The Accuracy of a Handheld Portable Spirometer*

David A. Rebuck; Nicola A. Hanania, MBBS; Anthony D. D'Urzo, MD; and Kenneth R. Chapman, MD, FCCP

Background: Objective measurement of lung function is considered essential in the management of patients with asthma and COPD. Many primary care practitioners lack the means necessary to obtain these measurements conveniently. To meet this need, electronic spirometers, offering portability, ease of operation, and timesaving readout options have been introduced. We compared the accuracy of a typical pneumotachograph-based device with a conventional volume displacement spirometer.

Methods: We compared indexes of pulmonary function (FVC, FEV\textsubscript{1}, mean forced expiratory flow during the middle half of FVC, [FEF\textsubscript{25-75\%}], and peak expiratory flow rate [PEFR]) measured by the handheld device with those measured by a conventional spirometer in 75 white subjects (33 men, 42 women) with a median age of 43 years (22 to 77 years) who were either healthy or were referred to the pulmonary function laboratory of a large tertiary care teaching hospital. The order of the instrument tested first was randomized and the patients were blinded to which instrument was being studied.

Results: There was a linear relationship between instruments for all indexes measured (r=0.97, 0.98, 0.94, 0.94 for FVC, FEV\textsubscript{1}, FEF\textsubscript{25-75\%}, and PEFR, respectively, for all p<0.001). The random error (precision) was within 5\% only for FEV\textsubscript{1}. The mean of the differences between the values obtained using both instruments (the bias ± limits of agreement (±2 SD) were 0.06±0.56 L for FVC (p=NS), 0.2±0.44 L for FEV\textsubscript{1} (p<0.05), 0.61±1.26 L/s for FEF\textsubscript{25-75\%} (p<0.05), and 0.44±1.9 L/s for PEFR (p<0.05).

Conclusion: Our data suggest that measurements obtained using the pneumotachograph device are closely related to those obtained by volume displacement spirometry and that the handheld device may be useful in clinical practice. However, because the limits of agreement are wide and the difference between the two instrument measurements are significant for FEV\textsubscript{1}, FEF\textsubscript{25-75\%}, and PEFR, the bias between them is not consistent nor is it insignificant. Therefore, the measurements made with the two types of machine cannot be used interchangeably. (CHEST 1996; 109:152-57)

Key words: COPD; pneumotachograph; spirometry

The physical examination is an insensitive means of detecting and monitoring obstructive lung disease. Reviews of asthma deaths point to underestimation of asthma severity by both patients and physicians as a contributing factor so that modern consensus guidelines on asthma management universally recommend objective measurement of expiratory flows to guide therapy.1,2 Similarly, COPD is usually diagnosed late leading to recommendations that screening spirometry be used in high-risk populations.3 Regrettably, surveys show that primary care physicians underuse spirometry often because they do not have ready access to pulmonary function test equipment.4 The availability of a low-cost and easy-to-use handheld spirometer could remedy this problem and improve primary practice management of obstructive lung disease.

A growing number of low-cost electronic spirometers have been developed for use in primary care but it is not clear that these devices are sufficiently reliable for routine clinical use. The American College of Physicians,5 the Association for the Advancement of Medical Instrumentation,6 and the American Thoracic Society (ATS)7 have each published criteria describing adequate spirometer performance. Using these criteria, previous studies of spirometric accuracy have identified several performance problems.8-11 For example, Fitzgerald and coworkers8 found that all seven "electronic" spirometers they tested in 1973 were incapable of measuring FVC and FEV\textsubscript{1} accurately or reproducibly when compared with water-sealed spirometers. As well, Gardner et al8 tested 12 volume-based spirometers and 7 flow-based spirometers. Two thirds of the volume-based devices performed acceptably, while all seven of the flow-based devices had performance difficulties.

The Welch-Allyn Pneumocheck is a comparatively low-cost automated pneumotachograph-based device.

