

Accuracy of Office Spirometry Performed by Trained Primary-Care Physicians Using the MIR Spirobank Hand-Held Spirometer

J. Degryse^{a, b} J. Buffels^a Y. Van Dijck^c M. Decramer^d B. Nemery^c

^aDepartment of Primary Health Care, ^bInstitut de Recherche Santé et Société, Université Catholique de Louvain,

^cDepartment of Occupational Medicine, and ^dDepartment of Pulmonology, KU Leuven, Leuven, Belgium

Key Words

Office spirometry • Validity • Reliability • Primary care

Abstract

Background: With the availability of compact, portable, effective microspirometers, pulmonary function tests no longer need to be performed only in specialized laboratories. However, the perception persists that small flow-sensing devices are less accurate than volume-sensing spirometers.

Objectives: To study the accuracy of spirometry performed with the MIR Spirobank[®] and to investigate how accurately trained primary-care physicians can perform spirometry using a portable electronic spirometer. **Methods:** Patients with suspected occupational asthma were submitted to specific bronchial challenge tests in the pulmonary function laboratory according to published recommendations. Serial measurements were performed with the Jaeger MasterScope device (reference standard) or the Spirobank device. Data were generated from 908 parallel measurements on 34 patients. Furthermore, 16 patients with documented moderate to severe COPD were examined in a carousel set-up by four trained physicians who each used his/her own Spirobank device coupled to a laptop computer. **Results:** The Spirobank spirometer performed very well compared with the Jaeger MasterScope in a laboratory environment, displaying an underestimation of the forced expiratory volume in 1 s (FEV₁)

and FEV₁/forced vital capacity (FVC) of 2–5%. High correlations were found for the pulmonary function parameters. The highest correlation was for FEV₁ ($r^2 = 0.949$) and the lowest for the maximum expiratory flow at 25% of FVC (MEF₂₅) ($r^2 = 0.864$). Only 2% of the observed variation in the measurement results could be explained by the type of device.

Conclusions: The Spirobank device seems to be appropriate for research purposes if the standardized protocol is used correctly and the acceptability criteria are respected.

Copyright © 2012 S. Karger AG, Basel

Introduction

With the availability of compact, portable, effective microspirometers, pulmonary function tests no longer need to be performed only in specialized laboratories. However, because early models of small flow-sensing spirometers were less accurate than volume-sensing spirometers, the perception persists that even the current fourth-generation models are less accurate [1]. Recently, questions have been raised again about the quality of ambulatory spirometry performed outside a pulmonary function laboratory [2]. The performance of new spirometers is generally evaluated using computer-generated waveforms for laboratory testing and using medical staff members for in vivo testing. Unlike the laboratory situa-

tion, the clinical setting allows spirometric assessment of the pathological flow and volume combinations that occur in various lung diseases [3]. Testing spirometers in the clinical setting is challenging because it adds noise to the measurement. In this paper, we will focus on whether the results of the measurements with modern, portable spirometers of one specific type are accurate as well: in other words, whether they are reliable and valid.

The accuracy of the measurements can be approached from three different angles. First, the *reproducibility* of the measurement results can be examined. This depends, among others, on the technical properties of the equipment used, the standardization of the execution, the possible quality controls included in the software and the feedback messages generated by algorithms. Second, the *validity* of the measurements can be examined. Volume calibration alone is insufficient. The validity of the measurements should be examined by organizing a series of parallel measurements compared with a reference standard [4, 5]. Third, the *reliability* of the measurements can also be tested from a broader perspective with regard to the kind of changes that need to be measured or the kind of decisions that can be taken based on the measurement.

The central question for the study of validity is how the compact, electronic spirometers can relate to a reference standard: for example, a pulmonary function test executed with a well-calibrated pneumotachograph under the direction of an experienced technician. Several researchers have already demonstrated that pulmonary function values show discrete and systematic underestimations [4, 6–8] when measured using compact turbine spirometers. Moreover, the observed differences increase with the increase in the value of the pulmonary function parameters [4, 7, 9]. van den Boom et al. [5] also reported on this nonlinearity of modern, compact spirometers.

An important condition for generating useful measurement results, regardless of the technical quality of the equipment being used, is the correct execution of the measurement. In this respect, the criteria of the American Thoracic Society (ATS) [10] are often referenced. When providing training and standardization, a high percentage of pulmonary function tests seem to meet the ATS criteria in pulmonary function laboratories whatever the technique applied. This is not as obvious in an ambulatory test environment. Research in 15 practices of general practitioners in New Zealand showed that after 4 months, only one third of all the performed pulmonary function tests met the ATS criteria, despite preliminary training of the general practitioners and their assistants [11]. Leuppi et al. [12] analyzed a total of 29,817 office spi-

rometries performed by 440 primary-care physicians and report an acceptable quality grade A, B or C of 60.1%. However, even in a hospital setting, the quality of spirometry performed outside the pulmonary function laboratories is not always adequate [2].

Another problem can involve the possibility and method of calibration. Poor volume calibration can lead to incorrect results and misleading conclusions [5]. Validation by means of computer-controlled simulation equipment as recommended by the ATS (standardized volume waveform testing) [10, 13] is hardly ever performed in Belgium.

