

# UCLA

## UCLA Previously Published Works

### Title

Validity and Clinically Meaningful Changes in the Psychosexual Daily Questionnaire and Derogatis Interview for Sexual Function Assessment: Results From the Testosterone Trials

### Permalink

<https://escholarship.org/uc/item/06g0w4qd>

### Journal

The Journal of Sexual Medicine, 15(7)

### ISSN

1743-6095

### Authors

Wang, Christina  
Stephens-Shields, Alisa J  
DeRogatis, Leonard R  
[et al.](#)

### Publication Date

2018-07-01

### DOI

10.1016/j.jsxm.2018.05.008

Peer reviewed



Published in final edited form as:

*J Sex Med.* 2018 July ; 15(7): 997–1009. doi:10.1016/j.jsxm.2018.05.008.

## Validity and Clinically Meaningful Changes in the Psychosexual Daily Questionnaire and Derogatis Interview for Sexual Function Assessment: Results From the Testosterone Trials

Christina Wang, MD<sup>1,\*</sup>, Alisa J. Stephens-Shields, PhD<sup>2,\*</sup>, Leonard R. DeRogatis, PhD<sup>3</sup>, Glenn R. Cunningham, MD<sup>4</sup>, Ronald S. Swerdloff, MD<sup>1</sup>, Peter Preston, MS<sup>2</sup>, David Cella, PhD<sup>5</sup>, Peter J. Snyder, MD<sup>6</sup>, Thomas M. Gill, MD<sup>7</sup>, Shalender Bhasin, MD<sup>8</sup>, Alvin M. Matsumoto, MD<sup>9</sup>, and Raymond C. Rosen, PhD<sup>10</sup>

<sup>1</sup>Division of Endocrinology, Department of Medicine, Harbor-University of California, Los Angeles Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA, USA

<sup>2</sup>Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>3</sup>Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA

---

**Corresponding Author:** Christina Wang, MD, Clinical and Translational Science Institute, Harbor-University of California, Los Angeles Medical Center and Los Angeles Biomedical Research Institute, 1000 West Carson Street, Torrance, CA 90509, USA; wang@labiomed.org.

\*Contributed equally.

### STATEMENT OF AUTHORSHIP

#### Category 1

##### (a) Conception and Design

Christina Wang; Alisa J. Stephens-Shields; Leonard R. DeRogatis; Glenn R. Cunningham; Ronald S. Swerdloff; David Cella; Peter J. Snyder; Raymond C. Rosen

##### (b) Acquisition of Data

Christina Wang; Alisa J. Stephens-Shields; Peter J. Snyder; Raymond C. Rosen

##### (c) Analysis and Interpretation of Data

Christina Wang; Alisa J. Stephens-Shields; Leonard R. DeRogatis; Glenn R. Cunningham; Ronald S. Swerdloff; Peter Preston; David Cella; Peter J. Snyder; Thomas M. Gill; Shalender Bhasin; Alvin M. Matsumoto; Raymond C. Rosen

#### Category 2

##### (a) Drafting the Article

Christina Wang; Alisa J. Stephens-Shields; Raymond C. Rosen

##### (b) Revising It for Intellectual Content

Christina Wang; Alisa J. Stephens-Shields; Leonard R. DeRogatis; Glenn R. Cunningham; Ronald S. Swerdloff; Peter Preston; David Cella; Peter J. Snyder; Thomas M. Gill; Shalender Bhasin; Alvin M. Matsumoto; Raymond C. Rosen

#### Category 3

##### (a) Final Approval of the Completed Article

Christina Wang; Alisa J. Stephens-Shields; Leonard R. DeRogatis; Glenn R. Cunningham; Ronald S. Swerdloff; Peter Preston; David Cella; Peter J. Snyder; Thomas M. Gill; Shalender Bhasin; Alvin M. Matsumoto; Raymond C. Rosen

**Conflicts of interest:** C.W. reports research funding from University of California, Los Angeles Clinical Science and Translation Institute grant UL1TR001881 and *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) and research support from Clarus Therapeutics and served as a consultant for Antares. G.R.C. reports receiving fees for serving as an advisor to AbbVie, Clarus Therapeutics, Ferring Pharmaceuticals, Eli Lilly, Lipocine, and Repros Therapeutics and has served as an expert witness for Repros Therapeutics and Merck. R.S.S. reports research funding from NICHD and support from Clarus Therapeutics and AETerna Zentaris Inc and served as a consultant for Clarus Therapeutics. T.M.G. is the recipient of an Academic Leadership Award (K07AG043587) from the National Institute on Aging and is supported by the Yale Claude D. Pepper Older Americans Independence Center (P30AG21342). S.B. reports research grants from National Institute on Aging, National Institute of Nursing Research, Foundation for the National Institutes of Health, Patient-Centered Outcomes Research Institute, AbbVie, Metro International Biotechnology, and Transition Therapeutics; served as a consultant for AbbVie and Novartis; and has equity interest in FPT LLC. A.M.M. reports research support from AbbVie and GlaxoSmithKline, and served as an advisor for AbbVie, Endo, Lipocine, and Aytu Bioscience. R.C.R. reports research support from Bayer Healthcare. A.J.S.-S., L.R.D., D.C., P.P., and P.J.S. report no disclosures.

<sup>4</sup>Division of Diabetes, Endocrinology, and Metabolism, Baylor College of Medicine and Baylor St Luke's Medical Center, Houston, TX, USA

<sup>5</sup>Department of Medical Social Science, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

<sup>6</sup>Division of Endocrinology, Diabetes, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>7</sup>Section of Geriatrics, Department of Medicine, Yale School of Medicine, New Haven, CT, USA

<sup>8</sup>Research Program in Men's Health: Aging and Metabolism; Boston Claude D. Pepper Older Americans Independence Center; Harvard Medical School; Brigham and Women's Hospital, Boston, MA, USA

<sup>9</sup>Geriatric Research, Education and Clinical Center, Department of Veterans Affairs Puget Sound Health Care System, and Division of Gerontology and Geriatric Medicine, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA

<sup>10</sup>HealthCore/New England Research Institutes Inc, Watertown, MA, USA

## Abstract

**Background:** Limited information is available on the performance characteristics of 2 questionnaires commonly used in clinical research, the Psychosexual Daily Questionnaire (PDQ) and the Derogatis Interview for Sexual Function (DISF)-II Assessment, especially in older men with low testosterone (T) and impaired sexual function.

**Aim:** To determine reliability of PDQ and DISF-II by assessing the correlation within and between domains in the questionnaires and to define clinically meaningful changes in sexual activity (PDQ question 4 [Q4]) and desire (DISF-II sexual desire domain [SDD]) domains.

**Methods:** Data from 470 men participating in the T Trials were used to calculate Spearman correlation coefficients of individual items and total score among questionnaires to determine convergent and construct validity. Clinically meaningful changes for sexual desire and activity were determined by randomly dividing the sample into training and validation sets. Anchor- and distribution-based clinically meaningful change criteria were defined in the training set, and selected changes were evaluated in the validation set.

**Outcomes:** Validity of the PDQ and DISF-II and clinically meaningful changes in sexual desire and activity were determined in older men in T Trials.

**Results:** Moderate to strong correlations were shown within and between domains from different questionnaires. Using Patient Global Impression of Change as an anchor, clinically meaningful change in PDQ sexual activity was 0.6, and in DISF-SDD was 5.0. Applying these change cut-points to the validation set, a greater proportion of T-treated men achieved clinically meaningful improvement in their sexual desire and activity compared to placebo-treated men.

**Clinical Implications:** The PDQ-Q4 and DISF-II-SDD can be used to reliably assess clinically meaningful changes in sexual activity and sexual desire in hypogonadal men treated with T.

