Year: 2013

Agreement between spirometers: a challenge in the follow-up of patients and populations?

Gerbase, M W ; Dupuis-Lozeron, E ; Schindler, C ; Keidel, D ; Bridevaux, P O ; Kriemler, S ; Probst-Hensch, N M ; Rochat, T ; Künzli, N

Abstract: BACKGROUND: Long-term cohort studies and lung function laboratories are confronted with the need for replacement of spirometers. Lack of agreement between spirometers might affect the longitudinal comparison of data, notably when replacing conventional by portable spirometers. OBJECTIVES: To compare the handheld EasyOne (EO) with the conventional SensorMedics (SM) spirometer, and to analyze the interdevice reproducibility of EO spirometers. METHODS: In total, 82 volunteers completed spirometry sessions with 1 SM and 2 of 3 EO spirometers following a Latin square design. Analyses of differences in forced vital capacity (FVC), forced expiratory flow in 1 s (FEV1), FEV1/FVC and mean forced expiratory flow calculated between 25 and 75% of the FVC between spirometers used a mixed effect model with a random intercept for each subject and the effect of the device as fixed effect adjusted for sex, age, height and order of spirometer tested. Bland-Altman plots show the 95% limits of agreement. RESULTS: Comparisons between EO and SM showed relatively small mean differences of <3%, but systematically lower values for FVC and FEV1 in all EO devices. The 95% agreement exceeded the limits for FEV1 by 50 ml in 2 EO spirometers. The EO interdevice comparisons showed mean differences and limits of agreement within established thresholds, thus indicating fair accuracy when comparing devices. Repeats with the same spirometer did not result in statistically significant differences. CONCLUSIONS: This study suggests fair agreement between the handheld and the conventional spirometer. Differences slightly exceeding limits for FEV1 in 2 EO devices might be considered mostly irrelevant for clinical practice. However, the systematically lower FVC and FEV1 observed with EO may be significant for epidemiological studies, thus justifying inspection before replacing devices.

DOI: https://doi.org/10.1159/000346649

Respiration

Agreement between Spirometers: A Challenge in the Follow-Up of Patients and Populations?

M.W. Gerbase[a]   E. Dupuis-Lozeron[a, b]   C. Schindler[c, d]   D. Keidel[c]
P.O. Bridevaux[a]   S. Kriemler[c, d]   N.M. Probst-Hensch[c, d]   T. Rochat[a]   N. Künzli[c, d]

[a]Division of Pulmonary Medicine, University Hospitals of Geneva, and [b]Research Center for Statistics and Department of Management Sciences, University of Geneva, Geneva, and [c]Swiss Tropical and Public Health Institute and [d]University of Basel, Basel, Switzerland

Abstract
Background: Long-term cohort studies and lung function laboratories are confronted with the need for replacement of spirometers. Lack of agreement between spirometers might affect the longitudinal comparison of data, notably when replacing conventional by portable spirometers. ***Objectives:*** To compare the handheld EasyOne (EO) with the conventional SensorMedics (SM) spirometer, and to analyze the interdevice reproducibility of EO spirometers. ***Methods:*** In total, 82 volunteers completed spirometry sessions with 1 SM and 2 of 3 EO spirometers following a Latin square design. Analyses of differences in forced vital capacity (FVC), forced expiratory flow in 1 s (FEV_1), FEV_1/FVC and mean forced expiratory flow calculated between 25 and 75% of the FVC between spirometers used a mixed effect model with a random intercept for each subject and the effect of the device as fixed effect adjusted for sex, age, height and order of spirometer tested. Bland-Altman plots show the 95% limits of agreement. ***Results:*** Comparisons between EO and SM showed relatively small mean differences of <3%, but systematically lower values for FVC and FEV_1 in all EO devices. The 95% agreement exceeded the limits for FEV_1 by 50 ml in 2 EO spirometers. The EO interdevice comparisons showed mean differences and limits of agreement within established thresholds, thus indicating fair accuracy when comparing devices. Repeats with the same spirometer did not result in statistically significant differences. ***Conclusions:*** This study suggests fair agreement between the handheld and the conventional spirometer. Differences slightly exceeding limits for FEV_1 in 2 EO devices might be considered mostly irrelevant for clinical practice. However, the systematically lower FVC and FEV_1 observed with EO may be significant for epidemiological studies, thus justifying inspection before replacing devices.

Copyright © 2013 S. Karger AG, Basel

Introduction

Spirometric lung function values are relatively easy to measure and have long been proven as excellent objective markers of respiratory morbidity and life expectancy. Hence, the use of spirometry is highly attractive and has often been included in cohort studies aiming to follow the same participants at the long term. However, such studies might be confronted at times with the need for spirometer replacement. The same applies to certified pulmonary

Margaret W. Gerbase, MD, PhD
Division of Pulmonary Medicine, University Hospitals of Geneva
4 rue Gabrielle-Perret-Gentil, CH–1211 Geneva 14 (Switzerland)
E-Mail margaret.gerbase@hcuge.ch

function laboratories which are required to adopt the best available technology in order to ensure quality control. Yet, the replacement of long-lasting spirometers may reveal challenging for longitudinal and retrospective comparisons of data both in epidemiological studies and in clinical practice. Evaluating the agreement between the new and the dated spirometer to be replaced is therefore essential for the correct interpretation of longitudinal results.

Several studies have assessed the comparability between spirometers meeting the American Thoracic Society (ATS)/European Respiratory Society (ERS) recommendations both in clinical and epidemiological settings [1–6]. Differences between spirometers reported in these studies varied from 0.07 to 0.41 liters (forced vital capacity, FVC) and from 0.05 to 0.11 liters (forced expiratory flow in 1 s, FEV_1). Variation between spirometers may relate to the characteristics of the devices but also to the design and the setting of the study, the underlying disease of patients enrolled in the assessment, the degree of compliance of subjects performing spirometric tests and the technicians. To date, no study has published the potential need for correction of spirometric values when differences between devices are detected. Yet, systematic errors due to intrinsic discrepancies found between spirometers might introduce bias when analyzing the intrasubject variation over time. For example, mild chronic obstructive pulmonary disease could be under- or overdiagnosed depending on the direction of the bias.