ATS=American Thoracic Society; BTPS=body temperature and pressure, saturated; FEF\textsubscript{25-75\%}=mean forced expiratory flow during the middle half of FVC; PEFR=peak expiratory flow rate

*From the Asthma Centre, The Toronto Hospital, and Division of Respiratory Medicine, University of Toronto, Ontario, Canada. Manuscript received November 3, 1994; revision accepted July 3, 1995.

Reprint requests: Dr. Chapman: Asthma Centre of The Toronto Hospital, 399 Bathurst Street, Suite 4-011 ECW, Toronto, Ontario M5T 2S8 Canada

Claudia C. C. Wu, MD; Robert A. Ter行动计划; Cho; and Dwayne E. Thomas, Jr., MD

Clinical Investigations

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21726/ on 06/27/2017
aimed at making spirometric results readily available in the primary practitioner's office. The specific objective of this study was to determine the accuracy of the Welch-Allyn Pneumocheck as compared with the P.K. Morgan SPIROFLOW 12, a conventional laboratory volume-based spirometer (see Discussion).

METHODS

Instrument

The Pneumocheck (Welch-Allyn Corp; Skaneateles Falls, NY), (received from distributor in June 1993), consists of two physical components, the pneumotachometer (model 61000) and the printer/charger (model 76100). The apparatus measures instantaneous flows that are integrated electronically into volumes. The following indexes are measured or derived: FVC, FEV₁, FEV₁/FVC, mean forced expiratory flow during the middle half of FVC, (FEF₂₅₋₇₅%), and peak expiratory flow rate (PEFR). The measured values, predicted normal values, measured values as percentages of predicted values, the best FVC, the best FEV₁, patient information, and the flow-volume curve are all printed on heat-sensitive paper.

The device will measure as many patient trials as needed according to ATS standard (at least three, two of which must be reproducible to within 5% or 100 mL, whichever is greater). The best trial is selected according to the ATS standard as the test that gives the largest sum of FVC and FEV₁. Four sets of predicted normals selected by dual in-line polarity (DIP) switch settings are Knudson et al (1983); Knudson et al (1976); Crapo et al (1983); or European (1983). The DIP switches also allow a selection of a long or short report as well as the plotting of a flow-volume graph or a volume-time graph. It is also possible to adjust the predicted volume values for different ethnic groups. All measured values are reported at body temperature and pressure, saturated (BTPS). Finally, the Pneumocheck spirometer is reported by the manufacturer to comply with the recommendations of the ATS as they apply to both the spirometer and the printer/charger.

Protocol

Healthy subjects and patients (referred for pulmonary function testing in the laboratory of The Toronto Hospital, Western Site) were considered eligible for participation in the study if they were medically stable, capable of performing repeated spirometric measurements, and capable of offering informed consent. The protocol was reviewed and approved by the Human Subjects Review Committee of the University of Toronto. The subjects were asked to perform two sets of tests, one on the established Morgan spirometer and one on the pneumotachograph-based device. The order of testing was randomized and conducted in single-blind fashion, the subjects being unaware as to which machine was under study. All measurements were taken with the patient in the seated posture with nose clips. No more than three subjects were tested per day with the subjects' test sessions being separated from one another by a time interval of at least 1 h (as the pneumotachograph-based device uses an unheated ceramic sensor, this procedure prevented the significant buildup of moisture within the sensor itself, and provided enough time to allow for the evaporation of any condensation present, to take place). Paired data were taken for the FVC, FEV₁, FEF₂₅₋₇₅%, and PEFR. The number of trials performed and the results recorded for each subject were consonant with ATS standards on number of patient trials and best pick criteria. Measurements taken from the SPIROFLOW 12 were made by computer (IBM PC with Wyvern PFT Software, version 2.33) and corresponding measurements taken from the Pneumocheck were analyzed by its own computer (using Pneumocheck Software, version 3.2). The results were obtained via printouts, as both devices provide this function automatically.

