



Published in final edited form as:

*J Am Geriatr Soc.* 2021 February ; 69(2): 517–523. doi:10.1111/jgs.16942.

## Clinically Important Differences for Mobility Measures Derived from The Testosterone Trials

**Alisa J. Stephens-Shields, PhD<sup>1,\*</sup>, John T. Farrar, MD, PhD<sup>1</sup>, Susan S. Ellenberg, PhD<sup>1</sup>, Thomas W. Storer, PhD<sup>2</sup>, Thomas M. Gill, MD<sup>3</sup>, Shehzad Basaria, MD<sup>2</sup>, Marco Pahor, MD<sup>4</sup>, Jane A. Cauley, DrPH<sup>5</sup>, Kristine E. Ensrud, MD<sup>6</sup>, Peter Preston, MS<sup>7</sup>, David Cella, PhD<sup>8</sup>, Peter J. Snyder, MD<sup>9</sup>, Shalender Bhasin, MD<sup>3</sup>**

<sup>1</sup>Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

<sup>2</sup>Research Program in Men's Health: Aging and Metabolism; Boston Claude D. Pepper Older Americans Independence Center, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

<sup>3</sup>Section of Geriatric Medicine, Yale School of Medicine, New Haven, CT.

<sup>4</sup>Department of Aging & Geriatric Research, University of Florida, Gainesville, FL

<sup>5</sup>Department of Epidemiology, University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA

<sup>6</sup>Department of Medicine, Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN; Minneapolis VA Health Care System, Minneapolis, MN

<sup>7</sup>Center for Clinical Epidemiology & Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

<sup>8</sup>Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL

<sup>9</sup>Division of Endocrinology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

### Abstract

---

\*Corresponding author: [alisaste@penmedicine.upenn.edu](mailto:alisaste@penmedicine.upenn.edu); Twitter: @DrSteShi.

Author Contributions

Category 1

A. Concept and design: Alisa Stephens-Shields and Shalender Bhasin

B. Acquisition of subjects and data: Susan Ellenberg, Thomas Storer, Thomas Gill, Shehzad Basaria, Marco Pahor, Jane Cauley, Kristine Ensrud, Peter Preston, David Cella, and Peter Snyder

C. Analysis and interpretation of data: Alisa Stephens-Shields, Susan Ellenberg, John Farrar, Thomas Gill, Peter Preston, David Cella, and Shalender Bhasin

Category 2

A. Drafting the article: Alisa J. Stephens-Shields and Shalender Bhasin

B. Revising the article for intellectual content: Susan Ellenberg, John Farrar, Thomas Storer, Thomas Gill, Shehzad Basaria, Marco Pahor, Jane Cauley, Kristine Ensrud, Peter Preston, David Cella, and Peter Snyder

Category 3

1. Final approval of the completed article: Alisa Stephens-Shields, Susan Ellenberg, John Farrar, Thomas Storer, Thomas Gill, Shehzad Basaria, Marco Pahor, Jane Cauley, Kristine Ensrud, Peter Preston, David Cella, Peter J. Snyder, and Shalender Bhasin

**Background:** Accurate estimates of clinically important difference (CID) are required for interpreting the clinical importance of treatments to improve physical function, but CID estimates vary in different disease populations.

**Aims:** We determined the CID of 6-minute walking distance (6MWD) and physical function component (PF10) of Medical Outcomes Study Short Form-36 in mobility-limited older men using the Testosterone Trials (TTrials) data.

**Methods:** TTrials participants (n=429) with mobility limitation and gait speed<1.2m/sec, were divided into training and test sets. Patient global impression of change was used as anchor. Three anchor-based methods – regression, receiver operator curves (ROC), and empirical cumulative distribution functions (eCDF) — were used to determine the CID.

**Results:** Baseline characteristics did not differ between the training and test sets. Mean changes from baseline, adjusting for time-in-intervention and site, were 29.6, 13.2, 12.5, -2.4, and -32.6 meters for 6MWD, and 15.4, 7.2, 2.1, -3.4, and -7.2 for PF10 in men who reported their mobility was ‘very/much better’, ‘little better’, ‘no change’, ‘little worse’, or ‘much worse,’ respectively. CID estimates using regression, ROC, and eCDF varied from 5.0—29.6 meters for 6MWD, and 5.0—15.2 points for PF10.

**Conclusions:** CID estimates vary by the population studied and by the method and precision of measurement. Increases of 16 to 30 meters for 6MWD and 5 to 15 points for PF-10 over 12 months appear to be clinically meaningful in mobility-limited, older hypogonadal men. These CID estimates may be useful in the design of efficacy trials of therapies to improve physical function.

