Eur Respir J, 1995, 8, 371–376
DOI: 10.1183/09031936.95.08030371
Printed in UK - all rights reserved

European Respiratory Journal
ISSN 0903 - 1936

Variability of FVC and FEV_1 due to technician, team, device and subject in an eight centre study:

N. Künzli[+], U. Ackermann-Liebrich[+], R. Keller[++], A.P. Perruchoud[**], C. Schindler[+], SAPALDIA team[*]

Variability of FVC and FEV_1 due to technician, team, device and subject in an eight centre study: three quality control studies in SAPALDIA. N. Künzli, U. Ackermann-Liebrich, R. Keller, A.P. Perruchoud, C. Schindler, SAPALDIA team. ©ERS Journals Ltd 1995.

ABSTRACT: Lung function testing of a random population sample in the eight SAPALDIA (Swiss study on air pollution and lung diseases in adults) centres had to be performed simultaneously, within one year, by eight teams and 23 technicians. We conducted quality control studies to test for technician, team and device related systematic measurement errors.

To assess technician effects, each centre conducted a study involving 12–19 subjects. Two studies with 13 participants each addressed team and device effects. In all studies, volunteers repeatedly performed spirometry with different technicians or devices. Effects due to technician, team or device were estimated (analysis of variance).

Neither "technician" within any of eight teams nor "team" accounted for significant differences of forced vital capacity (FVC) or forced expiratory volume in one second (FEV_1). The Device Effect Study revealed 10% lower FVC values for device No. 1 due to a technical problem occurring during the test day but not in the main SAPALDIA study. Further investigations revealed potential hardware and software sources of error which are not recognizable by trained technicians.

These studies gave no evidence for systematic errors due to technician, team or device during the main SAPALDIA study. However, they revealed potential sources of error in modern devices, which function as "black boxes". Manufacturers should improve spirometry software to further enhance the technicians' attempts at accurate assessment.

Eur Respir J., 1995, 8, 371–376.

*Swiss Study on Air Pollution and Lung Diseases in Adults. [+]Institute of Social and Preventive Medicine, University of Basle, Switzerland. [++]Pneumology Clinic, Barmelweid, Aarau, Switzerland. **Department of Pneumology, University of Basle, Switzerland.

Correspondence: N. Künzli, Institut für Sozial- und Präventivmedizin, Steinengraben 49, CH 4051, Basel, Switzerland

Keywords: Lung function variabilily, quality control, reproducibility, respiratory epidemiology

Received: December 19 1993
Accepted after revision November 24 1994

SAPALDIA is part of the Swiss National Research Programme 26A supported by the Swiss National Science Foundation (grant No. 4026-28099) and by the Federal Office of Education and Science. SAPALDIA Basle is part of the European Respiratory Survey. NK is supported by fellowship No. 823B-33245 of the Swiss National Research Fund Fellowship Program.

Forced expiratory vital capacity (FVC) and forced expiratory volume in one second (FEV_1) are a widely-used outcome in the assessment of the impact of ambient air pollution on respiratory diseases.

The Swiss Study on Air Pollution and Lung Disease in Adults (SAPALDIA) is assessing the impact of air pollutants and other environmental factors on respiratory health outcomes in the general population of eight areas [1]. The design of the cross-sectional part of the study corresponds to the approach common in air pollution epidemiology, assessing individual health measures across populations from different geographic areas with distinct levels of air pollution. The eight-centre design in a large population sample (n=17,300) required the hiring of eight teams, each consisting of 2–4 technicians, over a period of one year. Thus, systematic measurement errors between technicians within a team or between teams have to be considered as an important concern. Systematic errors in the pulmonary function assessment could bias the results in both directions. For example if technical errors in the most polluted area were producing consistently lower FVC, spurious effects due to air pollution are more likely to occur. Given the relatively low to moderate air pollution levels in Switzerland and the limited range of pollution across the eight SAPALDIA areas, random health measurement errors interfere with the power to detect environmental effects. High random measurement variability across areas could

