

Comparison of a New Desktop Spirometer (Spirospec) with a Laboratory Spirometer in a Respiratory Out-Patient Clinic

François Swart, Macé M Schuurmans MD, Johannes C Heydenreich, Clarissa H Pieper MD, and Chris T Bolliger MD

BACKGROUND: The performance of spirometers is often evaluated under ideal conditions with computer-generated waveforms or in vivo testing with healthy subjects. Real-life conditions are less ideal because of comorbidities, age of the subjects, and a variety of air flow limitations. Evaluation of new spirometry equipment can also be performed under these less favorable conditions. The Spirospec is a new desktop spirometer that is commercially available, but its accuracy has not been evaluated in a clinical setting. **OBJECTIVE:** Test the Spirospec with subjects with normal and pathologic pulmonary function. **METHODS:** A group of 45 patients (mean age 38.4 years, 27 male) booked for evaluation in the pulmonary function laboratory of a tertiary care university hospital were tested with both a Spirospec and a standard Jaeger Masterlab 4.0 spirometer, according to the guidelines of the American Thoracic Society. Three subgroups (normal spirometry, obstructive air flow limitation, and restrictive air flow limitation) of 15 consecutive subjects each underwent spirometry. **RESULTS:** Pulmonary function measurements from the Spirospec correlated closely ($r = 0.95-0.99$) with those from the Masterlab 4.0, showing good limits of agreement and differences between the 2 devices: forced vital capacity 0.03 L, forced expiratory volume in the first second (FEV_1) -0.01 L, peak expiratory flow -0.41 L/s, peak inspiratory flow 0.43 L/s, forced expiratory flow at 50% of total lung capacity 0.13 L/s, and forced expiratory flow at 75% of total lung capacity 0.12 L/s. With the exception of forced vital capacity and FEV_1 , these differences were statistically significant ($p < 0.05$). **CONCLUSION:** The Spirospec is comparable to the Masterlab 4.0, with high accuracy for FEV_1 and forced vital capacity and clinically acceptable differences in the measured flow variables. *Key words:* spirometry, Spirospec, Masterlab, chronic obstructive pulmonary disease, COPD, pulmonary function testing, PFT. [Respir Care 2003;48(6):591-595. © 2003 Daedalus Enterprises]

Introduction

Many new spirometers are being developed and marketed worldwide. The accuracy of low-cost and easy-to-

use portable spirometers has received more attention in recent years, since it is important to determine whether these devices are sufficiently reliable for routine clinical

SEE THE RELATED EDITORIAL ON PAGE 587

François Swart, Macé M Schuurmans MD, Johannes C Heydenreich, Clarissa H Pieper MD, and Chris T Bolliger MD are affiliated with the Pulmonary Function Laboratory, Tygerberg Hospital, Department of Internal Medicine, University of Stellenbosch, Cape Town, South Africa.

The authors presented a version of this report at the European Respiratory Society Annual Congress, September 22-26, 2001, in Berlin, Germany.

This research was supported in part by grants from the Novartis Foundation and the Swiss National Science Foundation (MMS).

Correspondence: Macé M Schuurmans MD, Lung Unit, Tygerberg Hospital, Department of Internal Medicine, Clinical Building, University of Stellenbosch, Francie van Zijl Drive, 7505 Tygerberg, Cape Town, South Africa. E-mail: mms2@sun.ac.za.

use, for example in the office setting.¹⁻³ A consensus statement from the National Lung Health Education Program recommended widespread use of office spirometry by primary-care providers, to screen patients ≥ 45 years old and who smoke.⁴ Implementing that recommendation requires

None of the persons involved in this study are shareholders of the manufacturer or distributors of the Spirospec. The device was provided free of charge by the distributor for the duration of the study. The study was supported by the distributor by compensating for the extra work involving the additional use of the Spirospec during the study.

accurate and reliable spirometers. Companies developing spirometers generally use computer-generated waveforms for laboratory testing and healthy staff members for in vivo testing. These testing methods assess the accuracy and reproducibility of spirometric measurements only under fairly ideal conditions. Unlike in the laboratory situation, the clinical setting allows spirometric assessment of various pathologic combinations of flow and volume that occur with various lung diseases.^{1,3} Testing spirometers in the clinical setting is challenging because it adds noise to the measurements. The total variability of the measurements is the sum of the individual patient variability (which can be influenced by disease) and the instrument variability.

