected during one session, the BTPS conditions were comparable.

Comparability in the laboratory

In a lung function laboratory, the accuracy and precision of the spirometers was tested, using a computer-driven syringe (CDS, custom-made by the Jaeger company, Würzburg, Germany). A total number of 110 clustered manoeuvres were simulated with chosen peak flow rates \((1-15 \text{ l s}^{-1})\) and total volumes \((1-5 \text{ l})\). The CDS could simulate 18 out of the possible 26 standard ATS waveforms (4; number 2, 5 14 and 18 24 from table C1, page 1130). The eight remaining waveforms could not be simulated as the specified flows or volumes were out of range.

Analysis

The results from the volume calibration are expressed as percentage differences from 3000 ml. Before assessing the comparability spirometric data, the influence of the randomized order in which the spirometers were presented was tested by means of one-way analysis of variance. The differences in paired observations were evaluated by means of Bland and Altman plots (5,6). The mean difference between, the two spirometers was tested with paired Student's t-tests. A least-square linear regression analysis (no-intercept model) was performed to describe possible systematic differences. Differences between the separate spirometers and the CDS were analysed identically.

Results

VOLUME CALIBRATION OF THE SPIROMETERS

The results from the calibration procedures on 8 consecutive test days are shown in Table 1. As Table 1 shows, all slow and fast procedures resulted in volumes within the ATS recommended 3% range. The plunger speed was of no significant influence with regard to the calibration results. In the last column, the mean deviation is given. The mean deviation of the first spirometer was 0.5% (or 15 ml). The second spirometer deviated by 1.5% (or 45 ml) on average in relatively slow procedures and close to 0% in relatively fast procedures.

COMPARABILITY IN PRACTICE

The forced expiratory volumes in 1 s (FEV\(_1\)) of 43 volunteers covered a wide range, as intended (forced vital capacity (FVC) range: 620-6050 ml, first spirometer).
average age of the volunteers was 37 years, the sex ratio of men/women was 20/23. The mean FEV₁ was 3680 ml (SD = 1347), the mean FVC 4702 ml (SD = 1588) and the mean peak expiratory flow rate (PEFR) was 8.65 l s⁻¹ (SD = 2.63). Analyses of variance showed no significant influence of the order in which the spirometers were presented. With respect to the FEV₁ results, a clinically relevant and statistically significant difference was observed between the two spirometers, indicating poor accuracy. The mean difference was 289 ml (P < 0.001). The differences appeared to be dependent on the average volume and were distributed normally by approximation (see Fig. 1). The regression equation indicated that a systematic 8.6% difference between the spirometers best fitted the data (coefficient P < 0.0001).

COMPARABILITY IN THE LABORATORY

In the laboratory situation, the first spirometer compared well with the CDS. On average, a difference of 33 ml was observed, which was statistically significant but clinically of little importance (Fig. 2). Only at relatively low simulated PEF rates (4 l s⁻¹), did the first spirometer deviate from the CDS volumes: the mean difference between the first spirometer and the CDS volumes were statistically significant and clinically significant but not different from the gold-standard. The spirometer underestimated the true FEV₁ by approximately 122 ml on average. A distinctive volume-dependent trend in the difference can be inferred from the Bland-Altman plot (Fig. 3). These results from the laboratory situation confirm our earlier results in volunteers; the volume-dependent systematic difference. Based on the deviations of the respective spirometers with the CDS (+1.2 and −5.4%), the deviation between the two spirometers would be 6.6%, a percentage approximately similar to the deviation found in volunteers. Evaluation of the simulated standard ATS waveforms to calibrated flows revealed that the first spirometer performed well, whereas the second performed poorly.

Table 1. Results from volume calibration procedures with a 3000 ml hand-held syringe: percentage deviations

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First¹ spirometer: slow³</td>
<td>−2.3⁵</td>
<td>−1.0</td>
<td>2.7</td>
<td>−1.0</td>
<td>−0.7</td>
<td>−0.3</td>
<td>−2.0</td>
<td>1.3</td>
<td>−0.42 (1.7)</td>
</tr>
<tr>
<td>First spirometer: fast⁴</td>
<td>1.3</td>
<td>−0.3</td>
<td>0.7</td>
<td>1.3</td>
<td>0.7</td>
<td>0.7</td>
<td>1.7</td>
<td>1.3</td>
<td>0.50 (1.0)</td>
</tr>
<tr>
<td>Second² spirometer: slow</td>
<td>−1.3</td>
<td>0.0</td>
<td>−1.7</td>
<td>−1.0</td>
<td>−2.7</td>
<td>−1.3</td>
<td>−1.3</td>
<td>−2.0</td>
<td>−1.42 (0.8)</td>
</tr>
<tr>
<td>Second spirometer: fast</td>
<td>0.3</td>
<td>1.0</td>
<td>1.0</td>
<td>0.3</td>
<td>−0.7</td>
<td>1.3</td>
<td>0.7</td>
<td>−3.0</td>
<td>0.04 (1.4)</td>
</tr>
</tbody>
</table>

¹Portable spirometer with heated Fleisch pneumotach sensor.
²Portable spirometer with swirl sensor.
³Slow plunger speed: approximately 500 ml s⁻¹.
⁴Fast plunger speed: approximately 3000 ml s⁻¹.
⁵Percentage deviation from 3000 ml: in this instance the spirometer underestimated the volume by 70 ml.
Fig. 3. Differences in FEV₁ between second spirometer and the computer-driven syringe: 110 paired observations. Mean difference = 122 ml (P < 0.001) underestimation of the second spirometer. Regression equation: second spirometer = 0.9458 × CDS and the computer-driven syringe: 110 paired observations.

