HSR Health Services Research

DOI: 10.1111/j.1475-6773.2012.01395.x
RESEARCH ARTICLE

Comparison of Distribution- and Anchor-Based Approaches to Infer Changes in Health-Related Quality of Life of Prostate Cancer Survivors

Ravishankar Jayadevappa, Stanley Bruce Malkowicz, Marsha Wittink, Alan J. Wein, and Sumedha Chhatre

Objective. To determine the minimal important difference (MID) in generic and prostate-specific health-related quality of life (HRQoL) using distribution- and anchor-based methods.

Study Design and Setting. Prospective cohort study of 602 newly diagnosed prostate cancer patients recruited from an urban academic hospital and a Veterans Administration hospital. Participants completed generic (SF-36) and prostate-specific HRQoL surveys at baseline and at 3, 6, 12, and 24 months posttreatment. Anchor-based and distribution-based methods were used to develop MID estimates. We compared the proportion of participants returning to baseline based on MID estimates from the two methods.

Results. MID estimates derived from combining distribution- and anchor-based methods for the SF-36 subscales are physical function = 7, role physical = 14, role emotional = 12, vitality = 9, mental health = 6, social function = 9, bodily pain = 9, and general health = 8; and for the prostate-specific scales are urinary function = 8, bowel function = 7, sexual function = 8, urinary bother = 9, bowel bother = 8, and sexual bother = 11. Proportions of participants returning to baseline values corresponding to MID estimates from the two methods were comparable.

Conclusions. This is the first study to assess the MID for generic and prostate-specific HRQoL using anchor-based and distribution-based methods. Although variation exists in the MID estimates derived from these two methods, the recovery patterns corresponding to these estimates were comparable.

Key Words. Prostate cancer, health-related quality of life, minimal important difference, anchor based, distribution based

In the era of comparative effectiveness, health status measures are assuming importance in the arena of patient-reported outcomes. Health-related quality of life (HRQoL) is an important component of patient-reported outcomes

that captures the patient's perspective and plays a vital role in decisions made during all phases of prostate cancer care, from screening to treatment choice. However, to realize this potential, we need to interpret the relevance of HRQoL outcomes for decisions about treatment. Thus, a major challenge in the field of patient-reported outcomes has to do with measuring and interpreting minimal important difference (MID). The purpose of this paper was twofold: first, we sought to summarize the published evidence regarding generic and disease-specific MID. We then developed estimates of generic and disease-specific MID using two methods in the context of prostate cancer.

BACKGROUND

The interpretation of changes in scores of patient-reported outcomes, such as HRQoL, is a challenging task and can hamper the integration of patient-reported outcomes in clinical and policy decisions (Jeschke, Singer, and Guyatt 1989; Juniper et al. 1994; van Walraven et al. 1999; Norman et al. 2001; Crosby, Kolotkin, and Williams 2003; Barrett et al. 2005). MID is defined as the smallest difference in score of a domain that patients perceive as a change and that would mandate, in the absence of troublesome side effects and excessive cost, modification in a patient's management (Jeschke, Singer, and Guyatt 1989). There is a lack of consistency in conceptualization and interpretation of MID (Jeschke, Singer, and Guyatt 1989; Juniper et al. 1994; Redelmeier, Guyatt, and Goldstein 1996; van Walraven et al. 1999; Norman et al. 2001; Beaton, Boers, and Wells 2002; Crosby, Kolotkin, and Williams 2003; Barrett et al. 2005; Wyrwich et al. 2005; Coeytaux et al. 2006; Turner et al. 2010). For a patient, a meaningful

Address correspondence to Ravishankar Jayadevappa, Ph.D., Department of Medicine, Perelman School of Medicine, University of Pennsylvania, 224, Ralston-Penn Center, 3615 Chestnut Street, Philadelphia, PA 19104-2676; e-mail: jravi@mail.med.upenn.edu. S. Bruce Malkowicz is with the Division of Urology, Department of Surgery, Perelman School of Medicine, Perelman Center for Advanced Medicine, University of Pennsylvania, Philadelphia Veterans Administration Medical Center, Philadelphia, PA. Marsha Wittink, M.D., M.B.E., is with the Department of Psychiatry, School of Medicine and Dentistry, University of Rochester Medical Center, Rochester, NY. Alan J. Wein, M.D., is with the Division of Urology, Department of Surgery,Perelman Center for Advanced Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA. Sumedha Chhatre, Ph.D., is with the Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

change in HRQoL may be one that reduces symptoms or improves function. For a clinician, a meaningful change may be one that signals change in the therapeutic treatment or disease prognosis. The societal perspective of such meaningful change takes into account population level with small differences, whereas from an institutional perspective, the focus is on the degree of change required to influence health care policies (Redelmeier, Guyatt, and Goldstein 1996; van Walraven et al. 1999; Beaton, Boers, and Wells 2002; Wyrwich et al. 2005; Coeytaux et al. 2006; Jordan et al. 2006).

