Eur Respir J 2011; 37: 784–790
DOI: 10.1183/09031936.00063810

The minimal important difference of exercise tests in severe COPD

M.A. Puhan*,#,¶,¶¶, D. Chandra+,¶¶, Z. Mosenifar§, A. Ries^f, B. Make, N.N. Hansel##, R.A. Wise## and F. Sciurba+ for the National Emphysema Treatment Trial (NETT) Research Group++**

ABSTRACT: Our aim was to determine the minimal important difference (MID) for 6-min walk distance (6MWD) and maximal cycle exercise capacity (MCEC) in patients with severe chronic obstructive pulmonary disease (COPD).

1,218 patients enrolled in the National Emphysema Treatment Trial completed exercise tests before and after 4–6 weeks of pre-trial rehabilitation, and 6 months after randomisation to surgery or medical care. The St George's Respiratory Questionnaire (domain and total scores) and University of California San Diego Shortness of Breath Questionnaire (total score) served as anchors for anchor-based MID estimates. In order to calculate distribution-based estimates, we used the standard error of measurement, Cohen's effect size and the empirical rule effect size.

Anchor-based estimates for the 6MWD were 18.9 m (95% CI 18.1–20.1 m), 24.2 m (95% CI 23.4–25.4 m), 24.6 m (95% CI 23.4–25.7 m) and 26.4 m (95% CI 25.4–27.4 m), which were similar to distribution-based MID estimates of 25.7, 26.8 and 30.6 m. For MCEC, anchor-based estimates for the MID were 2.2 W (95% CI 2.0–2.4 W), 3.2 W (95% CI 3.0–3.4 W), 3.2 W (95% CI 3.0–3.4 W) and 3.3 W (95% CI 3.0–3.5 W), while distribution-based estimates were 5.3 and 5.5 W.

We suggest a MID of 26±2 m for 6MWD and 4±1 W for MCEC for patients with severe COPD.

KEYWORDS: Chronic obstructive pulmonary disease, chronic obstructive pulmonary disease treatment, clinical trials, exercise tests, rehabilitation

The 6-min walk distance (6MWD) and symptom-limited cardiopulmonary cycle exercise testing are two commonly used measures of exercise capacity in patients with chronic obstructive pulmonary disease (COPD) in clinical practice and research [1, 2]. Performance in these exercise tests is strongly predictive of survival in patients with COPD [3–5].

Nevertheless, there is an ongoing debate about what constitutes a change over time in these measures that is important to patients. Specifically, it is unclear at what minimum amount of change in 6MWD or maximal cycle exercise capacity (MCEC) patients start to sense an actual change in their functional status (the minimal important difference (MID)) [6]. In assessing the course of disease in cohort studies and determining effectiveness of interventions in clinical trials, knowledge of the MID is critical.

A few studies have attempted to determine the MID of 6MWD, but their methodology and estimates differed substantially (35–87 m) [7–9]. These studies were unable to use both anchor- and distribution-based methods, which would have been an ideal approach, because neither method is perfect [10, 11]. The only prior study reporting a MID for maximum exercise capacity (4 W) used distribution and expert opinion-based estimates [12].

Our aim was to use both preferred anchor- and distribution-based methods using data from the National Emphysema Treatment Trial (NETT) to estimate the MID of 6MWD and MCEC in patients with COPD.

METHODS

We included all 1,218 patients enrolled in NETT, a multicentre, parallel-arm randomised controlled trial, the design and methods of which have been described in detail previously [13, 14]. In brief, NETT compared lung volume reduction surgery with optimal medical management in patients with advanced emphysema. Major enrollment

For editorial comments see page 755.
This article has supplementary material available from www.erj.ersjournals.com

AFFILIATIONS
*Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health,
##Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD,
+Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA,
§Division of Pulmonary and Critical Care Medicine, Cedars Sinai Medical Center, University of California, Los Angeles, CA,
^fDivision of Pulmonary and Critical Care Medicine, University of California, San Diego, CA,
**Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado, Denver, CO, USA.
#Horten Centre, University of Zurich, Zurich, Switzerland.
¶Dept of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Canada.
¶¶These authors contributed equally to the study.
++A full list of the NETT Research Group members and their affiliations can be found in the Acknowledgements section.

CORRESPONDENCE
M.A. Puhan, Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Room W5010, Baltimore, MD 21205, USA
E-mail: mpuhan@jhsph.edu

Received:
April 23 2010
Accepted after revision:
July 15 2010
First published online:
Aug 06 2010

European Respiratory Journal
Print ISSN 0903-1936
Online ISSN 1399-3003

criteria included bilateral emphysema judged suitable for lung volume reduction surgery, forced expiratory volume in 1 s (FEV_1) $\leqslant$45% predicted, residual volume $\geqslant$150% pred, arterial carbon dioxide tension $\leqslant$60 mmHg ($\leqslant$55 mmHg in Denver, CO, USA) and absence of clinical pulmonary hypertension. Screening began in October 1997 and randomisation in January 1998. NETT was reviewed and approved by the institutional review boards of the 17 clinical and one coordinating centre that were involved. All patients provided informed consent. All eligible patients were enrolled into a pulmonary rehabilitation programme prior to randomisation to lung volume reduction surgery or optimal medical management. This 6–10-week rehabilitation programme included 16–20 sessions of exercise, psychosocial and nutritional counselling, and patient education.