Recently, Liistro et al. [14] examined the user friendliness and the validity of 10 different microspirometers. It was a limited study with a limited number of measurements in a laboratory in which the most important differences between the tested devices were demonstrated. However, no comparative studies have been performed for the MIR Spirobank® model (www.spirometry.com). The manufacturers (MIR, Rome, Italy) have delivered a certificate of conformity based on research performed by R. Crapo [pers. commun.]. However, the results of this research have not been published or released. The goal of the present research was to examine the accuracy of the pulmonary function tests performed using the Spirobank microspirometer in real-life conditions and specifically to establish the contribution of different sources of measurement error in case data are gathered within the context of a multicenter study by different investigators using their own similar equipment examining patients at different times of the day.

Research Questions

- (1) What is the accuracy of spirometry performed with the Spirobank? How do the measurements obtained with this spirometer relate to those performed with a pneumotachograph (Jaeger MasterScope) in a pulmonary function laboratory?
- (2) How accurately can trained primary-care physicians perform spirometry with a portable electronic spirometer and how do the measurements obtained by four primary-care physicians using their own devices relate to each other?

Methods

Study 1

In 1999 and 2000, 42 patients with suspected occupational asthma were submitted to specific bronchial challenge tests in the pulmonary function laboratory of the UZ Gasthuisberg (Katho-

lieke Universiteit Leuven) according to published recommendations. During the examination, serial measurements were performed with the Jaeger MasterScope device (reference standard) or the Spirobank device. This test was carried out with the purpose of validating planned measurements of lung function at home. Data were generated from 908 parallel measurements on 34 patients. The complaints of all patients suggested obstructive pulmonary disease and all patients had been exposed to potential toxic agents at work. All measurements were performed serially, always starting with the reference measurement using the Jaeger device. The second measurement with the portable Spirobank was performed in the first 2 min after the first measurement. This was not a truly parallel measurement in which both devices would be connected to the same mouthpiece. All measurements were performed under standardized conditions by the same experienced operator according to the ATS recommendations [13].

The same Jaeger MasterScope (model XC) was used at all times and a volume calibration was performed daily in the morning. A heated Jaeger pneumotachograph was used to determine inspiratory flow and volumes accurately. The system was completed using a Roc type occlusion shutter resistance system. The reference values were calculated according to the European Respiratory Society prediction equations. MIR delivered four Spirobank spirometers and they were used at random for the tests. The devices were checked every 3 weeks with a 3-liter calibration pump and the deviation was never allowed to be higher than 5%. The Spirobank device is a pocket spirometer, which can work autonomously as well as in real time when coupled to a personal computer. This device is equipped with an infrared mini flow sensor to measure both the flow and the volume and an internal temperature sensor for BTPS. The flow sensor is a bidirectional digital turbine and uses and infrared interruption mechanism. The maximum volume that can be measured is 10 liters, maximum flow range ± 16 liters/s. The manufacturer reports a volume accuracy of $\pm 3\%$ or 50 ml/s whichever is greater and a flow accuracy of ± 200 ml/s, whichever is greater. The device is connected to a personal computer and the sophisticated software program (WinspiroPro[®]) generates immediate visual and numerical feedback on the acceptability and reproducibility of different tests using a series of internal algorithms.

Study 2

Sixteen patients with documented moderate to severe chronic obstructive pulmonary disease were examined in a carousel set-up by four trained physicians who each used his/her own Spirobank device coupled to a laptop computer. Within the carousel, all patients were examined by an experienced technician who used the Jaeger MasterScope device. To reduce the effect of fatigue as a systematic measurement error to a minimum, it was ensured that the patients started and ended in different places in the carousel. The partaking doctors were asked to deliver curves and measurements of the best quality and to respect the instructed ATS criteria. Using this design, both the interobserver variability and the validity of the measurements performed by four doctors were examined.

Statistical Analysis

The data were analyzed using MedCalc version 1.4 (www.medcalc.be) and SPSS V12 (SPSS Inc.). Pearson correlation coefficients were calculated to explore the relationships between the

obtained measurements. Subsequently, regression analyses were applied and the data obtained with the Spirobank were treated as independent data. The data were analyzed graphically using Bland and Altman plots [15]. A generalizability analysis [16, 17] was performed to analyze the contribution of the several potential sources of error in the measurements. A generalization study was run by means of the urGenova program [18] to estimate the contribution of the patient to total variance. For measurements in clinical settings, generalizability theory offers a framework to estimate the magnitude of multiple sources of error and to assess the reliability of measurements tailored to specific clinical applications. The theory offers a framework in which these different conditions can be related to each other to subsequently assess their impact or contribution to the reliability of the tests [19].

The study protocol was approved by the Ethical Board of the Medical School of the University of Leuven (26/10/2006/198).