Hospital-based pulmonary function laboratories are ideal for testing new spirometers, since these laboratories routinely examine patients with various types of air flow limitation and with normal spirometry. The present study compared the accuracy of a new model of desktop spirometer to a standard laboratory spirometer, in a clinical setting with a group of patients exhibiting a range of flow limitations and patients with normal spirometry.

Methods

Devices

We compared the Spirospec desktop spirometer (developed by Vivesco CC, Milnerton, Cape Town, South Africa, and distributed by TDH Medical, Goodwood, Cape Town, South Africa) to the Masterlab 4.0 (Jaeger AG, Würzburg, Germany), a standard diagnostic spirometer that is commercially available worldwide. The Spirospec used in this study was a standard model that is commercially available. Both devices are pneumotachograph-based.

The Spirospec, which is battery-powered, records the flow-volume loop while the patient breathes through a hand-held pneumotachometer that contains a silicone pressure transducer. The spirometer digitizes the data by means of an analogue-to-digital converter and sends the data signal via an RS232 port to a standard Windows-operating-system computer.

The Spirospec measures flow rates with an orifice plate pneumotachometer, and the flow-time curve is processed by the computer, allowing a real-time display of the flow-volume loop on the monitor. Some of the pulmonary function variables derived from the curve are: forced inspiratory vital capacity, forced vital capacity (FVC), forced expiratory volume in the first second (FEV_1), FEV_1/FVC , peak expiratory flow (PEF), peak inspiratory flow (PIF), and the forced expiratory flows at 50% and 75% of FVC (FEF_{50} and FEF_{75}). FEV_1 is determined by back-extrapolation, as recommended by the American Thoracic Society (ATS).⁵ The results of the best curve, after all ATS criteria

are met, are displayed as an absolute value (corrected for body temperature and pressure saturated) and as a percentage of predicted normal.^{5,6} The Spirospec's software program (we used version 1.32) requests a full calibration every day before any flow-time or flow-volume curves can be measured. The calibration procedure consists of delivering various flows with a calibrated 3-L syringe. The manufacturer claims that the device has a volume accuracy of less than $\pm 2\%$ or 50 mL and a flow accuracy of $\pm 2\%$.

The current market price of the Spirospec is 9,600 South African Rands (excluding the value-added tax), which is approximately U.S. \$1,000, for the pneumotachograph and software.

Subjects

The procedures used in this study were performed in accordance with the ethical standards of the World Medical Association Declaration of Helsinki.⁷ Verbal informed consent was obtained prior to inclusion into the study. Eligible subjects were routine patients from the respiratory out-patient clinic of a tertiary care university hospital, where routine pulmonary function testing is performed. Prior to study the subjects were divided into 3 groups, based on medical diagnosis and previous spirometry results: normal spirometry, obstructive air flow limitation (including bronchial asthma), and restrictive air flow limitation. To eliminate any bias in favor of or against either of the devices, subjects were enrolled only if they had *not* previously had any flow-volume determination with either device. Consecutive patients were enrolled until 15 patients were allocated to each group. To avoid selection bias, patients not meeting ATS reproducibility criteria were *not* excluded from the study.

Data Collection

Both spirometers were used with a Pulmoguard filter (SDI Diagnostics, Easton, Massachusetts) and both were calibrated daily with a 3-L syringe. Pulmonary function testing was conducted by respiratory technologists, using ATS testing guidelines.⁶ To achieve a balanced design, each subject performed alternate maneuvers with the Masterlab 4.0 and the Spirospec, resulting in 3 acceptable maneuvers on each device, for a total of 6 maneuvers. The spirometer used for the first maneuver was randomly selected, which distributed the learning effect evenly between the instruments.