## Keywords

clinically important difference (CID); mobility improvement; randomized controlled trials

## Introduction

The “clinically important difference” (CID) is a patient-centered concept that refers to the change in an outcome measure that the patient perceives as beneficial (1). Accurate estimates of CID are required for defining dichotomous treatment response and for interpreting the clinical relevance of treatment effects in randomized trials of interventions to improve physical function. An average difference between groups, even if statistically significant, does not provide sufficient information about whether the treatment effect is important to patients.

The six-minute walking distance (6MWD) is a commonly used performance-based measure of mobility in older adults with mobility limitation, chronic obstructive pulmonary disease (COPD), heart failure, and other diseases (2–4). Although the CID for 6MWD is often cited to be 50 meters, the estimates have varied substantially - from 19 to 63 meters - depending upon the condition, study population, and the estimation method (5–12).

Many pharmacologic therapies, such as testosterone, selective androgen receptor modulators, myostatin-activin antagonists, growth hormone secretagogues, and troponin activators are being developed to treat aging- or chronic disease-associated mobility limitation; efficacy trials have typically included community-dwelling older adults with

mobility limitation (12–20). To facilitate the design of future trials of these therapies, we derived CID estimates for 6MWD and self-reported physical function in mobility-limited older adults using data from the Testosterone Trials (TTrials) (13, 21–23).

## Methods

The design, eligibility criteria, and main findings of the TTrials - a coordinated set of trials designed to determine the efficacy of testosterone in improving sexual function, physical function, and vitality in community-dwelling older men with low testosterone - have been published (13, 21–23) ([NCT00799617](#)).

### Study Population.

The Physical Function Trial, one of three main trials of the TTrials, enrolled men  $\geq 65$  years, with an average of two testosterone concentrations  $< 275$ -ng/dL (13), self-reported difficulty walking one-quarter mile and/or walking up one flight of stairs, and walking speed  $< 1.2$  m/sec in the 6-minute walk test (21). Of the 788 men enrolled across all TTrials, our analytic sample consisted of the 429 who self-reported difficulty walking and met gait speed criteria, inclusive of the 387 men who consented to the Physical Function Trial.

### Outcome Measures

The 6MWD, an objective measure of performance, is the total distance walked in 6 minutes. Men were instructed to walk as far as possible at a speed that they could safely maintain. Remaining time was announced at predetermined intervals; other encouragement was prohibited. The PF10 is a self-reported 10-item questionnaire that asked about limitations in activities such as lifting, climbing stairs, and walking. The Patient Global Impression of Change (PGIC) asked participants whether their walking ability was ‘very much better’, ‘much better’, ‘little better,’ ‘no change’, ‘little worse’, ‘much worse’, or ‘very much worse’ since the study’s start. ‘Very much’ and ‘much’ categories were combined to create a 5-point PGIC variable (21, 23). Responders were defined as men reporting ‘very much’ or ‘much better’.

The 6MWD, PF10, and PGIC were assessed at months 3, 6, 9, and 12. 6MWD and PF10 were also assessed at baseline.

### Statistical Analyses:

Spearman correlations of PGIC and change in the 6MWD and PF10 were calculated. Men were randomly divided into a training set and test set in a 1:2 ratio (n=143 training; n=286 test).

Because CID estimates can vary with the estimation method, we used three anchor-based methods in the training set - regression, receiver operator curves (ROC), and empirical cumulative distribution functions (eCDF) to determine CID thresholds for 6MWD and PF10; response to PGIC was the anchor. Regression utilized Generalized Estimating Equations clustered on the participant to account for repeated measures of 6MWD and PF10. Models included change from baseline in 6MWD or PF-10 as the outcome and 5-point PGIC

as the predictor; models adjusted for site and month. Dichotomous PGIC (responder, non-responder) was used in ROC analyses to evaluate candidate CID thresholds across the range of changes in 6MWD and PF10. Candidate thresholds considered increments of 5 meters for 6MWD and 5 points for PF10. Sensitivity and specificity were calculated for each threshold. ROC curves were plotted and areas under the curve (AUC) calculated. eCDFs of change in the 6MWD and PF10 were generated for each level of 5-point PGIC, and the median change in the ‘very/much better’ category was used as a CID estimate. A final threshold for CID was selected based on consideration of all three analyses and the coefficient of variation of 6MWD.

Testosterone’s effect on mobility outcomes was then evaluated in the test set using mixed effects logistic regression of binary change in 6MWD and PF10 according to selected CID estimates. Models included intervention arm, month, and balancing factors used in treatment allocation as fixed effects and a random intercept to account for within-participant correlation over time. Analyses were conducted using SAS version 9.4.

