



Quality control of spirometry: a lesson from the BRONCUS trial

R. Pellegrino*, M. Decramer[#], C.P.O. van Schayck[†], P.N.R. Dekhuijzen⁺,
T. Troosters[#], C. van Herwaarden⁺, D. Olivieri[§], M. Del Donno^f, W. De Backer^{**},
I. Lankhorst^{##} and A. Ardia^{##}

ABSTRACT: This report describes the quality control programme used within the Bronchitis Randomized on N-acetylcysteine (NAC) Cost-Utility Study, a trial designed to assess the decline in lung function, exacerbation rate, health status, and cost-effectiveness with NAC or a placebo in 523 patients with chronic obstructive pulmonary disease over a 3-yr period.

Spirometry was scored from 0 (worst quality) to 6 (best quality). The mean score of 314 spirometries from 243 patients evaluated during the trial was 5.63 ± 0.83 . Linear regression analysis of the scores of 47 participating centres plotted against the time at which spirometries were performed yielded an intercept of 5.7 ± 0.5 and a slope of -0.0001 ± 0.001 , which suggests that the initial high quality was maintained over time.

Retrospective examination of a further 345 postbronchodilator spirometries from 208 patients with a forced expiratory volume at one second exceeding the mean individual value recorded over the study in excess of 20% revealed a slightly lower quality of the start-of-test manoeuvre compared with the 314 spirometries.

In conclusion, these findings would suggest that the quality control programme is likely to have helped achieve and maintain long-term spirometry performance in the Bronchitis Randomized on N-acetylcysteine (NAC) Cost-Utility Study trial. Special care should be paid to the spirometries whose forced expiratory volume in one second values exceed the mean value.

KEYWORDS: Chronic obstructive pulmonary disease, N-acetylcysteine, quality control, spirometry

In multicentre studies the outcomes are dependent upon the quality of the collected parameters. This is especially true when the latter are derived from spirometry examination, because factors such as patient cooperation, technician skill, instrument quality, centre organisation and motivation over time may eventually affect the final output [1–7]. According to the literature, performance and quality of lung function parameters can be maintained at high levels during trials of long duration only if the variability of the instruments, software, and organisation is minimised, and regular and strict spirometry quality-control programmes are applied [1–5]. However, the implicit limits of such measures are the increased cost of the trials.

The Bronchitis Randomized on N-acetylcysteine (NAC) Cost-Utility Study (BRONCUS) is a recently designed trial for assessing the decline in lung function, exacerbation rate, health status and its decline, and the cost-effectiveness of NAC or a placebo in chronic obstructive pulmonary disease (COPD) over a 3-yr period [8]. The study

was assisted by a Quality Control Committee (QCC) aimed at evaluating the quality of the respiratory function to help the centres achieve and maintain best practice standards. This purpose was accomplished by evaluating and scoring 243 postbronchodilator spirometries taken in random order from 47 centres. The spirometry scores of each centre, linearly regressed against time, gave an index of the initial quality (intercept) and maintenance over time (slope). In addition, after the trial was terminated a further 345 postbronchodilator spirometries, which had a forced expiratory volume in one second (FEV₁) exceeding the mean individual value recorded over the study in excess of 20%, were evaluated to see whether FEV₁ assessments higher than expected were associated with poor quality of the forced expiratory manoeuvre.

MATERIALS AND METHODS

General design

The study was designed as a phase III, two-arm double-blind, randomised placebo-controlled

AFFILIATIONS

*Centro di Fisiopatologia Respiratoria e di Studio della Dispnea, Azienda Ospedaliera S. Croce e Carle, Cuneo, and

[§]Respiratory Division, Ospedale Rasori Parma, Parma, and

^fServizio di Pneumologia, Azienda Ospedaliera "G. Rummo", Benevento, and

^{##}Zambon Group, Bresso-Milano, Milan, Italy.

[#]Respiratory Division, University Hospital, Katholieke Universiteit Leuven, Leuven, and

^{**}University Hospital, Antwerp, Belgium.

[†]Dept of General Practice, Maastricht University, Maastricht, and

[‡]Dept of Pulmonary Diseases, University Medical Centre Nijmegen, The Netherlands.