The instrument under study was also calibrated according to manufacturer's specifications. Calibration involved calibration checks and recalibrations according to ATS standards. Both of these procedures utilized a 3-L syringe stored at the same temperature as the instrument. Before either calibration checks or recalibrations, the instrument was flushed with the syringe in order to equilibrate any slight temperature gradient between the air in the cylinder of the syringe and the air within the pneumotachograph. When the device is in the calibration mode, it assumes room air will be used and an ambient temperature and pressure dry factor is applied. During recalibration, three maneuvers with the syringe are completed and the resulting volumes are each compared against the average. Only if all three volumes are within 0.5% of the calculated average does the machine automatically update the calibration value. For a calibration check procedure, a single maneuver with the syringe is completed and the resulting volume of ±3% of the 3-L syringe volume indicates an accurate instrument as specified by ATS.

Analysis

Two types of statistical analysis were applied: the determination of the correlation coefficient (r) to determine the strength of the relationship between the two measurements and the determination of agreement with the descriptive use of bias and limits of agreement.

Strength of Relationship: The paired data were graphed on an x-y plot with the Morgan spirometric results acting as the independent variable and the Welch-Allyn results as the dependent variable. The correlation coefficient was then calculated, and the regression line and 95% confidence intervals were drawn by the method of least squares. The deviation of the Pneumocheck values from the regression line was defined as the "random error," a measure of the instrument precision, and was calculated as 1-r².

Agreement: In the second method of analysis, a plot of the average of the measurements from the two machines, which represents an approximation to the true values, was the independent variable and the difference of the two measurements (volume-based value minus pneumotachograph-based value) was the dependent variable. The mean of the differences was regarded as the bias. The consistency of the bias, a measure of the variation of the differences in readings between the two instruments, was calculated from the mean difference and the SD, the upper limit of agreement (upper control line, UCL) is the mean plus 2 SDs and the lower limit of agreement (lower control line, LCL) is the mean minus 2 SDs. Any difference (or bias) in the values obtained using different instruments can be seen from these calculations, and, provided that the limits of agreement are small, a consistent bias can be taken into account in the interpretation of the results.

To assess whether the accuracy of the Pneumocheck changed over the expected clinical range of lung volumes and flows, the agreement between measured variables was tested over three ranges. The individual ranges for FVC were less than 2 L, 2 to 4 L, more than 4 L, and for FEV₁, less than 1.5 L, 1.5 to 3 L, and more than 3 L. These ranges were clinically based and each range

Table 1—Correlation and Agreement for Indexes Measured

<table>
<thead>
<tr>
<th>Index</th>
<th>Precision</th>
<th>Agreement, L</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVC</td>
<td>5.91</td>
<td>0.06±0.56 (p=NS)</td>
</tr>
<tr>
<td>FEV₁</td>
<td>3.96</td>
<td>0.20±0.44 (p&lt;0.05)</td>
</tr>
<tr>
<td>FEF₂₅₋₇₅%</td>
<td>11.64</td>
<td>0.61±1.26 (L/S) (p&lt;0.05)</td>
</tr>
<tr>
<td>PEFR</td>
<td>11.64</td>
<td>0.44±1.9 (L/S) (p&lt;0.05)</td>
</tr>
</tbody>
</table>
corresponded to low, intermediate, and high, respectively. The agreement of the values within a range was assessed and tested for significance. Significant differences in biases between ranges were tested similarly.

**RESULTS**

A total of 75 adult white subjects (33 men, 42 women) with a median age of 43 years (range, 22 to 77 years) participated. Thirty-five (47%) patients suffered from obstructive disease, 17 (23%) suffered from restrictive disease, and 23 (30%) were healthy. The range of volumes for FVC measured was 1.72 to 6.44 L, and for FEV₁ it was 0.88 to 5.14 L. The range of flows for FEF₂₅₋₇₅% measured was 0.34 to 8.62 L/s, and for PEFR it was 1.53 to 15.16 L/s. On analyzing the correlation coefficients, a strong linear relationship was found between the pneumotachograph-based device and the volume displacement spirometer for all indexes measured (Table 1 and Fig 1). The mean of the differences between instruments (bias) ± the limits of agreement are shown in Table 2 and Figure 2. The random error (precision) was within 5% only for the FEV₁ index (3.96%). FVC, FEF₂₅₋₇₅%, and PEFR had random error values of 5.91%, 11.64%, and 11.64%, respectively.

