Accuracy and Precision of Desktop Spirometers in General Practices

Tjard R.J. Schermer^{a, d} Eddy H.A. Verweij^a Riet Cretier^a Annelies Pellegrino^{a, b}
Alan J. Crockett^{d, e} Patrick J.P. Poels^{a, c}

^aCOPD and Asthma Research and Development Unit, Department of Primary and Community Care, and

^bDepartment of Epidemiology, Biostatistics, and Health Technology Assessment, Radboud University Nijmegen Medical Centre, Nijmegen, and ^cGeneral Practice Bles en Poels, Huissen, The Netherlands; ^dPrimary Care Respiratory Research Unit, Discipline of General Practice, School of Population Health and Clinical Practice, Faculty of Health Sciences, University of Adelaide, ^eSchool of Health Sciences, University of South Australia, Adelaide, S.A., Australia

Key Words

Accuracy • Diagnosis • Equipment evaluation • General practice • Precision • Safety • Spirometry

Abstract

Background: Spirometry has become an essential tool for general practices to diagnose and monitor chronic airways diseases, but very little is known about the performance of the spirometry equipment that is being used in general practice settings. The use of invalid spirometry equipment may have consequences on disease diagnosis and management of patients. **Objectives:** To establish the accuracy and precision of desktop spirometers that are routinely used in general practices. **Methods:** We evaluated a random sample of 50 spirometers from Dutch general practices by testing them on a certified waveform generator using 8 standard American Thoracic Society waveforms to determine accuracy and precision. Details about the brand and type of spirometers, year of purchase, frequency of use, cleaning and calibration were inquired with a study-specific questionnaire. **Results:** 39 devices (80%) were turbine spirometers, 8 (16%) were pneumotachographs, and 1 (2%) was a volume

displacement spirometer. Mean age of the spirometers was 4.3 (SD 3.7) years. Average deviation from the waveform generator reference values (accuracy) was 25 ml (95% confidence interval 12–39 ml) for FEV₁ and 27 ml (10–45 ml) for FVC, but some devices showed substantial deviations. FEV₁ deviations were larger for pneumotachographs than for turbine spirometers ($p < 0.0031$), but FVC deviations did not differ between the two types of spirometers. In the subset of turbine spirometers, no association between age and device performance was observed. **Conclusions:** On average, desktop spirometers in general practices slightly overestimated FEV₁ and FVC values, but some devices showed substantial deviations. General practices should pay more attention to the calibration of their spirometer.

Copyright © 2012 S. Karger AG, Basel

Introduction

Spirometry has become an essential tool for primary care doctors and nurses to diagnose and monitor chronic airways diseases [1–4]. In the Netherlands, spirometer ownership in general practices has increased substantial-

ly in the past decade (from 25% in 1998 to 62% in 2007) [5]. In this same period of time, the rate of spirometry tests performed in Dutch primary care has tripled [5]. Spirometry tests performed in general practice have shown to be valid compared to tests performed in a pulmonary function laboratory – provided that the tests are carried out by well-trained staff and that spirometers are regularly tested for errors and deviations [6]. Nonetheless, substandard quality of spirometry tests routinely performed in general practices is a reason for concern [7–10].

General practices use various brands and types of desktop spirometers [5, 11, 12], which are generally less advanced but more affordable than the equipment typically used in laboratories. A previous study has shown that most brands of spirometers used in general practices measure forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC) sufficiently accurate to diagnose chronic respiratory conditions [13]. However, the spirometers in this particular study had been obtained directly from the manufacturers. Investigating such ‘off-the-shelf’ equipment may give a biased view on the performance of devices that have routinely been used in general practices for a period of time, especially while maintenance and calibration of spirometers in general practices is often well below laboratory standards [5, 11, 12]. Lack of maintenance and calibration and wear of the equipment over time are likely to negatively impact the level of spirometer performance [14], which may have consequences for the interpretations that are based on the results of a spirometry test and the subsequent management of patients.

Several previous studies have looked at performance markers of particular spirometry devices when used in research or secondary care settings [15–27], but as far as we are aware there are no published reports on the performance of spirometers that are being used in general practices as part of routine patient care. In this study, we evaluated a random sample of spirometers from Dutch general practices with regard to their accuracy and precision. We applied recommendations from the American Thoracic Society (ATS) [28] to determine spirometer performance.