CORRESPONDENCE

R. Pellegrino

Centro di Fisiopatologia Respiratoria e di Studio della Dispnea

Azienda Ospedaliera S. Croce e Carle
12100 Cuneo, Italy

Fax: 39 0171616798

E-mail:

pellegrino.r@ospedale.cuneo.it

Received:

March 07 2005

Accepted after revision:

August 01 2005

European Respiratory Journal

Print ISSN 0903-1936

Online ISSN 1399-3003

parallel group trial, examining the effects of NAC 600 mg·day⁻¹ versus a placebo on annual decline in lung function, exacerbation rate, quality of life and cost-effectiveness. A total of 523 patients were tracked for 3 yrs.

Centres selection

University or general hospital centres were selected on indication of the country coordinators regarding the institutions' previous experience in national or international clinical trials, established pulmonary function test laboratories following international recommendations for spirometry guidelines, and willingness to perform the trial.

Patient selection

The patients came from 50 centres in 10 European countries. Inclusion criteria were smoking-related COPD between 40–70 yrs, a postbronchodilator FEV₁ between 40–70% of predicted, FEV₁ and/or forced vital capacity (FVC) reversibility of <12% pred and <200 mL 15 min after inhalation of 400 µg of salbutamol through an metered dose inhaler and a spacer, FEV₁/vital capacity (VC) ratio of <88% of pred in males and <89% of pred in females, and a history of at least two exacerbations in 1 yr over the last 2 yrs prior to enrolment.

Spirometry evaluations

All spirometries were done with the use of a pneumotachograph, except in one centre which used a dry spirometer. The same equipment was maintained for the entire duration of the study.

The centres were asked to follow the spirometry recommendations of the European Respiratory Society (ERS) [9]. Prior to the trial, a 3-h teaching session on the technical aspects of spirometry and calibration was attended by the investigators and the personnel who conducted the trial. In brief, recommendations were given for acceptability and repeatability criteria of the maximum forced expiratory manoeuvres as listed in table 1. Because only a few centres could provide the complete series of the classical quality control parameters [1, 2, 6], records of at least the three best acceptable forced

expiratory manoeuvres with relevant flow-volume loops (FVL) and corresponding values of FEV₁, FVC, and peak expiratory flow (PEF) were requested for future evaluation by the QCC. The centres were asked to calibrate the equipment before each study session with a syringe of 3 L according to the ERS guidelines [9]. No checks were performed by the QCC on calibration.

Spirometries were arbitrarily scored by one of the authors of this paper on the basis of both visual inspection of the FVL and evaluation of the FEV₁ and FVC repeatability on the delivered spirometries (table 2). With visual analysis of the FVL, a score of 2 was assigned to a spirometry fulfilling all the following start-of-manoeuve parameters: sharp PEF, steep upslope of the FVL, and absent back-extrapolation volume (BEV) (<100 mL or 5% FVC, if available in the printout). In contrast, a score of 0 was assigned to a spirometry with either a visually flat PEF, or a flat upslope of the FVL, or an evident BEV (>100 mL or >5% FVC, if available in the printout). For the assessment of the FEV₁ repeatability criteria, the scoring was as follows: 2, if the largest FEV₁ of three reported acceptable manoeuvres did not exceed the next highest value by >0.1 L; 1 if it was between 100–200 mL; and 0 >200 mL. The same held for the repeatability of the FVC. Thus, for each spirometry, the highest score was 6. Expiratory time was not considered in the present scoring system because it was not constantly shown in the final report or measurable on the volume-time plots. The terminal portion of the FVL was carefully inspected to see whether or not flow tended to gradually decrease to zero rather than being suddenly interrupted, thus helping exclude obvious cases of premature termination of the forced manoeuvre before 6 s.