## Results

### CID thresholds in the training set

Of 429 qualifying TTrials participants, 212 were allocated to testosterone and 217 to placebo. Baseline characteristics did not differ between the training and test sets (Table 1). Spearman correlation of PGIC with change in mobility measures during months 3, 6, 9 and 12 ranged from 0.15–0.20 for 6MWD and 0.15–0.27 for PF10.

Adjusted mean (95% CI) changes in 6MWD from baseline in the training set were 29.6 (18.6,40.6), 13.2 (4.2,22.2), 12.5 (5.0,20.0), –2.4 (–20.5,15.6), and –32.6 (–58.2,–7.0) meters in men who reported ‘very much/much better’, ‘little better’, ‘no change’, ‘little worse’, ‘much worse,’ respectively. Similarly, mean changes in PF10 from baseline were 15.2 (9.1,21.3), 7.2 (3.1,11.2), 2.1 (–0.9,5.2), –3.4 (–9.6,2.8), and –7.2 (–14.8,0.3) across ‘very much/much better’ to ‘very much/much worse’ PGIC categories (Figure 1).

Using ROC analysis, the threshold for change in 6MWD that maximized the sum of sensitivity and specificity was 5 meters (Sensitivity=0.73, Specificity=0.56). A PF10 change of 5 maximized the sum of sensitivity and specificity (Sensitivity=0.74, Specificity=0.54, Figure 2).

The eCDFs of change in PF10 separated along PGIC response categories (Figure 2), with greater increases among men with more favorable PGIC responses. Little difference was observed in the distribution of change of 6MWD for “no change” and “little better” responses. Median changes in 6MWD were 16.2, 4.5, –1.6, and –10.6 meters in participants who reported “very much better/much better”, “little better”, “no change”, or “little worse” walking ability, respectively. Median changes in PF10 scores were 10, 5, 0, –5, and –5 in men who reported “very much/much better”, “little better”, “no change”, “little worse”, and “very much/much worse” function, respectively.

Supplementary Table 1 shows the estimated CID threshold by method. The ROC-based threshold was lower than the threshold suggested by regression and eCDF methods, and also substantially lower than reported test-retest variability of the 6MWD. We therefore exclude the ROC estimate from our proposed 6MWD CID range of 16–30 meters. The estimate of CID for PF10 ranged from 5–15 points.

### Proportions of Men in Each Intervention Arm Who Achieved Clinically Meaningful Change

The percent of men achieving change in 6MWD  $\geq$  16 meters ranged from 42% to 50% across months 3 to 12 in the testosterone arm and 32 to 37% in the placebo arm (Supplementary Figure 1). The larger threshold of 30 meters was achieved by 28% to 39% in testosterone arm and 17% to 25% in placebo arm. Adjusted ORs (95% CI) for increases of  $\geq$  16 meters and  $\geq$  30 meters were 1.5 (95% CI 0.9 to 2.7) and 1.8 (1.0, 3.1) for testosterone-versus placebo-treated men, respectively.

PF10 increased by  $\geq$  5 points in similar proportions of men in testosterone and placebo arms at month 3 (61% testosterone, 58% of placebo); the proportion of men with this degree of improvement remained stable over time across both arms (Supplementary Figure 1). The adjusted OR was 1.2 (95% CI 0.8-to-1.9) averaging across the treatment period. PF10 results did not differ between arms for the CID threshold of  $\geq$  15 (OR 1.3, 95% CI 0.8–2.2). Empirical cumulative distribution function curves of the PF10 and 6MWD by arm display differences in the percent of men achieving additional thresholds of change (Supplementary Figure 2).

## Discussion

In randomized trials of anabolic therapies in mobility-limited older adults, consideration of CIDs of mobility measures, such as 6MWD and PF10, is important in interpreting treatment effects. Although 50 meters has been widely accepted as the CID for the 6MWD and used to guide the design of many randomized trials, including the TTrials, our analyses indicate that 50 meters should not be viewed as the definitive estimate of CID because CID estimates vary depending on the population and the estimation methods used. We used three different methods for estimating CID; the estimates for the 6MWD varied from 5.0 meters by the ROC to 29.6-meters by regression. The test re-test variability of the measure should also be considered in establishing the CID; if the coefficient of variation exceeds the CID estimate, the higher of the two values should be considered. Because the ROC-based CID estimate was substantially lower than test-retest variability of the 6MWT (25–27), we propose 16–30 meters, based on the regression and eCDF methods, as reasonable estimates of the CID in mobility-limited, community-dwelling older men. The maximum value of 30 represents a conservative estimate that is least likely to result in false positive responses.