In recent years, a number of different last-generation portable spirometers has been released and increasingly adopted for clinical diagnosis and field studies, mainly for their ease of handling and lower costs. Among them, the handheld ndd EasyOne™ (EO) flow-sensing spirometer is reputed for its reliability, no need for calibration, automated quality checks and electronic storage of results [7, 8]. This spirometer was chosen by the large multinational epidemiological BOLD (Burden of Obstructive Lung Disease) study, which confirmed its reliability by showing comparable quality results among widely geographically separated centers [9, 10]. However, cross-validation studies comparing EO with other standard spirometers showed contrasting results with insufficient agreement reported between devices for FVC [6] or FEV_1 [11]. Replication studies may therefore be needed in order to confirm the validity of this instrument. Furthermore, the interdevice reproducibility of the EO has not been reported yet.

In the context of the 3rd survey of the SAPALDIA (Swiss Study on Air Pollution and Health in Adults) study in 2010/2011, there was a need for the replacement of the SensorMedics (SM) 2200 flow-sensing spirometers after nearly 20 years since the first examination of the cohort participants. The decision was taken to replace the dated spirometers from the 8 centers of the study with the handheld EO spirometer, and each site was equipped with 2 or 3 such devices depending on the expected number of participants. In order to examine the comparability of spirometers across the two brands and between EO devices, we conducted a nested study in 82 healthy nonsmoking volunteers. The primary objectives of the study were twofold: (1) to compare the agreement between the conventional SM and the handheld EO spirometer and (2) to analyze the interdevice reproducibility of the EO spirometer using 3 devices randomly selected from the SAPALDIA centers. In addition, we also tested the intradevice reproducibility of all spirometers.

Material and Methods

Setting
The study was conducted as a nested project of SAPALDIA. Details on the SAPALDIA study were published previously [12, 13]. Briefly, the study started in 1991 (S1) with the cross-sectional survey of 9,651 participants in 8 different regions of Switzerland reflecting the geographic, climatic, linguistic and cultural diversity of the country. The primary study objectives were to assess respiratory health and its dependence on air pollution in a representative sample of the population. Spirometry was an essential outcome at baseline, and in the 2002/2003 (S2) and the 2010/2011 (S3) follow-up surveys. Eight teams of trained technicians conducted all examinations following the same methods.

The current study was carried out at the University Hospitals of Geneva, which is 1 of the 8 SAPALDIA centers.

Subjects
Eighty-five healthy nonsmoking volunteers answered to an advertisement posted at the University Hospital and the Faculty of Medicine. Eligibility criteria were absence of respiratory and cardiovascular disease, no regular intake of respiratory or cardiovascular medication and being a never-smoker (defined by SAPALDIA as 'not having smoked more than 20 packs of cigarettes in life'). Excluded from the analyses were 2 subjects with poor compliance and 1 subject presenting with abnormal spirometric results.

Design
The study tested 4 spirometers (1 SM and 3 EO). Subjects qualifying for examinations were invited for a 2-hour visit to perform 4 series of full spirometry sessions using 3 different spirometers and 1 repeat session using 1 of the 3 devices (duplicate session). The duplicate session served to test the reproducibility of each instrument, thus we included 1 of the 3 spirometers in each test sequence. The sequence of spirometers was predefined following a Latin square design, and each sequence included 1 SM and 2 EO spirometers.

The calibration of the SM device and the volume checking of the EO devices were performed daily using a 3-liter certified syringe according to the ATS/ERS and the manufacturer recommendations [14].

The protocol of the study was approved by the Ethical Committee of the University Hospitals of Geneva. Consent for participation was obtained from all subjects enrolled in the study.

Spirometers and Spirometric Parameters

The spirometers used in the study were the SM 2200 (Sensor-Medics, Yorba Linda, Calif., USA), an open system device equipped with a mass heated wire flow sensor, and the EO (ndd Medical Technologies, Zurich, Switzerland), a handheld device equipped with an ultrasonic sensor to measure flow, both meeting ATS/ERS standards [14, 15]. Recommended disposable mouthpieces with filters were used for all spirometries. The SM spirometer was retrieved from the spirometers used during the first 2 SAPALDIA surveys (1991 and 2002/2003). All EO devices belonged to the equipment used during the 3rd SAPALDIA survey (2010/2011). Two from these were used since the beginning of this study and a 3rd device was added later.

According to the standardized operational procedures of SAPALDIA, a minimum of 3 and a maximum of 8 tests per device were performed in each subject. The 3 best spirometric maneuvers were recorded for analyses. Accordingly, for a session to be reproducible, the difference between the 2 largest FVC and the 2 largest FEV_1 had to be within 0.100 liters or 5%. These criteria were based on the ATS recommendations published in 1987 and adopted by SAPALDIA in 1991 [16]. The 2 subsequent SAPALDIA surveys maintained the same criteria concerning the reproducibility of FVC and FEV_1; all EO spirometers used in S3 had their configuration changed by the manufacturer to be in agreement with these previous standards. In the present study, the criterion used for reproducibility of FVC and FEV_1 followed the same standards.

The quality control of all spirometric maneuvers was automatically displayed on the computer screen (SM) or the device (EO). Quality degrees from A (best) to E (worst) are provided by EO at the end of each test. We considered only spirometries with highest degrees A and B for this study.

The following lung function parameters were analyzed: FVC, FEV_1, FEV_1/FVC ratio and mean forced expiratory flow calculated between 25 and 75% of the FVC (FEF_{25-75}). The FVC and FEV_1 were recorded from the highest value among all accepted curves. The FEV_1/FVC ratio was calculated from the best FEV_1 and the best FVC, whereas the FEF_{25-75} was recorded from the best curve with the largest sum of FEV_1 and FVC. All spirometries were performed by one nurse trained as a fieldworker for S3 examinations in the center of Geneva. Supervision of the spirometry tests was assured by an experienced pulmonologist who is also part of the SAPALDIA team.

Analyses

Lung function results are reported using absolute and percent of predicted values based on the SAPALDIA reference equations developed by Brändli et al. [17]. Characteristics of subjects and spirometric results obtained for each spirometer are expressed as means ± SD. Differences between spirometers are expressed as means for all spirometric parameters. To test the effect of the sequence of spirometries performed by subjects using the Latin square design, a Wald test was applied to the mixed effect model with a random effect for subjects and with the sequence as the only fixed effect.