SAPALDIA team: Study director: Ph. Leuenberger (p); Programme director: U. Ackermann-Liebrich (e), P. Alean (am), K. Blaser (a), G. Bolognini (p), J.P. Bongard (p), O. Braendli (p), P. Braun (p), C. Bron (l), M. Brutsche (l), C. Defila (m), G. Domenighetti (p), S. Elsasser (l), P. Filliger (am), L. Grizé (s), P. Guldimann (l), P. Hufschmied (l), W. Karrer (p), H. Keller-Wossidlo (o), R. Keller (p), N. Künzli (e), J.C. Luthi (l), B.W. Martin (l), T. Medici (p), Ch. Monn (am), A.G. Peeters (pa), A.P. Perruchoud (p), A. Radaelli (l), Ch. Schindler (s), J. Schwartz (s), G. Solari (p), M. Schoeni (p), J.M. Tschopp (p), B. Villiger (p), B. Wuethrich (a), J.P. Zellweger (p), E. Zemp (e). a: allergology; am: air pollution monitoring; e: epidemiology; l: local assistant MD; m: meteorology; o: occupational medicine; p: pneumology; pa: palynology; s: statistics.

obscure effects due to pollutants, yielding false negative conclusions. To focus on systematic and random measurement errors in the assessment of FVC and FEV_1, separate quality control studies were designed and will be presented in this report. The studies addressed the following questions:

1) Technician Effect Studies. Within each team, is there systematic measurement bias across technicians (2–4 per team)?

2) Team Effect Study. Is there systematic measurement bias across eight teams?

3) Device Effect Study. Is there evidence for systematic errors across the eight SAPALDIA pulmonary function devices (one device per team)?

4) Within-subject variability. How does within-subject variability measured with different technicians or devices compare to the expected biological within-subject variability, given only one technician and one device [2]?

The last question focuses on random variability, the former on systematic errors.

Material and methods

An experiment to evaluate systematic errors due to technicians, team and device, would ideally ask each of the 23 technicians to measure lung function on the same standard population. It would require 23 lung function sessions per subject; furthermore, we would repeat the experiment once for each device. Given the obvious logistical problems, we set up a different approach. Systematic errors due to technicians were assessed independently within each team, applying repetitive measurements on local groups of volunteers. Given no significant technician effects within a team, lung function assessment by any randomly chosen technician is assumed to depart from the subjects "true" value only due to random variation, *i.e.* technicians within a team may be considered exchangeable. Based on this assumption, to be tested within each team by the Technician Effect Studies, it will be feasible to conduct a small study assessing team effects. A study group will be tested once by each "team", selecting technicians out of each team at random.

Subjects and design

Table 1 provides a descriptive summary of the study populations. Each study involved different populations described below. Subjects from the SAPALDIA sample could not participate in these quality control studies. Spirometry methods described in the next section were identical for each study. All studies involved repetitive lung function measurements, with at least a 10–15 min interval between sessions.

Technician Effect Studies. This part consisted of eight independent, but methodologically identical studies: each local team organized 13–20 healthy, nonsmoking volunteers during months 6–8 of SAPALDIA. The numbers of subjects were assigned based on sample size calculations, to achieve a power of ≥0.8 at an alpha level of 0.1. Each subject had to participate in one spirometry session with each local technician, assigned in random order. Subjects performed their sessions at the local device on the same day.

Team Effect Study. Thirteen healthy volunteers were recruited at the University of Basle during the fourth month of SAPALDIA. Within 24 h, each subject performed nine spirometry sessions. The first session, considered to be practice, was excluded from the analysis. Each of the remaining eight sessions was carried out by a different technician, one from each team. Order of team was randomly assigned to each subject. All measurements were performed with the device of SAPALDIA Basle.