Materials and Methods

Study Design

We first took a random sample of 700 general practices from all 4,533 practices registered in the national database of the Netherlands Institute for Health Services Research (www.nivel.nl) in

2007 [5]. Out of these 700 practices, we took random subsamples of 25 practices for each of the 12 provinces in our country. These 300 practices were contacted by telephone to inquire whether or not the practice owned one or more spirometers and if so, if they were willing to submit the device they most frequently used in their regular patient care for performance testing. Details about the brand and type of the spirometer, year of purchase, frequency of use, cleaning and calibration routines, and history of previous damage and repairs (if any) were inquired with a study-specific questionnaire. No formal sample size calculation was performed, but we aimed to include 50 spirometers in the evaluation. Ethics approval or internal review board approval were not required for this study.

Spirometer Testing and Outcomes

After shipment by express courier from the general practice to the testing site (PT Medical BV, Leek, The Netherlands), spirometers were tested by a trained operator on an ATS-certified Standard wPWG S/N F02 waveform generator by applying 8 standard waveforms (Nos. 1, 4, 11, 15, 17, 19, 21, and 24) [28] during a single session under ambient conditions which – in all cases – were in the operating temperature range of the devices [mean temperature during testing: 22.9°C (SD 2.2), range 20.0–28.5°C; air pressure 1,016 mb (SD 11.9), range 989–1,054 mb; humidity 50.3% (SD 8.9), range 33–64%]. The waveform generator system was maintained in accordance with the manufacturers’ operating manual. The 8 waveforms were selected from the full set of 24 ATS waveforms because together they represent a wide range of FEV₁ and FVC values. A selection of the complete set of 24 waveforms has previously been used to evaluate lung function equipment [29].

Each waveform was applied 3 consecutive times to each spirometer. Spirometer performance was expressed in terms of accuracy and precision. To determine accuracy, we calculated the mean of the 3 readings for the FEV₁, FVC and forced mid-expiratory flow (FEF_{25–75}) for each waveform, and compared these with the gold standard reference values for these indices produced by the waveform generator. For FEV₁ and FVC, we considered a spirometer to be inaccurate for a particular waveform in case of a ≥ 100 ml or $\geq 3.5\%$ deviation (either positive or negative) from the waveform generator reference value [28]. For FEF_{25–75}, we considered a ≥ 250 ml/s or $\geq 5.5\%$ deviation as inaccurate [28]. Precision of the spirometers was determined by calculating the difference between the highest and the lowest of the 3 FEV₁, FVC and FEF_{25–75} readings for each of the 8 waveforms. In order to determine whether or not a spirometer was imprecise for a particular waveform, the same cutoff points as for accuracy were used [28].

Analysis

We used SPSS (Statistical Package for Social Sciences), version 17.0, for all analyses (volumes and flows are expressed in ml and ml/s, respectively). Mean values and 95% confidence intervals (CI) were calculated for accuracy and precision estimates for the 8 separate waveforms as well as for all waveforms combined. After calculating accuracy and precision estimates for FEV₁, FVC and FEF_{25–75} for each of the spirometers tested, we plotted the differences between the respective waveform generator reference values and the spirometer readings against the mean of these values in Bland-Altman plots [30]. The upper and lower limits of agreement

Table 1. Accuracy and precision for FEV₁ and FVC per waveform and overall for the sample of 49 spirometers from general practices [means (95% CI)]. Results for FEF_{25–75} are published as online supplementary material.