The two methods of estimating MID, distribution and anchor based, are conceptually very different (Jeschke, Singer, and Guyatt 1989; Juniper et al. 1994; van Walraven et al. 1999; Norman et al. 2001; Crosby, Kolotkin, and Williams 2003; Barrett et al. 2005). The distribution approach to MID uses the effect size of the difference between groups to measure variability, standardized response mean, standard error of measurement, and responsiveness statistics (Anastasi and Urbina 1997; Bonniaud et al. 2008). Most common distribution-based methods are (a) standard error measurement, (b) one-half standard deviation estimate, and (c) one-third standard deviation estimate. Typically, it is argued that effect sizes used in distribution method are based entirely on standard deviation (SD), which in itself is sample specific. In contrast, most anchor-based approaches do not consider the measurement precision of the instrument but are instead based on external criteria like retrospective judgment of change, and are thus presumed to be sample-independent (Jeschke, Singer, and Guyatt 1989; Juniper et al. 1994; Redelmeier, Guyatt, and Goldstein 1996; van Walraven et al. 1999; Norman et al. 2001; Beaton, Boers, and Wells 2002; Crosby, Kolotkin, and Williams 2003; Barrett et al. 2005; Wyrwich et al. 2005; Coeytaux et al. 2006). The anchor-based method examines the relationship between an anchor (cross-sectional or longitudinal) as an independent measure and an HRQoL measure to elucidate the meaning of a particular degree of change. To establish MID, the individual patient-focused strategy uses a single anchor, while the population-focused strategy requires multiple anchors. Distribution methods alone do not provide information regarding the clinical relevance of the observed change. Also, there are few agreed-upon benchmarks to establish MID using distribution methods. On the other hand, the measurement precision of the instrument is not integrated in the MID estimates derived from anchor-based methods. Hence, a combination of distribution- and anchor-based methods is needed for

determining MID (Cella et al. 2002b; Crosby, Kolotkin, and Williams 2003, 2004; Eton et al. 2004; Wyrwich 2004; de Vet et al. 2007; Swigris et al. 2010; Yost et al. 2011).

To gain an insight in the HRQoL differences between disease stages, MID must be measured using generic as well as disease-specific instruments. Table 1 summarizes the MID (anchor- and distribution-based approaches) for generic HRQoL measures for various disease conditions. These measures demonstrate some variation in MID estimates by population, disease, and contextual characteristics. It is thus crucial to compute MID for various patient-reported outcome instruments to ascertain the range of values to use as a basis for sample size calculation for studies in the arena of comparative effectiveness. Unlike generic measures, disease-specific HRQoL measures are primarily designed for particular health conditions. Table 2 summarizes the MID and effect sizes for disease-specific HRQoL instruments using anchor and distribution methods. The results show convergence between distribution- and anchor-based estimates of MID and demonstrate that understanding MID can facilitate interpretation of disease-specific HRQoL measures.

Prostate Cancer and HRQoL

Prostate cancer is the leading cancer diagnosed among men in the United States (Siegal et al. 2011). In the absence of conclusive evidence regarding superiority of any one treatment for prostate cancer, measurement of patient-reported outcomes such as HRQoL remains important for treatment effectiveness evaluation, clinical decisions, and patient education (Jeschke, Singer, and Guyatt 1989; Quek and Penson 2005; Siegal et al. 2011). In order for researchers and clinicians to accurately plan, conduct, and interpret the results of prostate cancer studies, it is necessary to know which changes are clinically significant. In this study, we sought to characterize and compare the MID of generic and prostate-specific HRQoL using anchor-based and distribution-based approaches in the course of 24 months of follow-up of newly diagnosed prostate cancer patients. Further, we evaluated the recovery patterns corresponding to MID estimates derived from these two methods. We hypothesized that both anchor-based and distribution-based methods will lead to comparable conclusions regarding recovery patterns of generic and prostate-specific HRQoL in prostate cancer patients.

