ERJ Express. Published on August 6, 2010 as doi: 10.1183/09031936.00063810

The minimal important difference of exercise tests in severe COPD

Milo A. Puhan*, MD, PhD [1,2,3], Divay Chandra*, MD, MSc[4], Zab Mosenifar, MD[5], Andrew Ries, MD, MPH[6], Barry Make, MD[7], Nadia N. Hansel, MD, MPH[8], Robert A. Wise, MD[8], and Frank Sciurba, MD[4] for the NETT Research Group

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
[2] Horten Centre, University of Zurich, Switzerland
[3] Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Canada
[4] Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, PA, USA
[5] Division of Pulmonary and Critical Care Medicine, Cedars Sinai Medical Center, University of California, Los Angeles, CA, USA
[6] Division of Pulmonary and Critical Care Medicine, University of California, San Diego, CA, USA
[7] Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado, Denver, CO, USA
[8] Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD, USA
*Contributed equally

Correspondence and reprint requests to:
Milo A. Puhan, MD, PhD
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street, Room W5010, Baltimore, MD 21205 USA
phone: 443-287-8777; fax: 410-502-4621
mpuhan@jhsph.edu

Abstract

Our aim was to determine the minimal important difference for 6-minute walk distance and maximal cycle exercise capacity in patients with severe COPD.

1218 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 randomization to surgery or medical care. The St Georges Respiratory Questionnaire (domain and total scores) and UCSD Shortness of Breath Questionnaire (total score) served as anchors for anchor-based minimal important difference estimates. 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 6-minute walk distance were 18.9 meters (95% CI 18.1-20.1), 24.2 meters (23.4-25.4), 24.6 meters (23.4-25.7) and 26.4 meters (25.4-27.4), which were similar to distribution-based minimal important difference estimates of 25.7, 26.8 and 30.6 meters. For maximal cycle exercise capacity, anchor-based estimates for the minimal important difference were 2.2 Watts (2.0-2.4), 3.2 Watts (3.0-3.4), 3.2 Watts (3.0-3.4) and 3.3 (3.0-3.5) Watts while distribution-based estimates were 5.3 and 5.5 Watts

We suggest a minimal important difference of 26±2 meters for 6-minute walk distance and 4±1 Watts for maximal cycle exercise capacity for patients with severe COPD.

Keywords: Clinical trials, COPD, COPD treatment, Exercise tests, Rehabilitation

Introduction

6-minute walk distance 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 on 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 walk distance or maximal cycle exercise capacity do 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 as well as determining effectiveness of interventions in clinical trials knowledge of the MID is critical.

A few studies have attempted to determine the MID of 6-minute walk distance, but their methodology and estimates differed substantially (between 35 and 87 meters [m]).[7-9] These studies were unable to use both anchor- and distribution-based methods, an ideal approach because neither method is perfect.[10] [11] The only prior study reporting a MID for maximum exercise capacity (4 Watts) 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 6-minute walk distance and maximal cycle exercise capacity in patients with COPD.

Methods

We included all 1218 patients enrolled in NETT, which was a multi-center parallel arm randomized controlled trial whose design and methods 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 criteria included bilateral emphysema judged suitable for lung volume reduction surgery, FEV1 ≤ 45% predicted, residual volume ≥ 150% predicted, PaCO2 ≤ 60 mm Hg (≤ 55mm Hg in Denver, CO), and absence of clinical pulmonary hypertension. Screening began in October 1997 and randomization in January 1998. NETT was reviewed and approved by the institutional review boards of the 17 clinical centers and 1 coordinating center that were involved. All patients provided informed consent. All eligible patients were enrolled into a pulmonary rehabilitation program prior to randomization to lung volume reduction surgery or optimal medical management. This 6-10 week long rehabilitation program included 16-20 sessions of exercise, psychosocial and nutritional counseling, and patient education.