Discussion

We have demonstrated that satisfactory results after volume calibration are no guarantee for valid and interchangeable spirometric results. Therefore, relying on volume calibration alone is not recommended as the results of these standard procedures can be misleading. In this example we have shown that, despite meeting equipment criteria for volume calibration, a clinically relevant and statistically significant volume dependent difference in FEV₁ data was observed. The lack of accuracy and precision found in one of the two tested spirometers made the apparatus useless for diagnostic and monitoring purposes.

The ATS equipment recommendations for diagnostic spirometry recommend a ±3% tolerance for vital capacity (VC) as well as for forced manoeuvres (FVC, FEV₁) (4). The test signal for calibration of the VC is a 5000 ml syringe but forced manoeuvres should be tested using 24 well-described standard wave forms. In daily outpatient practice, the latter calibration is hardly feasible. We demonstrated using data from volunteers, measured with two spirometers, that non-compliance with the ATS criteria could be detected and the estimated extent and nature of the difference agreed well with the comparison with a gold-standard. Therefore, in settings where CDS equipment is not easily available, we recommend that spirometric results from volunteers should be regularly compared to results obtained with a second spirometer, on the same occasion. If these results differ repeatedly, one should suspect a lack of accuracy. However, once detected, this still leaves open the question of which spirometer is inaccurate. The only method to determine this remains calibration with a CDS.

Although this study was not designed to explain possible discrepancies between spirometers, there are a number of known underlying reasons which could explain the differences between volume calibration (VC) and forced manoeuvres in a given time (FEV₁). Such reasons include the frequency response of the spirometer, drift phenomena, correction for BTPS circumstances, polytropic conditions within the pump system and software algorithms for detection of the start and end of the test and interpolation of points. The latter reasons, especially may explain the difference in our study. The volume calibration and the CDS manoeuvres were both syringe-driven, yet there was a substantial difference in volume from calibration and FEV₁, produced by the CDS. Assuming that the syringes used were not a source of bias, the differences found were most probably time-related. If the malfunctioning spirometer had difficulties in detecting the right start and end points, this would be likely to have a greater impact in case of shorter observation periods. Delay of the signal will, in fact, have no impact on the total volume but will have a substantial impact on volume during the first second of the manoeuvre. Frequency response (without delay) was unlikely to be the reason. A high plunger speed would then result in an underestimation of the volume where there was no actual difference between the slow and fast calibrations. Drift was also unlikely, for two reasons. The flow-volume curve was displayed on the LCD screen and showed no recorded flow before the manoeuvres. Secondly, the START button of the spirometer was pressed shortly before each manoeuvre, with an approximately constant time interval between START and test. We could not measure the polytropic conditions within the syringes but we consider these conditions very unlikely to cause such a large difference, certainly given the different test sequences and the reliability of the syringes used. As we measured paired results shortly after each other, BTPS circumstances were constant. The Microspiro is BTPS correct, so no adjustments were required.

To obtain the data from the volunteers, we asked them to perform two repeated manoeuvres from maximal inspiration in each of the two spirometers. For diagnostic purposes, two repeated manoeuvres do not meet the ATS recommendations (which require at least three reproducible manoeuvres). Subsequently, the most adequate manoeuvre should be selected as the ‘true’ volume, in accordance with the ATS recommendations (Measurement Procedures: Test Result Selection, page 121). In this case, however, all the paired results were used for calibration purposes only. To ensure reproducibility, all volunteers were allowed a number of practice manoeuvres. When the subjects were able to produce reproducible FEV₁ results, the data collection started. This resulted in a reproducibility well within the recommended 5% margin. The decision to collect two manoeuvres per apparatus per patient was not entirely arbitrary; as a number of patients with obstructive airways disease were selected, more than four repeated forced manoeuvres could lead to invalid comparisons, simply for reason of exhaustion. A drawback of this test sequence is that within-subject variation (or between-manoeuvre variation) is not entirely controlled for. To decrease possible bias, the order in which spirometers are presented should always be random and a number of volunteers, preferably with a wide range of lung volumes, should participate. Our design, using 43 volunteers, showed that the estimated systematic difference agreed very well with the gold-standard. The problem of within-subject variation could be overcome by connecting two spirometers
in series. However, this may result in turbulence and back-pressure phenomena, thus invalidating the test-sequence. In our study, we tried to calibrate a test-sequence where the two spirometers were connected in series, using a rubber tube. Volume calibration produced results well out of the ATS range. Volume calibration may be misleading. The results from volume calibration may meet the ATS criteria, but this is no guarantee that data from forced manoeuvres are accurate. If CDS equipment to simulate standard wave forms is not available, it is recommended that biological calibration is performed regularly and, if possible, that paired data from two (or more) different spirometers are compared.

References