Results

Study 1

Patient Population

There were data for 908 parallel measurements on 34 different patients. Their ages ranged from 19 to 56 years (mean 42.2; SD ± 9.8). There were 7 women and 27 men. The numbers of measurements on any one patient ranged from 7 to 58 (mean 26 ± 13). Of the 34 subjects, 15 showed a positive reaction to the test, the others a negative one. A bronchial challenge test was considered positive if there was a decrease in the forced expiratory volume in 1 s (FEV₁) measurement of 20% or more. On the first (blank) test day we measured a mean FEV₁/forced vital capacity (FVC) ratio of 72.6% (SD ± 10.26) with the Jaeger device in 34 subjects. One patient showed severe airflow obstruction before bronchial provocation (FEV₁/FVC = 41.8%).

Intersubject Variation as a Source of Error

We tested to what extent the observed differences between the two devices were affected by differences between patients. Figure 1 shows the distribution of the differences in the observed measurements (FEV₁, FVC and FEV₁/FVC ratio) as box plots. On the horizontal axis ('Patient No.'), the subjects are shown along with the number of parallel measurements. The vertical axis (Diff_{FEV₁}, Diff_{FVC} and Diff_{FEV₁/FVC}) shows the magnitude of the differences between the measured values obtained with the Jaeger and Spirobank devices. The outliers are indicated by the sequence number of the test concerned. The patient factor was mostly negligible as there was a similar distribution for all 34 test subjects with a median of all measurements around zero.

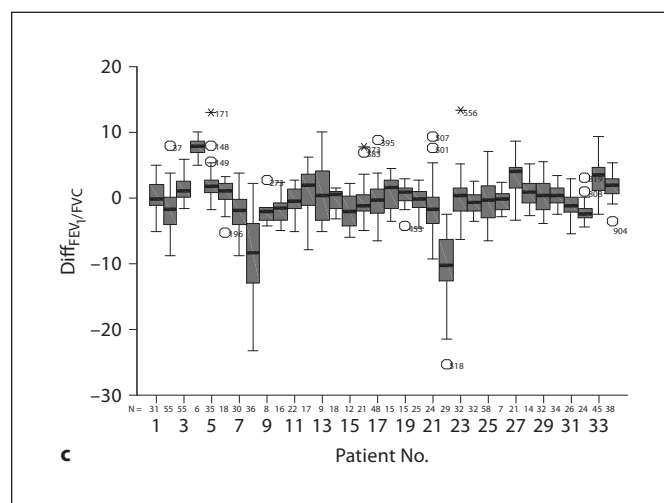
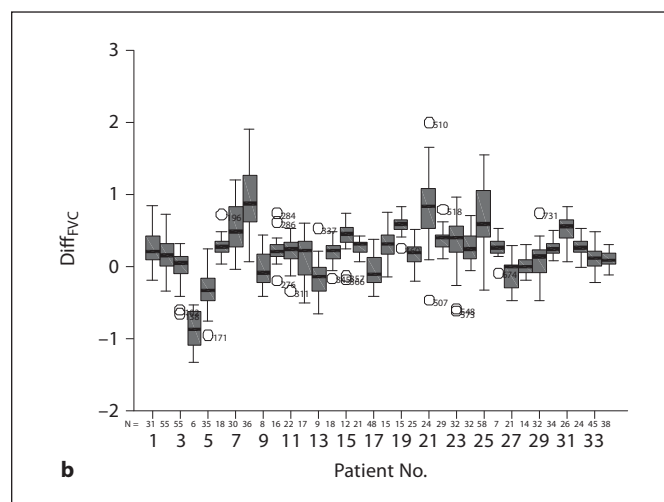
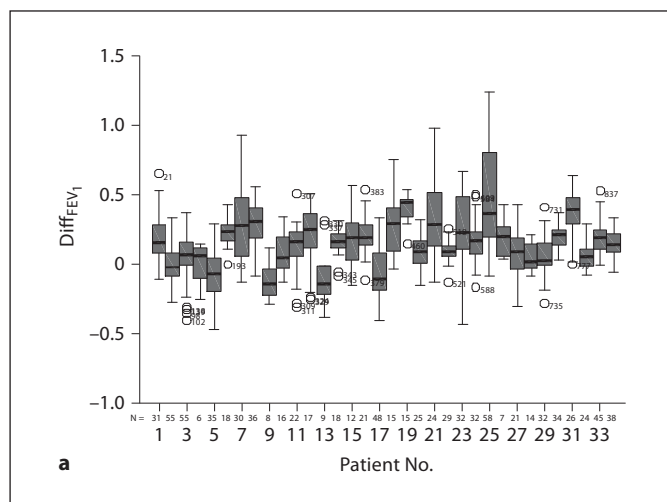


Fig. 1. Intersubject variation as a source of error. Distribution of the differences in the measurements (FEV₁, FVC and FEV₁/FVC). On the horizontal axis, the patients are shown with the number of parallel measurements. The vertical axis shows the magnitude of the differences between the measured values obtained with the Jaeger and Spirobank devices. **a** FEV₁. Diff_{FEV₁} = Difference between parallel measurements of FEV₁ (liters). **b** FVC. Diff_{FVC} = Difference between parallel measurements of FVC (liters). **c** FEV₁/FVC. Diff_{FEV₁/FVC} = Observed differences between parallel measurements in FEV₁/FVC(%).