These estimates of CID for the 6MWD are lower than the 50 meters used in the design of the TTrials and lower than the CIDs reported in some epidemiologic studies of older adults with COPD or stroke survivors. They are similar, however, to those reported in another randomized testosterone trial in mobility-limited older men, in some other observational studies in COPD patients (9, 14), and the estimates reported in a Cochrane review (14.0–

30.5 meters) (28). The CID estimates of PF10 are similar to the expert opinion at the time of TTriaIs' planning.

These findings have important implications for clinical trials to determine the efficacy of function promoting therapies, and for clinical practice. Consistent with the work of Enright et al, who showed that estimates of 6MWD are affected by individual's characteristics, our data highlight the importance of determining the CID in the patient population for which the intervention is developed (29). Our analyses also show that CID estimates vary with the method of estimation; it is advantageous to use multiple approaches.

These analyses have several strengths and limitations. The estimates were derived in the setting of one of the largest trials of testosterone's effects on physical function in mobility-limited, older hypogonadal men. The 6MWD was assessed by well-trained staff. Inclusion of a PGIC rendered it possible to corroborate whether participants perceived improvement in their walking speed, and to use it as an anchor. We used three different methods to estimate the CID. Limitations of our study include the modest correlation between the PGIC anchor and mobility measures and that we did not measure test-retest variability in the current population. The response to PGIC can also be affected by depressive symptoms; however, depressive symptom scores were similar between arms at baseline and were only minimally affected by testosterone.

The analyses presented were neither prespecified nor adjusted for multiple comparisons. These results do not replace the pre-specified primary endpoint analysis of the TTriaIs, nor should they be used to re-interpret the published results of the TTriaIs. Because CID estimates are specific for the context of use, these data do not imply that published estimates of CID for 6MWD derived in different populations are incorrect. Further, our CID estimates were derived in participants who were enrolled in an RCT; individuals participating in research may differ from those not participating in research. Our CID estimates, derived using multiple statistically rigorous approaches in the setting of a controlled trial, may be more appropriate for future studies of function-promoting therapies in mobility-limited older adults.

## Conclusion

We estimated the CID for 6MWD to be 16 to 30 meters and for PF10 5 to 15 points in older mobility-limited hypogonadal men. These findings emphasize the importance of context-specific CID estimates to determine the clinical significance of treatment effects in efficacy trials. Because the derivation of CID estimates is time consuming, embedding anchors in phase 2 trials can be efficient in deriving CIDs and guiding sample size estimates for phase 3 trials. Our analyses also show that multiple adjacent thresholds for response can be evaluated using a cumulative distribution function.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

### Conflict of Interest

The Testosterone Trials were supported by a grant from the National Institute on Aging, National Institutes of Health (U01 AG030644), supplemented by funds from the National Heart, Lung and Blood Institute, National Institute of Neurological Diseases and Stroke, and National Institute of Child Health and Human Development. AbbVie (formerly Solvay and Abbott Laboratories) provided funding and donated the study medication and the placebo gel. Dr. Bhasin was supported partially by the Boston Claude D. Pepper Older Americans Independence Center grant 5P30AG031679 (SB, PI).

TMG is the recipient of an Academic Leadership Award (K07AG3587) from the National Institute on Aging. The Yale Field Center was partially supported by the Claude D. Pepper Older Americans Independence Center (P30-AG021342) and CTSI (UL1TR000142).