Measurement of 6MWD and MCEC

We used measurements of 6MWD and MCEC from three different time-points: at the beginning of the pre-trial rehabilitation programme, at the end of the pre-trial rehabilitation and 6 months after randomisation if patients were alive and able to come to the clinic for the visit. Prior to the 6-minute walk test (6MWT), a treadmill test at 1.61–3.22 $km \cdot h^{-1}$(1–2 $miles \cdot h^{-1}$) was performed to determine supplemental oxygen requirements during testing. The 6MWT was performed using a standard protocol that included scripted prompts at 1-min intervals. If oxygen supplementation was required during testing, a staff member walked behind the participant to carry the oxygen. Course layout and length varied by participating institution. Until May 1999, the NETT protocol for the 6MWT included two tests per visit, performed on consecutive days; the maximum distance was used for the visit measure.

Measurement of maximum exercise capacity was performed on a bicycle ergometer using a step or ramp increase in workload of either 5 or 10 W (depending on whether the participant's resting maximum voluntary ventilation was $\leqslant$40 or $>$40 $L \cdot min^{-1}$, respectively) at 1-min intervals. Inspired oxygen fraction in all patients was 30% throughout testing, in order to limit hypoxaemia as a cause of exercise limitation. During exertion, the patient was instructed when the cadence fell out of the 40–70 $revolutions \cdot min^{-1}$ range and encouraged each minute with simple phrases, such as ''nice job'', ''keep it up'', ''you are doing fine'' or similar.

Statistical analysis

Anchor-based methods

Anchor-based methods utilise other measures that already have an established MID (anchors) to estimate the MID of the test of interest. Using linear regression analysis, the known MID of the anchor is used to determine the magnitude of change in the test of interest that corresponds to this established MID. This method requires that a reasonably strong linear relationship exists between the anchor and the test of interest [10].

We used patient-reported outcomes, namely the St George's Respiratory Questionnaire (SGRQ) and the University of California San Diego Shortness of Breath Questionnaire (SOBQ), as potential anchors because both instruments are responsive to change [15, 16] and have an established MID [17–19]. The SGRQ is a widely used self-administered questionnaire that measures health-related quality of life in COPD patients [20]. There are three domain scores (symptoms, activity and impacts) and a total score. The maximum score for the SGRQ is 100 points and higher scores indicate poorer health-related quality of life. The MID of the SGRQ domain and total scores has been established to be 4 points using anchor-based methods, where the domains of the Chronic Respiratory Questionnaire served as anchors with well established MIDs [17]. The SOBQ measures the patients' perceived severity of shortness of breath on a six-point scale during 21 activities of daily living associated with varying levels of exertion [21]. Three additional questions ask about fear of harm from overexertion, limitations and fear caused by shortness of breath, for a total of 24 items. If patients do not routinely perform an activity, they are asked to estimate their anticipated shortness of breath. A total sum score ranges 0–120 and the MID is 5 points, as determined by anchor- (Chronic Respiratory Questionnaire and Transition Dyspnoea Index) and distribution-based (Cohen's effect size and the standard error of measurement (SEM)) methods [18, 19].

Using change in the anchors as the independent variable and change in 6MWD as the dependent variable, we used linear regression to determine the change in the exercise test that was numerically equivalent to the MID of the anchor [17, 22]. We conducted identical analyses for MCEC. *A priori*, we considered anchor-based estimates to be robust if correlations between changes in the anchors and exercise capacity were $\geqslant$0.5 [17, 22]. If correlations were 0.3–0.5, we also performed these analyses, but were more cautious in the interpretation of estimates. If correlations were $<$0.3 we did not calculate any anchor-based MID estimates [9]. We confirmed that the assumptions of linear regression were met by assessing linearity of the relationship between dependent and independent variables (residuals *versus* predicted plots), homoscedasticity (constant variance) of the errors (residuals *versus* predicted plots) and normality of the error distribution (normal probability plot of the residuals), which is shown in the online supplementary material.

We first evaluated the time period between the beginning and the end of the pre-trial rehabilitation intervention. We expected these data to be most similar to previously reported values determined in patients participating in a respiratory rehabilitation programme [9]. We then evaluated the period between the end of the pre-trial rehabilitation (randomisation) and the 6-month follow-up. A greater variability in the change in exercise, health-related quality of life or dyspnoea was expected in this period, given the substantial variation in response to lung volume reduction surgery and the more modest changes associated with the course of disease in the medically treated control group, both of which were included in the analysis [14].

Distribution-based methods

Distribution-based methods are based on the effect estimate and its relationship to a measure of variability (*i.e.* variance of between- or within-person changes). They are commonly considered inferior to anchor-based methods because they rely solely on statistical criteria and depend heavily on the characteristics of a particular study [10].