Figure 2 shows the distribution of the observed differences between the FEV₁ values measured with the two devices, expressed as a percentage of the FEV₁ measured by the gold standard, i.e. the Jaeger device. Here, the overall difference was also around zero.

Correlation between the Observed Differences and the Basal FEV₁ Value

We tested whether there was any correlation between the extent of the observed differences and the basal FEV₁ measured with the Jaeger device. Figure 3 shows the results of a regression analysis with a 95% confidence interval (95% CI) between the amplitude of the observed differences and the basal pulmonary function as measured on day 1 with the Jaeger device. No significant correlation was observed ($r^2 = 0.00$).

Time of Measurement as a Source of Intrasubject Variability

In the database of 908 parallel measurements, 277 were basal measurements performed on day 1 (the day preceding the actual challenge tests) on each of the 34 patients at different times of the day. The variation in the observed measurements is caused by diurnal variations in pulmonary function and by errors in measurement, which can be the result of inaccuracies, and by individual variations in execution of the exhalation process. A learning effect of the tested individuals can play a role. Variation can also arise from the inherent properties of the devices. Figure 4 shows the distribution of the registered measurements with the Jaeger and the Spirobank devices for every patient. There was a significant difference only for 6 subjects in that the 95% CI did not overlap

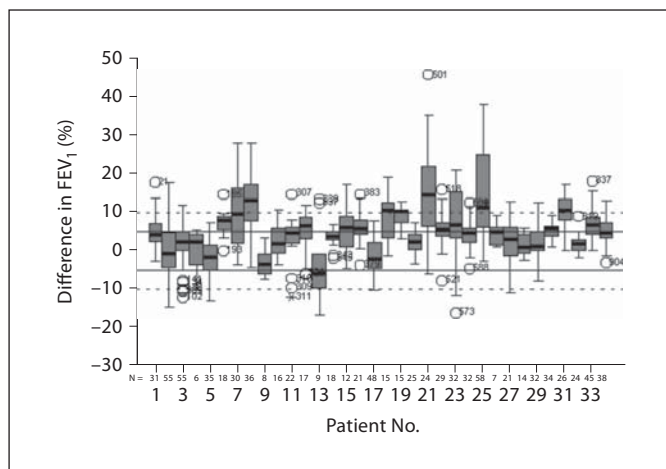


Fig. 2. Distribution of the observed differences between the FEV₁ values measured with the two different devices and expressed as a percentage of those measured with the Jaeger device. On the horizontal axis, the patients are shown with the number of parallel measurements. On the vertical axis the differences between the measurements of FEV₁ with the two devices are expressed in percentages of the FEV₁ measured with the Jaeger pneumotachograph.

(patients 11, 12, 16, 27, 30 and 31). The variation in measurements caused by differences in the devices was smaller than the intrasubject variation.

Validity Study: Correlations

Table 1 shows the Pearson correlations and the mean observed differences between measurements performed with the Jaeger and the Spirobank devices. High correlations were found for the pulmonary function parameters. The highest correlation was for FEV₁ ($r^2 = 0.949$) and the lowest for maximum expiratory flow at 25% of FVC (MEF₂₅) ($r^2 = 0.864$). Although statistically significant, the absolute values of the mean differences were relatively small.

Validity Study: Bland and Altman Plots

The data were also investigated using a Bland and Altman plot [15] as a statistical method to compare two measurement techniques. In this graphical method, the differences (or, alternatively, the ratios) between the two techniques are plotted against the means of the two techniques. Horizontal lines show the mean difference, and at the mean difference, ± 1.96 times the SD of the differences. If the differences within this range are not clinically important, the two methods may be used inter-

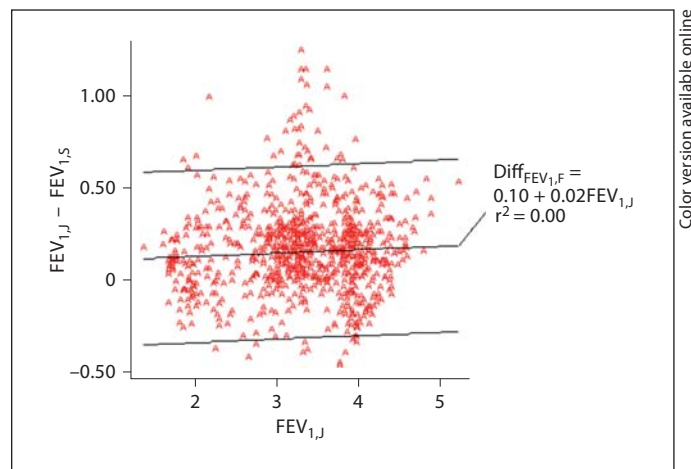


Fig. 3. Regression analysis (linear regression with 95% CI) between the amplitude of the observed differences in FEV₁ measured with the two devices and the basal FEV₁ measured with the Jaeger device (FEV_{1,J}). Vertical axis: differences in FEV₁ values measured with the Jaeger device and the Spirobank device FEV_{1,S} (liters) (FEV_{1,J} - FEV_{1,S}); horizontal axis: initial amplitude of FEV₁ measured with the Jaeger device in liters.