**Strengths & Limitations:** Strengths of this study include a large sample size, long trial duration, and inclusion of men with low libido and unequivocally low T levels. Limitations include using data from a single study that enrolled only older hypogonadal men, and only 1 anchor for both sexual desire and activity.

**Conclusion:** Moderate to strong correlations were demonstrated within and between different sexual domains of the PDQ and DISF-II confirming construct and convergent validity. Clinically meaningful improvement in elderly hypogonadal men was change of 0.6 score in the PDQ-Q4 and 5.0 in the DISF-SDD. Improvements in sexual activity and desire in the T Trials were modest but clinically meaningful.

### Keywords

Sexual Desire; Sexual Activity; Erectile Function; Clinically Meaningful Change; Sexual Function Assessment; Testosterone Deficiency; Testosterone Treatment in Older Men; Testosterone Trials

---

## INTRODUCTION

The most common manifestation of testosterone (T) deficiency in men is sexual dysfunction. In the Boston Area Community Health Survey of men with a mean age of  $47.3 \pm 12.5$  years, low libido and erectile dysfunction were the most common symptoms of T deficiency.<sup>1</sup> The European Male Aging Study, a population sample of 3,369 men between 40 and 70 years from 8 countries, showed that low T levels were associated with sexual symptoms, inability to perform vigorous activity, depression, and fatigue; however, only 3 sexual symptoms: poor morning erection, low sexual desire, and erectile dysfunction had a syndromic association with low T concentrations.<sup>2</sup> When sexual function was assessed over a 9-year period in 1,085 men in the Massachusetts Male Aging Study, decline in sexual intercourse, erection frequency, sexual desire, satisfaction with sex, and difficulty with orgasm were related to age.<sup>3</sup> In these studies, as well as many other studies, the assessment of sexual function varied considerably, from a single question to composite questionnaires.<sup>4-7</sup>

The T Trials are a coordinated set of placebo-controlled trials to determine if T gel treatment of older men with unequivocally low serum T concentrations and symptoms and objective evidence of impaired mobility and/or diminished libido and/or reduced vitality would be efficacious in improving sexual function, physical function, and vitality.<sup>8,9</sup> The Sexual Function Trial of the T Trials included only older men with sexual symptoms. The T Trials are the first study where 3 different questionnaires have been used to separately assess erectile dysfunction, sexual desire, and sexual activity. The International Index of Erectile Function (IIEF) is the most widely used extensively validated multi-dimensional instrument to assess male sexual function and to diagnose erectile dysfunction.<sup>6,7,10</sup> The Derogatis Interview for Sexual Function (DISF)-II is a multi-dimensional outcome measure of the quality of sexual functioning.<sup>11-13</sup> The Psychosexual Daily Questionnaire (PDQ) is a 7-day self-report questionnaire.<sup>14</sup> Improvements in PDQ items have been shown after T treatment of hypogonadal men.<sup>15-19</sup> The questionnaire also has been used to monitor sexual function after administration of exogenous T and progestins for male hormonal contraceptive development.<sup>20,21</sup>

After treatment with T gel for a year, older men with sexual symptoms at baseline had significant improvement in sexual activity assessed by the PDQ question 4 (Q4), sexual desire by DISF-II sexual desire domain (SDD), and erectile function by IIEF (IIEF erectile function domain [EF]) compared to men in the placebo group.<sup>9</sup> Men in the T gel group also reported improvement in Patient Global Impression of Change (PGIC) in sexual desire from the beginning of the trial.<sup>9</sup> In these men, incremental increases in serum total and free T and estradiol concentrations were associated with improvement in scores of sexual activity and sexual desire but not with erectile dysfunction.<sup>22</sup>

The responses of men to different sexual parameters have not been compared across the 3 questionnaires, nor have they been cross-validated with the PGIC. Additionally, the PDQ-Q4 and DISF-II-SDD do not have established thresholds to define clinically meaningful change. Rosen et al<sup>23</sup> has reported that a clinically important change in the question on satisfaction of sexual intercourse in the IIEF-EF is 4 points.

In this article, we used Spearman correlation coefficient to show that within and between domain associations of the PDQ and DISF-II were in the expected directions. We also assessed the magnitude of correlation among the 3 questionnaires (PDQ, DISF-II-SDD, and IIEF). We use PGIC scores to develop anchor-based thresholds for clinically meaningful change of sexual desire in DISF-II-SDD and sexual activity in PDQ-Q4 and additionally determined distribution-based estimates for further evaluation of the magnitude of clinically meaningful change.<sup>13,24,25</sup> We then used these thresholds to assess whether improvement in sexual activity and function was clinically meaningful in the T Trials.

## METHODS

### Study Design and Participants

Participants (n = 470) and data were from the Sexual Function Trial of the T Trials,<sup>9,22</sup> one of 7 highly coordinated, double-blind, placebo-controlled trials designed to determine the efficacy of T treatment of older hypogonadal men in improving several conditions thought to be related to low T, including sexual function, walking ability, and vitality. Participants were required to be at least 65 years old and have an average total T level <275 ng/dL over 2 early morning fasting screening assessments. Sexual Function Trial participants were additionally required to have self-reported decreased libido, decreased desire as indicated by a score  $\geq 20$  on the of the DISF-II-SDD,<sup>11</sup> and a partner willing to have intercourse at least twice per month. All T Trials participants were required to qualify for at least 1 of the 3 main trials, physical function, sexual function, or vitality, and they were able to enroll in multiple trials, if they qualified. Detailed inclusion and exclusion criteria have been described.<sup>8,9</sup> Men were allocated to receive active or placebo gel for 1 year via minimization, with men assigned to the optimally balancing treatment arm with 80% probability. Balancing factors were trial site, age  $\geq 75$  years, baseline serum total T  $\geq 200$  ng/dL, participation in each of the 3 main trials, use of anti-depressants, and use of phosphodiesterase-5 inhibitors. Outcomes were assessed at baseline and every 3 months during the treatment period.

## DISF-II

The DISF is a multi-item questionnaire that assesses sexual thoughts and activities.<sup>11</sup> It has 25 items and 5 domains: sexual cognition/fantasy, sexual arousal, sexual behavior/experience, orgasm, and sexual drive/relationship. Individual domains can be scored, and there is an aggregate DISF total score that summarizes the 5 domains. The DISF-II is an updated version of the self-report of DISF.<sup>11–13</sup> The DISF-II-SDD and sexual arousal domain each consists of 4 questions that are answered on a verbal response scale ranging from 0–7 and 1 question with responses valued 0–5. The total score is the sum of responses and ranges from 0–61 with larger values indicating greater desire or arousal.

## Psychosexual Daily Questionnaire

The PDQ is a 7-day diary used in multiple trials to evaluate both the frequency and the intensity of sexual desire and performance across sexual activities.<sup>14</sup> PDQ-Q4 asks participants to list how many of 12 specified activities (sexual daydreams, anticipation of sex, sexual interactions with partner, flirting by you, flirting toward you, orgasm, ejaculation, intercourse, masturbation, night spontaneous erection, day spontaneous erection, and erection in response to sexual activity) they engaged in each day of the week, and the daily count is averaged over 7 days for a total score ranging from 0–12. PDQ question 1 rates the overall sexual desire using a 0–7 numerical rating scale. PDQ question 5 is a self-assessment by the participant of the degree (%) of full erection using multiples of 10 from 0–100%. PDQ question 6 uses a numerical rating scale to assess satisfaction of duration of erection.