Waveform No.	FEV ₁					FVC				
	generator ml	accuracy ml	accuracy %	precision ml	precision %	generator ml	accuracy ml	accuracy %	precision ml	precision %
1	4,395	42 (21, 62)	0.9 (0.5, 1.4)	20 (15, 26)	0.5 (0.3, 0.6)	6,000	7 (–26, 40)	0.1 (–0.4, 0.7)	29 (18, 40)	0.5 (0.3, 0.7)
4	1,387	25 (17, 33)	1.8 (1.2, 2.4)	10 (6, 13)	0.7 (0.5, 0.9)	1,500	26 (15, 37)	1.7 (1.0, 2.4)	16 (12, 20)	1.1 (0.8, 1.3)
11	1,784	20 (9, 32)	1.1 (0.5, 1.8)	10 (6, 14)	0.6 (0.4, 0.8)	2,701	35 (18, 51)	1.3 (0.7, 1.9)	29 (20, 38)	1.1 (0.7, 1.4)
15	5,303	39 (15, 63)	0.7 (0.3, 1.2)	22 (17, 27)	0.4 (0.3, 0.5)	5,934	48 (22, 73)	0.8 (0.4, 1.2)	36 (22, 50)	0.6 (0.4, 0.8)
17	2,597	18 (1, 35)	0.7 (0.1, 1.4)	14 (8, 19)	0.5 (0.3, 0.7)	5,829	49 (3, 95)	0.8 (0.1, 1.6)	52 (31, 73)	0.9 (0.5, 1.2)
19	2,505	20 (7, 33)	0.8 (0.3, 1.3)	13 (10, 17)	0.5 (0.4, 0.7)	3,931	–9 (–39, 22)	–0.2 (–1.0, 0.6)	32 (21, 43)	0.8 (0.5, 1.1)
21	3,519	24 (8, 39)	0.7 (0.2, 1.1)	16 (10, 21)	0.4 (0.3, 0.6)	4,446	34 (12, 57)	0.8 (0.3, 1.3)	38 (16, 60)	0.8 (0.4, 1.3)
24	916	15 (7, 23)	1.6 (0.8, 2.5)	9 (7, 12)	1.0 (0.7, 1.3)	1,230	29 (11, 46)	2.3 (0.9, 3.7)	24 (5, 43)	1.7 (0.5, 2.9)
Overall		25 (12, 39)	1.0 (0.4, 1.6)	14 (11, 17)	0.6 (0.5, 0.7)		27 (10, 45)	1.2 (0.6, 1.9)	35 (24, 47)	1.0 (0.6, 1.4)

Accuracy = Mean of the 3 spirometer readings for the FEV₁ and FVC with the gold standard reference values for these indices as produced by the waveform generator; precision = difference between the highest and the lowest of the 3 FEV₁ and FVC readings for the applied waveform.

around the mean difference were set on ± 100 ml for FEV₁ and FVC, and ± 250 ml for FEF_{25–75}, respectively [28]. Based on the age distribution in the sample, we divided the subset of turbine spirometers (the predominant type of sensor in our sample as well as in Dutch general practices nationally [5]) into three categories: 1–2; 3–5, and ≥ 6 years old. We used analysis of variance (ANOVA) to analyze associations between spirometer type and age and FEV₁ and FVC accuracy and precision estimates. To correct for multiple comparisons in these analyses, statistical significance was defined as $p < 0.0031$ (Bonferroni correction: $\alpha = 0.05/2 \times 8$ waveforms).

Results

Spirometers in the Study

Of the random sample of 300 general practices, 269 (90%) practices could be contacted by telephone and 250 (83%) were willing to provide information about their

spirometry [5]. In total, 154 practices (62%) reported owning ≥ 1 spirometers. Of these 154 practices, 63 (41%) agreed to submit their spirometer for testing and to complete the questionnaire about its details and use. Eventually 50 (32%) spirometers were submitted for testing on the waveform generator. The main reasons for practices not to submit their spirometer after all were that they ‘could not miss their spirometer’ (35%) or ‘were too busy to arrange shipment of the spirometer to the testing site’ (21%).

One spirometer was sent to the testing site but could not be tested because the software necessary to operate the spirometer was not available. Fourteen of the tested spirometers could not measure FEF_{25–75}. Of the 49 spirometers tested, 38 (78%) had been manufactured by MicroMedical (now known as CareFusion), 6 (12%) by Vitalograph, 3 (6%) by Clement Clarke International, and 1

Table 2. Summary of accuracy (a) and precision (b) performance criteria for FEV₁ and FVC for the sample of 49 spirometers from general practices

a Accuracy

Number of waveforms not meeting accuracy ¹ criteria for FEV ₁	Number of waveforms not meeting accuracy criteria ¹ for FVC								
	0	1	2	3	4	5	6	7	8
0	21	10	4	1		1	1		
1			1	1					
2						1			
3									
4						1			
5		1							
6							1		
7				1		1			
8					1		1	1	

b Precision

Number of waveforms not meeting precision ¹ criteria for FEV ₁	Number of waveforms not meeting precision criteria ¹ for FVC								
	0	1	2	3	4	5	6	7	8
0	38	4	2	1			1		
1	2					1			
2									
3									
4									
5									
6									
7									
8									

Numbers in the cells indicate numbers of spirometers.
¹ ≥100 ml or ≥3.5% deviation (either positive or negative) from the waveform generator reference value.