Statistical Analysis

The sample size of 45 was based on previous experience with spirometer testing.³ The curves from both devices were assessed for acceptability and reproducibility accord-

COMPARISON OF A NEW DESKTOP SPIROMETER WITH A LABORATORY SPIROMETER

Table 1. Comparison of Measurements from Spirospec and Masterlab 4.0 for Total Group (n = 45)

	Spirospec (mean ± SD)	Masterlab 4.0 (mean ± SD)	Correlation Coefficient (r)	Differences (mean ± SD)*	Limits of Agreement
FVC (L)	3.45 ± 1.65	3.42 ± 1.60	0.99	0.03 ± 0.12	-0.221-0.284
FEV ₁ (L)	2.61 ± 1.46	2.62 ± 1.44	0.99	-0.01 ± 0.10	-0.192-0.213
PEF (L/s)	6.86 ± 3.41	7.27 ± 3.48	0.97	-0.41 ± 0.56†	-0.711-1.527
PIF (L/s)	6.39 ± 2.58	5.96 ± 2.89	0.95	0.43 ± 0.61†	-1.650-0.795
FEF ₅₀ (L/s)	2.86 ± 1.94	2.73 ± 1.97	0.98	0.13 ± 0.34†	-0.820-0.562
FEF ₇₅ (L/s)	1.06 ± 0.87	0.94 ± 0.83	0.98	0.12 ± 0.18†	-0.524-0.205

*Calculated as mean Spirospec reading minus mean Masterlab 4.0 reading.

FVC = forced vital capacity.

FEV₁ = forced expiratory volume in the first second.

PEF = peak expiratory flow.

†p < 0.05 (via paired t test).

PIF = peak inspiratory flow.

FEF₅₀ = forced expiratory flow at 50% of the forced vital capacity.

FEF₇₅ = forced expiratory flow at 75% of the forced vital capacity.

ing to ATS standards. We compared the best values for FVC, FEV₁, PEF, and PIF from all acceptable curves, for both devices. The FEF₅₀ and FEF₇₅ were also compared, but from the test with the highest sum of FVC and FEV₁ from each device. The mean difference of the best 2 efforts (FVC, FEV₁) on each device was calculated to quantify individual variability.

In order to test whether the measurements from the Spirospec correlated with those of the Masterlab 4.0, the correlation coefficient (r) was calculated. Differences were considered statistically significant when p < 0.05. The mean of the differences was analyzed

using the paired t test. To further elucidate whether the measurements from these spirometers could be used interchangeably, we calculated (as described by Bland and Altman) the mean of the differences between the variables measured by the devices.⁸ Results are reported as mean ± standard deviation.

Results

There were 45 subjects (18 female and 27 male) divided into 3 groups of 15 subjects each: (1) normal spirometry (mean age 33.3 y, age range 25-58 y, 12 male subjects),