Paired t tests were used to test the reproducibility of spirometers by comparing the results obtained from spirometry tests using the same device. To compare differences between spirometers in FVC, FEV_1, FEV_1/FVC and FEF_{25-75} results, we used a mixed effect model with a random intercept for each subject and the effect of the device as fixed effect adjusted for sex, age, height and order of spirometer tested. As suggested by the analysis of residuals of the models for FVC, FEV_1/FVC and FEF_{25-75}, we modeled heteroscedasticity by allowing the variance in the residuals to change with the fitted values of the model [18]. We also tested two mixed effect models, one with an additional random slope effect for each subject to model an eventual temporal trend due to the sequence order and one with a first order autoregressive covariance structure of errors within each subject. These two models gave similar estimates and standard errors as the one with a random intercept only.

The agreement between spirometers was assessed using the Bland-Altman method [19] with plots displayed by sex as a proxy of lung volume. For comparisons of absolute differences between instruments, we considered the ATS standards of accuracy for monitoring devices ≤0.100 liters or ±3% for FVC and FEV_1, and ≤0.200 l/s or ±5% for FEF_{25-75} [14]. The interpretation of the limits of agreement was based on criteria setting the threshold at 0.500 (FVC) and 0.350 liters (FEV_1), as reported by Liistro et al. [6].

All analyses were performed using R for Windows version 2.15.1 [20] with the packages NMLE [21] for the mixed effect models, and MethComp [22] and ggplot2 [23] for the Bland-Altman plots. The significance level was fixed at 5% (two tailed).

Results

The study was carried out from July 7 to November 3, 2011. A total of 82 subjects (mean age 24.1 ± 7.1 years) completed trials with at least 3 different spirometers following the preestablished sequence. Each session consisted of 3–8 spirometric maneuvers. In total, 357 sessions (average of 89 sessions per device) were performed (112 for SM, 105 for EO1, 102 for EO2 and 38 for EO 3, which was added later to the study). Table 1 summarizes the demographic characteristics and the lung function parameters of subjects enrolled in the study. Table 2 depicts the mean lung function results for each spirometer.

Quality Control of Tests and Design

The quality control of spirometries showed that 90% of the SM tests achieved the highest quality according to ATS/ERS guidelines [14]. For EO, highest quality control codes (level A = 2005 ATS/ERS quality criteria: 3 acceptable maneuvers and 2 highest FEV_1 and FVC within 150 ml) were achieved in 91% of tests performed with EO1, 86% with EO2 tests and 90% with EO3.

The order of spirometer testing had no detectable effect on FVC, FEV_1, FEV_1/FVC and FEF_{25-75} (all p > 0.7). Differences in results obtained by repeated measures with the

Table 1. Characteristics of the study participants (means ± SD)

	All (n = 82)	Men (n = 38)	Women (n = 44)
Age, years	23.6±7.4	24.6± 10.0	22.8± 3.9
Height, cm	172.2±9.1	178.7±7.4	166.5±6.2
Weight, kg	66.2±13.0	74.3±12.3	59.1±8.5
BMI	22.2±3.1	23.24±3.3	21.30±2.6
FVC, liters	4.6±1.0 (97.1±11.0)	5.5±0.8 (97.3±11.0)	4.0±0.5 (97.0±11.0)
FEV_1, liters	3.9±0.8 (97.1±11.0)	4.6±0.6 (97.0±10.0)	3.4±0.3 (97.3±10.5)
FEV_1/FVC, %	85.1±6.1 (100.7±6.6)	83.5±5.7 (100.6±6.4)	86.5±6.2 (100.8±6.7)
FEF_{25-75}, l/s	4.2±1.0 (100.1±20.4)	4.7±1.1 (99.9±20.0)	3.8±0.7 (101.1±20.8)

Percentages of predicted values for FVC, FEV_1, FEV_1/FVC and FEF_{25-75} are given in parentheses based on equations for the Swiss population from Brändli et al. [16].

Table 2. Spirometry results (means ± SD) obtained with SM and EO spirometers in 82 healthy, nonsmoking volunteers

Parameter	SM	EO1	EO2	EO3
FVC, liters	4.7±1.1	4.5±1.0	4.7±1.0	4.5±0.9
FEV_1, liters	4.0±0.8	3.9±0.8	4.0±0.8	3.8±0.6
FEV_1/FVC, %	84.7±6.3	85.4±5.9	85.7±6.0	84.0±6.8
FEF_{25-75}, l/s	4.2±1.1	4.2±1.0	4.4±1.0	4.0±0.9

same spirometer were also not statistically significant. The averaged absolute differences between the 1st and the 2nd measurements for all spirometers were: FVC = 0.001 liters (p = 0.928), FEV_1 = 0.001 liters (p = 0.917), FEV_1/FVC = 0.03% (p = 0.872) and FEF_{25-75} = 0.026 l/s (p = 0.839).

To take into account circadian fluctuations in lung function values, we also tested whether results depended on the time of the day of spirometric measurements. The mixed model with an indicator for morning/afternoon measurement time revealed no significant interaction between time of the day and device for FVC [coefficient of variation (CV) = 0.157, SE = 0.125, p value = 0.214], FEV_1 (CV = 0.093, SE = 0.100, p value = 0.357), FEF_1/FVC (CV = −0.723, SE = 1.258, p value = 0.567) or FEF_{25-75} (CV = 0.014, SE = 0.197, p value = 0.945).

Comparisons between Spirometers

The comparisons between the SM and the EO are shown in table 3. Overall, mean differences were relatively small, but systematically lower values were observed for FVC and FEV_1 in the 3 EO spirometers compared to the SM, with the difference slightly exceeding the 100-ml threshold for FVC when comparing EO1 to SM. All percent differences were within 3%. Mean differences in the FEV_1/FVC ratio were close to 0 and varied between 0.00 and 0.08 l/s in FEF_{25-75}.

Bland-Altman plots for limits of agreement stratified by sex in all considered spirometric parameters are displayed in figures 1 and 2. Limits of agreement (lower bound) exceeded the threshold criterion of 350 ml for FEV_1 in comparisons of the SM with the EO1 and the EO3. Wider variations were observed for FEF_{25-75} with mean relative differences of 2.9% comparing SM and EO2, and limits of agreement globally larger than those observed for FVC and FEV_1. To examine whether the agreement depended on lung volumes, we regressed the difference between devices on the spirometric parameters. Results were statistically significant for FVC [β: −0.048 (0.016), R^2: 0.147; p value 0.004] and FEV_1 [β: −0.046 (0.021), R^2: 0.144; p value 0.027] comparing SM and EO1, and FEF_{25-75} comparing SM and EO1 [β: −0.102 (0.038), R^2: 0.351; p value 0.010], and SM and EO2 [β: −0.110 (0.040), R^2: 0.369; p value 0.008].