### Sponsor's Role

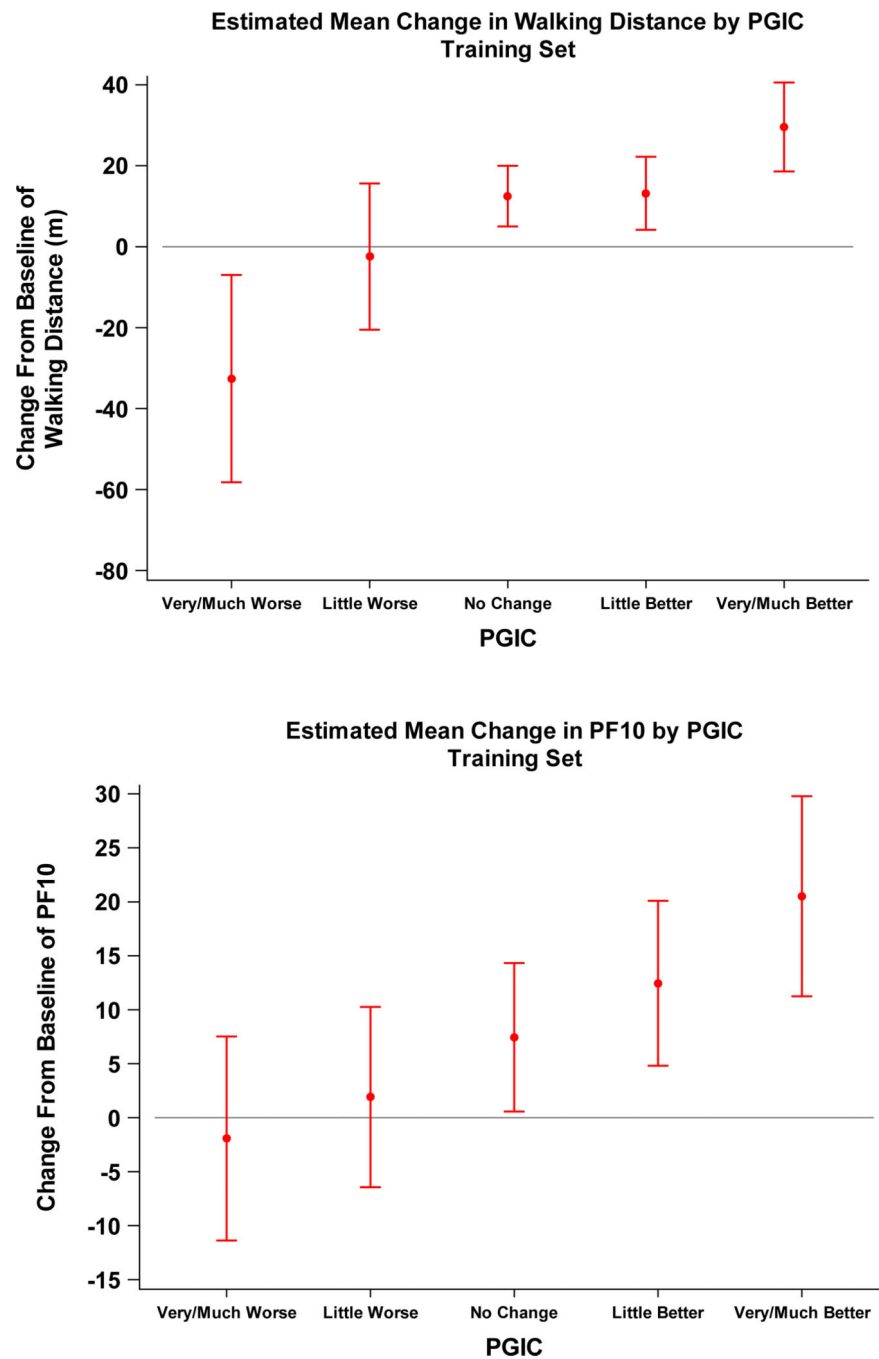
The funding agencies played no role in the design of the trial, analyses of data, preparation of the manuscript, or in the decision to publish.