Measurement of 6-minute walk distance and maximal cycle exercise capacity

We used measurements of 6-minute walk distance and maximal cycle exercise capacity from three different time points: at the beginning of the pre-trial rehabilitation program, at the end of the pre-trial rehabilitation and six months after randomization if patients were alive and able to come to the clinic for the visit. Prior

to the 6-minute walk test, a treadmill test at 1-2 miles per hour was performed to determine supplemental oxygen requirements during testing. The 6-minute walk test was performed using a standard protocol that included scripted prompts at 1-minute 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, NETT protocol for the 6-minute walk test included 2 tests per visit done 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 less than/equal to or greater than 40 L/min respectively) at 1-minute intervals. Inspired oxygen fraction in all patients was 30% throughout testing in order to limit hypoxemia as a cause of exercise limitation. During exertion, the patient was instructed when the cadence fell out of the 40-70 revolutions per minute 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 utilize other measures which 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 domains served as anchors with well established MIDs.[17] The SOBQ measures the patients' perceived severity of shortness on a 6-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 from 0 to 120 and the MID is 5 points as determined by anchor- (Chronic Respiratory Questionnaire and Transition Dyspnea Index serving as anchors) 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 6-minute walk distance as the dependent variable, we used linear regression to determine the change in the exercise test, which was numerically equivalent to the MID of the anchor.[17 22] We conducted identical analyses for maximal cycle exercise capacity. A priori we considered anchor-based estimates to be robust if correlations between changes in the anchors and exercise capacity were ≥0.5.[17 22] If correlations were between 0.3 and 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 appendix.

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 program.[9] We then evaluated the period between the end of the pre-trial rehabilitation (randomization) and the 6-months follow-up. A greater variability in the change in exercise, health-related quality of life or dyspnea 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 they depend heavily on the characteristics of a particular study.[10]

We used three commonly employed distribution-based methods to determine the MID of 6-minute walk distance: the SEM, Cohen's effect size, and the empirical rule effect size.[23-25] The SEM equals the baseline standard deviation (SD) times $\sqrt{(1-r)}$, where r is the test retest reliability coefficient (intra-class 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 6-minute walk test. 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 maximal cycle exercise capacity, we only calculated Cohen's effect size and the empirical rule effect size but not the standard error of measurement

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).

Results

1218 patients were included in the analysis. 472 were female (38.8%), the mean age was 66.4 ± 6.1 years, and mean post-bronchodilator FEV_1 26.9 ± 7.1 % predicted. Table 1 summarizes 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 randomization. As expected, there was greater variability in the changes (larger SD) 6 months after randomization 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 randomization to six month follow up with the exception of the symptoms domain of the SGRQ (table 2). .

MID of 6-minute walk distance

The anchor-based estimates of the MID were similar and around 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 standard error of measurement the MID was 30.6 m, which was based on the

test-retest reliability of two 6 minute walk distance measurements before - rehabilitation (intra-class correlation coefficient of 0.90 and mean difference between tests of 19.5 m [SD 44.6]) and a baseline standard deviation 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 maximal cycle exercise capacity

The anchor-based estimates of the MID of maximal cycle exercise capacity were similar and around 3.2 W again with the exception of the smaller MID based on the SGRQ impacts domain (2.2 Watts)(Table 3). The distribution-based methods yielded MID estimates that were larger and close to 5 Watts (Table 3).

Discussion

This is the first study to determine the MID for 6 minute walk distance and maximal cycle exercise capacity in patients with severe COPD using the optimal intra-subject multiple anchor-based methodology. Our estimates for the MID of 6-minute walk distance were around 26 m using anchor- and distribution-based methods. For maximum exercise capacity measured by cycle ergometry the MID estimates were 4 Watts. 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 6-minute walk distance and maximal cycle exercise capacity are patient-important.

We suggest a MID of 26 m for 6-Minute walk distance in patients with severe COPD with an uncertainty interval ±2 that reflects the variability (interquartile range) in estimates we found. For maximal cycle exercise capacity we suggest a MID of 4±1 Watts. Our MID for 6-Minute walk distance is smaller than the 35 m (95% CI 30-42) 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) reported by Redelmeier and colleagues over a decade ago[7], 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 former 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 Redelmeier study the mean FEV_1 was 975 ml (approximately 35% predicted) and mean 6-minute walk distance 371m. In our recent study using pooled data the mean FEV_1 was 39% predicted and mean baseline walk distance 361 m compared to a mean FEV_1 of 27 % predicted and mean 6-minute walk distance of 347 m in the current analysis. However, any inferences about the association of disease severity and MID estimates are currently limited by the focus on FEV_1, 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 or ADO index 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. Therefore, the following paragraphs will highlight some key methodological differences between prior studies and our analysis. .