Throughout the trial, the QCC reviewed and scored 314 postbronchodilator spirometries randomly selected by the Statistics Department of Zambon group from 243 patients at 50 centres. Selection of the tests to be evaluated was based on the criterion that 100% of the centres, and a minimum of 40% of the patients and 5% of the spirometries had to be included in the sample. In this current paper, only the results of 47 centres

TABLE 1 Main recommendations for spirometry

Procedure	Recommendations
Body posture	Upright position
Start-of-test criteria	Extrapolation volume <100 mL or 5% FVC, whichever is greater PEF sharp and close to maximum inflation. Steep upslope of the FVL
End-of-test criteria	Forced expiration lasting a minimum 6 s
Inspiration preceding the forced expiration	Fast and with no hesitation with breath-holding time at maximum inflation <1 s
Forced expiratory manoeuvre	No cough in the initial 1 s No tongue obstructing the mouthpiece No cessation of flow due to glottis closure or decreased effort No leaks, no drift
Manoeuvres n	A minimum of three to a maximum of eight
Repeatability	Three technically satisfactory manoeuvres with the largest FEV ₁ and FVC not exceeding the next highest values by >5% or 0.1 L, whichever is greater.

FVC: forced vital capacity; PEF: peak expiratory flow; FVL: flow-volume loop; FEV₁: forced expiratory volume in one second.

TABLE 2 Spirometry score system

Score	FVL visual analysis	FEV ₁ and FVC repeatability	
		FEV ₁ mL	FVC mL
2	Sharp PEF, and steep upslope, and absent BEV (or <100 mL or 5% FVC if available in the printout)	<100	<100
1		>100<200	>100<200
0	Either flat PEF, or flat upslope, or evident BEV (or >100 ml or 5% FVC if available in the printout)	>200	>200

FEV₁: forced expiratory volume in one second; FVC: forced vital capacity; FVL: flow-volume loop; BEV: back extrapolation volume; PEF: peak expiratory flow; upslope: upslope of the FVL.

are reported; this was because for three centres the spirometries were too few and insufficient to assess the quality standards over time. The QCC regularly reported on the quality of the spirometry at the biannual Steering Committee meetings and the results were subsequently sent to the participating centres together with the necessary persuasive recommendations on how to maintain or improve, if necessary, the standards of spirometric tests. When there was evidence that the quality of spirometry tended to decrease in a particular centre, the principal investigators were contacted in order to discuss the specific problems and potential actions to be taken to improve the quality.

The quality standards of the centres were estimated by linearly regressing the scores of all spirometries of each centre evaluated by the QCC *versus* the time at which the tests were performed. A representative example of the quality of a centre is shown in figure 1. The resulting intercept was taken as an index of the initial quality level, and the slope as an index of performance over time.

An additional 345 spirometries of 208 patients with a postbronchodilator FEV₁ exceeding the mean individual value recorded over the study in excess of 20% were evaluated after the study was completed.

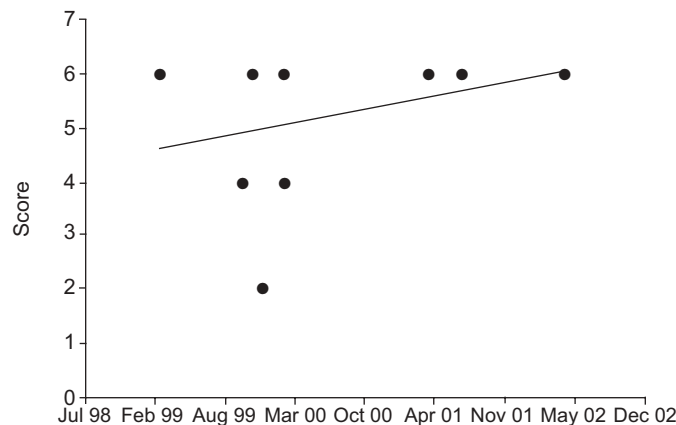


FIGURE 1. Typical example of linear regression analysis of the scores of the spirometries of a centre plotted *versus* the time at which the tests were performed.

STATISTICS

Mann-Whitney U-test, Wilcoxon signed-ranks test, and Chi-squared test were used whenever appropriate to test the differences between variables. Correlations were assessed by Pearson’s test. Values are reported as mean ± (SD). When p<0.05 it was considered to be statistically significant.

RESULTS

The main anthropometric and functional data of the patients studied during the trial and those with an FEV₁>20% of the individual mean were similar to the entire group that participated in the BRONCUS trial (table 3).