## REFERENCES

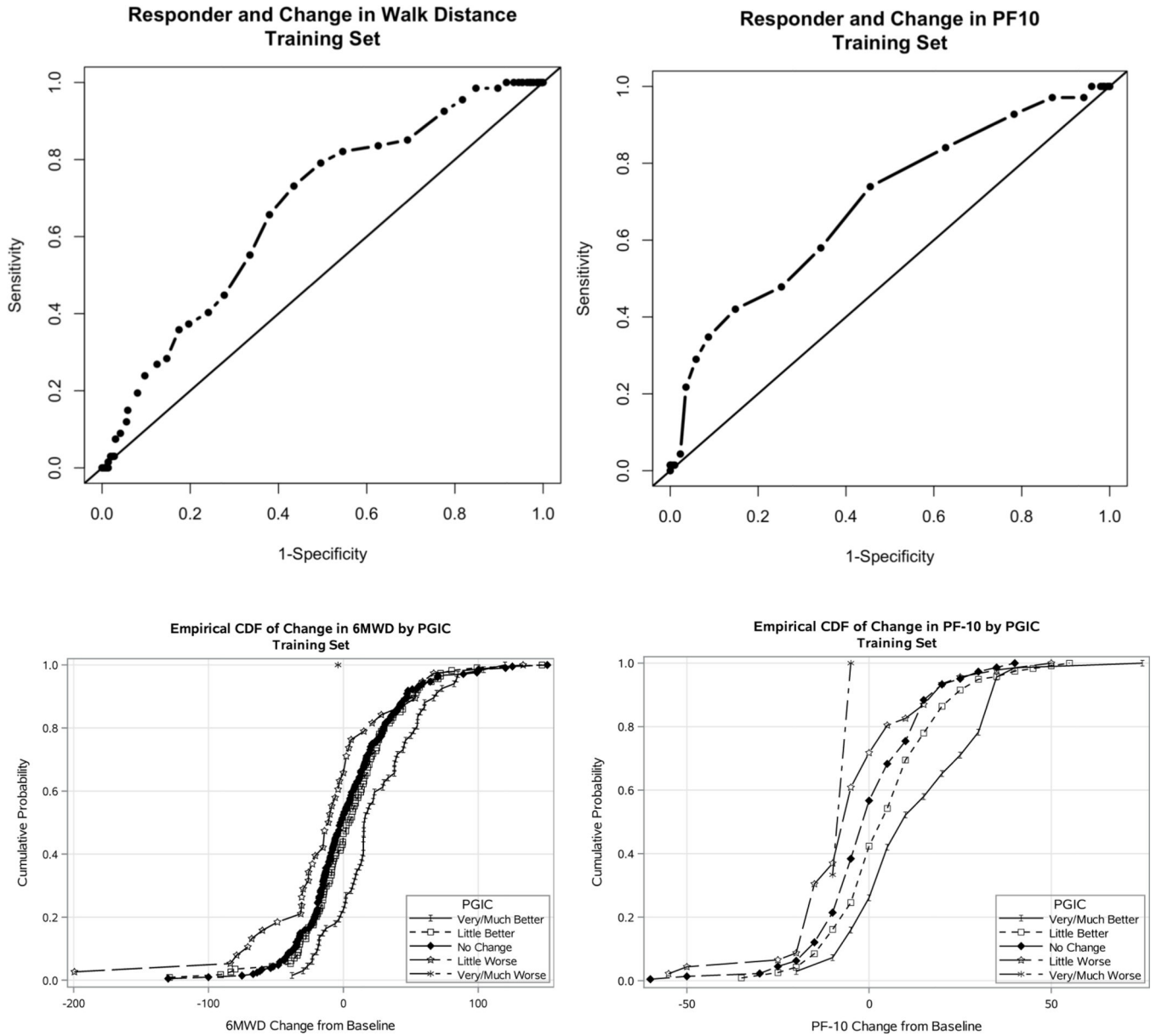
1. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA*. 2014 10 1;312(13):1342–3. [PubMed: 25268441]
2. Dajczman E, Wardini R, Kasymjanova G, Préfontaine D, Baltzan MA, Wolkove N. Six minute walk distance is a predictor of survival in patients with chronic obstructive pulmonary disease undergoing pulmonary rehabilitation. *Can Respir J*. 2015;22(4):225–229. [PubMed: 26252533]
3. Halliday SJ, Wang L, Yu C, et al. Six-minute walk distance in healthy young adults. *Respir Med*. 2020;165:105933.
4. Wolsk E, Kaye D, Borlaug BA, et al. Resting and exercise haemodynamics in relation to six-minute walk test in patients with heart failure and preserved ejection fraction. *Eur J Heart Fail*. 2018;20(4):715–722. [PubMed: 28949052]
5. Redelmeier D, Bayoumi A, Goldstein R and Guyatt G. 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. [PubMed: 9105067]
6. Perera S, Mody S, Woodman R and Studenski S. Meaningful change and responsiveness in common physical performance measures in older adults. *Journal of the American Geriatrics Society*. 2006; 54: 743–9. [PubMed: 16696738]
7. Kennedy D, Stratford P, Wessel J, Gollish J and Penney D. Assessing stability and change of four performing measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC musculoskeletal disorders*. 2005 6:3. [PubMed: 15679884]
8. 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;32:637–643. [PubMed: 18550610]
9. Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. *COPD*. 2005 3;2(1):125–9. [PubMed: 17136972]
10. Fulk GD, Ludwig M, Dunning K, Golden S, Boyne P, West T. Estimating clinically important change in gait speed in people with stroke undergoing outpatient rehabilitation. *J Neurol Phys Ther*. 2011 6;35(2):82–9. [PubMed: 21934363]
11. Tilson JK, Sullivan KJ, Cen SY, et al. Locomotor Experience Applied Post Stroke (LEAPS) Investigative Team. Meaningful gait speed improvement during the first 60 days poststroke: minimal clinically important difference. *Phys Ther*. 2010 2;90(2):196–208. [PubMed: 20022995]
12. Holland AE, Hill CJ, Rasekaba T, Lee A, Naughton MT, McDonald CF. Updating the minimal important difference for six-minute walk distance in patients with chronic obstructive pulmonary disease. *Arch Phys Med Rehabil*. 2010 2;91(2):221–5. [PubMed: 20159125]
13. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone on sexual function, physical function and vitality in older men. *N Engl J Med*. 2016;374(7):611–24. [PubMed: 26886521]