Device Effect Study. This study was conducted four months after the last SAPALDIA lung function test. All

Table 1. – Descriptive summary of the subjects for Technician, Team and Device Effect Studies, SAPALDIA 1991

Studies	Subject n	Sex F/M	Age yrs*	Height cm	FVC ml*	FEV_1 ml*
Technician Effect Studies						
Team No. 1	17	10/7	30 (10.0)	171 (7.1)	4728 (846)	3819 (618)
Team No. 2	15	7/8	36 (14.5)	171 (10.3)	4989 (1077)	3892 (883)
Team No. 3	13	13/0	25 (5.8)	169 (5.7)	3940 (520)	3467 (373)
Team No. 4	17	7/10	31 (8.5)	171 (6.8)	4802 (834)	3967 (608)
Team No. 5	12	4/8	36 (12.2)	170 (12.5)	5325 (1546)	4149 (1114)
Team No. 6	17	10/7	31 (8.3)	169 (6.1)	4627 (903)	3728 (600)
Team No. 7	17	13/4	37 (8.7)	168 (7.6)	4437 (888)	3513 (834)
Team No. 8	19	14/5	45 (11.5)	167 (8.5)	4224 (752)	3214 (461)
Team Effect Study	13	8/5	25 (3.5)	176 (11.4)	5260 (1036)	4430 (753)
Device Effect Study	13	6/7	24 (2.9)	175 (7.7)	5293 (1130)	4275 (743)

Data for age, height, FVC and FEV_1 are presented as mean (SD). *: ANOVA, 10 study populations, age, FVC, FEV_1: $p<0.01$; SAPALDIA: Swiss study on air pollution and lung diseases in adults; F: female; M: male; FVC: forced vital capacity; FEV_1: forced expiratory volume in one second; ANOVA: analysis of variance.

SAPALDIA devices, eight in total, were transported to the University of Basle. A water-sealed system from the same manufacturer was included in the test circuit (Sensormedics 2400). Thirteen healthy nonsmoking volunteers participated in 10 spirometry sessions each: after one practice session, excluded from the analysis, subjects performed one session on each device in randomly assigned order. All tests were carried out by the same technician on a single day.

Methods

All spirometry sessions were performed according to the SAPALDIA protocol, which corresponds to the European Community Respiratory Health Survey methods [3]. Sensormedics 2200, an open system equipped with a mass flow meter sensor, was used in each centre. The fully computerized device fulfils the American Thoracic Society (ATS) performance criteria [4]. Forced expiratory manoeuvres were performed in the sitting position wearing a noseclip. Subjects could not watch the screen or receive any personal results unless the last session was performed. Technicians were unaware of results from prior sessions. At least three and up to eight manoeuvres were required for each session, to provide a minimum of two acceptable results both for FVC and FEV_1 reproducible within 5% [5]. Immediate computerized feedback regarding the major acceptability criteria (including beginning, duration and end of the manoeuvre) and the required reproducibility supported the technicians attempt to standardize procedures. Device calibration with a 3 *l* syringe was performed whenever the device was switched on, at least once a day. Training of all technicians consisted of a 3 day workshop followed by 2 months of exercise on volunteers prior to the onset of SAPALDIA. Both technician and flow-volume charts were repeatedly supervised to enhance data quality.

Analysis

Our studies utilized repeated measure design with the same subject being repeatedly tested by different technicians, teams or devices. Thus, the subject serves as a "block" or as his own control, and the experimental unit within a block may be viewed as the test session provided by different technicians, teams or devices, *i.e.* our "main factors" of interest [6]. An advantage of the repeated measure design, in addition to economizing on subjects, is its good precision for comparing "main factor" effects, because all sources of variability between subjects are excluded from the experimental error. Analysis of variance (ANOVA) including "subject" and "main factor", *i.e.* technicians, team or device respectively, is an adequate model to test for the respective factor. From each test session, best values of FVC and FEV_1 were used as dependent variables in the analyses. "Subject" mean square captures between-variability due to gender, age, height, weight and other individual factors [2]. For example, for the Device Effect Study and FVC as dependent variable (Y), the single-factor repeated measure is described by the following model:

$$Y_{ij} = \mu_{..} + P_i + D_j + e_{ij}$$

where FVC of the ith person, tested at the jth device is estimated by the overall mean FVC $\mu_{..}$, a random effect P_i of the ith person, a random effect of the jth device ("main factor") and the error term e_{ij} [6].