Table 4 shows the pairwise comparisons of the EO spirometers. Though statistically significant, the mean differences between devices were within tolerated thresholds for all spirometric parameters, indicating fair accuracy when comparing devices. The limits of agreement were within the specified standards for FVC and FEV_1 in all EO spirometers (tables 3, 4). However, the mean difference and the upper bound of the 95% confidence interval reached limits for FVC when comparing EO2 with EO1.

Table 3. Comparisons between SM and EO spirometers derived from analyses of spirometric parameters measured in 82 healthy, non-smoking volunteers

Spirometry variables	EO1			EO2			EO3		
	mean difference	mean relative difference	Bland-Altman 95% limits of agreement	mean difference*	mean relative difference	Bland-Altman 95% limits of agreement	mean difference*	mean relative difference	Bland-Altman 95% limits of agreement
FVC	−0.13 liters[a]	−2.80%	−0.45 to 0.17	−0.02 liters	−0.54%	−0.33 to 0.27	−0.07 liters[a]	−1.34%	−0.37 to 0.23
FEV_1	−0.10 liters[a]	−2.33%	−0.39 to 0.20	−0.01 liters	0.03%	−0.28 to 0.26	−0.06 liters[a]	−1.26%	−0.41 to 0.29
FEV_1/FVC	0.01[c]	0.50%	−4.02 to 4.77	0.00[c]	0.60%	−4.21 to 5.13	0.00	0.08%	−4.74 to 4.74
FEF_{25-75}	−0.02 l/s	0.14%	−0.76 to 0.69	0.08 l/s[b]	2.87%	−0.69 to 0.84	−0.00 l/s	0.11%	−0.73 to 0.67

Mean difference = EO – SM readings. Analysis: mixed-effect models with a random intercept by subject using the device as fixed effect adjusted for sex, age, height and sequence of spirometric tests. Heteroscedasticity was taken into account except for FEV_1. Mean relative difference (%) = (EO – SM)/SM*100. [a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$. EO1 (serial No. 67598); EO2 (serial No. 74745) and EO3 (serial No. 74747) were used.

Discussion

Findings from this study show an overall good quality of the large number of spirometries performed on a single occasion by healthy and mostly young never-smokers. In this ideal context, comparisons between spirometers from different brands and within the same brand showed satisfactory agreement. Reading differences were lower than the generally tolerated thresholds for the majority of comparisons. However, systematic deviations in FVC and FEV_1 were observed between the SM and the EO devices.

Several studies reported on the quality of spirometric tests performed with the EO both in clinical [7, 24] and epidemiological [10, 25] settings. The excellent reproducibility of FVC and FEV_1 parameters within our subjects is in line with these reports. Unquestionably, reliable, high-quality tests are a prerequisite of a spirometer, and the vast experience gathered worldwide over the last years with the EO largely confirms its utility.

To further assess EO performance, we compared it to the SM spirometer used in SAPALDIA in the past. The latter had been tested in a comparison of lung function measurements made during ECRHS (European Community Respiratory Health Survey), and differences among equipments of different brands used in the study were not statistically significant [26]. SAPALDIA (Basel) is part of the ECRHS and used the same SM equipment as several other centers of the European study. Furthermore, the performance of the SM spirometers used in the 2 first surveys of SAPALDIA was validated and the variability across 8 spirometers was <3% for FVC, FEV_1 and FEF_{50} [27, 28]. Specifically for the SM used in the current study, the device deviation from the personal mean had been of the order of 1%. In all, these studies point to the fair reproducibility of the SM equipment.

Barr et al. [11] elegantly tested the EO using an in-line waveform generator and performed a clinical comparison with the SM 2130, a model similar to ours. As in our study, spirometric parameter values derived from the EO were lower than those of the SM. Furthermore, the authors report limits of agreement exceeding the preestablished criteria for the FEV_1, but not for the FVC, findings which we corroborate in this study.