14. Travison TG, Basaria S, Storer TW, et al. Clinical meaningfulness of the changes in muscle performance and physical function associated with testosterone administration in older men with mobility limitation. *J Gerontol Medical Sciences A*. 2011; 66:1090–9.
15. Blackman MR, Sorkin JD, Munzer T, et al. Growth hormone and sex steroid administration in healthy aged women and men: a randomized controlled trial. *JAMA*. 2002; 288: 2282–92. [PubMed: 12425705]
16. Basaria S, Harman SM, Travison TG, et al. Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: A Randomized Clinical Trial. *JAMA*. 2015;314:570–81. [PubMed: 26262795]
17. Page ST, Amory JK, Bowman FD, et al. Exogenous testosterone (T) alone or with finasteride increases physical performance, grip strength, and lean body mass in older men with low serum T. *J Clin Endocrinol Metab*. 2005; 90: 1502–10. [PubMed: 15572415]
18. Srinivas-Shankar U, Roberts SA, Connolly MJ, et al. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*. 2010; 95: 639–50. [PubMed: 20061435]
19. Becker C, Lord SR, Studenski SA, et al. Myostatin antibody (LY2495655) in older weak fallers: a proof-of-concept, randomised, phase 2 trial. *Lancet Diabetes Endocrinol*. 2015 12;3(12):948–57. [PubMed: 26516121]
20. Polkey MI, Praestgaard J, Berwick A, et al. Activin Type II Receptor Blockade for Treatment of Muscle Depletion in Chronic Obstructive Pulmonary Disease. A Randomized Trial. *Am J Respir Crit Care Med*. 2019;199(3):313–320. [PubMed: 30095981]
21. Bhasin S, Ellenberg SS, Storer TW, et al. Effect of testosterone replacement on measures of mobility in older men with mobility limitation and low testosterone concentrations: secondary analyses of the Testosterone Trials. *Lancet Diabetes Endocrinol*. 2018 11;6(11):879–890. [PubMed: 30366567]
22. Snyder PJ, Bhasin S, Cunningham GR, et al. Cohen HJ, Schrier S, Keaveny TM, Kopperdahl D, Lee D, Cifelli D, Ellenberg SS. Lessons From the Testosterone Trials. *Endocr Rev*. 2018 6 1;39(3):369–386. [PubMed: 29522088]
23. Snyder PJ, Ellenberg SS, Cunningham GR, et al. The Testosterone Trials: Seven coordinated trials of testosterone treatment in elderly men. *Clin Trials*. 2014 3 31;11(3):362–375. [PubMed: 24686158]
24. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30:473–83. [PubMed: 1593914]
25. Rikli RE, Jones CJ. The reliability and validity of a 6-minute walk test as a measure of physical endurance in older adults. *J Aging and Physical activity* 1998;6:363–375.
26. Alfonso-Rosa RM, del Pozo-Cruz B, del Pozo-Cruz J et al. Test retest reliability and minimal detectable change scores for fitness assessment in older adults with type 2 diabetes. *Rehabilitation Nursing* 2014; 39:260–268. [PubMed: 23780835]
27. King MB, Judge JO, Whipple R, Wolfson L. Reliability and responsiveness of two performance measures examined in the context of a functional training intervention. *Physical Therapy* 2000;80:8–16. [PubMed: 10623956]
28. Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *J Eval Clin Pract*. 2017 4;23(2):377–381. [PubMed: 27592691]
29. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. *Am J Respir Crit Care Med* 1998; 158: 1384–1387 [PubMed: 9817683]
30. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 94(2):149–58, 2001. [PubMed: 11690728]





**Figure 1.** Mean and 95% CI of estimated mean change in walking outcome: Six-minute walk distance (6MWD, top) and the physical function component of the Medical Outcomes Study Short Form-36 (PF10, bottom) for each level of Patient Global Impression of Change (PGIC) category. Mean change per PGIC response category was estimated by a marginal model and generalized estimating equations clustered by participant with walking outcome as the dependent variable and categorical PGIC response as the independent variable. The model adjusted for site and month.



**Figure 2.** Receiver Operating Characteristic Curves and Empirical Distribution Functions. Receiver Operating Characteristic Curves (ROC) of responder status as defined by dichotomized PGIC response and candidate thresholds of dichotomous change in walking outcome: 6MWD, PGIC (top left); PGIC PF10 (top right); Empirical cumulative distribution functions (eCDF) of change in 6MWD (bottom left, range –200 to 151 meters) and PF10 (bottom right, range –60 to 75) by each category of Patient Global Impression of Change response. Each point along the eCDF curve denotes the value such that the shown proportion of subjects increased by less than or equal to the indicated amount in respective walking outcome.