## International Index of Erectile Function

The IIEF has 15 items divided into 5 domains: EF, orgasmic function, sexual desire, intercourse satisfaction, and overall satisfaction. Scores from the IIEF and the abridged version can be used to diagnose erectile dysfunction and quantitate the severity.<sup>6,7,10</sup> We focused on the erectile function domain (IIEF-EF, 6 items) and the SDD (2 items) of the IIEF. The scores of each item range from 0–5. The maximum score for erectile function is 30 and for sexual desire is 10.<sup>6,10</sup>

## PGIC–Sexual Desire

Participants were asked at 3, 6, 9, and 12 months whether their sexual desire was “very much worse,” “much worse,” “a little worse,” “no change,” “a little better,” “much better,” or “very much better” since the start of the study. This 7-category response was first collapsed to 5 categories by combining the “very much better” and “much better”/“much worse” and “very much worse” responses since very few participants classified themselves as “very much better.” The PGIC was further reduced to 2 indicators of response: a category of PGIC stringent responders included “very much/much better”; non-responders included “a little better,” “no change,” “a little worse,” “much/ very much worse.” A category of PGIC more inclusive responders additionally included participants who were “a little better” (see Table 1 for the list of acronyms).

## Statistical Analysis

Spearman correlation coefficient as well as polychoric correlation were calculated for all pairwise combinations of each PDQ question, DISF domains, and IIEF domains, and PGIC categories at each month, to assess inter- and intra-domain correlations in 3 domains of sexual function: sexual activity, sexual desire, and erection. As shown in Table 1, the PDQ-Q4 was the sole item in the sexual activity domain; PDQ question 1, DISF-SDD, and SDD of IIEF comprised the SDD; and PDQ question 5, PDQ question 6, DISF-sexual arousal, and IIEF-EF defined the erection domain.

We first calculated the Spearman correlation between the 5-category PGIC and the PDQ-Q4 and DISF-SDD and also the proportion of responders according to PGIC stringent and PGIC inclusive. To identify clinically meaningful change, Sexual Function Trial participants<sup>22</sup> were first split into 2 sub-samples: a training set and a validation set. One-half was randomly assigned to training and the others to the validation sample. T and placebo treatment groups were pooled for the split. Treatment allocation, demographic, and clinical characteristics of the training and validation sets were compared by Student *t* test for continuous variables and the  $\chi^2$  test for categorical variables.

Anchor-based methods, including regression, receiver operating characteristic (ROC) curves, and empirical cumulative distribution function (CDF) analyses<sup>26,27</sup> were used to define the range and magnitude of clinically meaningful changes for the DISF-SDD and PDQ-Q4 in the training set. Distribution-based methods were employed to statistically characterize the magnitude of anchor-based clinically meaningful change. PGIC in sexual desire responses served as the mathematical anchor in this analysis. Linear regression models with the magnitude of change in sexual activity (PDQ-Q4) or sexual desire (DISF-II-SDD) as the outcome variable, and 5-category PGIC as the categorical predictor or independent variable, were used to determine the average difference in change in the respective measure for each PGIC response category during the treatment period. By this method, clinically meaningful change was defined as the average change among participants in PGIC stringent response category or more inclusive response category definitions. The models were adjusted for site and month of treatment. Models were fit by generalized estimating equations with independence working correlation to account for correlation among repeated measures within participant.

For ROC analysis, sensitivity and specificity were calculated for all points along the range of observed responses for the PDQ-Q4 and DISF-SDD. Increments of 0.05 were considered for the PDQ-Q4; increments of 0.1 were used for the DISF-SDD. To account for site and time, sensitivity and specificity were calculated separately for each combination of site and time and then averaged to obtain a single sensitivity and specificity measure for each candidate threshold that was then plotted on a ROC curve. The area under the curve (AUC) for each ROC curve was calculated. ROC-based clinically meaningful change was selected as the threshold that maximized the sum of sensitivity and specificity.

Finally, empirical CDF curves were generated for 5-category PGIC and the dichotomized PGIC stringent and PGIC inclusive. Median changes among categories of responders for PGIC stringent and PGIC inclusive were reported. For distribution-based measures, baseline

SD and reliability for each measure were calculated in the training set to determine the magnitude of change corresponding to medium effects as assessed by an effect size of 0.5 and SEM of 1.96.

Selected measures of clinically meaningful change were then evaluated in the validation set to confirm predictive accuracy and to assess for a clinically meaningful effect of T. In the validation set, the effect of T treatment on anchor- and distribution-based clinically relevant change was evaluated by logistic mixed models with dichotomous clinically meaningful change in the PDQ-Q4 or DISF-II-SDD as the outcome and treatment as the primary predictor. Models included a random intercept for participant and fixed effects for month and balancing factors (site, T = 200, participation in the Physical Function or Vitality Trial of the T Trials, age > 75 years, use of anti-depressants, and use of phosphodiesterase-5 inhibitors). Analyses were performed in software (SAS, Version 9.4; SAS Institute Inc, Cary, NC) and statistical significance was considered at the .05 level.

## RESULTS

### Convergent Validity Within and Across Domains

Spearman correlation coefficients at baseline and at all other time points reveal moderate to strong correlations ( $r > 0.50$  in 64% of correlations) across most measured concepts, although weaker correlations were observed for self-assessment of erectile function, both between and within measures (Table 2).

### Clinically Meaningful Change

The average Spearman correlation of the 5-point PGIC (anchor) with change in the PDQ-Q4 and DISF-SDD across visits was 0.29 and 0.27, respectively; average polyserial correlation, which accounts for the ordinal nature of PGIC, was identical to the Spearman correlation results. For the analyses of clinically meaningful change, men in the training set had a mean age of 71.2 (SD 5.1) years and the majority were Caucasian and overweight (Table 3). Nearly 89% were married or lived with a partner. Mean (SD) PDQ-Q4 response ranged from 1.4–1.8 at baseline and month 12, and mean DISF-SDD response ranged from 11.8–12.7. At each 3-month interval 24 (11.7%) to 29 (13.9%) men were classified as responders (1 or 2, very much or much better) according to PGIC stringent, and 54 (28.1%) to 66 (34.6%) according to PGIC inclusive category. Participants in the validation set did not differ significantly from the training set (Table 3).

### Clinically Meaningful Change Using Regression Analysis in the Training Set

We used the 5 PGIC categories to determine the mean change in the response of PDQ-Q4 and DISF-SDD. Mean change in PDQ-Q4 and DISF-II-SDD scores varied monotonically across 5-category PGIC responses with men who reported very much/ much improved PGIC scores demonstrating the greatest improvements in both PDQ-Q4 and DISF-II-SDD (Figure 1A and B). Responders (includes all participants of the Sexual Function Trial), based on the 5 category PGIC score, reported greater increases in PDQ-Q4 and DISF-II-SDD scores than non-responders (Figure 1C–F). The model-based average difference in change in the PDQ-Q4 was 0.80 (95% CI 0.20–1.40) for participants in the “very/much better” response



category of the PGIC and  $-0.05$  (95% CI  $-0.54-0.43$ ) for participants in the “little better” category (Figure 1A). Similarly, the DISF-II-SDD increased by 5.6 (95% CI 3.3–7.9) and 4.6 (95% CI 2.6–6.7) on average for participants in PGIC “very/much better” and “little better” categories, respectively (Figure 1B).