Mean differences in FEV_1 between SM and EO were on average 100 and 50 ml in the study by Barr et al. [11] and in ours, respectively. Standards for judging the magnitude of error and the potential need for correction of spirometric measurements derived from different equipments with limited agreement are inexistent. Systematic differences originating from the technology changes are, however, a major concern in longitudinal studies where the change in spirometric indices is the key health outcome. According to our findings and those observed by Barr et al. [11], the systematic deviations are different for FEV_1 and FVC, thus, the potential for systematic biases in the FEV_1/FVC ratio needs to be considered. The latter is crucial in the investigation of the incidence of chronic obstructive pulmonary disease and its determinants. While absolute differences observed in our comparison were generally 'small', we emphasize that many risks and treatments confer only 'small' but nevertheless relevant effects. Moreover, the shorter the follow-up time in longitudinal studies, the more influential are systematic differences that originate from the change in technologies. In investigating health effects of air pollution, SAPALDIA further faces the challenge that contrasts in exposure
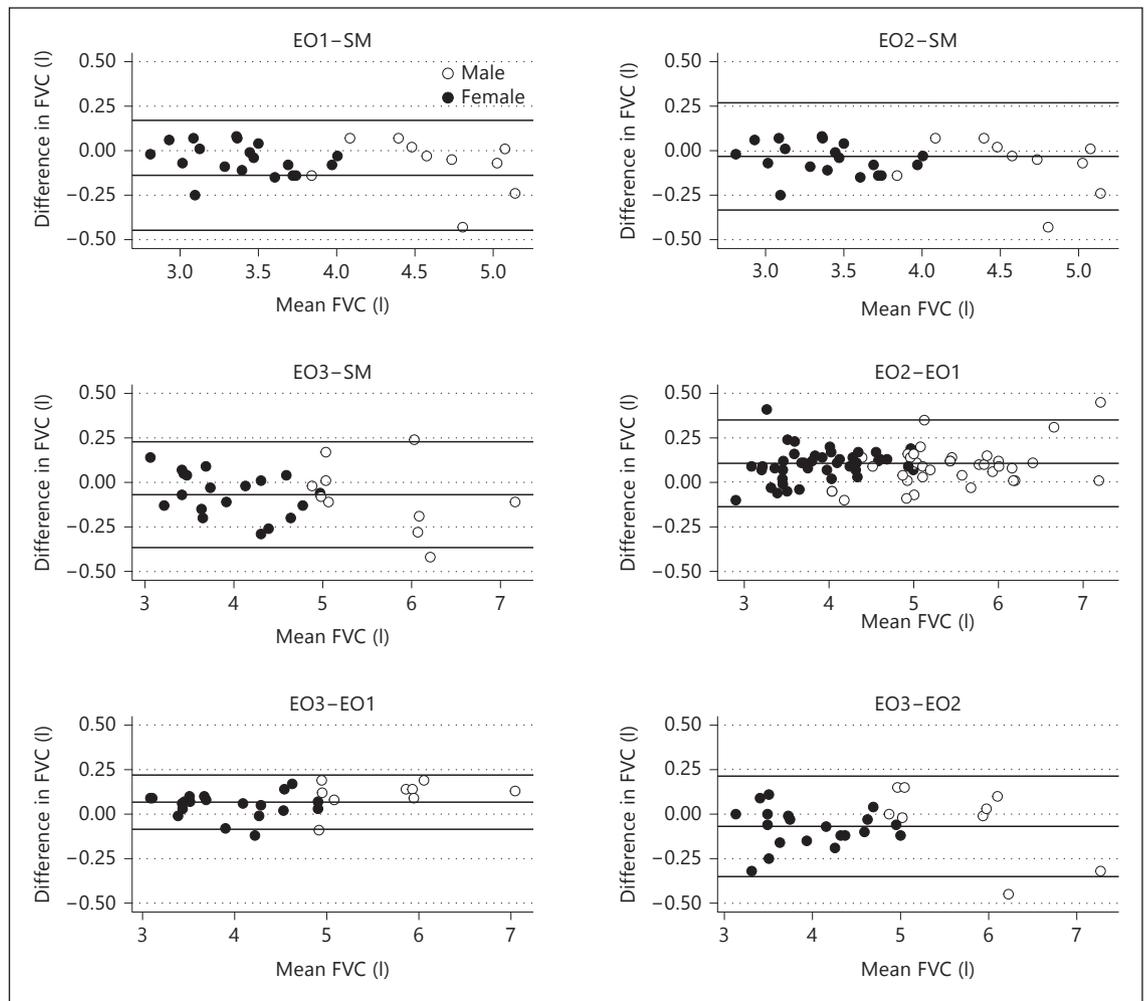
Fig. 1. Bland-Altman plots showing the agreement between SM (reference) and EO spirometers for FVC (this page) and FEV$_1$ (next page). The middle line is the mean difference and the two extreme lines are the limits of agreement ($\pm 1.96 \times$ SD) of differences between spirometers.

having been moderate to begin with further decrease due to clean air policies. Thus, it will be essential to find innovative solutions for correcting the systematic differences due to the inevitable change in technology. As shown in our assessment, the agreement also depended on the volume, thus, corrections of data collection with different devices may not be a fixed value but require some more complex models. This may involve the development of further methods.

Sources of error affecting the spirometric results can be multifold and linked to the device, the technician or the subject being examined. Several features differ between the two spirometer types that were compared. For example, recommended mouthpieces equipped with filters are rigid for the SM, but fairly soft pliable spirettes

were proposed for the EO in earlier models. Barr et al. [11] tested the effect of the mouthpiece compression, which had no significant effect on FEV$_1$. Moreover, our EO models were equipped with rigid spirettes, which are unlikely to have influenced the results. However, one disadvantage of handheld spirometers is the fact that it is the subject performing the test who handles the device, and during the SAPALDIA examinations we had observed a marked tendency of study participants to bend frontwards while performing the forced expiration. This could potentially introduce bias for lung function measurements. For example, patients with chronic obstructive pulmonary disease may be more likely to bend in order to fully exhale. On the other hand, bending during the test could potentially hold back expiration and decrease the
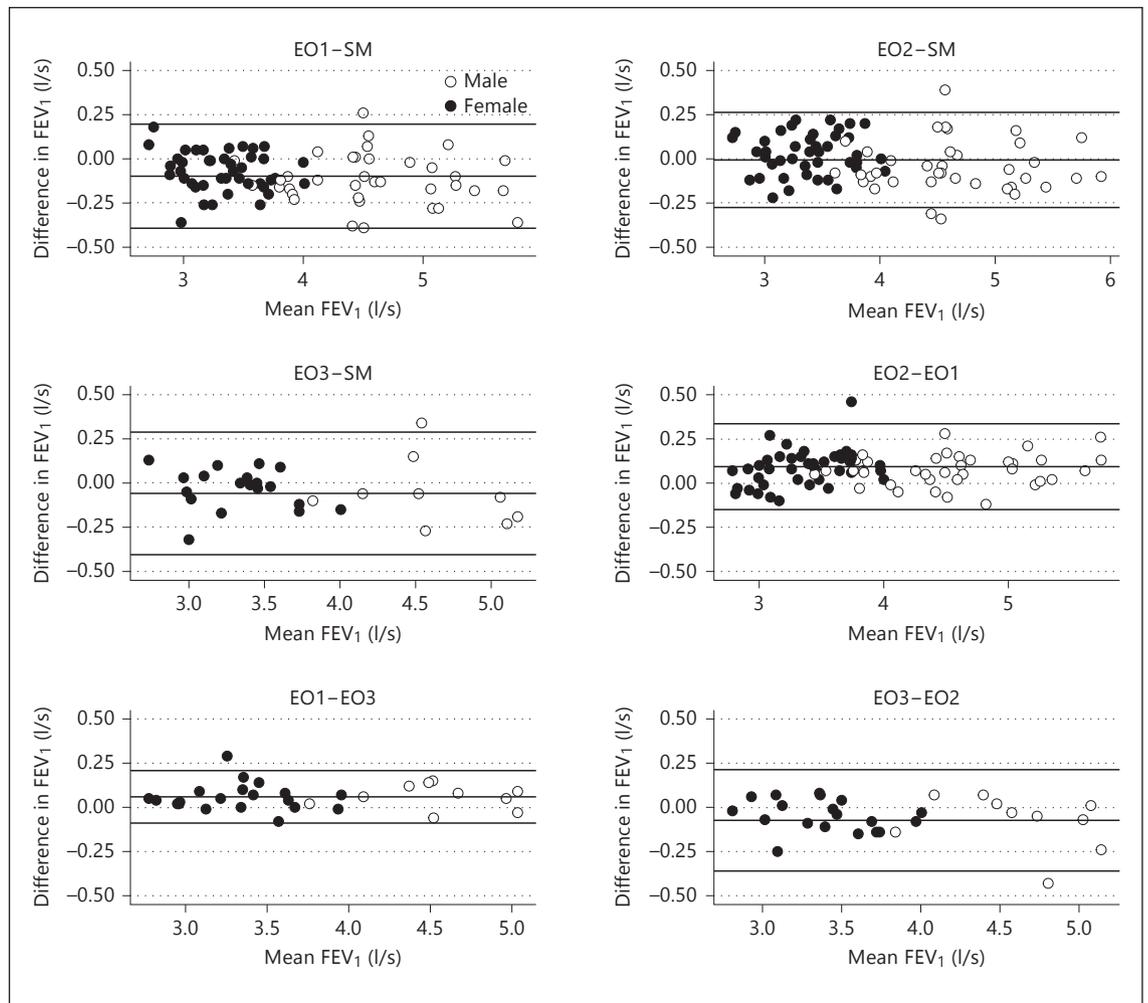
1