Table 1: Summary of Established Minimal Important Difference for Generic Measures

Instrument (No. Items)	*Disease Group*	*MID Approach*	*Effect Size (Small = 0.2–0.5; Moderate = 0.5–0.8; Large ≥ 0.8)*	*MID Values*	*Reference*
FACT-G	FACT-An-cancer patients	Anchor and distribution	0.25–0.67	4	Cella et al. (2002b)
FACT-G	FACT-B breast cancer	Anchor and distribution	0.31–0.54	5–6	Eton et al. (2004)
SF-12-PCS		Anchor based	Not reported	6.8	Schmitt and De Fabio (2004)
FACT-head and neck		Anchor based	Not reported	4–8	Ringash et al. (2004, 2007)
SF-6D EQ-5D	Multiple patient groups	Anchor and distribution	0.12–0.97 0.04–0.44		Walters and Brazier (2005)
SF-6D	Spinal cord	Anchor based	0.23 and 0.86	0.03–0.10	Lee et al. (2008)
SF-36 and SGRQ		Anchor and distribution	Not reported	SF-36 = 2–4 SGRQ = 5–8	Swigris et al. (2010)

Table 2: Summary of Established Minimal Important Differences for Disease-Specific Measures

Instrument and Disease Group	*MID Approach*	*Effect Size (Small = 0.2–0.5; Moderate = 0.5–0.8; Large ≥ 0.8)*	*MID Values*	*Reference*
EORTC cancer	Anchor	0.73–0.86	6.9–10.7	Osoba et al. (1998)
MacNew heart disease HRQoL	Distribution	0.50	7	Dixon, Lim, and Oldridge (2002)
FACT-L lung cancer	Anchor and distribution	LCS = 0.57 TOI = 0.59	2–3 5–7	Cella et al. (2002a)
FACT-anemia and fatigue	Anchor and distribution	Fatigue scale = 0.29–0.8 FACT-An = 0.29–0.59 TOI-fatigue = 0.31–0.71 TOI-anemia = 0.31–0.55	3 7 5 6	Cella et al. (2002b)
Transition dyspnea index	Anchor-based approach	Not reported	One-unit change	Witek and Mahler (2003)
King's health questionnaire overactive bladder	Anchor-based methods	0.2–0.5	5–10	Kelleher et al. (2004)
FACT-B breast cancer	Anchor and distribution	BCS = 0.22–0.50 TOI-PFB = 0.23–0.56 FACT-B = 0.25–0.57	2–3 5–6 7–8	Eton et al. (2004)
DASH	Anchor	1.06–1.67	12.6	Schmitt and De Fabio (2004)
FACT-C	Anchor and distribution	CCS = 0.19–0.82 TOI-C = 0.22–1.04 FACT-C = 0.28–1.15	2–3 4–6 5–8	Yost et al. (2005)
DLQI-chronic idiopathic urticaris	Anchor and distribution	Not reported	2.24–3.10	Shikiar et al. (2005)
I-QOL incontinence	Anchor	0.2–0.5	2.3–6.3	Yalcin et al. (2006)

continued

Table 2. *Continued*

Instrument and Disease Group	*MID Approach*	*Effect Size (Small = 0.2–0.5; Moderate = 0.5–0.8; Large ≥ 0.8)*	*MID Values*	*Reference*
FACT-head and neck	Anchor-based approach	0.2–0.3	6–12 4–8	Ringash et al. (2004)
HeRQoLEDv2 eating disorders	Anchor and distribution	0.30	0.7–14.5	Hayas et al. (2007)
HADS for patients with COPD	Anchor and distribution	0.50	1.5	Puhan et al. (2008)
Qualiveen, urinary disorder	Anchor-based method	0.50 (0.36–0.72)	Not reported	Bonniaud et al. (2008)
SF-36 and SGRQ	Anchor and distribution	Not reported	SF-36 = 2–4 SGRQ = 5–8	Swigris et al. (2010)
LCQ	Anchor-based method	0.2–0.5	1.3	Raj, Pavord, and Birring (2009)
FAST Flushing symptom	Anchor and distribution	Not reported	0.29–0.38 0.66–0.86	Kawata et al. (2008)