Redelmeier et al studied patients following a pulmonary rehabilitation.[7] Similar to us they could not apply the preferred intra-subject anchor based methodology because of a low correlation between changes in 6-minute walk distance and patient reported changes in exercise capacity ($r<0.20$). The authors hypothesized that this was because patients had little recollection of their previous exercise status. However, due to the availability of data on patients after randomization 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 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 chose to assess the MID using an inter-subject technique by asking patients to observe and compare their own exercise capacity with that of others in the same rehabilitation program. Such a method has two significant limitations. Firstly, it is human nature to be biased in one's own favor: in Redelmeier's study patients rated themselves the same as their peers when their 6-minute walk distance was in fact on the average 10 m less than 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 and colleagues arbitrarily calculated the average of the two, namely the widely quoted figure of 54 m. The second limitation of the inter-subject approach is that it is easier for patients to recognize 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 maximal cycle exercise capacity 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 approximately 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 6-minute walk distance or maximal cycle exercise capacity 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 generalizability 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 generalizability of distribution-based methods. In particular baseline standard deviations can be small and not representative of the heterogeneous COPD population. However, since we

used standard deviations 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 these measures in any case 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 6-minute walk test (26 m±2) is lower than the currently employed MID of 54 m, while the MID of maximal cycle exercise capacity is similar to previous estimates of 4±1 Watts. 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.

Acknowledgments

NETT Credit Roster

Source of funding

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).

Members of the NETT Research Group:

Office of the Chair of the Steering Committee, University of Pennsylvania, Philadelphia, PA: Alfred P. Fishman, MD (Chair); Betsy Ann Bozzarello; Ameena Al-Amin.

Clinical centers

Baylor College of Medicine, Houston, TX: Marcia Katz, MD (Principal Investigator); Carolyn Wheeler, RN, BSN (Principal Clinic Coordinator); Elaine Baker, RRT, RPFT; Peter Barnard, PhD, RPFT; Phil Cagle, MD; James Carter, MD; Sophia Chatziioannou, MD; Karla Conejo-Gonzales; Kimberly Dubose, RRT; John Haddad, MD; David Hicks, RRT, RPFT; Neal Kleiman, MD; Mary Milburn-Barnes, CRTT; Chinh Nguyen, RPFT; Michael Reardon, MD; Joseph Reeves-Viets, MD; Steven Sax, MD; Amir Sharafkhaneh, MD; Owen Wilson, PhD; Christine Young PT; Rafael Espada, MD (Principal Investigator 1996-2002); Rose Butanda (1999-2001); Minnie Ellisor (2002); Pamela Fox, MD (1999-2001); Katherine Hale, MD (1998-2000); Everett Hood, RPFT (1998 B 2000); Amy Jahn (1998-2000); Satish Jhingran, MD (1998-2001); Karen King, RPFT (1998-1999); Charles Miller III, PhD (1996-1999); Imran Nizami, MD (Co-Principal Investigator, 2000-2001); Todd Officer (1998-2000); Jeannie Ricketts (1998 -2000); Joe Rodarte, MD (Co-Principal Investigator 1996-2000); Robert Teague, MD (Co-Principal Investigator 1999-2000); Kedren Williams (1998-1999).

Brigham and Women's Hospital, Boston, MA: John Reilly, MD (Principal Investigator); David Sugarbaker, MD (Co-Principal Investigator); Carol Fanning, RRT (Principal Clinic Coordinator); Simon Body, MD; Sabine Duffy, MD; Vladmir Formanek, MD; Anne Fuhlbrigge, MD; Philip Hartigan, MD; Sarah Hooper, EP; Andetta Hunsaker, MD; Francine Jacobson, MD; Marilyn Moy, MD; Susan Peterson, RRT; Roger Russell, MD; Diane Saunders; Scott Swanson, MD (Co-Principal Investigator, 1996-2001).