Table 4. Pairwise comparisons of the 3 EO spirometers derived from analyses of spirometric parameters measured in 82 healthy, non-smoking volunteers

Spirometry variables	EO2-EO1			EO3-EO1			EO3-EO2		
	mean difference	mean relative difference	Bland-Altman 95% limits of agreement	mean difference	mean relative difference	Bland-Altman 95% limits of agreement	mean difference	mean relative difference	Bland-Altman 95% limits of agreement
FVC	0.10 liters[a]	2.39%	–0.14 to 0.35	0.06 liters[a]	1.50%	–0.09 to 0.22	–0.04 liters[b]	–1.48%	–0.35 to 0.21
FEV_1	0.09 liters[a]	2.50%	–0.15 to 0.34	0.03 liters	1.67%	–0.09 to 0.21	–0.04 liters[b]	–1.86%	–0.36 to 0.21
FEV_1/FVC	–0.04	0.10%	–2.84 to 3.01	–0.35	0.18%	–2.95 to 3.27	–0.31	–0.40%	–4.20 to 3.47
FEF_{25-75}	0.10 l/s[a]	2.94%	–0.38 to 0.60	0.02 l/s	2.04%	–0.40 to 0.53	–0.08 l/s[c]	–2.58%	–0.69 to 0.47

Mean difference = EO – SM readings. Analysis: mixed-effect models with a random intercept by subject using the device as fixed effect adjusted for sex, age, height and sequence of spirometric tests. Heteroscedasticity was taken into account except for FEV_1. Mean differences: EO1 is the reference in differences calculated with EO2 or EO3; EO2 is the reference in differences calculated with EO3. Analysis: mixed-effect models with a random intercept by subject using the device as fixed effect adjusted for sex, age, height and sequence of spirometric tests. Heteroscedasticity was taken into account except for FEV_1. Mean relative difference (%) = (EO2 – EO1)/EO1*100. [a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$.
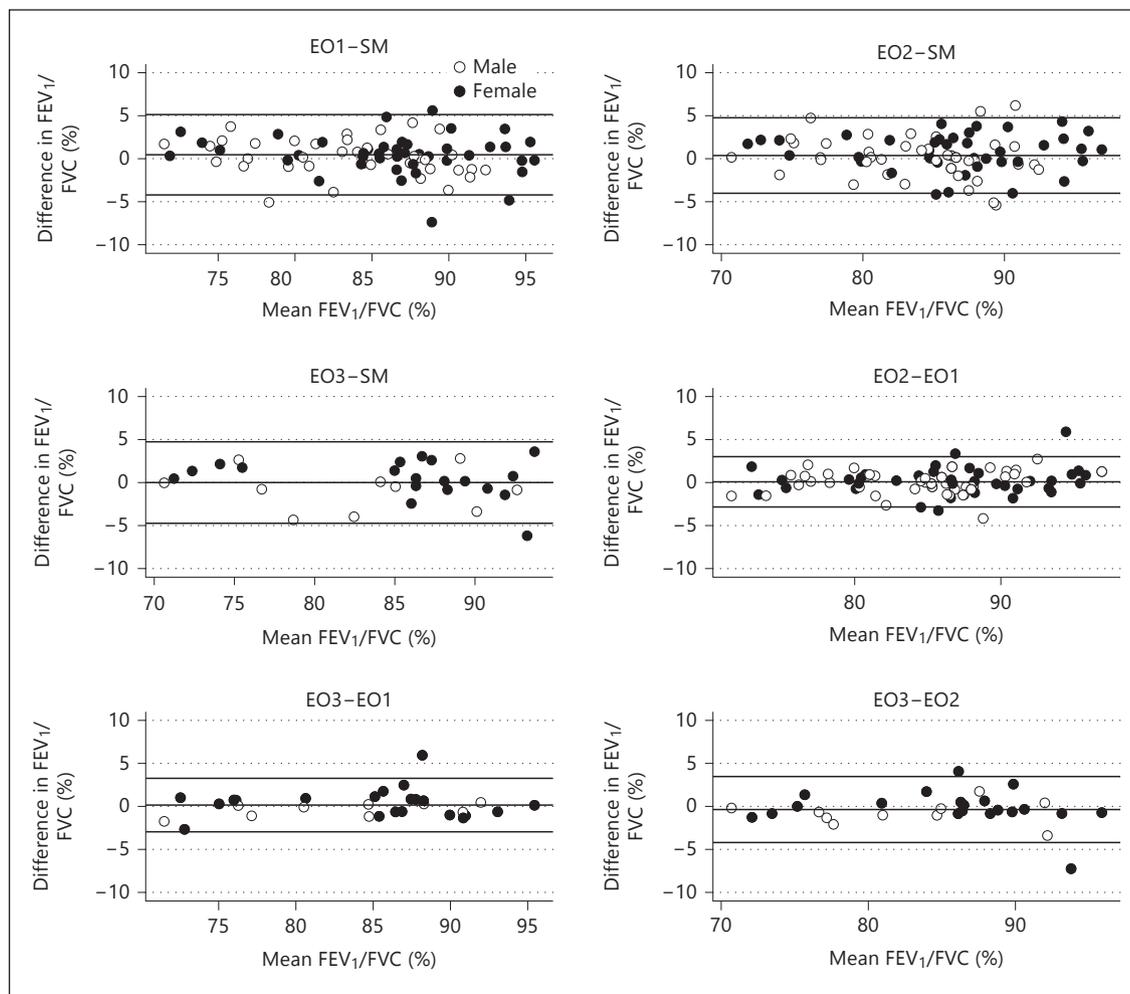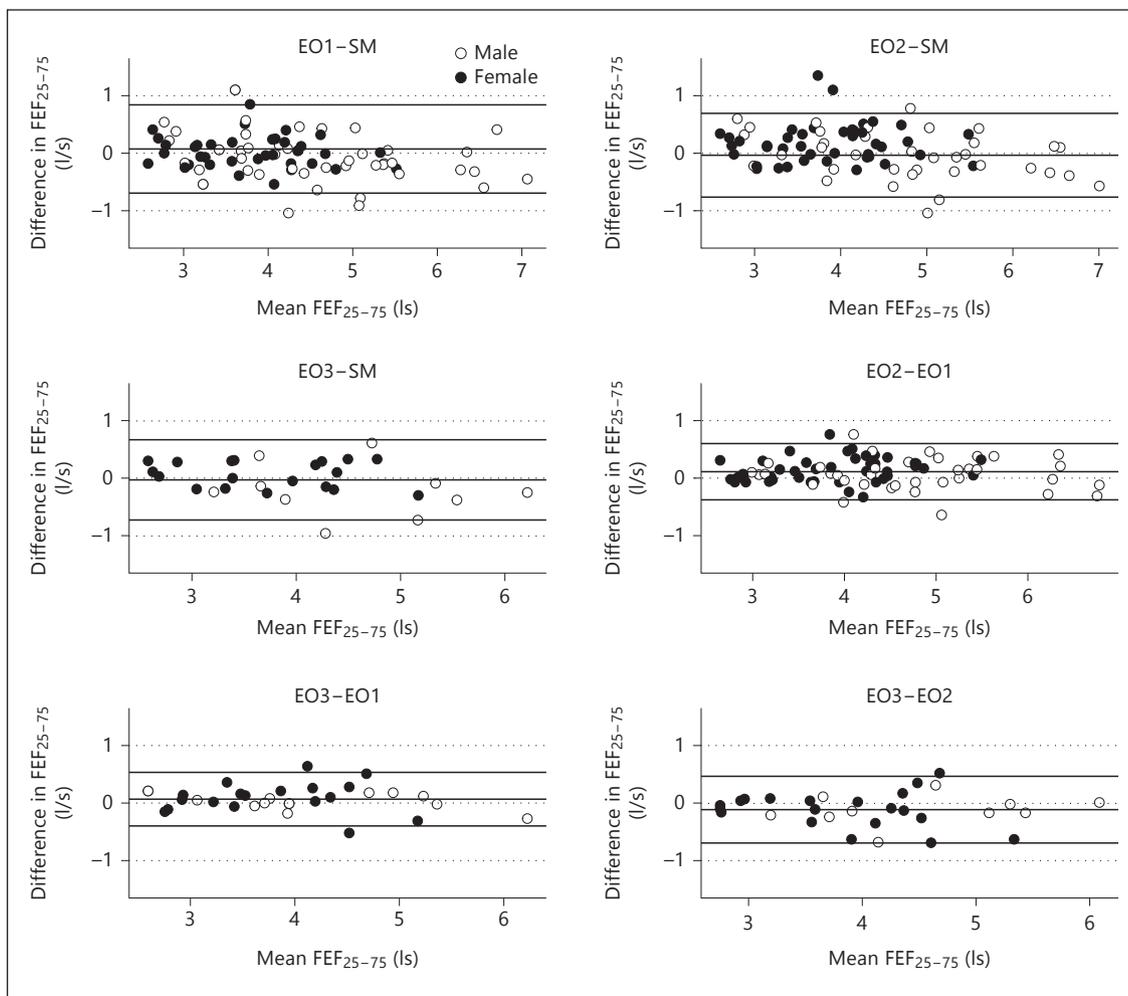
Fig. 2. Bland-Altman plots showing the agreement between SM (reference) and EO spirometers for FEV_1/FVC (this page) and FEF_{25-75} (next page). The middle line is the mean difference and the two extreme lines are the limits of agreement ($\pm 1.96 \times SD$) of differences between spirometers.