Mean squares of the main effect divided by the mean square error provides a small F-statistic, if the main effect does not explain within-subject variability in the lung function measure. The F-statistic in a balanced design, *i.e.* with equal number of test sessions per subject, is quite robust against deviations from normality and homoscedasticity of residuals. The null hypothesis of no main effect was rejected at a conservative level of p=0.1. Power calculations were based on the methods described by NETER *et al.* [6].

Table 2. – Results of Technician, Team and Device Effect Studies, SAPALDIA 1991

Main effect	Total number				FVC			FEV_1		
tested	Technicians	Devices	Subjects	Tests	Mean†	F#	p-value	Mean†	F#	p-value
Technician										
Team No. 1	3	1	17	51	111	1.49	0.24	105	0.76	0.48
Team No. 2	4	1	15	60	119	1.11	0.35*	110	0.7	0.56*
Team No. 3	2	1	13	26	99	0.37	0.56	100	0.97	0.34
Team No. 4	3	1	17	51	110	2.49	0.1	107	0.97	0.34
Team No. 5	2	1	12	24	124	1.29	0.28*	115	0.01	0.91*
Team No. 6	3	1	17	51	111	0.13	0.88	105	0.08	0.92
Team No. 7	4	1	17	68	118	1.3	0.28	109	0.62	0.6
Team No. 8	2	1	19	38	120	1.86	0.19	109	0.77	0.39
Team	8+	1	13	104	113	1.59	0.15	112	0.85	0.54
Device	1	9	13	117	110	48.7	<0.001	104	34.3	<0.001

+: per subject, 8 technicians out of 23 (one technician from each of eight teams); †: expressed as percentage predicted, ECSC reference values [8] (ANOVA across studies: FVC p<0.01; FEV_1 p=0.26); #: F-statistic for main effect (technician, team, or device, respectively) from ANOVA on subject and main effect; *: power to detect ±2.5% difference 0.7–0.8 at alpha=0.1 (other studies >0.8). ECSC: European Community for Steel and Coal. For further abbreviations see legend to table 1.

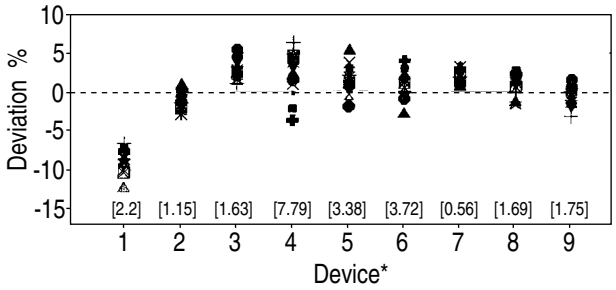

Fig. 1. – Percentage deviation of FVC from individual mean FVC in nine devices (13 subjects). Values in square brackets indicate between subject variance of the percent deviation from individual mean FVC. Each symbol represents one subject. FVC measured at leaking device No. 1 was significantly different from all others ($p<0.001$). Devices No. 1–8; SAPALDIA devices, Sensormedics 2200; Device No. 9: water-sealed spirometer, Sensormedics 2400. SAPALDIA: Swiss Study on Air Pollution and Lung Diseases in Adults; FVC: forced vital capacity.

From both the team and device effect study, within-subject variability was calculated using within-subject standard deviation across best values of repetitive sessions, expressed as percentage of the mean (coefficient of variation). The mean coefficient of variation across 13 subjects represents an overestimation of the within-subject variability, due to the involvement of different technicians or devices, respectively. All procedures were analysed using SAS statistical software run on a main frame [7].