Cedars-Sinai Medical Center, Los Angeles, CA: Rob McKenna, MD (Principal Investigator); Zab Mohsenifar, MD (Co-Principal Investigator); Carol Geaga, RN (Principal Clinic Coordinator); Manmohan Biring, MD; Susan Clark, RN, MN; Jennifer Cutler, MD; Robert Frantz, MD; Peter Julien, MD; Michael Lewis, MD; Jennifer Minkoff-Rau, MSW; Valentina Yegyan, BS, CPFT; Milton Joyner, BA (1996-2002).

Cleveland Clinic Foundation, Cleveland, OH: Malcolm DeCamp, MD (Principal Investigator); James Stoller, MD (Co-Principal Investigator); Yvonne Meli, RN,C (Principal Clinic Coordinator); John Apostolakis, MD; Darryl Atwell, MD; Jeffrey Chapman, MD; Pierre DeVilliers, MD; Raed Dweik, MD; Erik Kraenzler, MD; Rosemary Lann, LISW; Nancy Kurokawa, RRT, CPFT; Scott Marlow, RRT; Kevin McCarthy, RCPT; Priscilla McCreight, RRT, CPFT; Atul Mehta, MD; Moulay Meziane, MD; Omar Minai, MD; Mindi Steiger, RRT; Kenneth White, RPFT; Janet Maurer, MD (Principal Investigator, 1996-2001); Terri Durr, RN (2000-2001); Charles Hearn, DO (1998-2001); Susan Lubell, PA-C (1999-2000); Peter O'Donovan, MD (1998-2003); Robert Schilz, DO (1998-2002).

Columbia University, New York, NY in consortium with Long Island Jewish Medical Center, New Hyde Park, NY: Mark Ginsburg, MD (Principal Investigator); Byron Thomashow, MD (Co-Principal Investigator); Patricia Jellen, MSN, RN (Principal Clinic Coordinator); John Austin, MD; Matthew Bartels, MD; Yahya Berkmen, MD; Patricia Berkoski, MS, RRT (Site coordinator, LIJ); Frances Brogan, MSN, RN; Amy Chong, BS, CRT; Glenda DeMercado, BSN; Angela DiMango, MD; Sandy Do, MS, PT; Bessie Kachulis, MD; Arfa Khan, MD; Berend Mets, MD; Mitchell O=Shea, BS, RT, CPFT; Gregory Pearson, MD; Leonard Rossoff, MD; Steven Scharf, MD, PhD (Co-Principal Investigator, 1998-2002); Maria Shiau, MD; Paul Simonelli, MD; Kim Stavrolakes, MS, PT; Donna Tsang, BS; Denise Vilotijevic, MS, PT; Chun Yip, MD; Mike Mantinaos, MD (1998-2001); Kerri McKeon, BS, RRT, RN (1998-1999); Jacqueline Pfeffer, MPH, PT (1997-2002).

Duke University Medical Center, Durham, NC: Neil MacIntyre, MD (Principal Investigator); R. Duane Davis, MD (Co-Principal Investigator); John Howe, RN (Principal Clinic Coordinator); R. Edward Coleman, MD; Rebecca Crouch, RPT; Dora Greene; Katherine Grichnik, MD; David Harpole, Jr., MD; Abby Krichman, RRT; Brian Lawlor, RRT; Holman McAdams, MD; John Plankeel, MD; Susan Rinaldo-Gallo, MED; Sheila Shearer, RRT; Jeanne Smith, ACSW; Mark Stafford-Smith, MD; Victor Tapson, MD; Mark Steele, MD (1998-1999); Jennifer Norten, MD (1998-1999).