The relationship between the two instruments was not uniform across the tested range. The agreement for FVC ranges of less than 2 L, 2 to 4 L, and more than 4 L was $-0.13 \pm 0.57$ L ($p=NS$), $0.10 \pm 0.45$ L ($p<0.05$), and $0.05 \pm 1.25$ L ($p=NS$), respectively (Table 3). The bias for the 2 to 4 L range (0.10 L) was significantly higher ($p<0.05$) than the bias for the less than 2 L range ($-0.13$ L), while there was no significant difference ($p>0.05$) between the biases of the less than 2 L and more than 4 L ranges or the 2 to 4 L and more than 4 L ranges.

The agreement for the FEV₁ range of less than 1.5 L, 1.5 to 3 L, and more than 3 L was $0.08 \pm 0.43$ L ($p=NS$), $0.18 \pm 0.30$ L ($p<0.05$), and $0.26 \pm 0.47$ L ($p<0.05$), respectively. There was no significant difference between the bias of the 1.5 to 3 L range and the less than 1.5 L range or between the 1.5 to 3 L range and the more than 3 L range. There was a significant

Figure 1. Scatter plot of volume-based spirometer and corresponding pneumotachograph-based spirometer values (FVC [top left], FEV₁ [bottom left], FEF₂₅₋₇₅% [top right], and PEFR [bottom right]) with line of regression (---), line of identity (---), and 95% confidence intervals (......).
difference in the bias between the less than 1.5 L and more than 3 L ranges.

The agreement for the FEF_{25-75}\% ranges of less than 2.2 L/s, 2.2 to 4 L/s, and more than 4 L/s was 0.2±0.46 L/s (p<0.05), 0.49±0.93 L/s (p<0.05), and 1.15±1.42 L/s (p<0.05), respectively. There was a significant difference between all pairs of ranges.

The agreement for the PEFR ranges of less than 6 L/s, 6 to 8.3 L/s, and more than 8.3 L/s was 0.11±2.02 L/s (p=NS), 0.67±1.23 L/s (p=NS), and 0.55±2.23 L/s (p<0.05), respectively. The only significant difference in bias that existed occurred between the less than 6 L/s and 6 to 8.3 L/s ranges. The biases were insignificantly different between the more than 8.3 L/s and less than 6 L/s ranges as well as between the more than 8.3 L/s and 6 to 8.3 L/s ranges (Table 3).

**DISCUSSION**

We have assessed the performance of the Welch-Allyn Pneumocheck spirometer by comparing it against the widely used Morgan volume-based spirometer. Assessment of performance showed that, over the volume and flow ranges tested, the Pneumocheck was linear although tended to underestimate the Morgan spirometer for all four indexes measured. The instrument was precise to within 5% only for the FEV_{1} index (Table 1). The mean differences (biases) for FEF_{25-75}\% and PEFR were large and all indexes measured, except FVC, were significantly different (p<0.05) between the two instruments. This was particularly so for the PEFR, which is more dependent on the force of the initial portion of the FVC maneuver.20

The SPIROFLOW 12 (P.K. Morgan Limited; Chatham, Kent, United Kingdom), (purchased November 1988), was used as the conventional volume-based spirometer. Nelson et al.10 tested the accuracy of 62 spirometers, three of which were P.K. Morgan devices (the Pocket Respirometer, the SPIROFLOW 12, and the TTC). Dynamic waveform testing was performed using a computer-controlled air pump to inject the 24 ATS standard waveforms* into the spirometers while FVC, FEV_{1}, and FEF_{25-75}\% were measured. Of the P.K. Morgan devices tested, only the model we used, the SPIROFLOW 12, performed acceptably, while the TTC and Pocket Respirometer were judged as performing marginally and unacceptably, respectively. Furthermore, the SPIROFLOW 12 was 1 of 15 spirometers (of the 62 tested) that measured the three

---

**Figure 2.** Agreement of volume-based spirometer and pneumotachograph-based spirometer values (FVC [top left], FEV_{1} [bottom left], FEF_{25-75}\% [top right], and PEFR [bottom right]) with upper and lower control limits (UCL and LCL, respectively).
spirometric parameters on all 24 waveforms without error.