### Clinically Meaningful Change: ROC Analysis in the Training Set

In ROC curves predicting dichotomized PGIC in sexual function responses as a function of change in PDQ-Q4 and DISF-SDD, optimal candidate thresholds and respective sensitivity and specificity for the PDQ-Q4 and the DISF-SDD are shown in Figure 2. ROC-based clinically meaningful change was 0.6 (sensitivity 0.64, specificity 0.80) for the PDQ-Q4 (Figure 2A) and 2.0 (sensitivity 0.75, specificity 0.54) for the DISF-SDD (Figure 2C) in the training set. AUCs of respective ROC curves were 0.74 and 0.67 for the PDQ-Q4 and DISF-SDD with PGIC stringent as the dichotomized anchor. The same thresholds were selected based on PGIC inclusive (0.6 PDQ-Q4, sensitivity 0.48, specificity 0.84; 2.0 DISF-SDD, sensitivity 0.68, specificity 0.60), with AUCs of 0.67 and 0.66 for PDQ-Q4 and DISF-SDD, respectively (Figure 2B and D). An AUC of 0.70 indicates adequate performance.<sup>28</sup>

### Clinically Meaningful Change: Empirical CDF Analysis in the Training Set

Empirical CDFs for the PDQ-Q4 show that change scores were generally higher among participants who reported being very much/much better, although the maximum response did not fall within that PGIC stringent response category at every time point. Median change in very much/much better respondents was 0.6–1.3 across time points (eg, at month 3 and 12 selected for early and late responses) (Figure 3A and B). For the DISF-SDD, median change among very much/much better respondents was 4.0–6.0. Using the PGIC more inclusive definition of responder, median change among responders was 4.0–5.0 across times. Participants reporting no change generally had lower scores than participants reporting little or much/very much better and higher scores than participants reporting little or much/very much worse. However, at some time points, empirical CDF curves were similar between little and much/very much better or worse participants (Figure 3C and D).

### Test of Clinically Important Changes in the Validation Set

Considering all 3 anchor-based approaches, selected thresholds were 0.6 and 5.0 for the PDQ-Q4 and DISF-SDD, respectively, in the training set. Distribution-based medium-level changes were 0.65 and 1.05 for the PDQ-Q4 using effect size and SEM criteria, respectively, in the training set. The selected threshold is therefore nearly consistent with medium change. For the DISF-SDD, these changes were 3.37 and 4.45, respectively. The selected anchor-based threshold falls between medium and large change sizes (Table 4).

In the validation set, men allocated to receive T were significantly more likely to have an increase in their PDQ-Q4 of 0.6 (Figure 4A) (odds ratio [OR] 3.0, 95% CI 1.7–5.3,  $P < .001$ ) and an increase in their DISF-SDD of 5.0 (Figure 4B) (OR 4.3, 95% CI 2.5–7.4,  $P < .001$ ) when compared with the placebo group. T treatment also resulted in significantly improved outcomes for distribution-based definitions of medium change in sexual function and desire (Table 4).

## DISCUSSION

In this study, we utilize data from the Sexual Function Trial of the T Trials, where 3 instruments were used to assess sexual function: IIEF, DISF-II, and PDQ in symptomatic hypogonadal men. Half of the men were randomly assigned to the training set and the other half to the validation set. Using PGIC sexual desire as an anchor, clinically meaningful change in PDQ-Q4 sexual activity and in DISF-SDD were determined first in the training set and then confirmed in the validation set. Using these cut-points, we showed that the proportions of men with clinically meaningful changes in sexual activity and sexual desire were significantly higher in the T-treated compared to the placebo-treated men, strengthening the previously reported findings of the T Trials.<sup>9</sup>

The reliability of each measure of sexual function was confirmed by the strong correlations in the expected directions for specific sexual activity, sexual desire, and erectile function questions and domains among the 3 questionnaires. Even though sexual activity was measured in 1 instrument (PDQ) only, this parameter showed strong inter-domain correlation with sexual desire, sexual arousal, and erectile function in both the DISF-II-SDD and IIEF. Overall, we demonstrated moderate to strong correlations between measures of the same or similar construct (eg, sexual desire, activity), and lower correlations with conceptually independent constructs (eg, erection ability). Taken together, our findings strongly support the construct validity of our main sexual function measure (PDQ), in addition to supporting our findings regarding the effect of T on sexual function in older men.

Thresholds have been established for the IIEF-EF and can be used to define clinically meaningful change.<sup>6,10,23</sup> A score of <25 in the IIEF-EF indicates erectile dysfunction, and an increase of 4 points is clinically meaningful.<sup>10</sup> The DISF-II-SDD and PDQ-Q4 to date do not have established thresholds for sexual dysfunction or for clinically meaningful changes.<sup>13</sup> To validate and determine the clinically meaningful changes, we utilized the data from the T Trials and PGIC as our anchor to separate responders from non-responders after T replacement. We selected change of 0.6 in PDQ sexual activity score and 5.0 in DISF-II-SDD as anchor-based thresholds considering the regression, ROC, and empirical CDF anchor-based analyses. The magnitude of change from the anchor-based analyses were supported by distribution analyses, by which these changes were classified as medium or medium to large. Using these thresholds, we then demonstrated that the proportions of men with a change in score 0.6 for the PDQ sexual activity and DISF-SDD of 5.0 were significantly higher in the T gel-treated group compared to the placebo-treated group. The results were similar when distribution-based thresholds were used.

Overall, the T Trials showed that all 3 sexual outcome scores were significantly ( $P < .001$ ) improved in the older men treated with T gel compared to the placebo group.<sup>9</sup> Moreover, the improvement in each sexual outcome was correlated with both serum total and free T concentrations as well estradiol.<sup>22</sup> However, except for the IIEF-EF the clinical importance of these outcome measure changes was not known. Using the cutoff for PDQ-Q4 and DISF-II-SDD determined in the training set (randomly selected half of the men in the Sexual Function Trial of the T Trials) of this study, we then reexamined the responses in a

validation set of older men with hypogonadism to T replacement in the T Trial (composed of the other randomly selected half of matched men in the Sexual Function Trial).

In the validation set, men in the T-treated vs placebo-treated group had a greater likelihood of achieving a change of  $\geq 0.6$  in PDQ-Q4, suggesting a greater likelihood of patient-assessed clinically important improvement with T treatment. Similarly, using DISF-II-SDD, the adjusted OR for an increase of  $\geq 5.0$  points was 4.3 (95% CI 2.5–7.4) for men allocated to T treatment compared to men allocated to placebo. Our findings for the magnitude of clinically meaningful improvement in sexual activity and desire may inform end-point selection and interpretation for future studies of older men. Previously reported studies have shown improvement in sexual activity and desire in younger men in response to T treatment.<sup>16,18,19</sup> Future studies in younger men may need to adjust our suggested thresholds to increase either sensitivity or specificity in identifying treatment responders.

The strengths of this study are the large study size and that data were obtained from older men with symptoms of decreased sexual desire and consistently low T levels. 3 questionnaires were used to assess sexual function allowing demonstration of convergent validity within and between domains and instruments. Other studies showing mean improvement scores between 0.9 to about 1.5 in PDQ-Q4 in younger hypogonadal men (average age of about 50 years) after T treatment recruited men who were withdrawn from prior T replacement or who may not have had sexual symptoms,<sup>19</sup> or were without a placebo treated arm for comparisons.<sup>17,18</sup>

Limitations in our analyses include our restrictive population of older men with hypogonadism. Younger hypogonadal men may have higher baseline sexual activity, or different patterns of sexual function or impairment than older men with hypogonadism, though recent studies of placebo-controlled T replacement in men with decreased sexual drive showed improvement in sexual desire in middle-aged men as well as older men.<sup>29</sup> Our use of a single anchor of sexual desire that had modest correlation with both sexual activity and desire measures is also a limitation; a separate anchor for sexual activity may or may not be more accurate, but an anchor of sexual activity was not measured in T Trials. Interestingly, the PGIC sexual desire cut-point had very similar correlations with the PDQ-Q4 and DISF-SDD, despite the different measurement domains. Finally, we performed analyses with data from 1 large trial of T gel treatment using multiple instruments. Therefore, we did not evaluate the generalizability of our results to other hypogonadal men treated with other T replacement modalities (eg, injectable T).

## CONCLUSION

Our study in symptomatic older men with T deficiency demonstrated excellent convergent validity of our instruments, with strong internal consistency and correlation both within and between domains of sexual function. Cutoffs for clinically meaningful change in sexual activity and desire were defined for the PDQ-Q4 ( $\geq 0.6$ ) and DISF-II-SDD ( $\geq 5$ ), respectively. When these cutoffs were applied to the Sexual Function Trial of the T Trial, clinically meaningful improvement was observed for sexual activity and desire in response to T treatment.