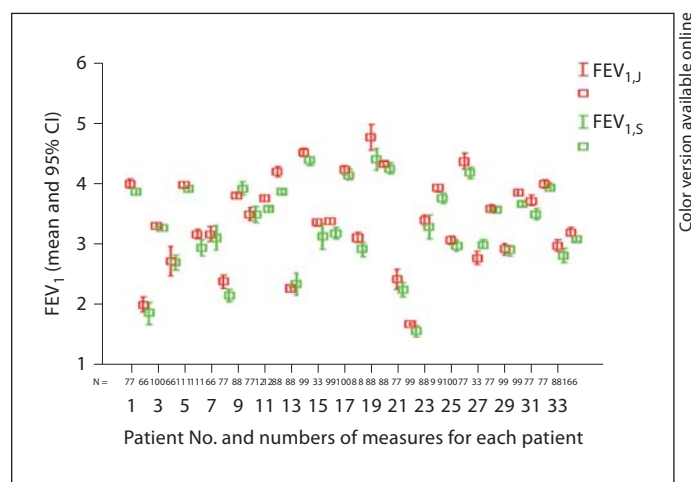


Fig. 4. Distribution of the measurements performed at different times of the day before the bronchial challenge test (means and 95% CI). FEV_{1,J} = Measurements with the Jaeger device; FEV_{1,S} = measurements with the Spirobank device. On the horizontal axis the patients are shown with the number of parallel measurements. On the vertical axis, mean FEV₁ is represented as a box with a graphical plot of the 95% CI; measurements with the Jaeger device are shown in red, whereas measurements with the Spirobank device are shown in green. Colors refer to the online version only.

Table 1. Comparison of the results from the Jaeger and Spirobank equipment (908 measurements on 34 patients): correlations and mean observed differences

Parameter	Jaeger mean \pm SD	Spirobank mean \pm SD	Correlation coefficient	Difference mean \pm SD	p
FVC, liters	4.61 \pm 0.88	4.36 \pm 0.98	0.918*	0.24 \pm 0.39	0.000
FEV ₁ , liters	3.28 \pm 0.72	3.12 \pm 0.75	0.949*	0.15 \pm 0.23	0.000
PEF, liters/s	8.00 \pm 1.74	7.82 \pm 1.72	0.901*	-0.48 \pm 4.16	0.000
MEF ₂₅ , liters/s	1.01 \pm 0.47	1.03 \pm 0.43	0.864*	-0.020 \pm 0.23	0.010
MEF _{25/75} , liters/s	2.41 \pm 0.95	2.48 \pm 0.91	0.918*	-0.071 \pm 0.38	0.000
FEV ₁ /FVC, %	71.66 \pm 11.24	72.15 \pm 10.86	0.929*	0.17 \pm 0.77	0.000

* p < 0.01 (two-tailed) statistically significant.

Table 2. Generalizability analysis of the data in an (i:p)•m design (p = persons, i = instant of measurement, m = measurement method)

Effect	Degrees of freedom	Uncorrected sums of squares	Sums of squares	Mean squares	Variance component	Variance component, %
p	33	19,573.20	895.47	27.13	0.50483	86.4
i:p	874	19,654.62	81.42	0.09	0.03657	6.3
m	1	18,688.80	11.08	11.08	0.01177	2.0
pm	33	19,594.67	10.39	0.31	0.01113	1.9
im:p	874	19,693.61	17.50	0.02	0.02003	3.4

Two percent of the observed variation in the measurement results can be explained by the type of device, whereas 6.3% of the variation arose because subjects were examined at different times of the day.

changeably. The plot is useful to reveal a relationship between the differences and the averages, to look for any systematic biases and to identify possible outliers (fig. 5a, b).

Only a few values fell outside the 95% CI (fewer than 5%). Thus, the suggested nonlinearity of the microspirometers [5] was not confirmed. For FVC, the differences appeared to be greater (between 3 and 4 liters), but the dots in this range are only represented by the values of 2 subjects. Compared with the Jaeger device, the FEV₁ and FVC values were underestimated by the Spirobank device by 5–6%. For FEV₁/FVC, there was a mean difference of less than 1%.

Generalizability

A generalizability analysis was performed using the urGenova program on the asymmetrical dataset (the numbers of parallel measurements per patient varied).

An (i:p)•m design was used, in which p = person, m = method and i = instance (time) of measurement. Table 2 shows the results of this analysis. Less than 2% of the observed variation in the measurement results could be explained by the type of device. In other words, 98% of the variation was not related to the use of two different types of devices, whereas 6.3% of the variation arose because subjects were examined at different times of the day.

Study 2

Table 3 shows the relevant characteristics of the 16 patients examined by four doctors. The patients' ages ranged from 56 to 81 years (mean 69.9 \pm 7.8). There were 1 woman and 15 men. The mean FEV₁ value as measured by the reference standard was 2.069 \pm 0.508 liters.