Mayo Foundation, Rochester, MN: James Utz, MD (Principal Investigator); Claude Deschamps, MD (Co-Principal Investigator); Kathy Mieras, CCRP (Principal Clinic Coordinator); Martin Abel, MD; Mark Allen, MD; Deb Andrist, RN; Gregory Aughenbaugh, MD; Sharon Bendel, RN; Eric Edell, MD; Marlene Edgar; Bonnie Edwards; Beth Elliot, MD; James Garrett, RRT; Delmar Gillespie, MD; Judd Gurney, MD; Boleyn Hammel; Karen Hanson, RRT; Lori Hanson, RRT; Gordon Harms, MD; June Hart; Thomas Hartman, MD; Robert Hyatt, MD; Eric Jensen, MD; Nicole Jenson, RRT; Sanjay Kalra, MD; Philip Karsell, MD; Jennifer Lamb; David Midthun, MD; Carl Mottram, RRT; Stephen Swensen, MD; Anne-Marie Sykes, MD; Karen Taylor; Norman Torres, MD; Rolf Hubmayr, MD (1998-2000); Daniel Miller, MD (1999-2002); Sara Bartling, RN (1998-2000); Kris Bradt (1998-2002).

National Jewish Medical and Research Center, Denver, CO: Barry Make, MD (Principal Investigator); Marvin Pomerantz, MD (Co-Principal Investigator); Mary Gilmartin, RN, RRT (Principal Clinic Coordinator); Joyce Canterbury; Martin Carlos; Phyllis Dibbern, PT; Enrique Fernandez, MD; Lisa Geyman, MSPT; Connie Hudson; David Lynch, MD; John Newell, MD; Robert Quaife, MD; Jennifer Propst, RN; Cynthia

Raymond, MS; Jane Whalen-Price, PT; Kathy Winner, OTR; Martin Zamora, MD; Reuben Cherniack, MD (Principal Investigator, 1997-2000).

Ohio State University, Columbus, OH: Philip Diaz, MD (Principal Investigator); Patrick Ross, MD (Co-Principal Investigator); Tina Bees (Principal Clinic Coordinator); Jan Drake; Charles Emery, PhD; Mark Gerhardt, MD, PhD; Mark King, MD; David Rittinger; Mahasti Rittinger.

Saint Louis University, Saint Louis, MO: Keith Naunheim, MD (Principal Investigator); Robert Gerber, MD (Co-Principal Investigator); Joan Osterloh, RN, MSN (Principal Clinic Coordinator); Susan Borosh; Willard Chamberlain, DO; Sally Frese; Alan Hibbit; Mary Ellen Kleinhenz, MD; Gregg Ruppel; Cary Stolar, MD; Janice Willey; Francisco Alvarez, MD (Co-Principal Investigator, 1999-2002); Cesar Keller, MD (Co-Principal Investigator, 1996-2000).

Temple University, Philadelphia, PA: Gerard Criner, MD (Principal Investigator); Satoshi Furukawa, MD (Co-Principal Investigator); Anne Marie Kuzma, RN, MSN (Principal Clinic Coordinator); Roger Barnette, MD; Neil Brister, MD; Kevin Carney, RN, CCTC; Wissam Chatila, MD; Francis Cordova, MD; Gilbert D'Alonzo, DO; Michael Keresztury, MD; Karen Kirsch; Chul Kwak, MD; Kathy Lautensack, RN, BSN; Madelina Lorenzon, CPFT; Ubaldo Martin, MD; Peter Rising, MS; Scott Schartel, MD; John Travaline, MD; Gwendolyn Vance, RN, CCTC; Phillip Boiselle, MD (1997-2000); Gerald O=Brien, MD (1997-2000).

University of California, San Diego, San Diego, CA: Andrew Ries, MD, MPH (Principal Investigator); Robert Kaplan, PhD (Co-Principal Investigator); Catherine Ramirez, BS, RCP (Principal Clinic Coordinator); David Frankville, MD; Paul Friedman, MD; James Harrell, MD; Jeffery Johnson; David Kapelanski, MD; David Kupferberg, MD, MPH; Catherine Larsen, MPH; Trina Limberg, RRT; Michael Magliocca, RN, CNP; Frank J. Papatheofanis, MD, PhD; Dawn Sassi-Dambron, RN; Melissa Weeks.