Of the 314 spirometries randomly evaluated throughout the trial, 42 (13%) had BEV, and 123 (39%) had an expiratory time reported in the printout. Visual analysis of these FVLs was capable of identifying the lack of excessive BEV in 40 of the 42 spirometries. BEVs of the remaining cases were 0.13 and 0.11 L. Expiratory time >6 s for the forced expiratory manoeuvres occurred in 117 out of the 123 cases (10.9 ± 4.6 s). Average total score of all 314 spirometries was 5.63 ± 0.83 (table 4). Both scores of start-of-test manoeuvre and FVC repeatability were significantly less than that of FEV₁ repeatability (p<0.001 for both). Start-of-test manoeuvre and FEV₁ and FVC repeatability fulfilled the ERS requirements in

TABLE 3 Main anthropometric and lung function data

	Random control	Spirometry with FEV ₁ >20%	BRONCUS trial
Patients n	243	208	523
Spirometries	314	345	5042
Sex M/F	187/56	171/37	413/110
Smokers former or current/non	143/100	102/106	284/239
Age yr	62 ± 7	62 ± 8	63 ± 8
FEV ₁ % pred	53 ± 10	52 ± 9	53 ± 9

Data presented as n or mean ± SD unless otherwise stated. M: male; F: female; FEV₁ % pred: forced expiratory flow in one second percentage predicted; BRONCHUS: Bronchitis Randomized on N-acetylcysteine (NAC) Cost-Utility Study.

TABLE 4 Quality scores			
	Random control	p-value	Spirometries with FEV ₁ >20%
Patients n	314		345
Start-of-test	1.87±0.50*	<0.001	1.69±0.72 [#]
FEV ₁ repeatability	1.96±0.21*, [¶]	NS	1.94±0.30 ^{#, +}
FVC repeatability	1.81±0.49 [¶]	NS	1.73±0.60 ⁺
Total score	5.63±0.83	<0.002	5.36±1.13

FEV₁: forced expiratory volume in one second; FVC: forced vital capacity. Differences between groups were tested by the Mann-Whitney U-test. Differences within groups were tested by the Wilcoxon test and are indicated with pairs of symbols (*, #, ¶, + <0.001).

93, 96, and 81% of the patients respectively. Both start-of-test and FEV₁ criteria were satisfied in 90% of the patients. The linear regression analysis of the spirometry scores plotted against time was constructed with an average of 7±3 observations per centre and over a period of 899±344 days (mean $r^2=0.51\pm0.44$). The intercept was 5.7±0.5, thus suggesting high initial performance. The slope was not significantly different from zero (-0.0001 ± 0.001), thus suggesting that the overall quality of spirometry remained quite stable throughout the 3 yr trial. Intercept and slopes of the 47 centres are shown in figure 2. The average total score was not correlated with age, body mass index (BMI), education, FEV₁ % pred, FVC % pred, or FEV₁/FVC. Males and females performed similarly, whereas former smokers performed slightly, but significantly, better than current smokers (5.5 ± 1.0 and 5.2 ± 1.2 , respectively; $p<0.05$).

Of the 345 spirometries with a postbronchodilator FEV₁>20% than the mean individual value recorded over the study, 35 had BEV and 123 had expiratory times reported in the printout. Visual analysis of these FVLs was capable of identifying the absence of BEV in 32 cases. Expiratory time of the expiratory manoeuvre was >6 s on 117 occasions

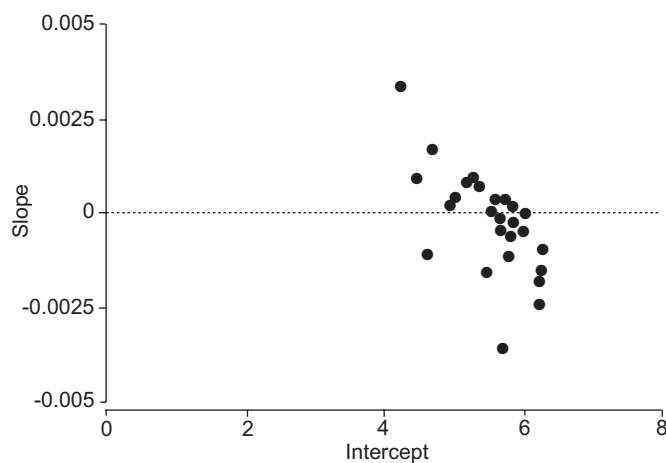


FIGURE 2. Intercepts and slopes of total score values plotted against time for the 47 centres.