Figure 6 shows the FEV₁, FVC and FEV₁/FVC values as recorded by four trained examiners in relation to the standard of reference (in this case, the values generated

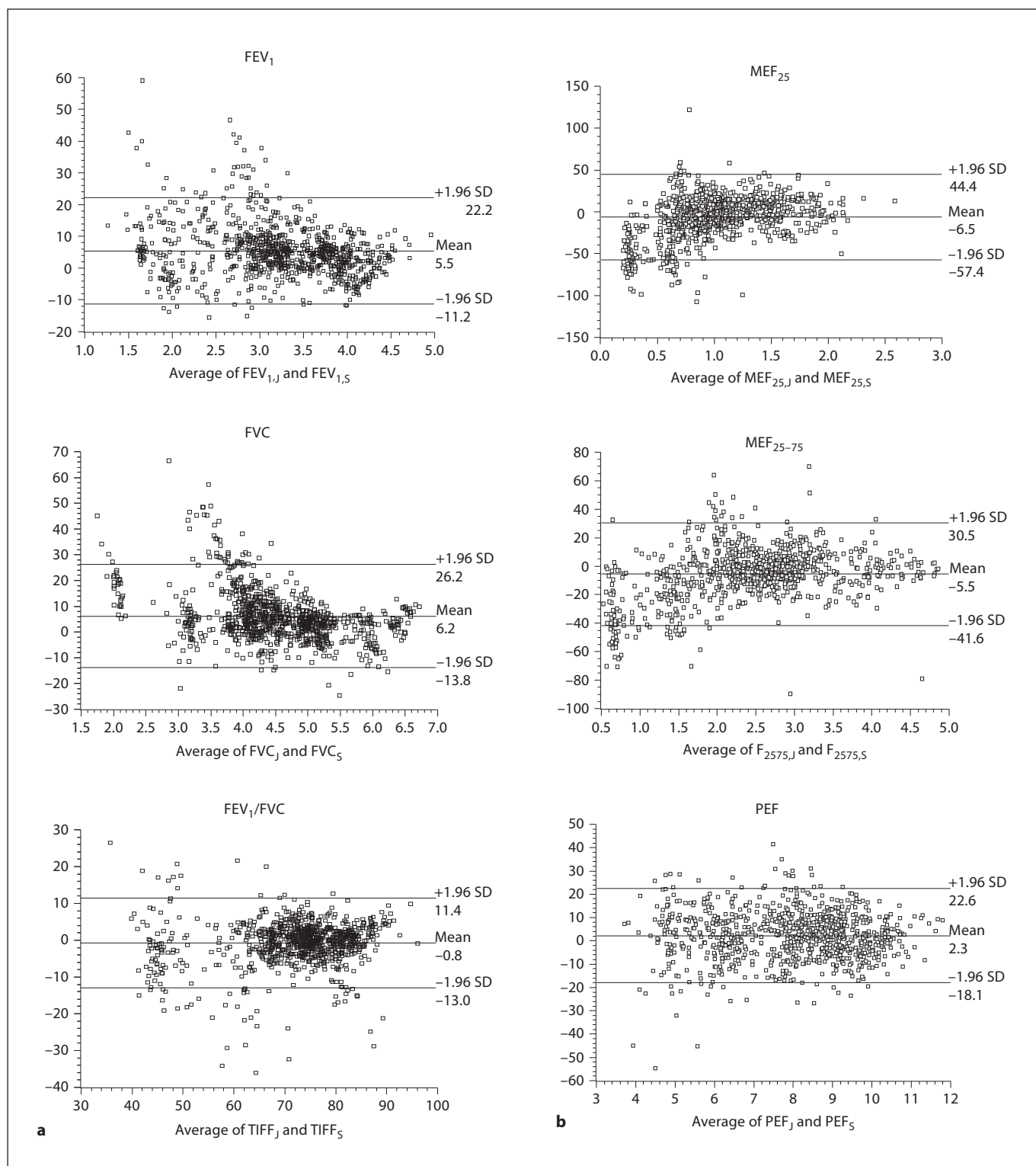


Fig. 5. a Bland and Altman plots showing differences in measurements obtained with the Jaeger and the Spirobank devices, expressed as a percentage of the mean difference (vertical axis), versus the mean of the two measurements for the FEV₁, FVC and FEV₁/FVC measures (horizontal axis). **b** Bland and Altman plots

showing differences in measurements obtained with the Jaeger and the Spirobank devices, expressed as a percentage of the mean difference (vertical axis), versus the mean of the two measurements for the MEF₂₅, MEF₂₅₋₇₅ and peak expiratory flow (PEF) measures (horizontal axis).