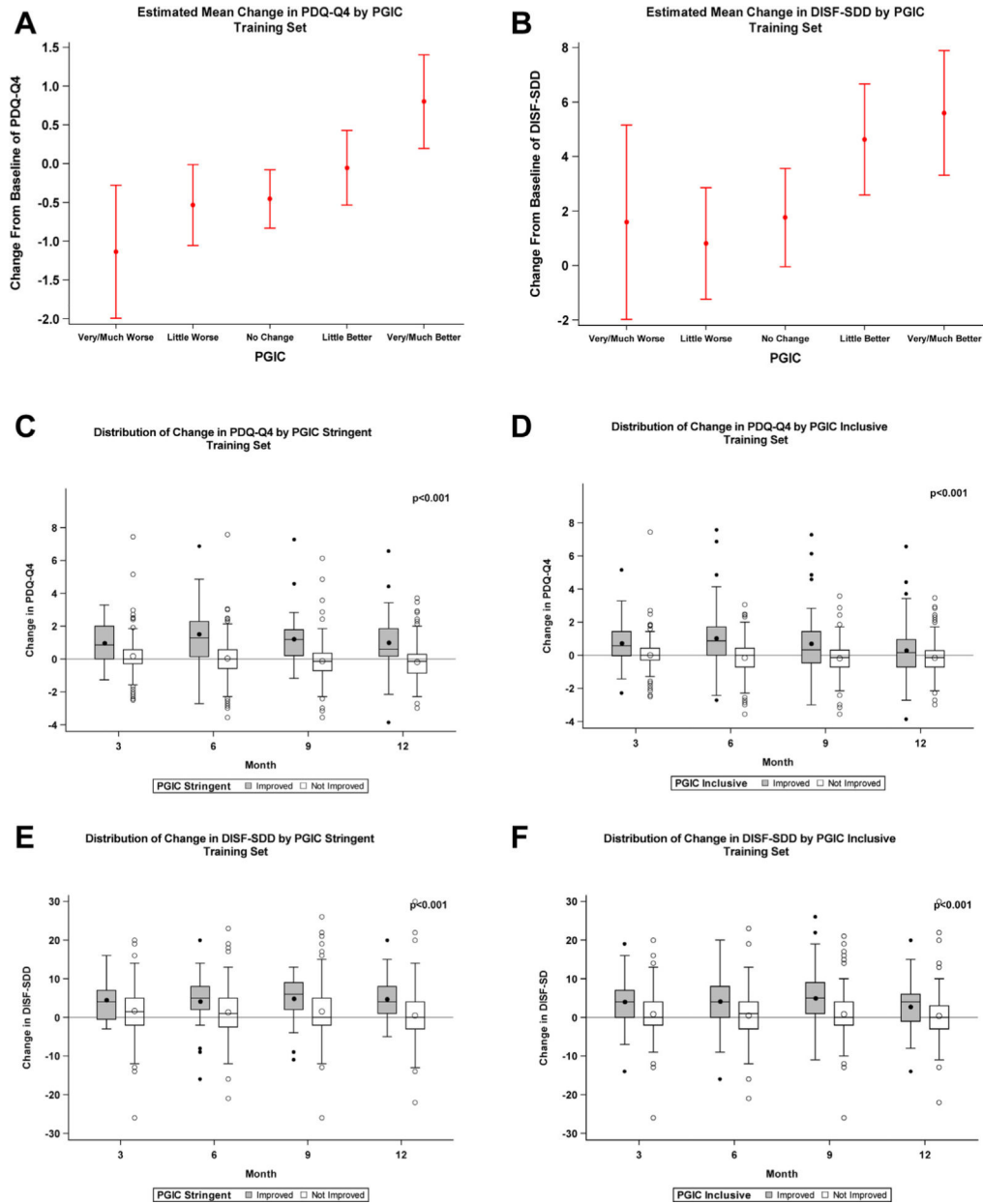
## Acknowledgments

*Funding:* 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 Eunice Kennedy Shriver National Institute of Child Health and Human Development. AbbVie (formerly Solvay and Abbott Laboratories) generously provided funding, AndroGel and placebo gel.

## REFERENCES

1. Araujo AB, Esche GR, Kupelian V, et al. Prevalence of symptomatic androgen deficiency in men. *J Clin Endocrinol Metab* 2007;92:4241–4247. [PubMed: 17698901]
2. Wu FC, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med* 2010;363:123–135. [PubMed: 20554979]
3. Araujo AB, Mohr BA, McKinlay JB. Changes in sexual function in middle-aged and older men: longitudinal data from the Massachusetts Male Aging Study. *J Am Geriatr Soc* 2004; 52:1502–1509. [PubMed: 15341552]
4. O'Connor DB, Corona G, Forti G, et al. Assessment of sexual health in aging men in Europe: development and validation of the European male ageing study sexual function questionnaire. *J Sex Med* 2008;5:1374–1385. [PubMed: 18331267]
5. O'Donnell AB, Araujo AB, Goldstein I, et al. The validity of a single-question self-report of erectile dysfunction. Results from the Massachusetts Male Aging Study. *J Gen Intern Med* 2005;20:515–519. [PubMed: 15987326]
6. Rosen RC, Riley A, Wagner G, et al. The International Index of Erectile Function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology* 1997;49:822–830. [PubMed: 9187685]
7. Rosen RC, Cappelleri JC, Smith MD, et al. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. *Int J Impot Res* 1999;11:319–326. [PubMed: 10637462]
8. Snyder PJ, Ellenberg SS, Cunningham GR, et al. The testosterone trials: seven coordinated trials of testosterone treatment in elderly men. *Clin Trials* 2014;11:362–375. [PubMed: 24686158]
9. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *N Engl J Med* 2016; 374:611–624. [PubMed: 26886521]
10. Rosen RC, Cappelleri JC, Gendrano N, III. The International Index of Erectile Function (IIEF): a state-of-the-science review. *Int J Impot Res* 2002;14:226–244. [PubMed: 12152111]
11. Derogatis LR. The Derogatis Interview for Sexual Functioning (DISF/DISF-SR): an introductory report. *J Sex Marital Ther* 1997;23:291–304. [PubMed: 9427208]
12. Derogatis LR, Melisaratos N. The DSFI: a multidimensional measure of sexual functioning. *J Sex Marital Ther* 1979; 5:244–281. [PubMed: 513144]
13. DeRogatis LR. Assessment of sexual function/dysfunction via patient reported outcomes. *Int J Impot Res* 2008;20:35–44. [PubMed: 17703221]
14. Lee KK, Berman N, Alexander GM, et al. A simple self-report diary for assessing psychosexual function in hypogonadal men. *J Androl* 2003;24:688–698. [PubMed: 12954659]
15. Wang C, Alexander G, Berman N, et al. Testosterone replacement therapy improves mood in hypogonadal men—a clinical research center study. *J Clin Endocrinol Metab* 1996; 81:3578–3583. [PubMed: 8855804]
16. Wang C, Swedloff RS, Iranmanesh A, et al. Transdermal testosterone gel improves sexual function, mood, muscle strength, and body composition parameters in hypogonadal men. Testosterone Gel Study Group. *J Clin Endocrinol Metab* 2000;85:2839–2853. [PubMed: 10946892]
17. Wang C, Cunningham G, Dobs A, et al. Long-term testosterone gel (AndroGel) treatment maintains beneficial effects on sexual function and mood, lean and fat mass, and bone mineral density in hypogonadal men. *J Clin Endocrinol Metab* 2004;89:2085–2098. [PubMed: 15126525]