(11.3 ± 3.8 s). An average total score of the spirometries (5.36 ± 1.13) was slightly, but significantly, less than that of the 314 spirometries evaluated throughout the trial ($p<0.002$), thus suggesting a slightly worse performance. This was due to a slightly, but significantly, worse start-of-test quality (1.69 ± 0.72) ($p<0.001$). Both start-of-test and FVC repeatability scores were significantly worse than FEV₁ repeatability ($p<0.001$ for both). The start-of-test and FEV₁ and FVC repeatability criteria fulfilled the ERS requirements in 85, 94, and 73% of the patients respectively. Both start-of-test and FEV₁ criteria were satisfied in 83% of the patients. Average data are presented in table 4. Average total scores were correlated with BMI ($r^2=0.14$, $p<0.01$), but not with age, education, FEV₁ % pred, FVC % pred, and FEV₁/FVC.

DISCUSSION

The main findings of this study suggest that a regular quality control programme based on random supervision was apparently capable of achieving and maintaining an adequate level of performance over the 3-yr period in the BRONCUS trial. However, the quality of the spirometries with an FEV₁ exceeding the mean individual value recorded over the study in excess of 20% was slightly, but significantly, worse than the spirometries below that threshold.

The rate of spirometries failing to meet the criteria of acceptability and repeatability for the classical functional parameters reported in the literature is highly variable. For example, a variability rate as low as 2% to as high as 80% of the spirometries had been reported as failing to meet the American Thoracic Society (ATS) standard for the FEV₁ [1–5, 10]. This rate would be slightly higher if the stricter ERS criteria were applied. Worse results have been reported for FVC [2, 4, 7]. Even if assessed with different methods, the rate of performance in the present trial was apparently well within the average range of the studies assisted by very strict QCCs [1–7]. Many potential factors could have contributed to the success of the trial. Firstly, the trial was conducted in well-recognised European centres of excellence with long-standing experience in basic and clinical research. Secondly, rather than standardising the devices in all centres, the investigators were allowed to use their own equipment in the hope that working with their familiar spirometers could largely overcome the disadvantages of using new homogeneous equipment. Whether these two factors contributed to maintaining the standard of the current results cannot be inferred from these data, even though they do represent a reasonable presupposition.

QCCs have been used repeatedly in recent multicentre trials without any clear indication regarding the extent of effort and resources needed to be deployed in order to guarantee the desired results. Some studies have utilised top technical and human resources including effective training of the investigators, the supply of the same spirometers to the investigating centres together with software capable of collecting and transmitting all data to a coordinating centre along with personnel dedicated to visiting, to data processing, and statistical analysis [1–5]. Notwithstanding all this, the rate of spirometries not fulfilling the international guidelines never fell to zero [1, 5]. It is difficult to say whether our QCC was as effective in maximising the quality of the spirometric tests as those of previous trials, mainly due to the differences in the

quality assessment of the spirometries and the effort spent encouraging the investigators. From a technical point of view, the current authors acknowledge the numerous methodological differences with the previous trials [1–7]. Yet, for the reasons listed below, it is believed that the current approach was at least acceptable for the purpose of the trial. First, visual analysis of the start-of-test manoeuvre has already been used in research [11]. Secondly, visual assessment of the initial part of the FVL in the current subsets of 42 and 35 cases yielded similar results to the numerical analysis. Thirdly, a comprehensive gold standard for proper evaluation of the quality of spirometry is still lacking, for important parameters such as time and volume history effects of the inspiratory manoeuvre preceding the forced expiration have never been considered [12]. The current authors accept that simple analysis of FVC repeatability, without measuring the expiratory time, could have missed some unacceptable tests. However, the current findings in the subsets of spirometries with reported expiratory time in the printout would demonstrate that the duration of the best manoeuvre was, in most cases, above 6 s. Thus, the problem of the low repeatability of the FVC must have accounted for the difference in time of the three acceptable manoeuvres.