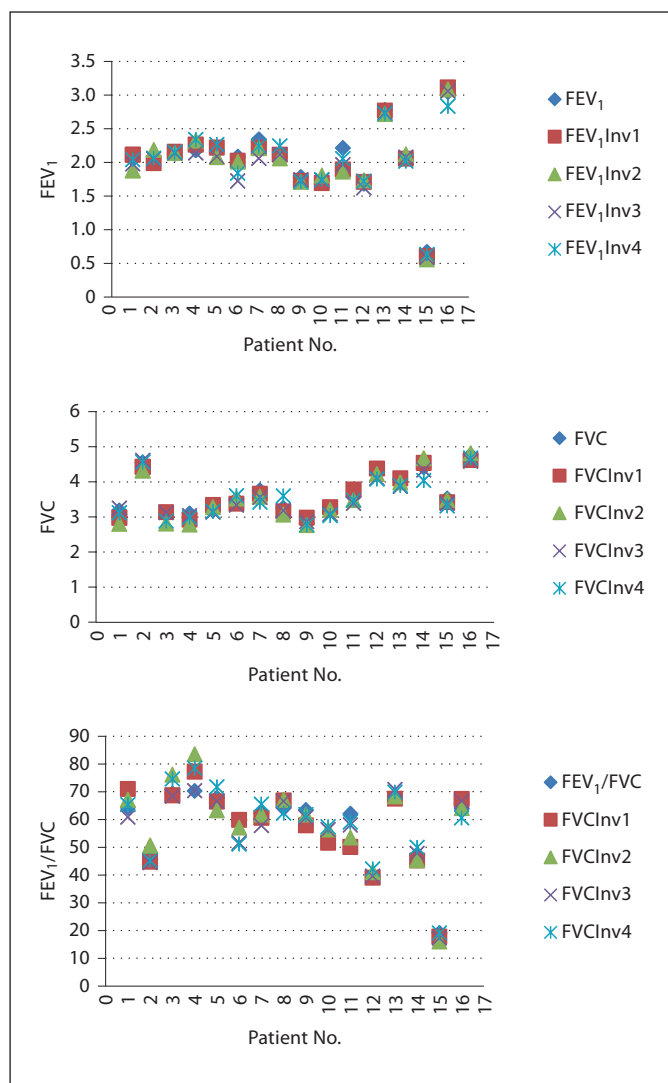


Fig. 6. FEV₁, FVC and FEV₁/FVC values as measured by the four different clinicians (Inv1–Inv4), respectively, in relation to the reference standard. On the horizontal axis: 16 patients. On the vertical axis: FEV₁ and FVC (liters); FEV₁/FVC (%). ♦ = Reference standard.

by the Jaeger device). The mean difference in the values obtained was -0.044 ± 0.068 liters for FEV₁, -0.071 ± 0.138 liters for FVC and $-0.07 \pm 2.98\%$ for the FEV₁/FVC ratio.

Interobserver Reliability

For the FEV₁, a generalizability analysis gave an intra-class coefficient of 0.992 (an index for the extent of agreement between the four examiners). The ϕ -coefficient for four parallel measurements (the reproducibility index of

Table 3. Demographics and lung function parameters of 16 patients examined by four general practitioners (standard of reference)

Patient No.	Age years	Sex	FEV ₁ liters	FVC liters	FEV ₁ /FVC %
1	54	M	2.05	3.18	64
2	73	M	2.08	4.55	46
3	64	M	2.12	3.06	69
4	69	M	2.18	3.10	70
5	63	M	2.16	3.18	68
6	76	M	2.08	3.49	60
7	67	M	2.34	3.75	62
8	83	M	2.08	3.25	64
9	65	M	1.78	2.81	63
10	74	M	1.72	3.18	54
11	79	M	2.21	3.56	62
12	66	F	1.68	4.30	39
13	59	M	2.78	4.06	68
14	76	M	2.08	4.41	47
15	78	M	0.67	3.52	19
16	72	M	3.04	4.76	64

the absolute measurements) was 0.991 with a standard error of the mean (SEM) of 0.035 liters. This gave a 95% CI of approximately 70 ml when the mean values of the four measurements were used. A D study enabled us to estimate the SEM and the corresponding 95% CI when the measurement was performed by a single clinician: $\phi = 0.991$, SEM 0.088 liters, giving a 95% CI of 172 ml for FEV₁.

Discussion

We found that the Spirobank showed acceptable validity compared with the Jaeger MasterScope when pulmonary function tests were performed in a laboratory under the supervision of an experienced technician. Nonlinearity, which has been claimed for many other portable spirometers in numerous other studies [5–7], was not found.

With regard to the standard of reference, the FEV₁ values measured with the Spirobank device were underestimated by up to 5% and the FEV₁/FVC by 3–4%. This underestimation may have been caused by the inaccuracy of the device itself. However, individual variations can also play a role. In the first study, the pulmonary function tests with the two devices were executed one after the other (first with the Jaeger and then with the Spirobank), which therefore constituted two different exhalation ma-

neuvres. Ideally, the devices should be connected in series to obtain results generated by a single manoeuvre to avoid bias caused by intrasubject variability. On the other hand, our test set-up seems relevant because it is representative of the way in which microspirometers are used in clinical practice. Furthermore, it is a moot point whether a mean difference of up to 5% for FEV₁ and of up to 4% for the FEV₁/FVC is clinically relevant. The absolute values of the mean differences in the correlation study turned out to be small. The core criterion of relevance is probably the degree of diagnostic mismatch due to measurement errors. The present study design did not permit us to estimate this aspect.