Results

Technician Effect Study

Table 2 reports the results for each of eight local studies (teams No. 1–8). Lung function mean values are presented as percentage predicted [8]. Across the eight local substudies assessing technicians effects, 10 out of 137 volunteers did not fulfil the acceptability or reproducibility criteria in more than one session. These subjects were excluded from the analysis, limiting random variability, thus increasing the likelihood to reject the null hypothesis of no team effect. Exclusions were unrelated to team or technician. None of the F-statistics for FVC and FEV_1 were significant at $p<0.1$. In all but two studies, power to detect a ±2.5% difference was ≥0.8. Group mean difference between technicians with lowest and highest mean ranged 0.5–2.9% for FVC and 0.2– 2.5% for FEV_1 across eight teams.

Team Effect Study

F-statistics for both, FVC and FEV_1, were not statistically significant (table 2). Lowest and highest team-wise assessed mean values differed by 2.8% or 146 ml for FVC, and 2.3% or 98 ml for FEV_1. Four of the 104 tests did not meet the ATS reproducibility criteria, and two tests violated the criteria for an acceptable start and end of test, respectively. To maintain the balanced design, these values were included in the analysis. Failure to achieve ATS criteria was unrelated to team.

Device Effect Study

Surprisingly, the analysis showed significant device effects both for FVC and FEV_1, with 10% lower FVC for device No. 1 compared to the mean values across the remaining eight devices (table 2). As shown in figure 1, each subjects lowest performance happened to be with device No. 1. The results of other devices ranged over 3.9% or 208 ml around their average FVC of 5,290 ml. For FEV_1, similar results were obtained with an 11% lower mean assessed by device No. 1, and a 2.8% or 124 ml range around the mean of 4,285 ml for the other eight devices. Results from the "reference device" No. 9, a water-sealed spirometer, were similar to those measured on devices No. 2–8, used in SAPALDIA.

Within-subject variability

Mean within-subject coefficient of variation over a series of eight tests assessed by eight technicians were 2.7% (95% confidence interval (95% CI: 1.93–3.46%)

Table 3. – Within-subject variability of FVC and FEV_1 in healthy subjects: SAPALDIA quality tests compared to other published studies

Study	Subject n	Tests per subject n	Technicians per subject n	Retest within days n	Age yrs	Coefficient of variation % FVC	Coefficient of variation % FEV_1
SAPALDIA Quality Studies:							
With 8 teams	13	8	8	2	24 (20–31)	2.7 (SE 0.39)	3.3 (SE 0.5)
With 8 devices	13	8	1	1	24 (19–29)	2.0 (SE 0.14)	2.2 (SE 0.3)
McCarthy *et al.* [9]	12	10	1	1	27 (22–34)	2.5	2.5
Cochrane *et al.* [10]	9	10	1	1	24 (19–30)	1.8	2.3
Nickerson *et al.* [11]	15	5–12	1	1	18 (8–34)	3.5	3.6
Lebowitz *et al.* [12]	10	60	1	25–35	34 (25–50)	3.5	3.6
Rozas and Goldman [13]	15	5	1	5	- (25–52)	2.8 (SE 0.55)	2.8 (SE 0.4)
Groth *et al.* [14]	112	2	1–2	15–180	47 (30–70)	4.9	4.7

Data for age are presented as mean, and range in parenthesis. Coefficient of variation: standard deviation in % of mean. For abbreviations see legend to table 1.

and 3.3% (2.29–4.31%) for FVC and FEV_1, respectively. The corresponding results with eight tests assessed by one technician on eight different devices were 2.0% (1.75–2.25%) in FVC and 2.2% (1.6–2.8%) in FEV_1, excluding measurements on device No. 1 due to its calibration error discussed below. Table 3 compares our within-subject variability estimates with studies involving only one technician and device per subject [9–13].