University of Maryland at Baltimore, Baltimore, MD in consortium with Johns Hopkins Hospital, Baltimore, MD: Mark Krasna, MD (Principal Investigator); Henry Fessler, MD (Co-Principal Investigator); Iris Moskowitz (Principal Clinic Coordinator); Timothy Gilbert, MD; Jonathan Orens, MD; Steven Scharf, MD, PhD; David Shade; Stanley Siegelman, MD; Kenneth Silver, MD; Clarence Weir; Charles White, MD.

University of Michigan, Ann Arbor, MI: Fernando Martinez, MD (Principal Investigator); Mark Iannettoni, MD (Co-Principal Investigator); Catherine Meldrum, BSN, RN, CCRN (Principal Clinic Coordinator); William Bria, MD; Kelly Campbell; Paul Christensen, MD; Kevin Flaherty, MD; Steven Gay, MD; Paramjit Gill, RN; Paul Kazanjian, MD; Ella Kazerooni, MD; Vivian Knieper; Tammy Ojo, MD; Lewis Poole; Leslie Quint, MD; Paul Rysso; Thomas Sisson, MD; Mercedes True; Brian Woodcock, MD; Lori Zaremba, RN.

University of Pennsylvania, Philadelphia, PA: Larry Kaiser, MD (Principal Investigator); John Hansen-Flaschen, MD (Co-Principal Investigator); Mary Louise Dempsey, BSN, RN (Principal Clinic Coordinator); Abass Alavi, MD; Theresa Alcorn, Selim Arcasoy, MD; Judith Aronchick, MD; Stanley Aukberg, MD; Bryan Benedict, RRT; Susan Craemer, BS, RRT, CPFT; Ron Daniele, MD; Jeffrey Edelman, MD; Warren Gefter, MD; Laura Kotler-Klein, MSS; Robert Kotloff, MD; David Lipson, MD; Wallace Miller, Jr., MD; Richard O=Connell, RPFT; Staci Opelman, MSW; Harold Palevsky, MD; William Russell, RPFT; Heather Sheaffer, MSW; Rodney Simcox, BSRT, RRT; Susanne Snedeker, RRT, CPFT; Jennifer Stone-Wynne, MSW; Gregory Tino, MD; Peter Wahl; James Walter, RPFT; Patricia Ward; David Zisman, MD; James Mendez, MSN, CRNP (1997-2001); Angela Wurster, MSN, CRNP (1997-1999).

University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD (Principal Investigator); James Luketich, MD (Co-Principal Investigator); Colleen Witt, MS (Principal Clinic Coordinator); Gerald Ayres; Michael Donahoe, MD; Carl Fuhrman, MD; Robert Hoffman, MD; Joan Lacomis, MD; Joan Sexton; William Slivka; Diane Strollo, MD; Erin Sullivan, MD; Tomeka Simon; Catherine Wrona, RN, BSN; Gerene Bauldoff, RN, MSN (1997-2000); Manuel Brown, MD (1997-2002); Elisabeth George, RN, MSN (Principal Clinic Coordinator 1997-2001); Robert Keenan, MD (Co-Principal Investigator 1997-2000); Theodore Kopp, MS (1997-1999); Laurie Silfies (1997-2001).

University of Washington, Seattle, WA: Joshua Benditt, MD (Principal Investigator), Douglas Wood, MD (Co-Principal Investigator); Margaret Snyder, MN (Principal Clinic Coordinator); Kymberley Anable; Nancy Battaglia; Louie Boitano; Andrew Bowdle, MD; Leighton Chan, MD; Cindy Chwalik; Bruce Culver, MD; Thurman Gillespy, MD; David Godwin, MD; Jeanne Hoffman; Andra Ibrahim, MD; Diane Lockhart; Stephen Marglin, MD; Kenneth Martay, MD; Patricia McDowell; Donald Oxorn, MD; Liz Roessler; Michelle Toshima; Susan Golden (1998-2000).

Other participants

Agency for Healthcare Research and Quality, Rockville, MD: Lynn Bosco, MD, MPH; Yen-Pin Chiang, PhD; Carolyn Clancy, MD; Harry Handelsman, DO.