We used three commonly employed distribution-based methods to determine the MID of 6MWD: the SEM, Cohen's effect size and the empirical rule effect size [23–25].

$$\text{SEM} = \text{baseline SD} \times \sqrt{(1\text{-r})}$$

where r is the test–retest reliability coefficient (intraclass correlation coefficient) [23]. The pairs of tests from the pre-rehabilitation visit available in 437 patients provided an estimate of the test–retest reliability coefficient for the 6MWT. For Cohen's effect size and the empirical rule effect size, we used the SD of change scores (within-patient variability) between the beginning and end of the pre-trial rehabilitation rather than baseline SD (between-person variability), as we believe the within-patient analysis is most applicable to the interpretation of individual response.

For MCEC, we calculated Cohen's effect size and the empirical rule effect size but not the SEM, because we did not have data on test–retest reliability of maximum exercise capacity. We performed all analysis using STATA for Windows (version 10.1; StataCorp, College Station, TX, USA).

RESULTS

1,218 patients were included in the analysis. 472 (38.8%) patients were female, the mean±SD age was 66.4±6.1 yrs and post-bronchodilator FEV_1 was 26.9±7.1 % pred. Table 1 summarises the results for exercise tests and anchors at different time-points, and changes between these time-points. There were modest improvements in all parameters listed, with rehabilitation and 6 months after randomisation. As expected, there was greater variability in the changes (larger SD) 6 months after randomisation than during the pre-trial rehabilitation.

Table 2 shows that correlations between changes in the anchors and exercise tests were weak for the pre-trial rehabilitation period (all coefficients <0.2). Therefore, we did not use these data for estimating the MID with the anchor-based approach. In contrast, correlations were much stronger for change from randomisation to 6-month follow-up, with the exception of the symptoms domain of the SGRQ (table 2).

MID of 6MWD

The anchor-based estimates of the MID were similar (~25 m) with the exception of the smaller MID based on the SGRQ impacts domain (18.9 m) (table 3). The distribution-based methods yielded similar MID estimates. Based on the SEM the MID was 30.6 m, which was based on the test–retest reliability of two 6MWD measurements before rehabilitation (r=0.90; mean±SD difference between tests 19.5±44.6 m) and a baseline SD of 95.1 m. The MID was 26.8 m based on Cohen's effect size and 25.7 m based on the empirical rule effect size.

MID of MCEC

The anchor-based estimates of the MID of MCEC were similar (~3.2 W), again with the exception of the smaller MID based on the SGRQ impacts domain (2.2 W) (table 3). The distribution-based methods yielded MID estimates that were larger (close to 5 W) (table 3).

DISCUSSION

This is the first study to determine the MID for 6MWD and MCEC in patients with severe COPD using the optimal intrasubject, multiple anchor-based methodology. Our estimates for the MID of 6MWD were ~26 m using anchor- and distribution-based methods. For MCEC, the MID estimates were 4 W. The consistency of the several anchor- and distribution-based estimates in the largest cohort studied to date supports the validity of our findings. The correlations between patient-reported outcomes and exercise tests also provides evidence that changes in 6MWD and MCEC are patient-important.

We suggest an MID of 26 m for 6MWD in patients with severe COPD, with an uncertainty interval ±2 that reflects the variability (interquartile range) in estimates we found. For MCEC, we suggest an MID of 4±1 W. Our MID for 6MWD is smaller than the 35 m (95% CI 30–42 m) reported by us using distribution-based methods in a cohort of 460 patients pooled across nine studies from North America and Europe [9]. It is substantially lower than the MID of 54 m (95% CI 37–71 m) reported by REDELMEIER *et al.* [7] over a decade ago, which has gained wide acceptance in clinical trials determining the efficacy of new treatments for COPD, and other lung and heart diseases. It is also smaller than the results of a previous

TABLE 1 Changes in exercise capacity and patient-reported outcomes during pre-trial rehabilitation and from randomisation to 6-month follow-up

Measurement	Beginning of rehabilitation	End of rehabilitation	6 months after randomisation	Change from beginning to end of rehabilitation	Change from randomisation to 6-month follow-up
Patients n	1217	1217	1001	1217	1001
6MWD m	348.0±95.1	370.8±96.2	377.0±99.9[#]	22.8±53.6	-4.6±68.7[#]
MCEC W	36.0±21.1	38.9±21.6	42.2±22.6[¶]	2.9±11.0	0.9±13.9[¶]
SGRQ total score	56.5±13.0	53.0±12.7	48.4±16.8	-3.5±9.9	-4.3±14.3
SGRQ symptoms	58.3±19.3	56.0±19.9	54.3±20.9	-2.3±16.2	-1.4±19.2
SGRQ activities	81.9±12.7	79.4±13.4	71.7±20.0	-2.5±10.8	-7.2±18.6
SGRQ impact	41.4±16.7	36.9±15.6	33.1±18.6	-4.4±13.4	-3.5±16.1
SOBQ total score	65.6±19.0	62.7±18.4	55.2±24.1	-3.0±13.6	-6.7±20.8

Data are presented as mean±SD, unless otherwise stated. 6MWD: 6-min walk distance; MCEC: maximal cycle exercise capacity; SGRQ: St George's Respiratory Questionnaire; SOBQ: University of California San Diego Shortness of Breath Questionnaire. [#]: n=907; [¶]: n=905.