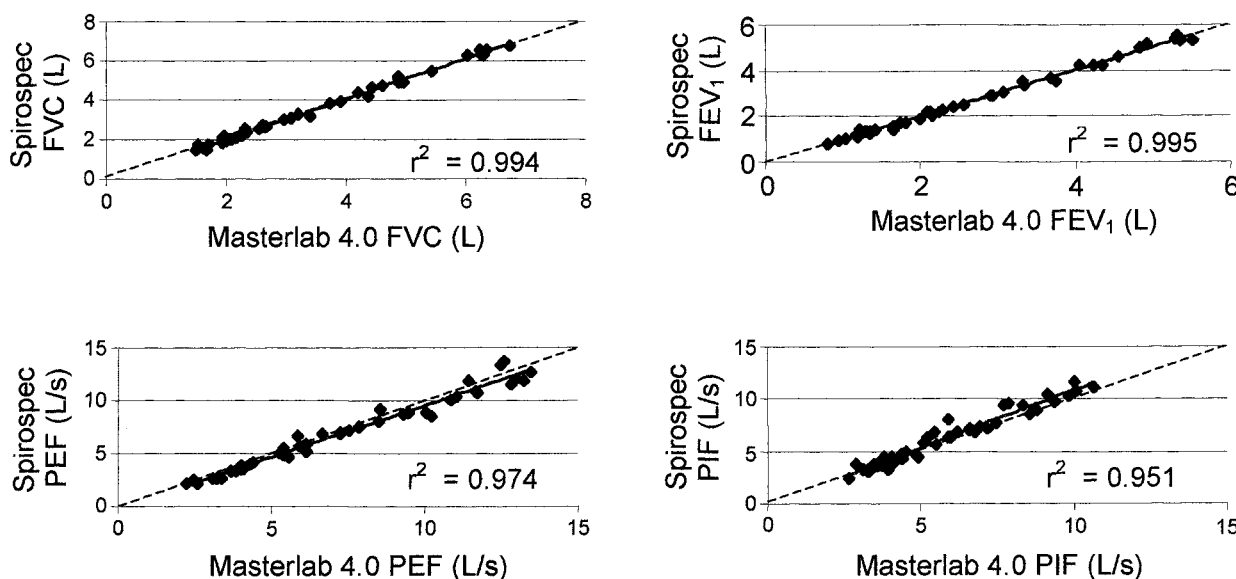


Fig. 1. Correlations between pulmonary function measurements made with the Spirospec and Masterlab 4.0 spirometers. For the total group (all 45 subjects): forced vital capacity (FVC) r = 0.99, regression equation y = 1.0244x - 0.0519; forced expiratory volume in the first second (FEV₁) r = 0.99, y = 1.0129x - 0.0442; peak expiratory flow (PEF) r = 0.97, y = 0.9678x - 0.178; peak inspiratory flow (PIF) r = 0.95, y = 1.1007x - 0.1722. The solid lines are the regression lines. The dashed lines are the identity lines.

COMPARISON OF A NEW DESKTOP SPIROMETER WITH A LABORATORY SPIROMETER

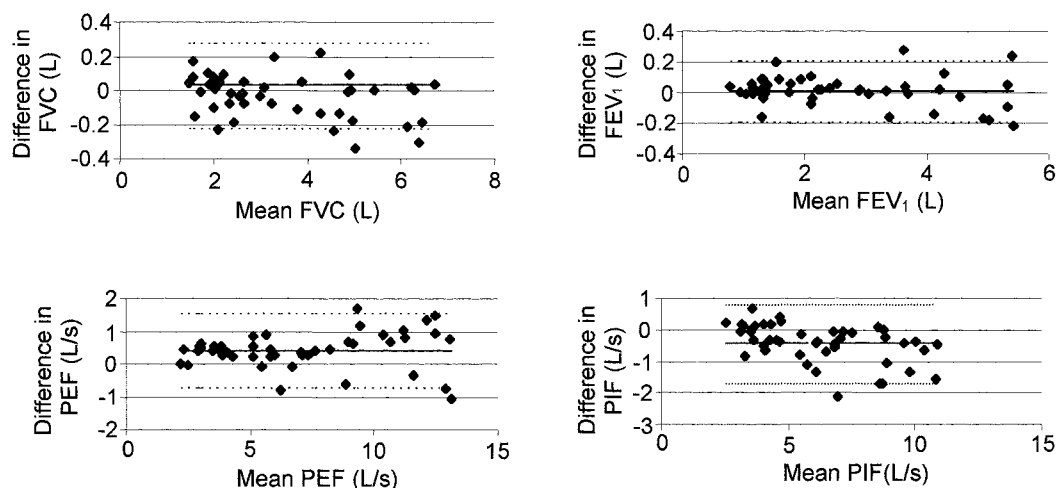


Fig. 2. Differences versus mean values for pulmonary function measurements made with the Spirospec and Masterlab 4.0 spirometers. The solid lines represent the mean differences. The dotted lines represent the upper and lower limits of agreement (± 2 SD). FVC = forced vital capacity. FEV₁ = forced expiratory volume in the first second. PEF = peak expiratory flow. PIF = peak inspiratory flow.

(2) obstructive air flow limitation (7 with chronic obstructive airway disease and 8 with bronchial asthma, mean age 43 y, age range 19–57 y, 8 male subjects), and (3) restrictive air flow limitation (mean age 39 y, age range 23–69 y, 7 male subjects).