FEV_1, an artifact that would be less likely with the SM due to its voluminous and standstill format. Technicians in our study were trained to visually control the subjects being tested and coach them to prevent bending; however, uncontrollable minor bending may occur and systematically affect spirometric values. Finally, sources of error related to compliance during spirometry are crucial. Repeated spirometries done with the same device did not reveal significant differences, thus suggesting that the discrepancies observed between devices were rather related to the specific technical characteristics of the 2 spirometers than to compliance during tests.

In this study, we were able to expand previous work evaluating the EO spirometer by assessing the interdevice reproducibility of this instrument. Our pairwise compari-

sons analyzing 3 EO devices show relatively small average differences between devices of the same brand with limits of agreement generally found within accepted limits for the FVC and the FEV_1. Nonetheless, the significant statistical differences found between devices with values reaching the considered limits notably for FVC warrants caution when interpreting individual results from spirometries performed with different devices of the same brand. While less relevant in epidemiological analyses – where adjustment for 'device' can be done – this may be relevant in clinical settings and decision making. In addition, as described above for comparisons between SM and EO, generally higher limits of agreement were found for FEF_{25-75}, but the absence of specific limits precludes further interpretation of the results obtained for this parameter.

2

The figure shows six Bland-Altman plots titled EO1–SM, EO2–SM, EO3–SM, EO2–EO1, EO3–EO1, and EO3–EO2, each plotting Difference in FEF_{25-75} (l/s) against Mean FEF_{25-75} (ls), with legend ○ Male and ● Female.