TABLE 2 Correlations between changes in exercise capacity and changes in patient-reported outcomes

Measurement	Correlation with change in 6MWD from beginning to end of rehabilitation	Correlation with change in MCEC from beginning to end of rehabilitation	Correlation with change in 6MWD from end of rehabilitation to 6-month follow-up	Correlation with change in MCEC from end of rehabilitation to 6-month follow-up
SGRQ total score	-0.15	-0.11	**-0.46**	**-0.49**
SGRQ activities	-0.11	-0.10	**-0.46**	**-0.48**
SGRQ impact	-0.15	-0.08	**-0.40**	**-0.43**
SGRQ symptoms	-0.02	-0.06	-0.18	-0.19
SOBQ total score	-0.13	-0.14	**-0.51**	**-0.50**

6MWD: 6-min walk distance; MCEC: maximal cycle exercise capacity; SGRQ: St George's Respiratory Questionnaire; SOBQ: University of California San Diego Shortness of Breath Questionnaire. Bold represents correlations $\geqslant$0.3, justifying anchor-based estimates.

analysis of a smaller cohort from the NETT data that yielded estimates between 40 and 87 m [8].

Available studies do not suggest that differences in MID estimates across studies can be explained by differences in COPD disease severity. In the study of REDELMEIER *et al.* [7], the mean FEV1 was 975 mL (~35% pred) and mean 6MWD was 371 m. In our recent study using pooled data, the mean FEV1 was 39% pred and mean baseline 6MWD was 361 m, compared to a mean FEV1 of 27% pred and mean 6MWD of 347 m in the present study. However, any inferences about the association of disease severity and MID estimates are currently limited by the focus on FEV1, which represents only one marker of disease severity that is reported consistently across studies. The use of more comprehensive disease severity indices, such as the modified BODE (body mass index, airflow obstruction, dyspnoea, exercise capacity) or ADO (age, dyspnoea, airflow obstruction) indices, could offer more insights into the association of disease severity and MID [3]. Instead, it appears likely that some differences in the estimates between studies arose from the methods used to estimate the MID.

TABLE 3 Minimal important difference (MID) of 6-min walk distance (6MWD) and maximal cycle exercise capacity (MCEC)

Method used	6MWD m	MCEC W
Anchor-based		
SGRQ total score#	24.6 (23.4–25.7)	3.3 (3.0–3.5)
SGRQ impact#	18.9 (18.1–20.1)	2.2 (2.0–2.4)
SGRQ activities#	24.2 (23.4–25.4)	3.2 (3.0–3.4)
SOBQ total score¶	26.4 (25.4–27.4)	3.2 (3.0 –3.4)
Distribution-based		
SEM	30.6	
Cohen's effect size	26.8	5.5
Empirical rule effect size	25.7	5.3

Data are presented as MID (95% CI) or MID. All regression equations multiplied by -1 to facilitate interpretation (improvements of St George's Respiratory Questionnaire (SGRQ) and University of California San Diego Shortness of Breath Questionnaire (SOBQ) if change score is negative). SEM: standard error of measurement. #: MID 4 points; ¶: MID 5 points.

Therefore, the following paragraphs will highlight some key methodological differences between prior studies and our analysis.

REDELMEIER *et al.* [7] studied patients following pulmonary rehabilitation. Similarly to us, they could not apply the preferred intrasubject anchor-based methodology because of a low correlation between changes in 6MWD and patient-reported changes in exercise capacity ($r<0.20$). The authors hypothesised that this was because patients had little recollection of their previous exercise status. However, due to the availability of data on patients after randomisation in the NETT cohort, we can shed more light on the likely reason for this finding. The more uniform response of patients to rehabilitation, as opposed to lung volume reduction surgery, after which some patients experience great improvement but others experience deterioration of health status, results in smaller variability in change of health status and, thus, smaller correlation coefficients.

Because of the lack of correlation between measured and patient-reported changes in exercise capacity, REDELMEIER *et al.* [7] chose to assess the MID using an intersubject technique, by asking patients to observe and compare their own exercise capacity with that of others in the same rehabilitation programme. Such a method has two significant limitations. First, it is human nature to be biased in one's own favour: in the study of REDELMEIER *et al.* [7], patients rated themselves the same as their peers when their 6MWD was, in fact, on average 10 m less than that of their peers. Also, they were more sensitive to determining when they were a little better than their peers (+30 m) as opposed to when they felt they were worse (-80 m). To reconcile this almost three-fold difference in their two MID estimates, REDELMEIER *et al.* [7] arbitrarily calculated the average of the two, namely the widely quoted figure of 54 m. The second limitation of the intersubject approach is that it is easier for patients to recognise change in themselves rather than change in a third person. If we assume this to be true, then 54 m would be an overestimate of the MID and indicate that the use of between-person differences may not validly inform what constitutes an important within-person change.