When assessing the agreement of the instrument under study at different ranges of lung volumes, estimated by FVC, it was found that a bias of $-0.13 \pm 0.57$ L indicated an insignificant overestimation at low lung volumes. Conversely, at the intermediate and high lung volume ranges, agreements of $0.10 \pm 0.45$ L and $0.05 \pm 1.25$ L were found indicating significant underestimation. When assessing the agreement of the instrument at different ranges of flows, estimated by FEV$_1$, it was found that a bias of $0.18 \pm 0.30$ L and $0.26 \pm 0.47$ L significantly underestimated the standard in the intermediate- and high-flow ranges, respectively, while a bias of $0.08 \pm 0.43$ L insignificantly underestimated the standard in the low-flow range. Our data found no consistent trend in errors of volume measurement. For measures of flow, however, bias increased with increasing flow rates.

Some possible sources of error pertaining to the observed underestimation of values were the following: the method of the BTPS correction used, the possibility of moisture buildup within the flow sensor, and the possibility of premature termination of flows. Each is addressed below.

The Pneumocheck utilizes a static BTPS correction factor. Hankinson et al$^{16}$ have shown that ceramic sensors in flow-based spirometers are warmed from their initial room temperature with each successive FVC maneuver, resulting in smaller required BTPS correction factors as maneuver order increases. Because of this, spirometers that use static BTPS correction factors are prone to yield results that increasingly overestimate true values. The Pneumocheck does not avoid this erroneous source by taking into consideration sensor temperature before correcting to BTPS; however, the device does contain a thermometer contained within the housing of the pneumotachograph that measures ambient room temperature and can vary the BTPS correction factor from this source only. Regardless, the predicted overestimation of values, made by Hankinson et al.$^{16}$ for a device that uses a static BTPS correction factor was opposite to the observed underestimation of values for all indexes measured.

The buildup of moisture within the flow-sensor is another possible source of error that can be ruled out. However, the resistive element of the flow sensor Pneumocheck utilizes a porous material (Celcor) that absorbs moisture to the extent that any condensation present would not have decreased the cross-sectional area. Significant moisture buildup within the sensor of the pneumotachograph would change the unit’s resistance and either increase the pressure differential (resulting in overestimation of values) or cause a decrease in its linearity.$^{17}$ Neither of these was observed.

Finally, it is possible that a large proportion of patients with obstructive disorders and corresponding low terminal flows could explain some of our observations. However, we doubt that failure to sense low terminal flows accounted for our observation of underestimated indices. Premature cutoff of the flow volume loop was not seen by trained technicians conducting the tests (with either spirometer). Also, the FVC was in closer agreement between devices than other variables; this was particularly true for the lowest range of FVC values measured. ATS standards for end-of-test criteria are the 15-s mark of an FVC maneuver and/or the point in the maneuver in which there is no volume change for at least 2 s with a minimal

### Table 2—Mean, SD, Limits of Agreement, and 95% Confidence Intervals for the Difference Between the Devices

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean, L</th>
<th>SD, L</th>
<th>Limits of Agreement, L</th>
<th>95% Confidence Intervals, L</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVC</td>
<td>0.06</td>
<td>0.29</td>
<td>-0.05 to 0.62</td>
<td>-0.005 to 0.13</td>
</tr>
<tr>
<td>FEV$_1$</td>
<td>0.20</td>
<td>0.22</td>
<td>0.24 to 0.64</td>
<td>0.15 to 0.26</td>
</tr>
<tr>
<td>FEF$_{25,75%}$, L/s</td>
<td>0.61</td>
<td>0.64</td>
<td>-0.65 to 1.87</td>
<td>0.46 to 0.76</td>
</tr>
<tr>
<td>PEFR, L/s</td>
<td>0.44</td>
<td>0.96</td>
<td>-1.46 to 2.34</td>
<td>0.22 to 0.67</td>
</tr>
</tbody>
</table>