METHODS

Data Source

We used data from our prospective cohort study of newly diagnosed prostate cancer patients. Participants were recruited from an urban academic hospital and a VA medical center between 2002 and 2006. The study was approved by the local institutional review boards. All study personnel completed human subject protection training and met the Health Information Portability and Accountability Act (HIPAA) requirements before engaging in this research. Study inclusion criteria were self-identified African American or Caucasian men of age $\geq$ 45 years at the time of diagnosis, newly diagnosed for prostate cancer in the prior 4 months, and yet to initiate treatment. Diagnosis of prostate cancer was based on prostate-specific antigen level, prostate biopsy, and staging. Patients were excluded if they were diagnosed with metastatic cancer (10 percent), had visited the urology clinics for a second opinion only (18 percent), were unable to communicate in English (0 percent), or were cognitively impaired (<1 percent).

Subject Selection and Recruitment

Potential participants received study information from their urologists during clinic visits or during the weekly prostatectomy classes. Study research assistants then contacted those who had expressed an interest in participating. Enrolled participants provided written informed consent and HIPAA consent prior to data collection. The baseline data were obtained prior to treatment initiation, and the mean time between diagnosis of prostate cancer and baseline survey was 12 days with a median value of 10 days.

Retention plan and follow-up: After providing baseline data, participants received follow-up surveys via mail at 3, 6, 12, and 24 months. Nonrespondents were contacted via telephone after 10 days, and a second mailing was sent to them within 4 weeks of the first mailing. Of the 602 participants who completed baseline assessment, 517 completed 3-month follow-up assessment, 546 completed 6-month follow-up assessment, and 528 completed 12-month follow-up assessment. The attrition was highest (14.11 percent) at 3-month follow-up. Reasons for attrition were nonresponse to two consecutive mailings and a follow-up telephone contact (11.05 percent), withdrawal from study (2 percent), and incorrect mailing address (1.06 percent). Comparable attrition pattern was observed at other follow-up intervals.

Outcome Measures and Data Collection

Participants provided self-reported information on ethnicity, education, marital status, living arrangement, and income. Structured medical chart review was used to obtain data on patient age, date of prostate cancer diagnosis, health insurance, treatment, prostate-specific antigen level, Gleason score, tumor, lymph nodes, and metastasis (TNM) stage of cancer, and comorbidity. Primary curative treatment for prostate cancer was radical prostatectomy or external beam radiation therapy.

Participants completed self-administered surveys at baseline and at 3, 6, 12, and 24 months of follow-up. Prostate-specific HRQoL was assessed using the UCLA prostate cancer index (PCI), a comprehensive self-administered 20-item questionnaire that quantifies prostate-specific HRQoL in six domains (Litwin et al. 1998). The PCI has performed well in the general population, has demonstrated good psychometric properties with an internal consistency reliability ranging from 0.82 to 0.94, and is easy to understand and complete (Litwin et al. 1998). Generic HRQoL was assessed by the Medical Outcome Study Short Form (SF-36), a single multi-item scale that assesses eight health concepts (Ware and Sherbourne 1992). The SF-36 has exhibited high psychometric properties with internal consistency reliability ranging from 0.80 to 0.93 (Ware and Sherbourne 1992). Data on following eight patient-reported physical signs and symptoms were obtained: (a) difficulty or discomfort urinating (passing water), (b) having to urinate too often (frequent urination), (c) weak urinary stream, (d) infection of bladder or prostate, (e) blood in urine, (f) pains or aches in back, hips, or legs, (g) more tired or worn out than usual, and (h) other physical symptoms. These dichotomous items yield a homogeneous estimate of physical signs and symptoms and are extensively studied in the arena of prostate cancer care. Baseline Charlson comorbidity index score was computed using ICD 9 codes for inpatient and outpatient events in the 6 months prior to prostate cancer diagnosis (Charlson et al. 1987). These data were obtained from Pennsylvania Integrated Clinical and Research Database for the non-VA patients and from Patient Treatment File and the Outpatient Care Files for the VA patients (Barnett 1999).