Another study, also based on NETT, using only distribution-based methods, reported MID estimates between 44 and 86 m [8].

The higher MID estimates of that study were due to a cross-sectional analysis of variation in the NETT cohort at baseline, rather than following the response to an intervention, such as lung volume reduction surgery or continued medical management over time, as in the current analysis. Because the MID is used to interpret changes in the patients' exercise capacity over time, we believe that a longitudinal analysis is preferred [9, 22]. Since variation of baseline assessments are almost always larger than of the variation of change over time, the use of baseline SD will lead to larger distribution-based MID estimates.

Prior evidence on MID estimates for MCEC is scarce. One study, also based on NETT, found MID estimates between 4 and 10 W, based on baseline distribution-based and expert opinion methods [12]. In our analysis, we determined anchor-based estimates to be ~3 W and distribution-based methods to be 5 W; thus, we offer the intermediate value of 4 ± 1 W to be a reasonable estimate.

Our study has implications for the design of clinical trials in the future that use change in 6MWD or MCEC as outcome measures for the evaluation of new therapies for severe COPD and other common exercise-limiting diseases, such as pulmonary hypertension and heart failure. A lower MID will imply that newer treatments will need to have a smaller effect to be classified as effective, while the clinical trials that test these interventions will need to enroll more patients to detect this smaller change with the same statistical power.

Our results may also have implications for the design of future studies evaluating MIDs using anchor-based strategies. Our study showed that it is not only the type of anchors that drive correlations between the outcome of interest and the anchors, but also the type of intervention and the magnitude of variation in response: this is the between-person variablity. This is particularly important if the outcome of interest and the anchors do not capture closely related constructs, as is the case for exercise capacity, and patient-reported symptoms and health-related quality of life, respectively. In contrast, if the anchor was, for example, a measure of physical activity with established MID, stronger correlations could be expected even if the intervention did not yield a highly variable response.

A limitation of our study is that NETT only included patients with advanced COPD, which may limit the generalisability of our findings to a broader group of COPD patients. A restricted sample of the entire COPD population is less of a problem for anchor-based methods but may limit the generalisability of distribution-based methods. In particular, baseline SD can be small and not representative of the heterogeneous COPD population. However, since we used SDs of change, the limitations of our distribution-based approach may be less problematic. Another limitation in our determination of MID for exercise capacity is the nature of our anchors. Ideally, anchors should measure the same construct. We did not have anchors available to capture constructs specifically related to exercise capacity, such as physical activity measured by activity monitors or questionnaires, and, in any case, these measures do not have validated MID values [26]. Also, we did not have patient ratings available that would reflect perceived changes in exercise capacity and that could have been used as anchors as in previous studies [7]. By using patient-reported symptoms and health-related quality of life as the anchors in our approach we accepted a compromise between the strength of evidence regarding the MID of our anchors and the limited correlations of these anchors with the exercise measures of interest.

Conclusion

Our MID for the 6MWT (26 ± 2 m) is lower than the currently employed MID of 54 m, while the MID of MCEC is similar to previous estimates of 4 ± 1 W. These estimates provide the strongest evidence-based estimates available to plan and interpret the results of clinical trials and cohort studies in patients with severe COPD.

SUPPORT STATEMENT

The National Emphysema Treatment Trial (NETT) is supported by contracts with the National Heart, Lung and Blood Institute (N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118 and N01HR76119), the Centers for Medicare and Medicaid Services (CMS), and the Agency for Healthcare Research and Quality (AHRQ).

STATEMENT OF INTEREST

A statement of interest for N.N. Hansel can be found at www.erj.ersjournals.com/site/misc/statements.xhtml