(2%) by Mijnhardt. Information about the brand and type of the sensor was missing for 1 spirometer. Thirty-nine spirometers (80% of all) had turbine flow sensors, 8 (16%) were pneumotachographs, and 1 (2%) was a volume displacement spirometer. Six spirometers (12%) had been repaired after having been purchased by the practice, and 1 practice reported that the transducer of their spirometer had once been replaced.

Mean and median age for the total sample of spirometers were 4.3 (SD 3.7) and 3 (interquartile range: 4.5) years. The ‘youngest’ spirometer was 4 months old, the

oldest 18 years. Mean and median age for the subset of turbine spirometers were 3.7 (SD 2.9) and 3.0 (interquartile range: 4.0) years, respectively, with a minimum age of 4 months and a maximum age of 12 years.

Online supplementary appendix 1 provides further details for each of the tested spirometers in the sample (for all online supplementary material, see www.karger.com/doi/10.1159/000334320).

Accuracy and Precision of the Spirometers

The total average deviation from the waveform generator reference values for the 8 applied waveforms (i.e. accuracy) was 25 ml (95% CI 12–39 ml) for FEV₁, 27 ml (10–45 ml) for FVC, and –24 ml/s (–48 to 1 ml/s) for FEF_{25–75}, respectively. For precision, the respective mean ranges were 14 (11–17) and 35 (24–47) ml and 45 (23–67) ml/s. Table 1 shows the accuracy and precision results per waveform (in ml as well as in %).

Figure 1 shows the difference against mean (Bland-Altman) plots for FEV₁ and FVC accuracy in milliliter by type of spirometer. The plot for FEV₁ (fig. 1a) shows non-linearity of accuracy: the deviations between the generator reference values and the spirometer readings became larger as the FEV₁ value of the waveform increased. The vast majority of the out-of-range FEV₁ readings were attributable to 5 of the 8 pneumotachographs in the sample (for the deviations above the upper limit of agreement) and to the single volume displacement spirometer (for the deviations below the lower limit). Non-linearity was also observed for the deviations in FVC readings (fig. 1b), with deviations up to +590 and –490 ml. The deviations above the upper limit of agreement were attributable to 19 turbine spirometers, 7 pneumotachographs, and the volume displacement spirometer. Online supplementary figure 1c shows the Bland-Altman plot for FEF_{25–75} accuracy for the subset of 35 spirometers that could measure this index.

With regard to accuracy, ANOVA testing showed that FEV₁ deviations were statistically significantly larger in the pneumotachographs than in the turbine spirometers for all waveforms; FVC deviations did not differ between the two types of spirometers for any of the waveforms. With regard to precision, the ranges in FEV₁ readings were statistically significantly larger in the pneumotachographs than in the turbine spirometers for 5 waveforms (Nos. 1, 4, 11, 17, and 19). For the ranges in FVC readings, this was true for 3 waveforms (Nos. 1, 4, and 19).

Table 2 provides a summary of the combined accuracy and precision performance criteria for FEV₁ and FVC for the total spirometer sample. With regard to ac-

Fig. 1. Difference against mean plots for accuracy of FEV₁ (a) and FVC (b) for all 49 spirometers in the study sample. Each symbol represents the result for 1 of the 8 applied waveforms (WF) (mean of 3 readings for WF 1, 4, 11, 15, 17, 19, 21, and 24 [28]) for 1 spirometer.

accuracy, 21 (42%) devices fulfilled the FEV₁ and FVC performance criteria for all 8 waveforms, 11 (22%) devices failed to meet the criteria for only 1 waveform, and the remaining 17 (36%) devices failed to meet the criteria for ≥ 1 waveform. With regard to precision, 42 (86%) devices

fulfilled the FEV₁ and FVC performance criteria for all 8 waveforms or failed to meet the criteria for only 1 waveform. If the precision performance criteria were not met, this was mostly due to FVC values being imprecise (table 2b).