Centers for Medicare and Medicaid Services, Baltimore, MD: Steven M Berkowitz, PhD; Tanisha Carino, PhD; Joe Chin, MD; JoAnna Baldwin; Karen McVearry; Anthony Norris; Sarah Shirey; Claudette Sikora Steven Sheingold, PhD (1997-2004).

Coordinating Center, The Johns Hopkins University, Baltimore, MD: Steven Piantadosi, MD, PhD (Principal Investigator); James Tonascia, PhD (Co-Principal Investigator); Patricia Belt; Amanda Blackford, ScM; Karen Collins; Betty Collison; Ryan Colvin, MPH; John Dodge; Michele Donithan, MHS; Vera Edmonds; Gregory L. Foster, MA; Julie Fuller; Judith Harle; Rosetta Jackson; Shing Lee, ScM; Charlene Levine; Hope Livingston; Jill Meinert; Jennifer Meyers; Deborah Nowakowski; Kapreena Owens; Shangqian Qi, MD; Michael Smith; Brett Simon, MD; Paul Smith; Alice Sternberg, ScM; Mark Van Natta, MHS; Laura Wilson, ScM; Robert Wise, MD.

Cost Effectiveness Subcommittee: Robert M. Kaplan, PhD (Chair); J. Sanford Schwartz, MD (Co-Chair); Yen-Pin Chiang, PhD; Marianne C. Fahs, PhD; A. Mark Fendrick, MD; Alan J. Moskowitz, MD; Dev Pathak, PhD; Scott Ramsey, MD, PhD;

Steven Sheingold, PhD; A. Laurie Shroyer, PhD; Judith Wagner, PhD; Roger Yusen, MD.

Cost Effectiveness Data Center, Fred Hutchinson Cancer Research Center, Seattle, WA: Scott Ramsey, MD, PhD (Principal Investigator); Ruth Etzioni, PhD; Sean Sullivan, PhD; Douglas Wood, MD; Thomas Schroeder, MA; Karma Kreizenbeck; Kristin Berry, MS; Nadia Howlader, MS.

CT Scan Image Storage and Analysis Center, University of Iowa, Iowa City, IA: Eric Hoffman, PhD (Principal Investigator); Janice Cook-Granroth, BS; Angela Delsing, RT; Junfeng Guo, PhD; Geoffrey McLennan, MD; Brian Mullan, MD; Chris Piker, BS; Joseph Reinhardt, PhD; Blake Wood; Jered Sieren, RTR; William Stanford, MD.

Data and Safety Monitoring Board: John A. Waldhausen, MD (Chair); Gordon Bernard, MD; David DeMets, PhD; Mark Ferguson, MD; Eddie Hoover, MD; Robert Levine, MD; Donald Mahler, MD; A. John McSweeny, PhD; Jeanine Wiener-Kronish, MD; O. Dale Williams, PhD; Magdy Younes, MD.

Marketing Center, Temple University, Philadelphia, PA: Gerard Criner, MD (Principal Investigator); Charles Soltoff, MBA.

Project Office, National Heart, Lung, and Blood Institute, Bethesda, MD: Gail Weinmann, MD (Project Officer); Joanne Deshler (Contracting Officer); Dean Follmann, PhD; James Kiley, PhD; Margaret Wu, PhD (1996-2001).

Other acknowledgments

Arthur Gelb, MD, Lakewood Regional Medical Center, Lakewood, CA.

References

1. Lacasse Y, Brosseau L, Milne S, Martin S, Wong E, Guyatt GH, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane.Database.Syst.Rev.* 2004(4):CD003793.
2. Peytremann-Bridevaux I, Staeger P, Bridevaux PO, Ghali WA, Burnand B. Effectiveness of chronic obstructive pulmonary disease-management programs: systematic review and meta-analysis. *Am J Med* 2008;121:433-43 e4.
3. Puhan MA, Garcia-Aymerich J, Frey M, ter Riet G, Anto JM, Agusti AG, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009;374:704-11.
4. Pinto-Plata VM, Cote C, Cabral H, Taylor J, Celli BR. 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, Miyazato I, Yoshizawa Y, Miyasaka N. Pulmonary exercise testing predicts prognosis in patients with chronic obstructive pulmonary disease. *Intern Med* 2005;44:20-5.
6. Schunemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;40:593-7.
7. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. *Am.J Respir Crit Care Med.* 1997;155:1278-82.
8. Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. *COPD* 2005;2:125-9.
9. Puhan MA, Mador MJ, Held U, Goldstein R, Guyatt GH, Schunemann HJ. Interpretation of treatment changes in 6-minute walk distance in patients with COPD. *Eur Respir J* 2008;32:637-43.
10. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371-83.
11. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. 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;2:137-41.