ACKNOWLEDGEMENTS

The members of the National Emphysema Treatment Trial (NETT) Research Group are: A.P. Fishman (Chair), B.A. Bozzarello and A. Al-Amin (all Office of the Chair of the Steering Committee, University of Pennsylvania, Philadelphia, PA, USA); M. Katz (Principal Investigator), C. Wheeler (Principal Clinic Coordinator), E. Baker, P. Barnard, P. Cagle, J. Carter, S. Chatziioannou, K. Conejo-Gonzales, K. Dubose, J. Haddad, D. Hicks, N. Kleiman, M. Milburn-Barnes, C. Nguyen, M. Reardon, J. Reeves-Viets, S. Sax, A. Sharafkhaneh, O. Wilson, C. Young, R. Espada (Principal Investigator 1996–2002), R. Butanda (1999–2001), M. Ellisor (2002), P. Fox, (1999–2001), K. Hale (1998–2000), E. Hood, (1998–2000), A. Jahn (1998–2000), S. Jhingran (1998–2001), K. King (1998–1999), C. Miller III (1996–1999), I. Nizami (Co-Principal Investigator, 2000–2001), T. Officer (1998–2000), J. Ricketts (1998–2000), J. Rodarte, (Co-Principal Investigator 1996–2000), R. Teague (Co-Principal Investigator 1999–2000) and K. Williams (1998–1999) (all Baylor College of Medicine, Houston, TX, USA); J. Reilly (Principal Investigator), D. Sugarbaker (Co-Principal Investigator), C. Fanning (Principal Clinic Coordinator), S. Body, S. Duffy, V. Formanek, A. Fuhlbrigge, P. Hartigan, S. Hooper, A. Hunsaker, F. Jacobson, M. Moy, S. Peterson, R. Russell, Diane Saunders and S. Swanson (Co-Principal Investigator, 1996–2001) (Brigham and Women's Hospital, Boston, MA, USA); R. McKenna (Principal Investigator), Z. Mohsenifar (Co-Principal Investigator), C. Geaga (Principal Clinic Coordinator), M. Biring, S. Clark, J. Cutler, R. Frantz, P. Julien, M. Lewis, J. Minkoff-Rau, V. Yegyan and M. Joyner (1996–2002) (all Cedars-Sinai Medical Center, Los Angeles, CA, USA); M. DeCamp (Principal Investigator), J. Stoller, (Co-Principal Investigator), Y. Meli (Principal Clinic Coordinator), J. Apostolakis, D. Atwell, J. Chapman, P. DeVilliers, R. Dweik, E. Kraenzler, R. Lann, N. Kurokawa, S. Marlow, K. McCarthy, P. McCreight, A. Mehta, M. Meziane, O. Minai, M. Steiger, K. White, J. Maurer (Principal Investigator, 1996–2001), T. Durr (2000–2001), C. Hearn (1998–2001), S. Lubell (1999–2000), P. O'Donovan (1998–2003) and R. Schilz (1998–2002) (all Cleveland Clinic Foundation, Cleveland, OH, USA); M. Ginsburg (Principal Investigator), B. Thomashow (Co-Principal Investigator), P. Jellen (Principal Clinic