**Table 1.**

Comparison of Characteristics of the Participants in the Training and Test Sets

Parameter	Levels	Training Set	Test Set	P-value
N		143 (100%)	286 (100%)	
Arm	A	72 (50.3%)	140 (49.0%)	0.7848
Arm	B	71 (49.7%)	146 (51.0%)	
Age		143 73.8 ± 6.4	286 73.3 ± 6.3	0.4313
Race	Black/African American	7 (4.9%)	18 (6.3%)	0.8390
Race	White/Caucasian	127 (88.8%)	251 (87.8%)	
Race	Other	9 (6.3%)	17 (5.9%)	
Ethnicity	Hispanic/Latino	7 (4.9%)	9 (3.2%)	0.3715
Ethnicity	Not Hispanic/Latino	136 (95.1%)	276 (96.8%)	
Education	College graduate	72 (50.3%)	135 (47.2%)	0.5386
BMI		142 31.1 ± 3.3	286 31.7 ± 3.5	0.0711
PHQ-9	4	61 (43.3%)	140 (50.2%)	0.2019
PHQ-9	5 – 14	72 (51.1%)	131 (47.0%)	
PHQ-9	>14	8 (5.7%)	8 (2.9%)	
History of Smoking		102 (71.8%)	195 (68.9%)	0.5351
Diabetes		66 (46.2%)	113 (39.5%)	0.1884
Hypertension		114 (79.7%)	207 (72.4%)	0.0986
Sleep Apnea		28 (19.7%)	54 (18.9%)	0.8489
6MWD - Month 0		139 347 ± 67.4	284 345 ± 69.5	0.7628
6MWD - Month 3		128 351 ± 74.0	262 357 ± 70.4	0.4279
6MWD - Month 6		123 355 ± 77.9	247 354 ± 71.6	0.9169
6MWD - Month 9		114 350 ± 80.2	242 352 ± 80.2	0.7814
6MWD - Month 12		116 363 ± 77.2	245 358 ± 76.7	0.5465
PF10 - Month 0		143 64.7 ± 22.0	286 64.4 ± 20.4	0.8832

Parameter	Levels	Training Set	Test Set	P-value
PF10 - Month 3		124 70.1 ± 21.8	255 69.7 ± 19.6	0.8593
PF10 - Month 6		122 69.9 ± 22.7	248 70.8 ± 20.6	0.6940
PF10 - Month 9		125 69.3 ± 22.7	248 70.0 ± 21.6	0.7715
PF10 - Month 12		127 68.2 ± 23.1	261 69.0 ± 21.3	0.7496
PGIC - Month 3	Very/Much better	20 (17.7%)	45 (19.7%)	0.6530
PGIC - Month 3	Little better	33 (29.2%)	66 (28.8%)	
PGIC - Month 3	No change	51 (45.1%)	107 (46.7%)	
PGIC - Month 3	Little worse	9 (8.0%)	10 (4.4%)	
PGIC - Month 3	Very/Much worse		1 (0.4%)	
PGIC - Month 6	Very/Much better	15 (13.4%)	37 (16.4%)	0.5448
PGIC - Month 6	Little better	34 (30.4%)	71 (31.4%)	
PGIC - Month 6	No change	57 (50.9%)	101 (44.7%)	
PGIC - Month 6	Little worse	6 (5.4%)	13 (5.8%)	
PGIC - Month 6	Very/Much worse		4 (1.8%)	
PGIC - Month 9	Very/Much better	21 (17.4%)	44 (18.7%)	0.0753
PGIC - Month 9	Little better	23 (19.0%)	67 (28.5%)	
PGIC - Month 9	No change	58 (47.9%)	104 (44.3%)	
PGIC - Month 9	Little worse	17 (14.0%)	15 (6.4%)	
PGIC - Month 9	Very/Much worse	2 (1.7%)	5 (2.1%)	
PGIC - Month 12	Very/Much better	13 (11.3%)	35 (15.0%)	0.6073
PGIC - Month 12	Little better	28 (24.3%)	66 (28.3%)	
PGIC - Month 12	No change	59 (51.3%)	106 (45.5%)	
PGIC - Month 12	Little worse	14 (12.2%)	22 (9.4%)	
PGIC - Month 12	Very/Much worse	1 (0.9%)	4 (1.7%)	