Strengths of the study were the randomized design and the systematic sequential spirometries effectuated by a single experienced technician and performed by a relatively large number of highly motivated healthy never-smokers equally distributed according to sex. However, the study has also several limitations. First, our findings derive from examinations of healthy volunteers, mostly young adults, so the results might not apply to other groups, e.g. elderly patients with respiratory diseases and impaired lung function. We found indeed that the level of agreement significantly depends on the lung volumes – also a proxy for age – as shown in the Bland-Altman plots. This was in particular the case in the comparisons between SM and 1 EO for FVC, FEV_1 and FEF_{25-75}. Second, only 1 of the formerly used SM of SAPALDIA was available for testing. Although previous comparisons had not revealed major differences across the 8 SM devices, we cannot assure that results would be as reproducible 10 years later for

all devices. Third, we did not test spirometers using a waveform generator, a reproducibility testing procedure of equipment that does not involve test persons [29]. This was not a feasible option in our case. This test cannot take into account differences that may occur due to subject-related factors (e.g. ability to comply with instructions or bending). Previous work had tested the same spirometer brands as those used in our study with good reproducibility of in-line testing using a mechanical generator [11]; moreover, after testing 24 subjects, the authors found similar differences between spirometers as in our study. We were able to confirm these results in a larger study cohort.

In conclusion, our findings show overall good agreement between the conventional SM and the handheld EO spirometers, as well as between EO devices. For clinical diagnosis and follow-up of individual patients, differences between the two spirometer types and among EO spirometers can be considered as clinically mostly irrelevant. For

cohort studies, however, measuring for instance systematically lower FVC and FEV_1 at follow-up, even in case of very small amounts, may be falsely interpreted as accelerated lung function decline of the population under study. This may lead to erroneous conclusions about the effect of environmental, biologic or life-style factors on lung function changes. In this case, the development of methods to adjust for systematic differences will be needed.

Acknowledgments

References

1 Caras WE, Winter MG, Dillard T, Reasor T: Performance comparison of the handheld MicroPlus portable spirometer and the SensorMedics Vmax22 diagnostic spirometer. Respir Care 1999;44:1465–1473.

2 Viegi G, Simoni M, Pistelli F, Englert N, Salonen R, Niepsuj G, et al: Inter-laboratory comparison of flow-volume curve measurements as quality control procedure in the framework of an international epidemiological study (PEACE project). Respir Med 2000; 94:194–203.

3 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.

4 Korhonen H, Remes ST, Kannisto S, Korppi M: Hand-held turbine spirometer: agreement with the conventional spirometer at baseline and exercise. Pediatr Allergy Immunol 2005; 16:254–257.

5 Akhtar R, Wilson A: A comparison of spirometry in general practice and a pulmonary function laboratory. Prim Care Respir J 2005; 14:215–220.

6 Liistro G, Vanwelde C, Vincken W, Vandevoorde J, Verleden G, Buffels J, et al: Technical and functional assessment of 10 office spirometers: a multicenter comparative study. Chest 2006;130:657–665.

7 Walters JAE, Wood-Baker R, Walls J, Johns DP: Stability of the EasyOne ultrasonic spirometer for use in general practice. Respirology 2006;11:306–310.

8 Skloot GS, Edwards NT, Enriight PL: Four-year calibration stability of the EasyOne portable spirometer. Respir Care 2010;55:873–877.

9 Perez-Padilla R, Vasquez-Garcia JC, Marquez MN, Menezes AMB, PLATINO Group: Spirometry quality-control strategies in a multinational study of the prevalence of chronic obstructive pulmonary disease. Respir Care 2008;53:1019–1026.

10 Enright P, Vollmer WM, Lamprecht B, Jensen R, Jithoo A, Tan W, et al: Quality of spirometry tests performed by 9893 adults in 14 countries: the BOLD study. Respir Med 1995; 105:1507–1515.

11 Barr RG, Stemple KJ, Mesia-Vela S, Basner RC, Derk SJ, Hennenberger PK, et al: Reproducibility and validity of a handheld spirometer. Respir Care 2008;53:433–441.

12 Martin BW, Ackermann-Liebrich U, Leuenberger P, Künzli N, Stutz EZ, Keller R, et al: SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. Soz Präventivmed 1997;42:67–84.

13 Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, et al: Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991–2003: methods and characterization of participants. Soz Präventivmed 2005;50:245–263.

14 Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al: Standardisation of spirometry. Eur Respir J 2005;26:319–338.

15 Standardization of spirometry, 1994 update. American Thoracic Society. Am J Respir Crit Care Med 1995;152:1107–1136.

16 Standardization of spirometry – 1987 update. Statement of the American Thoracic Society. Am Rev Respir Dis 1987;136:1285–1298.

17 Brändli O, Schindler C, Künzli N, Keller R, Perruchoud AP: Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. Thorax 1996;51:277–283.

18 Pinheiro JC, Bates DM: Mixed-Effects Models in S and S-PLUS. Statistics and Computing Series. New York, Springer, 2000.

19 Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.

20 R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing, 2012.

21 Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team: NLME: linear and nonlinear mixed effects models. R package version 3.1-104, 2012. R-core@R-project.org.

22 Carstensen B, Gurrin L, Ekstrom C: MethComp: functions for analysis of method comparison studies. R package version 1.15, 2012. http://CRAN.R-project.org/package=MethComp.

23 Wickham H: ggplot2: Elegant Graphics for Data Analysis. New York, Springer, 2009.

24 Schoh RJ, Fero LJ, Shapiro H, Aslor JP, Kaelin OJ, Rollins DR, et al: Performance of a new screening spirometer at a community fair. Respir Care 2002;47:1150–1157.

25 Perez-Padilla R, Vasquez-Garcia JC, Marquez MN, Jardim JRB, Pertuzé J, Lisboa C, et al: The long-term stability of portable spirometers used in a multinational study of the prevalence of chronic obstructive pulmonary disease. Respir Care 2006;51:1167–1171.

26 Chinn S, Jarvis D, Svanes C, Burney P: Sources of variation in forced expiratory volume in one second and forced vital capacity. Eur Respir J 2006;27:767–773.

27 Künzli N, Ackermann-Liebrich U, Keller R, Perruchoud AP, Schindler C: Variability of FVC and FEV1 due to technician, team, device and subject in an eight centre study: three quality control studies in SAPALDIA. Swiss Study on Air Pollution and Lung Disease in Adults. Eur Respir J 1995;8:371–376.

28 Künzli N, Kuna-Dibbert B, Keidel D, Keller R, Brändli O, Schindler C, et al: Longitudinal validity of spirometers – a challenge in longitudinal studies. Swiss Med Wkly 2005;135:503–508.

29 Schermer TRJ, Verweij EHA, Cretier R, Pellegrino A, Crockett AJ, Poels PJP: Accuracy and precision of desktop spirometers in general practices. Respiration 2012;83:344–352.