Coordinator), J. Austin, M. Bartels, Y. Berkmen, P. Berkoski (Site coordinator; all Long Island Jewish Medical Center (LIJ)), F. Brogan, A. Chong, G. DeMercado, A. DiMango, S. Do, B. Kachulis, A. Khan, B. Mets, M. O'Shea, G. Pearson, L. Rossoff, S. Scharf, (Co-Principal Investigator, 1998–2002), M. Shiau, P. Simonelli, K. Stavrolakes, D. Tsang, D. Vilotijevic, C. Yip, M. Mantinaos (1998–2001), K. McKeon, (1998–1999) and J. Pfeffer (1997–2002) (all Columbia University, New York, NY in consortium with LIJ, New Hyde Park, NY, USA); N. MacIntyre (Principal Investigator), R.D. Davis (Co-Principal Investigator), J. Howe, (Principal Clinic Coordinator), R.E. Coleman, R. Crouch, D. Greene, K. Grichnik, D. Harpole Jr., A. Krichman, B. Lawlor, H. McAdams, J. Plankeel, S. Rinaldo-Gallo, S. Shearer, J. Smith, M. Stafford-Smith, V. Tapson, M. Steele, (1998–1999) and J. Norten MD (1998–1999) (all Duke University Medical Center, Durham, NC, USA); J. Utz (Principal Investigator), C. Deschamps (Co-Principal Investigator), K. Mieras (Principal Clinic Coordinator), M. Abel, M. Allen, D. Andrist, G. Aughenbaugh, S. Bendel, E. Edell, M. Edgar, B. Edwards, B. Elliot, J. Garrett, D. Gillespie, J. Gurney, B. Hammel, K. Hanson, L. Hanson, G. Harms, J. Hart, T. Hartman, R. Hyatt, E. Jensen, N. Jenson, S. Kalra, P. Karsell, J. Lamb, D. Midthun, C. Mottram, S. Swensen, A-M. Sykes, K. Taylor, N. Torres, R. Hubmayr (1998–2000), D. Miller (1999–2002), S. Bartling (1998–2000) and K. Bradt (1998–2002) (all Mayo Foundation, Rochester, MN, USA); B. Make (Principal Investigator), M. Pomerantz (Co-Principal Investigator), M. Gilmartin (Principal Clinic Coordinator), J. Canterbury, M. Carlos, P. Dibbern, E. Fernandez, L. Geyman, C. Hudson, D. Lynch, J. Newell, R. Quaife, J. Propst, C. Raymond, J. Whalen-Price, K. Winner, M. Zamora, R. Cherniack, (Principal Investigator, 1997–2000) (all National Jewish Medical and Research Center, Denver, CO, USA); P. Diaz (Principal Investigator), P. Ross (Co-Principal Investigator), T. Bees (Principal Clinic Coordinator), J. Drake, C. Emery, M. Gerhardt, M. King, D. Rittinger and M. Rittinger (all Ohio State University, Columbus, OH, USA); K. Naunheim (Principal Investigator), R. Gerber (Co-Principal Investigator), J. Osterloh (Principal Clinic Coordinator), S. Borosh, W. Chamberlain, S. Frese, A. Hibbit, M.E. Kleinhenz, G. Ruppel, C. Stolar, J. Willey, F. Alvarez (Co-Principal Investigator, 1999–2002) and C. Keller (Co-Principal Investigator, 1996–2000) (all Saint Louis University, Saint Louis, MO, USA); G. Criner, (Principal Investigator), S. Furukawa (Co-Principal Investigator), A.M. Kuzma (Principal Clinic Coordinator), R. Barnette, Neil Brister, K. Carney, W. Chatila, F. Cordova, G. D'Alonzo, M. Keresztury, K. Kirsch, C. Kwak, K. Lautensack, M. Lorenzon, U. Martin, P. Rising, S. Schartel, J. Travaline, G. Vance, P. Boiselle (1997–2000) and Gerald O'Brien (1997–2000) (all Temple University, Philadelphia, PA, USA); A. Ries (Principal Investigator), R. Kaplan (Co-Principal Investigator), C. Ramirez (Principal Clinic Coordinator), D. Frankville, P. Friedman, J. Harrell, J. Johnson, D. Kapelanski, D. Kupferberg, C. Larsen, T. Limberg, M. Magliocca, F.J. Papatheofanis, D. Sassi-Dambron and Melissa Weeks (all Univerversity of California, San Diego, CA, USA); M. Krasna (Principal Investigator), H. Fessler (Co-Principal Investigator), I. Moskowitz (Principal Clinic Coordinator), T. Gilbert, J. Orens, S. Scharf, D. Shade, S. Siegelman, K. Silver, C. Weir and C. White (all University of Maryland at Baltimore, in consortium with Johns Hopkins Hospital, Baltimore, MD, USA); F. Martinez (Principal Investigator), M. Iannettoni (Co-Principal Investigator), C. Meldrum (Principal Clinic Coordinator), W. Bria, K. Campbell, P. Christensen, K. Flaherty, S. Gay, P. Gill, P. Kazanjian, E. Kazerooni, V. Knieper, T. Ojo, L. Poole, L. Quint, P. Rysso, T. Sisson, M. True, B. Woodcock and L. Zaremba (all University of Michigan, Ann Arbor, MI, USA); L. Kaiser (Principal Investigator), J. Hansen-Flaschen (Co-Principal Investigator), M.L. Dempsey (Principal Clinic Coordinator), A. Alavi, T. Alcorn, S. Arcasoy, J. Aronchick, S. Aukberg, B. Benedict, S. Craemer, R. Daniele, J. Edelman, W. Gefter, L. Kotler-Klein, R. Kotloff, D. Lipson, W. Miller Jr, Richard O'Connell, S. Opelman, H. Palevsky, W. Russell, H. Sheaffer, R. Simcox, S. Snedeker, J. Stone-Wynne, G. Tino, P. Wahl, J. Walter, P. Ward, D. Zisman, J. Mendez (1997–2001) and A. Wurster (1997–1999) (all University of Pennsylvania, Philadelphia, PA, USA); F. Sciurba (Principal Investigator), J. Luketich (Co-Principal Investigator), C. Witt (Principal Clinic Coordinator), G. Ayres, M. Donahoe, C. Fuhrman, R. Hoffman, J. Lacomis, J. Sexton, W. Slivka, D. Strollo, E. Sullivan, T. Simon, C. Wrona, G. Bauldoff (1997–2000), M. Brown (1997–2002), E. George (Principal Clinic Coordinator, 1997–2001), R. Keenan (Co-Principal Investigator, 1997–2000), T. Kopp (1997–1999) and L. Silfies (1997–2001) (all University of Pittsburgh, Pittsburgh, PA, USA); J. Benditt (Principal Investigator), D. Wood (Co-Principal Investigator), M. Snyder (Principal Clinic Coordinator), K. Anable, N. Battaglia, L. Boitano, A. Bowdle, L. Chan, C. Chwalik, B. Culver, T. Gillespy, D. Godwin, J. Hoffman, A. Ibrahim, D. Lockhart, S. Marglin, K. Martay, P. McDowell, D. Oxorn, L. Roessler, M. Toshima and S. Golden (1998–2000) (all University of Washington, Seattle, WA, USA); L. Bosco, Y-P. Chiang, C. Clancy and H. Handelsman (all Agency for Healthcare Research and Quality, Rockville, MD, USA); S.M. Berkowitz, T. Carino, J. Chin, J. Baldwin, K. McVearry, A. Norris, S. Shirey, C. Sikora and S. Sheingold (1997–2004) (all Centers for Medicare and Medicaid Services, Baltimore, MD); S. Piantadosi (Principal Investigator), J. Tonascia (Co-Principal Investigator), P. Belt, A. Blackford, K. Collins, B. Collison, R. Colvin, J. Dodge, M. Donithan, V. Edmonds, G.L. Foster, J. Fuller, J. Harle, R. Jackson, S. Lee, C. Levine, H. Livingston, J. Meinert, J. Meyers, D. Nowakowski, K. Owens, S. Qi, M. Smith, B. Simon, P. Smith, A. Sternberg, M. Van Natta, L. Wilson and R. Wise (all Coordinating Center, The Johns Hopkins University, Baltimore, MD); R.M. Kaplan (Chair), J.S. Schwartz (Co-Chair), Y-P. Chiang, M.C. Fahs, A.M. Fendrick, A.J. Moskowitz, D. Pathak, S. Ramsey, S. Sheingold, A.L. Shroyer, J. Wagner and R. Yusen (all Cost Effectiveness Subcommittee); S. Ramsey (Principal Investigator), R. Etzioni, S. Sullivan, D. Wood, T. Schroeder, K. Kreizenbeck, K. Berry and N. Howlader (all Cost Effectiveness Data Center, Fred Hutchinson Cancer Research Center, Seattle, WA); E. Hoffman (Principal Investigator), J. Cook-Granroth, A. Delsing, J. Guo, G. McLennan, B. Mullan, C. Piker, J. Reinhardt, B. Wood, J. Sieren and W. Stanford (all CT Scan Image Storage and Analysis Center, University of Iowa, Iowa City, IA); J.A. Waldhausen (Chair), G. Bernard, D. DeMets, M. Ferguson, E. Hoover, R. Levine, D. Mahler, A.J. McSweeny, J. Wiener-Kronish, O.D. Williams and M. Younes (all Data and Safety Monitoring Board); G. Criner (Principal Investigator) and C. Soltoff (both Marketing Center, Temple University, Philadelphia, PA, USA); and G. Weinmann (Project Officer), J. Deshler (Contracting Officer), D. Follmann, J. Kiley and M. Wu (1996–2001) (all Project Office, National Heart, Lung, and Blood Institute, Bethesda, MD).