Discussion

The tests presented here are, to our knowledge, the first to report FVC and FEV_1 variability due to technician, team, and device within an eight centre study involving 23 technicians. Our comparisons showed homogeneous lung function assessment across 23 experienced SAPALDIA technicians and eight teams. To the extent that these studies are representative of the technicians' performance throughout 1991, systematic measurement errors related to technicians or teams are unlikely to have occurred in SAPALDIA. This generalization is hampered by the limited power of 0.7–0.8 among two Technician Effect Studies (teams No. 2 and 5), and by potential interactions between subjects and technicians in the fieldwork. Power differences across the eight independent within-team studies are due mainly to demographic differences in the volunteers selected (table 1). Furthermore, despite the requirement to engage "healthy nonsmoking volunteers", health was not explicitly assessed. Less healthy subjects could increase total variability.

In our Team Effect Study, mean results across eight teams, measuring the same 13 subjects, varied randomly within a small range among the most extreme team values. The analysis is based on the assumption of no technician effects within a team. To simultaneously address team and technician effects, a constrained regression model was applied, resulting in slightly smaller F-values, confirming our results of no team effect.

The magnitude of the within-subject between-session variability was comparable with other studies, despite our involvement of eight different technicians or eight devices (table 3) [14]. The most comparable studies regarding number of healthy participants, tests and time intervals reported within-subject coefficients of variation ranging 2.3–3.6%. We also estimated within-subject variability by partitioning the variance in the above-mentioned constraint regression model, getting very similar results. Technician training and on-line quality control are probable explanations for these homogeneous results. As shown in the Lung Health Study, continuous quality control procedures considerably decrease measurement variability [15]. Computerized devices, including immediate acceptability and reproducibility check, support standardization of the procedures and may limit within-subject variability across best values of several sessions. This statement is supported by an analyses including the first lung function session of each subject, performed as a training session. In this young healthy population, best values from the first session were not different from the others. The only sign for a "learning effect" was given by the tendency toward more manoeuvres needed in the first session to meet the criteria (3.54 *vs* 3.25 attempts; p=0.1).

Device effect

Performance of all eight devices was compared 4 months after the last SAPALDIA assessment. This test showed a systematic measurement error, with one device 10% removed from the average of others. Each subject's lowest value was measured with device No. 1 (fig. 1). To investigate this unexpected result, we organized an additional two device comparison with the same volunteers. In this test, results of device No. 1 were no more different from those measured on the reliable devices a month previously. An explanation for the error observed previously could be detected. During the Device Effect Study, device No. 1 was equipped with a mouthpiece with two holes of 0.5 cm diameter each. As reported by Townsend [16] for a rolling-seal spirometer and reproduced with our open system, such holes yield lower FVC and FEV_1. Fortunately, these adapted mouthpieces did not regularly circulate in the local SAPALDIA teams.

Another potential source of device related errors could be detected during our last investigation. Accuracy of electronic spirometry systems depends both on "external calibration", performed with a syringe, and "internal" hardware/software performance. Control of "internal calibration", including the potentiometer status, was not considered a regular procedure by the manufacturer and, fortunately, remains highly constant under normal circumstances. However, during the last two device comparison tests we were able to reproduce unreliable results without getting any error message from the software, *i.e.* technicians could not be aware of important internal error sources. Such unrecognizable technical problems might introduce errors on any day, with any device. To prevent unrecognized hardware or software errors, manufacturer should: 1) improve on-line messages for system errors; 2) extend the usual calibration procedures, including accuracy checks for hardware and software; and 3) provide software which allows storage of daily calibration data as part of each individual spirometry session. The former would enable technicians to clearly recognize technical errors that may otherwise influence the results. The latter would allow researchers to retrospectively check technical reliability of the data, or locate errors that had occurred in time. Lacking such software options requires researchers to statistically control for technician, team and/or device effects on the stage of analysis.

Conclusion

The potential for systematic errors and increased random error is a trade-off in multicentre study designs involving several local teams and technicians. Our tests indicate that the application of widely-used guidelines for lung function testing, including teaching, training,

on-line quality control and regular supervision, are effective tools to prevent systematic technician effects and to limit random measurement error. However, our device comparison tests highlight potential device error sources that are not easily recognized in the daily work. Even for well-trained technicians, technical details of the device remain a "black box". In addition to published guidelines for pulmonary laboratories [17], we recommend software adaptations that enhance the technicians' attempt at accurate unbiased assessment. Whilst focusing on technicians' performance, we recommend researchers not to assume but to test reliability and accuracy of hardware and software performance in the fieldwork. Comparison tests should be organized prior to onset of the study. Enforcement of quality on all levels of the data collection process remains a primary goal.