Influence of Age on Turbine Spirometer Performance

Figure 2 shows the difference against mean plots for FEV_1 and FVC accuracy by age category for the subset of turbine spirometers. Only 4 devices (all aged >2 years) showed out-of-range deviations for 1 or 2 waveforms in FEV_1 readings (fig. 2a), whereas 17 devices (equally distributed over all 3 age categories) showed out-of-range deviations for ≥ 1 waveforms in FVC readings (fig. 2b). Apart from the fact that for FVC the most extreme deviations below the lower limit of agreement were attributable to several devices from the oldest age category (fig. 2b), the plots show no clear associations between age and performance in terms of accuracy. Online supplementary figure 2c shows the Bland-Altman plot for FEF_{25-75} accuracy for the subset of 27 turbine spirometers that could measure this index. ANOVA testing did not show any differences in FEV_1 or FVC accuracy or precision estimates between the three age categories.

Discussion

In this study, we tested a random sample of 49 spirometers from Dutch general practices with regard to their accuracy and precision using an ATS waveform generator and a selection of standard ATS waveforms to determine spirometer performance. Turbine spirometers dominated the sample, which is consistent with observations from our larger survey in Dutch general practices [5]. With regard to accuracy, for FEV_1 as well as FVC the overall average deviation from the waveform generator reference values was approximately +25 ml. With regard to precision, the average difference between the highest and the lowest of the three readings was 14 ml for FEV_1 and 35 ml for FVC. The single (18-year-old) volume displacement spirometer in the sample should be considered an exception, as nowadays only very few practices will use such a spirometer. Although the number of pneumotachographs in the sample was small ($n = 8$) and pneumotachographs are more susceptible to environmental temperature and humidity than turbine spirometers are [31], any comparison between the different types of devices should be interpreted with caution; our observations do suggest that lack of accuracy of both FEV_1 and FVC may be a problem when general practices choose to use a pneumotachograph for their spirometry. In general, accuracy seemed to be more of a problem than precision in our spirometer sample. The subset of 38 turbine spirometers showed a wide age range (4 months to 12 years) but despite the limited attention towards calibration and maintenance in the

Fig. 2. Difference against mean plots for accuracy of FEV_1 (a) and FVC (b) by age category for the subset of turbine spirometers ($n = 37$). Each symbol represents the result for 1 of the 8 applied waveforms (WF) for 1 spirometer. The dotted lines represent the limits of agreement (± 100 ml).

practices who owned them, we did not observe an association between the age of the devices and their actual performance.

Strengths and Limitations

Initially, 63 practices agreed to submit their spirometer for testing, but only 49 practices eventually did. The 14 'missing' spirometers may have caused selection bias because the main reason for not submitting was that the practices could not miss them. This might indicate that these spirometers were more frequently used than the ones that we were able to test, and frequency of use may cause faster wear and declining performance of the device [32]. We do not know the actual motives of the 91 practices that did own a spirometer but were not able or willing to submit it for testing, but this may also have caused selection bias. We cannot exclude the possibility that practices that owned a rather old spirometer were more likely to cooperate because they would like to see their spirometer tested and be reassured that it still performed sufficiently well.

Budget restraints did not allow us to test each spirometer according to the full procedure as recommended by the ATS (i.e. 5 consecutive maneuvers for all 24 waveforms) [28]. However, we did use the full procedure in a random subsample of 10 spirometers, and for these devices applying the full procedure did not change our interpretation regarding their performance.

Another limitation of our study is that there were no ultrasonic spirometers in our sample, because these were only recently introduced in the Netherlands. Previous reports have shown that this type of spirometer seems rather robust in the long term [20, 25], and may therefore be a relevant addition to the current assortment of spirometers Dutch general practices can choose from. In other countries, ultrasonic spirometers already seem to be rather common in primary care practices [8, 11].

Comparison with Previous Studies

From their evaluation of 10 brand-new desktop spirometers in a large sample of subjects who were also measured with standard diagnostic laboratory spirometers, Liistro et al. [13] concluded that accuracy of FEV₁ and FVC was generally sufficient for turbine spirometers as well as for pneumotachographs, but their study also showed that some brands of pneumotachographs seemed to perform not as good as turbine spirometers did in terms of linearity and precision. Our observations seem to support this, and also suggest that once a spirometer has been used in a general practice for a period of time, a

pneumotachograph may be more prone to lose its accuracy. In the past decade, there have been technological advancements in the design and durability of turbine sensors. A previous (1996) report about the performance of turbine sensors over time is therefore rather outdated [21], but nonetheless it supports our observation that these sensors seem to retain their accuracy for a number of years.