The authors would like to thank A. Gelb (Lakewood Regional Medical Center, Lakewood, CA, USA).

REFERENCES

1 Lacasse Y, Brosseau L, Milne S, *et al.* Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2004; 4: CD003793.

2 Peytremann-Bridevaux I, Staeger P, Bridevaux PO, *et al.* Effectiveness of chronic obstructive pulmonary disease-management programs: systematic review and meta-analysis. *Am J Med* 2008; 121: 433–443.

3 Puhan MA, Garcia-Aymerich J, Frey M, *et al.* Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009; 374: 704–711.

4 Pinto-Plata VM, Cote C, Cabral H, *et al.* The 6-min walk distance: change over time and value as a predictor of survival in severe COPD. *Eur Respir J* 2004; 23: 28–33.

5 Tojo N, Ichioka M, Chida M, *et al.* Pulmonary exercise testing predicts prognosis in patients with chronic obstructive pulmonary disease. *Intern Med* 2005; 44: 20–25.

6 Schunemann HJ, Guyatt GH. Commentary: goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005; 40: 593–597.

7 Redelmeier DA, Bayoumi AM, Goldstein RS, *et al.* Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. *Am.J Respir Crit Care Med* 1997; 155: 1278–1282.
8 Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. *COPD* 2005; 2: 125–129.
9 Puhan MA, Mador MJ, Held U, *et al.* Interpretation of treatment changes in 6-minute walk distance in patients with COPD. *Eur Respir J* 2008; 32: 637–643.
10 Guyatt GH, Osoba D, Wu AW, *et al.* Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77: 371–383.
11 Revicki DA, Cella D, Hays RD, *et al.* Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006; 4: 70.
12 Sutherland ER, Make BJ. Maximum exercise as an outcome in COPD: minimal clinically important difference. *COPD* 2005; 2: 137–141.
13 Rationale and design of the National Emphysema Treatment Trial (NETT): a prospective randomized trial of lung volume reduction surgery. *J Thorac Cardiovasc Surg* 1999; 118: 518–528.
14 Fishman A, Martinez F, Naunheim K, *et al.* A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003; 348: 2059–2073.
15 Puhan MA, Guyatt GH, Goldstein R, *et al.* Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respir Med* 2007; 101: 308–316.
16 Verrill D, Barton C, Beasley W, *et al.* The effects of short-term and long-term pulmonary rehabilitation on functional capacity, perceived dyspnea, and quality of life. *Chest* 2005; 128: 673–683.
17 Schunemann HJ, Griffith L, Jaeschke R, *et al.* Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *J Clin Epidemiol* 2003; 56: 1170–1176.
18 Kupferberg DH, Kaplan RM, Slymen DJ, *et al.* Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil* 2005; 25: 370–377.
19 Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *COPD* 2005; 2: 105–110.
20 Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Eur Respir J* 1991; 85: Suppl. B, 25–31.
21 Eakin EG, Resnikoff PM, Prewitt LM, *et al.* Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113: 619–624.
22 Puhan MA, Frey M, Buchi S, *et al.* The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. *Health Qual Life Outcomes* 2008; 6: 46.
23 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861–873.
24 Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003; 1: 4.
25 Sloan JA, Vargas-Chanes D, Kamath CC, *et al.* Detecting worms, ducks, and elephants: a simple approach for defining clinically relevant effects in quality-of-life measure. *Cancer Integrative Med* 2003; 2: 41–47.
26 Pitta F, Troosters T, Probst VS, *et al.* Quantifying physical activity in daily life with questionnaires and motion sensors in COPD. *Eur Respir J* 2006; 27: 1040–1055.