18. Wang C, Ilani N, Arver S, et al. Efficacy and safety of the 2% formulation of testosterone topical solution applied to the axillae in androgen-deficient men. *Clin Endocrinol (Oxf)* 2011; 75:836–843. [PubMed: 21689131]
19. Steidle C, Schwartz S, Jacoby K, et al. AA2500 testosterone gel normalizes androgen levels in aging males with improvements in body composition and sexual function. *J Clin Endocrinol Metab* 2003;88:2673–2681. [PubMed: 12788872]
20. Gonzalo IT, Swerdloff RS, Nelson AL, et al. Levonorgestrel implants (Norplant II) for male contraception clinical trials: combination with transdermal and injectable testosterone. *J Clin Endocrinol Metab* 2002;87:3562–3572. [PubMed: 12161475]
21. Ilani N, Roth MY, Amory JK, et al. A new combination of testosterone and nesterone transdermal gels for male hormonal contraception. *J Clin Endocrinol Metab* 2012;97:3476–3486. [PubMed: 22791756]
22. Cunningham GR, Stephens-Shields AJ, Rosen RC, et al. Testosterone treatment and sexual function in older men with low testosterone levels. *J Clin Endocrinol Metab* 2016; 101:3096–3104. [PubMed: 27355400]
23. Rosen RC, Allen KR, Ni X, et al. Minimal clinically important differences in the erectile function domain of the International Index of Erectile Function scale. *Eur Urol* 2011;60:1010–1016. [PubMed: 21855209]
24. DeRogatis LR, Graziottin A, Bitzer J, et al. Clinically relevant changes in sexual desire, satisfying sexual activity and personal distress as measured by the profile of female sexual function, sexual activity log, and personal distress scale in postmenopausal women with hypoactive sexual desire disorder. *J Sex Med* 2009;6:175–183. [PubMed: 19170847]
25. Gerstenberger EP, Rosen RC, Brewer JV, et al. Sexual desire and the Female Sexual Function Index (FSFI): a sexual desire cutpoint for clinical interpretation of the FSFI in women with and without hypoactive sexual desire disorder. *J Sex Med* 2010;7:3096–3103. [PubMed: 20561170]
26. Cappelleri JC, Bushmakina AG. Interpretation of patient-reported outcomes. *Stat Methods Med Res* 2014;23:460–483. [PubMed: 23427226]
27. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res* 2018;27:33–40. [PubMed: 28620874]
28. Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Stat Med* 2002;21:3093–3106. [PubMed: 12369084]
29. Brock G, Heiselman D, Maggi M, et al. Effect of testosterone solution 2% on testosterone concentration, sex drive and energy in hypogonadal men: results of a placebo controlled study. *J Urol* 2016;195:699–705. [PubMed: 26498057]



**Figure 1.** Average change in Psychosexual Daily Questionnaire (PDQ) question 4 (Q4) (A) and Derogatis Interview for Sexual Function (DISF)-sexual desire domain (SDD) (B) score in each Patient Global Impression of Change (PGIC) category (very much better or much better, little better, no change, little worse, and much worse and very much worse). Box and whiskers plot of the change in PDQ-Q4 score (sexual activity) in responders, improved (closed circles, shaded boxes) compared to non-responders, not improved (open circles, open boxes) for PGIC stringent and PGIC inclusive (C and D) and change in DISF-SDD (sexual desire) for PGIC stringent and inclusive (E and F). The PGIC had a 7-point response that was reduced to 2 indicators of response: improved PGIC stringent responders included very much better or much better vs not improved responders that included little better, no

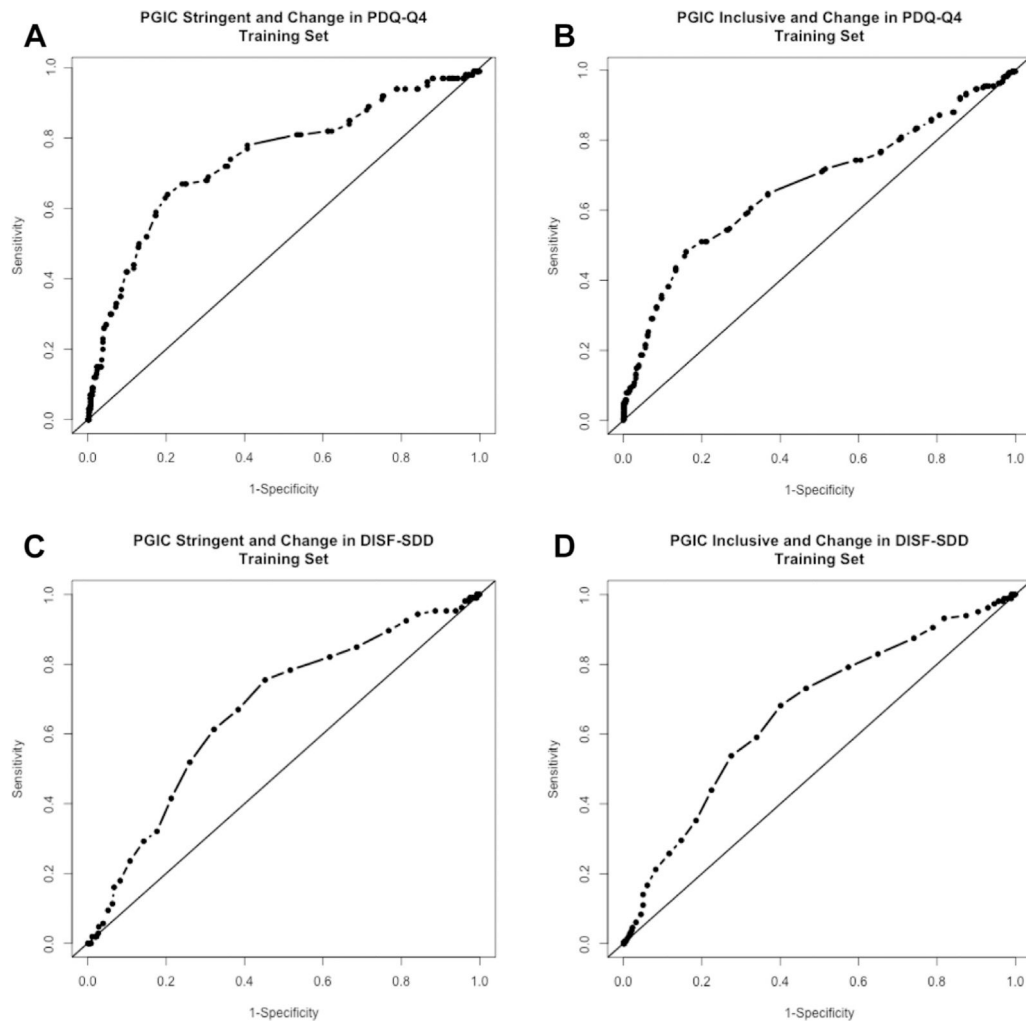
change, little worse, and much worse and very much worse. Improved PGIC more inclusive responders included very much better or much better, or little better, vs not improved responders that included no change, little worse, much worse or very much worse. Figure 1 is available in color online at [www.jsm.jsexmed.org](http://www.jsm.jsexmed.org).