### Table 3—Correlation and Agreement for Ranges of FVC, FEV$_1$, FEF$_{25,75\%}$, and PEFR

<table>
<thead>
<tr>
<th></th>
<th>Agreement (Bias±Limits), L</th>
<th></th>
<th>Agreement (Bias±Limits), L</th>
<th></th>
<th>Agreement (Bias±Limits), L/s</th>
<th></th>
<th>Agreement (Bias±Limits), L/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range, L</td>
<td></td>
<td>Range, L</td>
<td></td>
<td>Range, L</td>
<td></td>
<td>Range, L</td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>$-0.13 \pm 0.57$ (p=NS)</td>
<td>&lt;1.5</td>
<td>$0.08 \pm 0.43$ (p=NS)</td>
<td>&lt;2.2</td>
<td>$0.24 \pm 0.46$ (p=NS)</td>
<td>&lt;6</td>
<td>$0.11 \pm 0.02$ (p=NS)</td>
</tr>
<tr>
<td>2-4</td>
<td>$0.10 \pm 0.45$ (p=NS)</td>
<td>1.5-3</td>
<td>$0.18 \pm 0.30$ (p=NS)</td>
<td>2.2-4</td>
<td>$0.49 \pm 0.93$ (p=NS)</td>
<td>6-8.3</td>
<td>$0.67 \pm 1.23$ (p=NS)</td>
</tr>
<tr>
<td>&gt;4</td>
<td>$0.05 \pm 1.25$ (p=NS)</td>
<td>&gt;3</td>
<td>$0.26 \pm 0.47$ (p=NS)</td>
<td>&gt;4</td>
<td>$1.15 \pm 1.42$ (p=NS)</td>
<td>&gt;8.3</td>
<td>$0.55 \pm 2.23$ (p=NS)</td>
</tr>
</tbody>
</table>
We recommend the use of the Pneumocheck spirometer. The Pneumocheck, however, has a flow resolution of 0.016 L, a flow sampling rate of 2,304 samples per second, and a volume resolution of 0.00045 L. Thus, the early termination of flows as a source of error seems improbable.

Our data suggest that measurements obtained using the Pneumocheck are closely related to those obtained with the standard Morgan spirometer and that the former may be useful in clinical practice for pulmonary function assessment. However, the limits of agreement between the two instruments are wide (ie, the bias between them is not consistent) and the difference between their measurements for FVC, FEF_{25-75%}, and PEFR is significant (p<0.05). Thus, they cannot be used interchangeably and results obtained using one cannot be compared directly with those obtained by the other.

This limitation will have practical implications for clinicians who make use of this and similar devices in office practice. Although the handheld device might be useful as a screening tool and monitoring tool, if used repeatedly, its measurements might not be directly comparable to those obtained in a conventional laboratory setting. This would be particularly true when testing patients with marked airflow limitation. A COPD patient undergoing an oral steroid trial would be best assessed by the same device both before and after therapy; office-based measurements and laboratory-based measurements may differ significantly in the absence of physiologic change. Similarly, measurements made with the handheld device cannot be used interchangeably with laboratory-based measurements in clinical trials where the portability of the former device is exploited to allow serial ambulatory measurements.

Despite such limitations, the Pneumocheck handheld spirometer may offer reasonable clinical accuracy for use in the office setting. This may be particularly true in primary practice where the lack of objective pulmonary function measurement is a common reason for underdiagnosis and undertreatment of obstructive airways disease.

ACKNOWLEDGMENTS: We would like to thank Dr. P. Enright of the University of Arizona Health Sciences Center and Dr. J. P. Szalai of the University of Toronto Department of Biostatistics for their recommendations and assistance in the preparation of this manuscript.

REFERENCES