Implications for Clinical Practice

The majority of general practices in our study seemed to pay no or only limited attention to calibration of their spirometers, which has previously been shown to be rather common practice in the Netherlands [5]. For some practices, this lack of attention seemed not justified, as their spirometers showed substantial inaccuracy and/or imprecision when tested on the waveform generator. On average, the spirometers slightly (~25 ml) overestimated both FEV₁ and FVC values, but some devices in the sample – especially the pneumotachographs – showed substantial deviations from the waveform generator values. We do not have patient-related data from the general practices to confirm this, but it is quite likely that devices that showed large measurement errors will have caused misinterpretation (especially overestimation) of FEV₁ and FVC values in patients who have been assessed with these devices.

One of the current issues in primary care is how spirometer performance can be checked in this setting, with limited resources, expertise and access to laboratory facilities. Clearly, regular use of a calibration syringe would be desirable, even while this does not guarantee that the spirometer readings from patients' forced maneuvers are indeed accurate [15, 33]. Regular biological calibration (by a healthy test operator in the practice) [34, 35] is probably the most feasible option. When purchasing a spirometer, general practices should consider the robustness of the different types of sensors and the frequency of maintenance and calibration they require, next to other important factors like hygiene, inclusion of the appropriate reference equations, user friendliness, and price. Regarding accuracy and precision, the limited evidence currently available suggests that when attention to equipment maintenance and calibration is likely to be limited, a turbine or ultrasonic spirometer may be the better option for a general practice.

In conclusion, our evaluation of a sample of spirometers from Dutch general practices showed that on average these spirometers may slightly (~25 ml) overestimate FEV₁ and FVC values, with larger deviations being more

likely at higher values for these indices. In general, accuracy seemed to be more of an issue than precision was. The vast majority of the sample consisted of turbine spirometers, but lack of accuracy was mainly attributable to the small number of pneumotachographs, especially with regard to FEV₁. The subset of turbine spirometers showed a wide age range, from only a few months up to 12 years, but we did not observe an association between the age of the devices and their actual performance, despite the limited attention towards calibration and maintenance in many of the practices that owned them. General practices should pay more attention to the calibration of their spirometer, either by using a calibration syringe or a biological control. When obtaining a spirometer, they should take the robustness of the different types of sensors and the frequency of maintenance and calibration they require into consideration.

Acknowledgments

This study was sponsored by the Radboud University Nijmegen Medical Centre. The authors appreciate the work put in the project by Jorn de Vreede, who conscientiously tested the spirometers on the waveform generator. We also appreciate the time and effort it took the general practice staffs to submit their spirometers for testing.

Disclosure Statement

None of authors has any conflict of interest to declare.