In practice, diurnal variations in pulmonary function always need to be taken into account, especially for subjects between the ages of 9 and 12 years. However, large differences can be found in young adults, smokers and people with pulmonary diseases [20]. One of the strengths of study 1 is that the measurements with the Spirobank and the Jaeger device were executed one after the other and that the measurements were spread over the day (from 8 a.m. to 5 p.m.). The generalizability analysis of the present study confirmed that the variability caused by diurnal variation is more than three times higher than the variability caused by the use of different devices.

Study 2 showed that trained general practitioners who used the Spirobank devices according to a standardized protocol could measure FEV₁/FVC values accurately. With an SEM of 88 ml and a corresponding 95% CI of 172 ml, it seems that the FEV₁ values can be considered acceptable and this makes the device fit for both epidemiological and clinical research when used by trained general practitioners.

The Spirobank is a practical, compact and valid device to use in the daily routine. It could be useful to perform additional research to further examine the value of home self-monitoring by patients. The device is then no longer connected to a personal computer and visual feedback is thus not possible. However, the built-in software does generate messages regarding the acceptability of the curves.

Conclusion

The Spirobank spirometer performed very well compared with the Jaeger MasterScope in a laboratory environment, and trained primary-care physicians managed to generate accurate measurements with this equipment. The study also showed that in practice other sources of errors such as the timing of the test will be much more important than the small measurement errors by the device itself. Within a research design, patients are rarely or never examined at the same moment of the day. Thus, the Spirobank device seems to be appropriate for research purposes if the standardized protocol is used correctly and the acceptability criteria are respected.

Acknowledgements

We would like to thank F. Rochette for the excellent cooperation with the Lung Function Laboratory of the University Hospital of the KU Leuven as well as the participating general practitioners and their patients.

References

- Enright PL: Spirometers' technological evolution spurs increased confidence in results. Four generations of advances enable more useful diagnoses for asthma, emphysema or tuberculosis. *Occup Health Saf* 1994;63:81–82.
- Stoller JK, Odhe, McCarthy K: Bedside spirometry in a tertiary care hospital: The Cleveland Clinic experience. *Respir Care* 2002;47:578–682.
- Swart F, Schuurmans MM, Heydenreich JC, Pieper CH, Bolliger CT: Comparison of a new desktop spirometer (Spirospec) with a laboratory spirometer in a respiratory out-patient clinic. *Respir Care* 2003;48: 591–595.
- Rebuck DA, Hanania NA, D'Urzo AD, Chapman KR: The accuracy of a handheld portable spirometer. *Chest* 1996;109:152–157.
- van den Boom G, van der Star LM, Folgering H, et al: Volume calibration alone may be misleading. *Respir Med* 1999;93:643–647.
- Barr RG, Stemple KJ, Mesia-Vela S, Basner RC, Derk SJ, Henneberger PK, et al: Reproducibility and validity of a handheld spirometer. *Respir Care* 2008;53:433–441.
- Johns DP, Abramason M, Bowes G: Evaluation of a new ambulatory spirometer for measuring forced expiratory volume in one second and peak expiratory flow rate. *Am Rev Respir Dis* 1993;147:1245–1250.
- Wilthshire N, Kendrick AH: Evaluation of a new electronic spirometer: the Vitalograph 'Escort' spirometer. *Thorax* 1994;49:175–178.
- Malmberg KLP, Hedman J, Soviarivi AR: Accuracy and reliability of a pocket turbine spirometer: comparison with a rolling seal flow-volume spirometer. *Clin Physiol* 1993; 13:89–98.
- Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A: Standardisation of spirometry. *Eur Respir J* 2005;26: 319–338.
- Eaton T, Withy S, Garret JE: The importance of quality assurance and the impact of spirometry workshops. *Chest* 1999;116:416–423.

- 12 Leuppi JD, Miedinger D, Chhajed PN, Buess C, Schafroth S: Quality of spirometry in primary care for case finding of airway obstruction in smokers. *Respiration* 2010;79:469–474.
- 13 American Thoracic Society: Standardization of spirometry. *Am J Respir Crit Care Med* 1999;144:1202–1218.
- 14 Liistro G, Vanwelde C, Vincken W, Vandevoorde J, Verleden G, Buffels J: Technical and functional assessment of 10 office spirometers: a multicenter comparative study. *Chest* 2006;130:657–665.
- 15 Blant JM, Altman DG: Statistical method for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–310.
- 16 Brennan RL: Elements of generalizability theory. Iowa: American College Testing Program, 1983.
- 17 Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N: The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York, Wiley, 1972.
- 18 Brennan RL: Generalizability Theory. New York, Springer, 2002.
- 19 Mitchell S: Interobserver agreement: reliability and generalizability of data collected in observational studies. *Psychol Bull* 1979; 86:376–390.
- 20 Enright PL, Connet JE, Kanner RE: Spirometry in the Lung Health study II. Determinants of short-term intraindividual variability. *Am J Respir Crit Care Med* 1995;151:406–411.