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;118:518-28.
14. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059-73.
15. Puhan MA, Guyatt GH, Goldstein R, Mador J, McKim D, Stahl E, 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;101:308-16.
16. Verrill D, Barton C, Beasley W, Lippard WM. The effects of short-term and long-term pulmonary rehabilitation on functional capacity, perceived dyspnea, and quality of life. *Chest* 2005;128:673-83.
17. Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003;56:1170-76.
18. Kupferberg DH, Kaplan RM, Slymen DJ, Ries AL. Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil* 2005;25:370-7.
19. Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *COPD* 2005;2:105-10.
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, Ries AL, Kaplan RM. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998;113:619-24.
22. Puhan MA, Frey M, Buchi S, Schunemann HJ. 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;52:861-73.
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, Sargent DJ, Novotny P, Atherton P, et al. Detecting worms, ducks, and elephants: A simple approach for defining

clinically relevant effects in quality-of-life measure. *Cancer Integrative Med* 2003;41-47.

26. Pitta F, Troosters T, Probst VS, Spruit MA, Decramer M, Gosselink R. Quantifying physical activity in daily life with questionnaires and motion sensors in COPD. *Eur Respir J* 2006;27:1040-55.

Table 1: Changes in exercise capacity and patient-reported outcomes during pre-trial rehabilitation and from randomization to 6-months follow-up

Measurement	Beginning of rehabilitation (n=1217)	End of rehabilitation (n=1217)	6-months after randomization (n=1001)	Change from beginning to end of rehabilitation (n=1217)	Change from randomization to 6 month follow-up (n=1001)
6MWD[1] (meters)	348.0 ± 95.1	370.8 ± 96.2	377.0 ± 99.9[3]	22.8 ± 53.6	-4.6 ± 68.7[3]
MCEC[2] (Watts)	36.0 ± 21.1	38.9 ± 21.6	42.2 ± 22.6[4]	2.9 ± 11.0	0.9 ± 13.9[4]
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

[1] 6MWD: 6-minute walk distance; [2] MCEC: Maximal Cycle Exercise Capacity; [3] n=907; [4] n= 905

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

Measurement	Correlation with change in 6MWD[1] from beginning to end of rehabilitation	Correlation with change in MCEC[2] 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**

[1] 6MWD: 6-minute walk distance; [2] MCEC: Maximal Cycle Exercise Capacity

In bold, correlations ≥0.3 and thus justifying anchor-based estimates

Table 3: Minimal important difference (MID) of 6-minute walk distance and maximal cycle exercise capacity

Method Used	MID of 6MWD (95% CI)	MID of MCEC[2] (95% CI)
Anchor Based		
SGRQ total score (MID=4)	24.6 (23.4-25.7)	3.3 (3.0-3.5)
SGRQ impact (MID=4)	18.9 (18.1-20.1)	2.2 (2.0-2.4)
SGRQ activities (MID=4)	24.2 (23.4-25.4)	3.2 (3.0-3.4)
SOBQ total score (MID=5)	26.4 (25.4-27.4)	3.2 (3.0 -3.4)
Distribution Based		
Standard error of measurement	30.6	-
Cohen's effect size	26.8	5.5
Empirical rule effect size	25.7	5.3

[1] 6MWD: 6-minute walk distance; [2] MCEC: Maximal Cycle Exercise Capacity; all regression equations multiplied by -1 to facilitate interpretation (improvements of SGRQ and SOBQ if change score is negative)