The present results show that start-of-manoevure failed to meet standard criteria significantly more often than the repeatability of the FEV₁. It is common experience that many patients consistently hesitate to initiate the forced expiratory manoeuvre or lose some air before blowing out. Others in contrast, are able to do submaximal efforts with impressive repeatability of FEV₁ and FVC. Because the beginning of forced expiration may seriously affect the FEV₁ and its interpretation [13], especially after medical interventions or over time, these findings suggest the importance of visual inspection of the FVL together with the back extrapolation volume and time-to-PEF whenever available as part of the quality control parameters of the instruments. Also the higher rate of failure for FVC to meet standard criteria than FEV₁ is in line with the data from the literature [2, 4, 7], and suggests that when this parameter is not the main outcome of the trial, it is too often neglected.

The fact that the current QCC was effective in maintaining the standards of spirometry throughout the study is suggested by the slope of the linear regression of the overall score of spirometry over time not being significantly different from zero. Data from the literature and common experience show that performance in pulmonary function testing tends to worsen with time due to lack of interest, motivation, and control [1, 5]. To explain the success of the BRONCUS trial in this regard, the authors deemed it necessary that random monitoring of lung function should never be interrupted and sharing the QCC results with all investigators at least twice a year, in combination, could contribute to minimising, or at least avoiding, the commonest pitfalls, and that in turn best maintained the desired quality standards.

Interestingly, spirometries with an FEV₁ exceeding the mean individual value recorded over the study in excess of 20% showed significantly poorer quality than those with an FEV₁ below that threshold. Most of the problems were observed in the start-of-test phase as a result of submaximal efforts and/or hesitation. Because poorer effort is associated

with a greater FEV₁ as a result of less thoracic gas compression in ~25% of the cases [13], this most likely contributed to explain the observed overestimation of the FEV₁. The current authors doubt that such a small, though statistically significant difference affected the results of the BRONCUS trial on the decay of lung function over time, either because the current study's scoring system was not intended to address this point or the number of spirometries with an FEV₁>20% of mean value was too small compared with the total number used to construct the curves of lung function decay (6.8%). Yet, these findings suggest once again the importance of strict quality programmes when clinical questions have to be addressed with lung function tests in multi-centre trials.

In summary, random evaluation of the spirometries of the Bronchitis Randomized on N-acetylcysteine (NAC) Cost-Utility Study trial was apparently helpful to achieve and maintain the desired levels of performance. Factors in the mechanism of this success were likely to involve the awareness and need for adequate support of the investigators in order to minimise or avoid the commonest pitfalls in lung function testing, and provide encouragement to them in ways that best maintained the desired standards. Special attention should be paid to the start-of-test phase because most of the overestimation of the forced expiratory volume in one second was often the result of patient submaximal effort or hesitating expiratory manoeuvres.

ACKNOWLEDGEMENTS

The Steering Committee of the BRONCUS trial was represented by M. Decramer (Leuven, Belgium; Chairman), A. Ardia (Zambon, Italy), W. De Backer (Antwerp, Belgium), D. Olivieri (Parma, Italy), M. Rutten-van Mölken (Rotterdam, the Netherlands), T. Troosters (Leuven, Belgium), and C. van Herwaarden (Nijmegen, The Netherlands), C.P. Onno van Schayck (Maastricht, The Netherlands).

The following investigators participated in the study: M. Decramer (Leuven, Belgium), W. De Backer (Antwerp, Belgium), P.M. Mengeot (Montignies sur Sambre, Belgium), J. Verhaert (Lanaken, Belgium), B. De Wispelaere (Turnhout, Belgium), L. Delaunois (Mont-Godinne, Belgium), R.J. Broux (Liège, Belgium), J. Bruart (La Louvière, Belgium), A. Nusch (Essen, Germany), H. Reichert (Essen, Germany), H. Steffen (Landsberg am Lech, Germany), B. Kromer (Augsburg, Germany), D. Rost (Augsburg, Germany), V. Sobradillo (Bilbao, Spain), R. Rodriguez-Roisin (Barcelona, Spain), H. Vera Hernando (La Coruna, Spain), J.L. Viejo Banelos (Burgos, Spain), J. Martinez Gonzalez del Rio (Oviedo, Spain), N. Gonzalez-Mangado (Madrid, Spain), J.M. Rodriguez Gonzalez-Moro (Madrid, Spain), J. Castillo Gomez (Sevilla, Spain), T. Veidebaum (Tallinn, Estonia), E. Leesik (Tartu, Estonia), J.F. Muir (Rouen, France), B. Blaive (Nice, France), P. Guerin (Lyon, France), P. Camus (Dijon, France), E. Weitzemblem (Strasbourg, France), J.C. Pujet (Paris, France), M. Del Donno (Parma, Italy), M. Luisetti (Pavia, Italy), L. Pesce (Padova, Italy), C. Sanguinetti (Osimo-An, Italy), T. Todisco (Perugia, Italy), A.J. Neve (Sittard, the Netherlands), D.S. Postma (Groningen, The Netherlands), A.M.P. Greefhorst (Hengelo, The Netherlands), H.M.M. Pouwels (Venlo, The Netherlands), H.E.J. Sinnighe Damste