References

- Schirnhöfer L, Lamprecht B, Firlei N, Kaiser B, Buist AS, Halbert RJ, Allison MJ, Studnicka M: Using targeted spirometry to reduce non-diagnosed chronic obstructive pulmonary disease. *Respiration* 2011;81:476–482.
- Levy ML, Fletcher M, Price DB, Hausen T, Halbert RJ, Yawn BP: International Primary Care Respiratory Group (IPCRG) Guidelines: diagnosis of respiratory diseases in primary care. *Prim Care Respir J* 2006;15:20–34.
- Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;176:532–555.
- National Collaborating Centre for Chronic Conditions: Chronic obstructive pulmonary disease: national clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. *Thorax* 2004;59(suppl 1):1–232.
- Schellekens D, Poels P, Pellegrino A, Cretier R, Smele I, Schermer T: Spirometry in Dutch general practice. Results from a national survey (in Dutch). *Huisarts Wet* 2008;51:434–439.
- Schermer TR, Jacobs JE, Chavannes NH, Hartman J, Folgering HT, Bottema BJ, et al: Validity of spirometric testing in a general practice population of patients with chronic obstructive pulmonary disease (COPD). *Thorax* 2003;58:861–866.
- Schermer T, Crockett A, Poels P, van Dijke J, Akkermans R, Vlek H, et al: Quality of routine spirometry tests in Dutch general practices. *Br J Gen Pract* 2009;59:e376–e382.
- Walters JA, Hansen EC, Johns DP, Blizzard EL, Walters EH, Wood-Baker R: A mixed methods study to compare models of spirometry delivery in primary care for patients at risk of COPD. *Thorax* 2008;63:408–414.
- Leuppi JD, Miedinger D, Chhajed PN, Buess C, Schafroth S, Bucher HC, Tamm M: Quality of spirometry in primary care for case finding of airway obstruction in smokers. *Respiration* 2010;79:469–474.
- Carr R, Telford V, Waters G: Impact of an educational intervention on the quality of spirometry performance in a general practice: an audit. *Prim Care Respir J* 2011;20:210–213.
- Johns DP, Burton D, Walters JA, Wood-Baker R: National survey of spirometer ownership and usage in general practice in Australia. *Respirology* 2006;11:292–298.
- Dowson LJ, Yeung A, Allen MB: General practice spirometry in North Staffordshire. *Monaldi Arch Chest Dis* 1999;54:186–188.
- 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.
- Townsend MC, Hankinson JL, Lindesmith LA, Slivka WA, Stiver G, Ayres GT: Is my lung function really that good? Flow-type spirometer problems that elevate test results. *Chest* 2004;125:1902–1909.
- Nelson SB, Gardner RM, Crapo RO, Jensen RL: Performance evaluation of contemporary spirometers. *Chest* 1990;97:288–297.
- Dompeling E, van Schayck CP, Folgering H, van den Hoogen HJ, van Weel C: Accuracy, precision and linearity of the portable flow-volume meter Microspiro HI-298. *Eur Respir J* 1991;4:612–615.
- Rode A, Shephard RJ: Accuracy of an electronic spirometer – a field trial. *Respiration* 1986;50:66–69.
- Maree DM, Videler EA, Hallauer M, Pieper CH, Bolliger CT: Comparison of a new desktop spirometer (Diagnosa) with a laboratory spirometer. *Respiration* 2001;68:400–404.
- Rebuck DA, Hanania NA, D'Urzo AD, Chapman KR: The accuracy of a handheld portable spirometer. *Chest* 1996;109:152–157.
- Walters JA, Wood-Baker R, Walls J, Johns DP: Stability of the EasyOne ultrasonic spirometer for use in general practice. *Respirology* 2006;11:306–310.
- Dirksen A, Madsen F, Pedersen OF, Vedel AM, Kok-Jensen A: Long-term performance of a hand held spirometer. *Thorax* 1996;51:973–976.
- 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.
- Perez-Padilla R, Vázquez-García JC, Márquez MN, Jardim JR, 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.

- 24 Malmberg LP, Hedman J, Sovijärvi AR: Accuracy and repeatability of a pocket turbine spirometer: comparison with a rolling seal flow-volume spirometer. *Clin Physiol* 1993; 13:89–98.
- 25 Skloot GS, Edwards NT, Enright PL: Four-year calibration stability of the EasyOne portable spirometer. *Respir Care* 2010;55:873–877.
- 26 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.
- 27 Jensen RL, Teeter JG, England RD, White HJ, Pickering EH, Crapo RO: Instrument accuracy and reproducibility in measurements of pulmonary function. *Chest* 2007;132:388–395.
- 28 American Thoracic Society. Standardization of spirometry, 1994 Update. *Am J Respir Crit Care Med* 1995;152:1107–1136.
- 29 Hankinson JL, Filios MS, Kinsley KB, Petsonk EL: Comparing MiniWright and spirometer measurements of peak expiratory flow. *Chest* 1995;108:407–410.
- 30 Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1: 307–310.
- 31 Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J: Standardisation of spirometry. *Eur Respir J* 2005;26:319–338.
- 32 Johns DP, Side E, Kendrick AH, Williams TJ, Walters EH: The effect of physiologic and mechanical aging on the performance of peak flowmeters. *Chest* 1998;113:774–779.
- 33 van den Boom G, van der Star LM, Folgering H, van Schayck CP, van Weel C: Volume calibration alone may be misleading. *Respir Med* 1999;93:643–647.
- 34 Ninaber MK, Schot R, Fregonese L, Stolk J: A syringe simulation of biological controls for quality assessment of prospective lung volume measurements. *Respiration* 2008;76: 187–192.
- 35 Poels PJ, Schermer TR: Value of recommended spirometer accuracy checks on office spirometers in primary care unknown. *Respirology* 2007;12:151.