All pulmonary function tests recorded met ATS acceptability criteria with both devices. The ATS-reproducibility criteria (FEV₁ and FVC of the 2 best curves of each device differing < 5% or < 200 mL) were met by 43 subjects (95%) with the Masterlab 4.0 and by 43 subjects (95%) with the Spirospec. All patients were able to fulfill reproducibility criteria on at least one of the 2 devices. The mean differences between the best 2 attempts on each device for FVC were 70.6 \pm 68 mL for the Masterlab 4.0, and 78.4 \pm 78 mL for the Spirospec (a nonsignificant

difference). For FEV₁ the mean differences were 55.1 \pm 60 mL for the Masterlab 4.0, and 56.7 \pm 51 mL for the Spirospec (also a nonsignificant difference). The Spirospec correlated well with the Masterlab 4.0 (*r* value range 0.95–0.99) for FVC, FEV₁, PEF, PIF, FEF₅₀, and FEF₇₅ (Table 1). Figure 1 shows the correlations and differences between the readings for FVC, FEV₁, PEF, and PIF for all subjects. Figure 2 plots (with the method described by Bland and Altman⁸) the differences versus the mean values for FVC, FEV₁, PEF, and PIF.

Significant differences of the means were found for PEF, PIF, FEF₅₀, and FEF₇₅, using the paired *t* test. All measurements except PEF read lower with the Masterlab 4.0 (see Table 1). There were statistically significant differences for some variables in all 3 subgroups (Table 2).

Table 2. Comparison of Measurements from Spirospec and Masterlab 4.0 for Subgroups (*n* = 15 each)

	Normal		Obstructive		Restrictive	
	Spirospec	Masterlab 4.0	Spirospec	Masterlab 4.0	Spirospec	Masterlab 4.0
FVC (L)	5.31 \pm 1.03	5.26 \pm 0.99	2.82 \pm 1.20	2.76 \pm 1.11	2.23 \pm 0.56	2.23 \pm 0.54
FEV ₁ (L)	4.39 \pm 0.84	4.37 \pm 0.82	1.66 \pm 0.84*	1.68 \pm 0.84	1.79 \pm 0.49	1.81 \pm 0.48
PEF (L/s)	10.30 \pm 2.62*	10.71 \pm 2.68	4.07 \pm 1.94*	4.39 \pm 1.95	82.22 \pm 5.95*	82.71 \pm 6.90
PIF (L/s)	9.06 \pm 1.86*	8.19 \pm 1.82	4.59 \pm 1.63	4.39 \pm 1.47	6.20 \pm 2.13	6.72 \pm 2.30
FEF ₅₀ (L/s)	5.13 \pm 1.24	4.94 \pm 1.43	1.56 \pm 0.79*	1.00 \pm 0.82	2.28 \pm 0.84	2.24 \pm 0.89
FEF ₇₅ (L/s)	2.13 \pm 0.56*	1.90 \pm 0.62	0.41 \pm 0.28*	0.32 \pm 0.31	0.65 \pm 0.35*	0.50 \pm 0.32

FVC = forced vital capacity.
 FEV₁ = forced expiratory volume in the first second.
 **p* < 0.05 (via paired *t* test).
 PEF = peak expiratory flow.
 PIF = peak inspiratory flow.
 FEF₅₀ = forced expiratory flow at 50% of the forced vital capacity.
 FEF₇₅ = forced expiratory flow at 75% of the forced vital capacity.

Discussion

The Spirospec correlated very well with the Masterlab 4.0 for all the measured variables. Although the difference of the means (Spirospec minus Masterlab 4.0) was significant for 4 variables, they were small in absolute values, the largest being 0.43 L/s for PIF. In Figure 2 the panel showing PIF shows a clustering of values above the mean difference in the lower range of the mean PIF. This suggests a range-dependent error, but the magnitude of this error appears to be small, since all the values fall within 2 standard deviations of the mean. At higher PIF values the readings of the test device are more accurate. The analysis of the limits of agreement for all 6 variables showed that, on average, only 3 of 45 measurements (4%) lay outside 2 standard deviations, which is generally accepted as evidence of good agreement.

Performance of the pulmonary function test maneuvers by patients and the conducting of the tests by the technologists did not show any particular shortcoming of the Spirospec or its software.