Acknowledgements: The authors thank the SAPALDIA team technicians at Aarau, Basle, Davos, Genève, Lugano, Montana, Payerne and Wald for their fieldwork. They are indebted to I. Tager for his review of an early version of the manuscript.

References

1. Ackermann-Liebrich U. Künzli N. Leuenberger Ph, SAPALDIA Team. Air pollution exposure and asthma. The Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA). Brussels and Budapest. *In*: EC (COST). East Europe Workshop on Air Pollution Epidemiology, 1993.
2. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991; 144: 1202–1218.
3. Burney P, Luczynska Ch, Chinn S, Jarvis D. The European Respiratory Health Survey. *Eur Respir J* 1994; 7: 954–960.
4. Nelson ST, Gardner RM, Crapo RO, Jensen RL. Performance evaluation of contemporary spirometers. *Chest* 1990; 97(2): 288–296.
5. American Thoracic Society. Standardization of spirometry: 1987 update. *Am Rev Respir Dis* 1987; 136: 1285–1298.
6. Neter J, Wasserman W, Kutner MH. Applied Linear Statistical Models; Regression, Analysis of Variance, and Experimental Designs. 3rd edn. IL, Boston MA, Irwin, Homewood; 1990.
7. SAS Institute Inc. SAS/STAT User's Guide, Version 6, 4th edn. Cary, NC, SAS Institute Inc., 1989.
8. Quanjer PhH, Tammeling GJ, Pederson OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report working party standardization of lung function tests. European Community for Steel and Coal. Official statement of the European Respiratory Society. *Eur Respir J* 1993; 6 (Suppl.16): 5–40.
9. McCarthy DS, Craig DB, Cherniak RM. Intraindividual variability in maximal expiratory flow-volume and closing volume in asymptomatic subjects. *Am Rev Respir Dis* 1975; 112: 407–411.
10. Cochrane GM, Prieto F, Clark TJH. Intrasubject variability of maximal expiratory flow-volume curve. *Thorax* 1977; 32: 171–176.
11. Nickerson BG, Lemen TJ, Gerdes CB, Wegmann MJ, Robertson G. Within-subject variability and percent change for significance of spirometry in normal subjects and in patients with cystic fibrosis. *Am Rev Respir Dis* 1980; 122: 859–866.
12. Lebowitz MD, Knudson RJ, Robertson G, Burrows B. Significance of intraindividual changes in maximum expiratory flow volume and peak expiratory flow measurement. *Chest* 1982; 81: 566–570.
13. Rozas CJ, Goldman AL. Daily spirometric variability; normal subjects and subjects with chronic bronchitis with and without airflow obstruction. *Arch Intern Med* 1982; 142: 1287–1291.
14. Groth S, Dirksen A, Dirksen H, Rossing N. Intraindividual variation and effect of learning in lung function examinations: a population study. *Bull Eur Physiopathol Respir (Clin Respir Physiol)* 1986; 22: 35–42.
15. Enright PL, Johnson LR, Connett JE, Voelker H, Buist SA. Spirometry in the Lung Health Study. 1. Methods and quality control. *Am Rev Respir Dis* 1991; 143: 1215–1223.
16. Ferris BG, Speizer FE, Bishop Y, Weener PJ. Spirometry for an epidemiologic study: deriving optimum summary statistics for each subject. *Bull Eur Physiopathol Respir* 1978; 14: 145–166.
17. Gardner RM, Clausen JL, Cotton DJ, *et al*. Computer guidelines for pulmonary laboratories. *Am Rev Respir Dis* 1986; 134: 628–629.