Author Manuscript

Author Manuscript

Author Manuscript

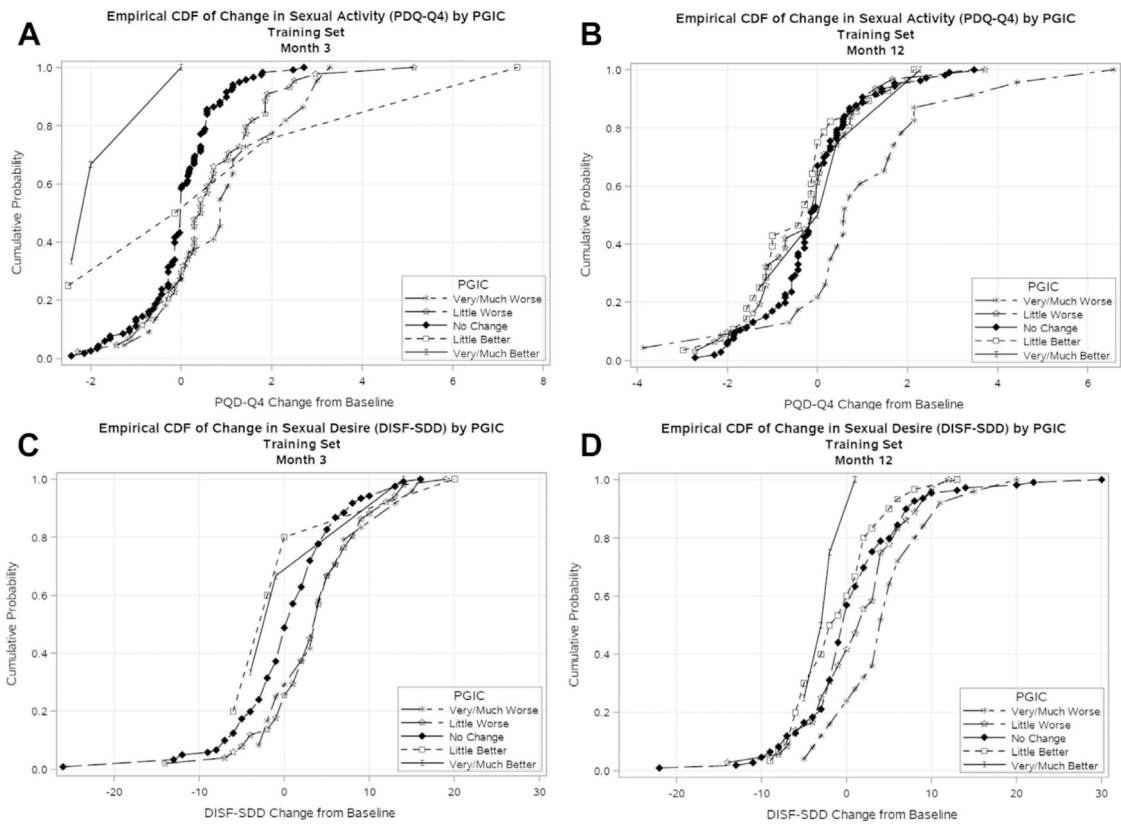
Author Manuscript



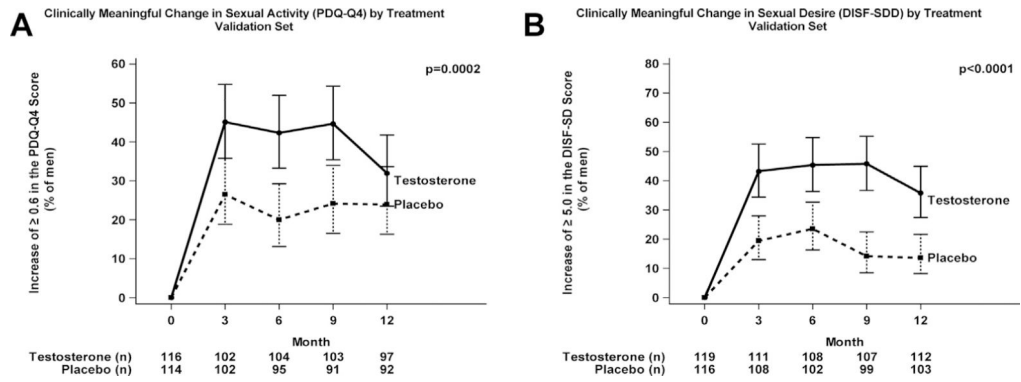
**Figure 2.**

Receiver operator characteristics curves where sensitivity and specificity were calculated for all points along the range of observed responses for the Psychosexual Daily Questionnaire (PDQ) question 4 (Q4) and Derogatis Interview for Sexual Function (DISF)-sexual desire domain (SDD) to predict dichotomized Patient Global Impression of Change (PGIC) in sexual function (stringent and more inclusive) responses as a function of PDQ-Q4 (A and B) and DISF-SDD (C and D). See PGIC stringent and more inclusive definitions in Figure 1.





**Figure 3.** Empirical cumulative distribution function (CDF) curves for each of the 5 Patient Global Impression of Change (PGIC) categories (see PGIC definitions in Figure 1) for the clinical outcome assessment scores from Psychosexual Daily Questionnaire (PDQ) question 4 (Q4) (A and B) and Derogatis Interview for Sexual Function (DISF)-sexual desire domain (SDD) (C and D) at months 3 and 12 of testosterone gel treatment.



**Figure 4.** Percent of men in testosterone (T) gel (solid line) and placebo (dotted line)-treated groups over 12 months in the validation set of men in the T Trials with a clinically meaningful improvement in sexual activity, defined as an increase in Psychosexual Daily Questionnaire (PDQ) question 4 (Q4) score  $\geq 0.6$  (A) and in sexual desire, as defined as an increase Derogatis Interview for Sexual Function-II (DISF)-sexual desire domain (SDD) score  $\geq 5.0$  (B). The number of subjects in each group at each time point is at the bottom of each graph.

Table 1.

## List of acronyms

Complete scale or item title	Abbreviation	Domain
Derogatis Interview for Sexual Function II	DISF-II	
Derogatis Interview for Sexual Function II sexual desire domain	DISF-SDD	Desire
Derogatis Interview for Sexual Function II sexual arousal domain	DISF-SA	Erection
International Index of Erectile Dysfunction	IIEF	
International Index of Erectile Dysfunction—erectile function	IIEF-EF	Erection
International Index of Erectile Dysfunction—sexual desire	IIEF-SD	Desire
Patient Global Impression of Change Sexual Desire	PGIC	
Patient Global Impression of Change Sexual Desire Dichotomization 1 (stringent): Responder defined as very much or much better	PGIC1	Anchor
Patient Global Impression of Change Sexual Desire Dichotomization 2 (inclusive): Responder defined as very much, much, or little better	PGIC2	Anchor
Psychosexual Daily Questionnaire	PDQ	
Psychosexual Daily Questionnaire — Question 1	PDQ-Q1	Desire
Psychosexual Daily Questionnaire — Question 4	PDQ-Q4	Activity
Psychosexual Daily Questionnaire — Question 5	PDQ-Q5	Erection
Psychosexual Daily Questionnaire — Question 6	PDQ-Q6	Erection
Statistical analysis		
Cumulative distribution function	CDF	
Receiver operator characteristic	ROC	

**Table 2.** Spearman correlation matrix of Psychosexual Daily Questionnaire, International Index of Erectile Function, and Derogatis Interview for Sexual Function items

	PDQ			DISF			IIEF		
	Q1 (desire) (467)	Q4 (activity) (459)	Q5: (erection) (275)	Q6: (erection) (275)	SDD (desire) (467)	SA (erection) (466)	EF (erection) (467)	SDD (desire) (467)	SDD (desire) (467)
Q1 (desire)	1.00 (467)	0.73 <.001 (459)	0.16 .0058 (275)	0.31 <.001 (275)	0.67 <.001 (467)	0.60 <.001 (466)	0.61 <.001 (467)	0.68 <.001 (467)	0.68 <.001 (467)
Q4 (activity)	—	1.00 (459)	0.18 .0024 (272)	0.27 <.001 (272)	0.62 <.001 (459)	0.62 <.001 (458)	0.60 <.001 (459)	0.63 <.001 (459)	0.63 <.001 (459)
Q5 (erection)	—	—	1.00 (275)	.52 <.0001 (275)	.22 <.0001 (275)	.37 <.0001 (274)	.31 <.0001 (275)	.20 0.0015 (275)	.20 0.0015 (275)
Q6 (erection)	—	—	—	1.00 (275)	.20 <.0001 (275)	.55 <.0001 (274)	.55 <.0001 (275)	.25 <.0001 (275)	.25 <.0001 (275)
DISF									
SDD (desire)	—	—	—	—	1.00 (470)	.56 <.0001 (469)	.53 <.0001 (470)	.70 <.0001 (470)	.70 <.0001 (470)
SA (erection)	—	—	—	—	—	1.00 (469)	.83 <.0001 (469)	.60 <.0001 (469)	.60 <.0001 (469)
IIEF									
EF (erection)	—	—	—	—	—	—	1.00 (470)	.64 <.0001 (470)	.64 <.0001 (470)
SDD (desire)	—	—	—	—	—	—	—	1.00 (470)	1.00 (470)

Each cell shows the Spearman correlation, *P* value, and number of responders.