Statistical Analysis

We used two distribution-based and two anchor-based methods to derive MID estimates in prostate cancer patients. In the first distribution method, SD

of generic and prostate-specific HRQoL scores and change scores were reported as 1/2 SD estimate and 1/3 SD estimate (Eton et al. 2004). In the second distribution method, we used baseline values of the HRQoL outcome measures to calculate means, SDs, internal consistency reliability (Cronbach's α), and standard error of measurement (Cohen 1988; Allison 1999; Taris 2000). Although the definition of standard error measurement as $\sigma_x (1 - \Gamma_{xx})^{1/2}$ is well accepted, multiple methods are available to estimate reliability (Γ_{xx}) and variability (σ_x). To be consistent with prior research, we used Cronbach's α as the measure of reliability and SD as the measure of variability. A single unit change in the standard error measurement was the MID estimate for each of the eight scales of generic HRQoL (SF-36) and six scales of prostate-specific HRQoL (PCI) scores.

For the anchor-based method, we derived MID estimates using one cross-sectional anchor and one longitudinal anchor. Our cross-sectional anchor was baseline global health derived from the SF-36 question "Compared to one year ago, how would you rate your health in general now?" The longitudinal anchor was one of the patient-reported physical signs/symptoms: "more tired or worn out than usual." Based on the responses to this question at baseline and at 12 months, three categories of change were created: no change in tiredness, improved (i.e., less tired), and not improved (i.e., more tired). We observed product-moment correlation coefficients between our anchors (longitudinal and cross-sectional) and generic and prostate-specific HRQoL items. We then fitted linear regression models to examine the relationship between anchors and HRQoL subscales (Cohen 1988; Allison 1999; Taris 2000). For the cross-sectional anchor of global health, we examined the relationship between baseline scores on global health (independent variable) and baseline generic and prostate-specific HRQoL scores (dependent variables). For the longitudinal anchor, we examined the relationship between the change in tiredness and change in generic and prostate-specific HRQoL scores. The three categories of change in tiredness were operationalized as three dummy variables and two were entered in the regression as independent variables. Point estimates (for cross-sectional anchor) and average of point estimates (for longitudinal anchor) of the regressions yielded the MID estimates.

Finally, we explored the recovery pattern of each generic and prostate-specific HRQoL subscale in terms of "return to baseline." A change greater than or equal to the MID is considered "return to baseline." We compared the proportion of patients returning to baseline using MID values corresponding to the distribution method and anchor method (Cohen 1988).

Missing data were minimal, spread evenly within and between instruments, and were handled using simple mean statistical imputation method (Oakes 1990; Engels and Diehr 2003). Analyses were performed using *SAS* version 9.2 (SAS Institute., Cary, NC, USA), and statistical significance was set at $p \leq .05$.

RESULTS

Demographic and clinical characteristics of the study population are presented in Table 3. Baseline generic and prostate-specific (SF-36 and PCI) HRQoL values are presented in Table 4. Mean baseline scores on generic HRQoL scales of physical function (mean = 62.6, CI = 60.7, 64.5), vitality (mean = 65.0, CI = 63.0, 67.0), and general health (mean = 67.3, CI = 65.3, 69.3) were lower compared with other items. Among prostate-specific HRQoL subscales, sexual function (mean = 51.8, CI = 49.2, 54.5) and sexual bother (mean = 61.3, CI = 57.8, 64.8) scores were lower compared with other scales. Each subscale of generic and prostate-specific HRQoL exhibited robust internal consistency reliability (Cronbach's α).

Distribution-Based Approach

Minimal important difference estimates from the distribution analyses are reported in Table 5. The 1/2 SD, 1/3 SD, and 1 SEM based MID estimates for generic and prostate-specific HRQoL subscales are presented for baseline, 12-month follow-up, and change from baseline to 12-month follow-up scores. The mean values of the 1/2 SD estimates for the generic SF-36 subscales where the total score can range between 0 and 100 points are physical function = 10.0 (10 items), role physical = 19 (four items), role emotional = 18.2 (three items), vitality = 10.9 (four items), mental health = 8.4 (five items), social function = 11.8 (two items), bodily pain = 11.4 (two items), and general health = 10.5 (five items). Mean values for the prostate-specific scales are urinary function = 11.8 (five items), bowel function = 7.9 (four items), sexual function = 13.5 (eight items), urinary bother = 13.6 (one item), bowel bother = 11.7 (one item), and sexual bother = 19.6 (one item). The last column of Table 5 presents the mean values for the 1/2 SD, 1/3 SD, and 1 SEM based estimates for generic and prostate-specific subscales. These estimates will be refined using anchor-based approach.