(Amsterdam, the Netherlands), A.J.M. Schreurs (Amsterdam, The Netherlands), P.N.R. Dekhuijzen (Nijmegen, The Netherlands), J.A. van Noord (Heerlen, The Netherlands), F.J.J. van den Elshout (Arnhem, The Netherlands), C.A.R. Groot (Oss, The Netherlands), V.D. Graaf (Utrecht, The Netherlands), P. Bresser (Amsterdam, The Netherlands), H. Gooszen and J. Creemers (Eindhoven, The Netherlands), R. Avila (Lisboa, Portugal), R. de Almeida (Vila Nova de Gaia, Portugal), L. Oliveira (Coimbra, Portugal), I. Gomes (Porto, Portugal), D. Nowak (Lodz, Poland), J. Zielinski (Warsaw, Poland), G. Riise (Gothenburg, Sweden).

REFERENCES

- 1 Enright PL, Johnson LR, Connett JE, Eoelker H, Buist SA. Spirometry in the Lung Health Study. 1. Methods and quality control. *Am Rev Respir Dis* 1991; 143: 1215–1223.
- 2 Bellia V, Pistelli R, Catalano F. Quality control of spirometry in the elderly. The SA.R.A. study. *Am J Respir Crit Care Med* 2000; 161: 1094–1100.
- 3 Künzli N, Ackermann-Liebrich U, Keller R, Perruchoud AP, Schindler C, and SAPALDIA team. Variability of FVC and FEV₁ due to technician, team, device and subject in an eight centre study: Three quality control studies in SAPALDIA. *Eur Respir J* 1995; 8: 371–376.
- 4 Stoller JK, Buist SA, Burrows B, *et al*, and the α_1 -Antitrypsin Deficiency Registry Study Group. Quality control of spirometry testing in the registry for patients with severe alpha 1-antitrypsin deficiency. *Chest* 1997; 111: 899–909.
- 5 Eaton T, Withy S, Garrett JE, Mercer J, Whitlock RML, Rea HH. Spirometry in primary care practice. The importance of quality assurance and the impact of spirometry workshops. *Chest* 1999; 116: 416–423.
- 6 Enright PL, Linn WS, Avol EL, Morgolis HG, Gong H, Peters JM. Quality of spirometry test performance in children and adolescents. Experience in a large field study. *Chest* 2000; 118: 665–671.
- 7 Schermer TR, Jacobs JE, Chavannes NH. Validity of spirometric testing in a general practice population of patients with chronic obstructive pulmonary disease (COPD). *Thorax* 2003; 58: 861–866.
- 8 Decramer M, Rutten-van Mülken M, Dekhuijzen PNR, *et al*. Effects of N-acetyl-cysteine on outcomes in COPD. The Bronchitis Randomized on NAC Cost-Utility Study (BRONCUS). *Lancet* 2005; 365: 1552–1560.
- 9 Quanjer Ph. Standardized lung function testing. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.
- 10 American thoracic Society. Standardization of spirometry. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
- 11 Sparrow D, O'Connor GT, Weiss ST, DeMolles D, Ingram RH Jr. Volume history effects and airway responsiveness in middle-aged and older men. The Normative Aging Study. *Am J Respir Crit Care Med* 1997; 155: 888–892.
- 12 D'Angelo E, Prandi E, Milic-Emili J. Dependence of maximal flow-volume curves on time-course of the preceding inspiration. *J Appl Physiol* 1993; 75: 1155–1159.
- 13 Krowka MJ, Enright PJ, Rodarte JR, Hyatt RE. Effect of effort on measurement of FEV₁. *Am Rev Respir Dis* 1987; 136: 829–833.