The number of responders was lower for PDQ-Q5 and PDQ-Q6 because if there were no erections in PDQ-Q4, PDQ-Q5 and PDQ-Q6 were left blank.

DISF = Derogatis Interview for Sexual Function; EF = erectile function domain; IIEF = International Index of Erectile Function; PDQ = Psychosexual Daily Questionnaire; Q = question; SA = sexual arousal domain; SDD = sexual desire domain.

Table 3.

Characteristics of participants in the training and validation sets

Study sample	Training	Validation	P
N	235	235	
Arm			
Testosterone	115 (49.2%)	119 (50.8%)	.782
Placebo	120(50.8%)	116 (49.2%)	
Age, y	71.2 ± 5.1	72.0 ± 5.5	.087
Race			
Caucasian (%)	206 (87.7%)	201 (85.5%)	.588
African-American (%)	14 (6.0%)	13 (5.5%)	1.0
Other (%)	15 (6.4%)	21 (8.9%)	.386
Ethnicity			
Hispanic (%)	9 (3.8%)	3 (1.3%)	.141
Non-Hispanic (%)	226 (96.2%)	232 (98.7%)	
College graduate (%)	123 (52.3%)	126 (53.6%)	.853
Married or living with partner (%)	209 (88.9%)	211 (89.8%)	.881
BMI, kg/m <sup>2</sup>	31.1 ± 3.4	30.8 ± 3.7	.349
Psychosexual Daily Questionnaire question 4 by month			.083
0	1.5 ± 1.3	1.4 ± 1.2	.202
3	1.8 ± 1.5	1.8 ± 1.7	.720
6	1.8 ± 1.7	1.7 ± 1.6	.629
9	1.6 ± 1.7	1.8 ± 1.7	.337
12	1.4 ± 1.6	1.6 ± 1.8	.303
Derogatis Interview for Sexual Function—sexual desire by month			.115
0	11.8 ± 6.8	11.7 ± 6.5	.945
3	14.0 ± 7.0	13.9 ± 7.1	.896
6	13.6 ± 6.7	14.4 ± 7.2	.226
9	13.7 ± 7.3	14.9 ± 7.7	.111

Study sample	Training		Validation		P	Patient Global Impression of Change by month							
	Very/much better	Little better	Little worse	No change		Little worse	Very/ much worse	Very/much better	Little better	No change	Little worse	Very/ much worse	
12	12.7 ± 6.9	13.5 ± 6.9	.215										
3	24 (10.2%)	51 (21.7%)	123 (52.3%)	5 (2.1%)	3 (1.3%)	29 (12.3%)	58 (24.7%)	110 (46.8%)	6 (2.6%)	1 (0.4%)			
6	29 (12.3%)	41 (17.4%)	117 (49.8%)	19 (8.1%)	3 (1.3%)	34 (14.5%)	42 (17.9%)	109 (46.4%)	8 (3.4%)	8 (3.4%)			
9	28 (11.9%)	30 (12.8%)	124 (52.8%)	15 (6.4%)	6 (2.6%)	39 (16.6%)	40 (17.0%)	96 (40.9%)	14 (6.0%)	5 (2.1%)			
12	25 (10.6%)	36 (15.3%)	109 (46.4%)	30 (12.8%)	4 (1.7%)	27 (11.5%)	42 (17.9%)	100 (42.6%)	16 (6.8%)	9 (3.8%)			

BMI = body mass index.

Table 4.

Effect of testosterone treatment on anchor-based and distribution-based clinically meaningful end points in validation set

Outcome	n	Baseline Mean $\pm$ SD	Month 3 N/total N (%)	Month 6 N/total N (%)	Month 9 N/total N (%)	Month 12 N/total N (%)	Odds ratio (95% CI)	P*
Anchor-based thresholds								
PDQ-Q4 ( 0.6)								
Testosterone	116	1.28 $\pm$ 1.14	45/102 (45.1%)	44/104 (42.3%)	46/103 (44.7%)	31/97 (32.0%)	3.00 (1.69–5.34)	<.001
Placebo	114	1.43 $\pm$ 1.30	27/102 (26.5%)	19/95 (20.0%)	22/91 (24.2%)	22/92 (23.9%)		
DISF-SDD ( 5.0)								
Testosterone	119	11.40 $\pm$ 6.49	48/111 (43.2%)	49/108 (45.4%)	49/107 (45.8%)	40/112 (35.7%)	4.35 (2.55–6.36)	<.001
Placebo	116	12.08 $\pm$ 6.47	21/108 (19.4%)	24/102 (23.5%)	14/99 (14.1%)	14/103 (13.6%)		
Distribution-based thresholds								
PDQ-Q4 ( 0.65) <sup>‡</sup>								
Testosterone	116	1.28 $\pm$ 1.14	46/102 (45.1%)	44/104 (42.3%)	46/103 (44.7%)	31/97 (32.0%)	3.11 (1.75–5.51)	<.001
Placebo	114	1.43 $\pm$ 1.30	27/102 (26.5%)	19/95 (20.0%)	21/91 (23.1%)	22/92 (23.9%)		
PDQ-Q4 ( 1.05) <sup>‡</sup>								
Testosterone	116	1.28 $\pm$ 1.14	33/102 (32.4%)	32/104 (30.8%)	36/103 (35.0%)	23/97 (23.7%)	3.27 (1.77–6.04)	<.001
Placebo	114	1.43 $\pm$ 1.30	17/102 (16.7%)	13/95 (13.7%)	15/91 (16.5%)	16/92 (17.4%)		
DISF-SDD ( 3.37) <sup>‡</sup>								
Testosterone	119	11.40 $\pm$ 6.49	55/111 (49.5%)	55/108 (50.9%)	59/107 (55.1%)	44/112 (39.3%)	4.16 (2.52–6.86)	<.001
Placebo	116	12.08 $\pm$ 6.47	26/108 (24.1%)	31/102 (30.4%)	16/99 (16.2%)	21/103 (20.4%)		
DISF-SDD ( 4.45) <sup>‡</sup>								
Testosterone	119	11.40 $\pm$ 6.49	48/111 (43.2%)	49/108 (45.4%)	49/107 (45.8%)	40/112 (35.7%)	4.42 (2.59–7.56)	<.001
Placebo	116	12.08 $\pm$ 6.47	21/108 (19.4%)	24/102 (23.5%)	14/99 (14.1%)	14/103 (13.6%)		

DISF = Derogatis Interview for Sexual Function; PDQ = Psychosexual Daily Questionnaire; Q = question; SDD = sexual desire domain.

\* P-value was determined by a logistic mixed model with random intercepts for participants, adjusted for age ( 75 y), baseline total testosterone ( 200 ng/dL), site, primary efficacy trial involvement, use of anti-depressants, and use of phosphodiesterase-5 inhibitors.

<sup>‡</sup>This threshold corresponds to a moderate effect, defined by 0.5 baseline SD units.<sup>‡</sup>This threshold corresponds to a moderate effect, defined by 1.96 SEM units.