Table 3: Baseline Patient Characteristics ($n = 602$)

Demographic and Clinical Variables	
Age (mean ± SD)	63.3 ± 8.0
Race (%)	
African American	32.26
Caucasian	67.74
Education (%)	
College or more	64.80
High School or less	35.20
Marital status (%)	
Single/widowed/divorced	27.47
Married	72.53
Employment status (%)	
Part-time/other	61.99
Full-time	38.01
Income level (%)	
>$40,000	60.30
$40,000	39.70
Signs and symptoms (%)	
Difficulty or discomfort urinating	21.62
Having to urinate too often	45.53
Weak urinary stream	34.76
Infection of bladder or prostate	8.42
Blood in urine	7.55
Pain or aches in back, hips, or legs	29.8
More tired or worn out than usual	25.36
Charlson comorbidity (mean ± SD)	1.3 ± 2.4
Prostate-specific antigen at diagnosis (ng/ml) (mean ± SD)	7.6 ± 8.0
Gleason score (total) (mean ± SD)	6.3 ± 0.9
TNM stage (%)	
T1a–T1c	68.32
T2a	15.08
T2b	2.57
T2c	3.31
T3a	7.35
T3b	3.31
Treatment type (%)	
Radical prostatectomy	61.70
External beam radiation therapy	32.30
Hormonal therapy	12.09
Watchful waiting	2.27

Anchor-Based Approach

To identify appropriate cross-sectional and longitudinal anchors, we examined the strength of correlations between several candidate anchors (D'Amico

Table 4: Baseline Generic and Prostate-Specific (SF-36 and Prostate Cancer Index, PCI) Scores

Scale	*Mean*	*SD*	*Confidence Interval*		*Internal Consistency Reliability**
Generic HRQoL-SF-36					
Physical function	62.6	21.8	60.7	64.5	0.84
Role physical	74.7	39.0	71.2	78.2	0.84
Role emotional	76.7	37.9	73.4	80.1	0.85
Vitality	65.0	22.7	63.0	67.0	0.84
Mental health	76.4	18.2	74.8	78.0	0.85
Social function	82.6	24.0	80.5	84.7	0.84
Bodily pain	81.7	24.4	79.6	83.9	0.84
General health	67.3	22.6	65.3	69.3	0.84
PCI					
Urinary function	89.2	18.5	87.5	90.8	0.85
Bowel function	87.5	14.7	86.2	88.8	0.85
Sexual function	51.8	30.0	49.2	54.5	0.85
Urinary bother	85.3	23.3	83.2	87.4	0.85
Bowel bother	88.7	21.3	86.8	90.6	0.85
Sexual bother	61.3	38.5	57.8	64.8	0.85

*Cronbach's α.

risk criteria, bodily pain, TNM stage, global health, various patients reported signs, and symptoms) and generic and prostate-specific HRQoL subscales. Of these, global health and sign and symptoms question about tiredness (more tired or worn out than usual) were observed to have the strongest correlations with generic and prostate-specific HRQoL subscales (range 0.23–0.62). We also compared the magnitude and direction of the correlation between baseline-HRQoL scores and the longitudinal anchor of tiredness with that between 12-month HRQoL scores and the longitudinal anchor. The magnitude of these two sets of correlations for all generic and prostate-specific HRQoL subscales was comparable (range 0.18–0.31). Except for the subscales of general health, sexual function, and sexual bother, the direction of the two sets of correlation for other subscales was opposite (range −0.13 to −0.28), indicating the transitional nature of the anchor. We thus selected global health as the cross-sectional anchor and change in tiredness as our longitudinal anchor. The results of the anchor-based analyses are presented in Table 6. Table 7 shows mean MID values of the distribution method, mean MID values of the anchor method, and the grand mean from both methods. The grand mean values for the generic SF-36 scales are physical function = 7, role physical = 14, role emotional = 12, vitality = 9, mental health = 6, social

Table 5: Distribution-Based Minimally Important Differences for Generic SF-36 and Prostate Cancer Index (PCI)