The significant differences of the means found for PEF, PIF, FEF₅₀, and FEF₇₅ (with all measurements except PEF reading lower with the Masterlab 4.0) indicate that the PEF, PIF, FEF₅₀, and FEF₇₅ results cannot be used interchangeably, and the values obtained with one of the devices cannot be directly compared with those obtained by the other. The differences in PEF, PIF, FEF₅₀, and FEF₇₅ have only limited practical implications for clinicians, because, with the exception of PEF, the values concerned are not often compared or used for monitoring patients.

A limitation of this study was the small sample size ($n = 45$), which resulted in a low power and increased the possibility of committing a type II error. Another limitation was the inclusion of 4 subjects who did not meet ATS reproducibility criteria, which may have introduced additional noise, increasing the variability of the results. Other studies with similar objectives have compared 2 devices with sample sizes of 45–75.^{1–3} A comparison of a turbine-based portable spirometer (MicroPlus) with a pneumotachograph-based diagnostic spirometer (SensorMedics Vmax22) in normal subjects and patients with obstructive airway limitation showed that the MicroPlus underestimated FVC and that there was a greater mean difference between the instruments for FEV₁ and PEF with obstructed-airway subjects.² The present study also shows a higher mean difference between the instruments for FEV₁ in obstructed-airway patients, whereas the difference between instruments is significant for PEF in all subgroups (see Table 2).

Rebuck et al compared a pneumotachograph-based device (Pneumocheck) and a volume displacement spirometer (Spiroflow 12) and showed good correlation between the devices, with r values ranging from 0.94 to 0.98 for all

variables measured, but there were significant mean differences between the 2 devices for FEV₁, PEF, and forced expiratory flow in the middle half of the FVC (FEF_{25–75}).¹

Maree et al compared a new desktop spirometer (Diagnosa) and a laboratory diagnostic spirometer (Masterlab 4.0) with normal subjects and patients with obstructive and restrictive air flow limitation.³ These 2 devices showed a close correlation (r range 0.92–0.99) for all variables measured, but there were significant differences between the devices for FVC, FEV₁, FEF₂₅, and FEF₇₅.

The latter 3 studies all found good correlation of the spirometry results, but because of significant mean differences (biases) between the devices, the results from one of the spirometers cannot be used interchangeably with those of the other. This also applies to the present study for all the studied pulmonary function variables except FEV₁ and FVC.

The clinically important measurements, for both screening and monitoring, are predominantly FEV₁ and FVC, and the Spirospec and Masterlab 4.0 showed excellent correlation ($r = 0.99$) and very good limits of agreement for FEV₁ and FVC. For FEV₁ and FVC the Spirospec and the Masterlab 4.0 could be used interchangeably. Nevertheless, it is generally recommended to retest a patient with the same apparatus for values that may have therapeutic implications. For example, it is advisable to use the same device before and after prednisone treatment, with a patient undergoing an oral steroid trial.¹

Conclusion

The Spirospec is comparable to the Masterlab 4.0 for the determination of FEV₁ and FVC in clinical practice.

REFERENCES

1. Rebuck DA, Hanania NA, D'Urzo AD, Chapman KR. The accuracy of a handheld portable spirometer. *Chest* 1996;109(1):152–157.
2. Caras WE, Winter MG, Dillard T, Reasor T. Performance comparison of the hand-held MicroPlus portable spirometer and the SensorMedics Vmax22 diagnostic spirometer. *Respir Care* 1999;44(12):1465–1473.
3. Maree DM, Videler EA, Hallauer M, Pieper CH, Bolliger CT. Comparison of a new desktop spirometer (Diagnosa) with a laboratory spirometer. *Respiration* 2001;68(4):400–404.
4. Ferguson GT, Enright PL, Buist AS, Higgins MW. Office spirometry for lung health assessment in adults: a consensus statement from the National Lung Health Education Program. *Respir Care* 2000;45(5):513–530.
5. Standardization of spirometry, 1994 update. American Thoracic Society. *Am J Respir Crit Care Med* 1995;152(3):1107–1136.
6. Turney SZ, Blumenfeld W. Heated Fleisch pneumotachometer: a calibration procedure. *J Appl Physiol* 1973;34(1):117–121.
7. World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *Respir Care* 1997;42(6):635–636.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between 2 methods of clinical measurement. *Lancet* 1986;1(8476):307–310.