	Baseline			*12 Months*			*Baseline–12 Months*			*Mean*		
Scale	*1/2 SD*	*1/3 SD*	*1 SEM*	*1/2 SD*	*1/3 SD*	*1 SEM*	*1/2 SD*	*1/3 SD*	*1 SEM*	*1/2 SD*	*1/3 SD*	*1 SEM*
Generic HRQoL-SF 36												
Physical function	10.9	7.3	8.7	11.2	7.5	7.7	7.9	5.3	8.7	10.0	6.7	8.4
Role physical	19.5	13	15.6	20.2	13.5	14.3	17.3	11.6	19.6	19.0	12.7	16.5
Role emotional	18.9	12.6	14.7	17.6	11.8	12.2	18.1	12.1	19.8	18.2	12.2	15.5
Vitality	11.3	7.6	9.1	12.1	8.1	8.4	9.4	6.3	10.5	10.9	7.3	9.5
Mental health	9.1	6.1	7.1	8.3	5.5	5.7	7.7	5.2	8.5	8.4	5.6	7.1
Social function	12	8.0	9.6	12.8	8.5	8.9	10.7	7.1	12.1	11.8	7.8	10.2
Bodily pain	12.2	8.1	9.8	12.4	8.3	8.6	9.7	6.5	10.9	11.4	7.6	9.8
General health	11.3	7.5	9.1	12.3	8.2	8.6	7.9	5.3	8.7	10.5	7.0	8.8
PCI												
Urinary function	9.2	6.2	7.2	12.7	8.5	8.5	13.6	9.1	15.4	11.8	7.9	10.4
Bowel function	7.3	4.9	5.7	8.3	5.6	5.5	8.3	5.6	9.1	7.9	5.2	6.8
Sexual function	15	10.0	11.6	12.3	8.2	8.2	13.1	8.8	14.4	13.5	9.2	11.4
Urinary bother	11.6	7.8	9.1	14.5	9.7	9.6	14.8	9.9	16.8	13.6	9.1	11.8
Bowel bother	10.6	7.1	8.2	12.5	8.3	8.7	12.1	8.10	13.5	11.7	7.8	10.1
Sexual bother	19.2	12.8	14.9	18.1	12.1	11.4	21.5	14.4	13.2	19.6	13.1	13.2

Table 6: Anchor-Based Minimally Important Difference Estimates for Generic SF-36 and Prostate Cancer Index

	Cross-Sectional Anchor			*Longitudinal Anchor*		
Scale	*General Health*	*95% Confidence Interval*	*p Value*	*More Tired or Worn Out Than Usual*	*95% Confidence Interval*	*p Value*
Physical function	5.46	2.93–7.97	<.0001	5.27	5.10–6.75	.0409
Role physical	12.3	7.81–16.78	<.0001	12.56	2.58–23.49	.0007
Role emotional	5.06	1.62–9.49	.0255	14.61	9.46–18.02	.0076
Vitality	6.94	4.34–9.55	<.0001	11.36	1.46–13.25	<.0001
Mental health	3.93	1.82–6.04	.0003	6.04	1.97–10.05	.0003
Social function	5.92	3.13–8.70	<.0001	9.96	3.85–11.92	.0007
Bodily pain	8.96	6.21–11.71	<.0001	6.74	6.32–8.29	.0295
General health	6.90	4.35–9.46	<.0001	6.71	1.89–12.52	<.0001
Urinary function	7.82	1.64–4.98	.0112	5.94	4.35–16.22	.0193
Bowel function	4.73	3.02–6.45	<.0001	6.60	5.48–7.19	.0492
Sexual function	3.78	1.20–7.36	.0383	5.51	1.03–10.26	.0495
Urinary bother	6.65	3.94–9.35	<.0001	4.66	1.57–15.9	.0282
Bowel bother	5.23	2.73–7.74	<.0001	8.81	1.37–17.99	.0237
Sexual bother	7.98	3.27–12.67	.0009	6.19	10.15–23.08	.0510

function = 9, bodily pain = 9, and general health = 8; and for the prostate-specific scales are urinary function = 8, bowel function = 7, sexual function = 8, urinary bother = 9, bowel bother = 8, and sexual bother = 11.

Recovery Pattern (Return to Baseline) Analysis

Return to baseline was defined as a change in score of magnitude equal to or greater than the MID for the subscales of generic and prostate-specific HRQoL. For each subscale of SF-36 and PCI, we determined the proportion of participants returning to baseline at 12 months. Table 7 presents the comparison of the return to baseline patterns corresponding to the distribution-based methods and anchor-based method. The recovery patterns corresponding to MID estimate from both methods demonstrated "good concordance" as indicated by the kappa statistics that ranged between 0.76 and 1.00 for the generic and prostate-specific HRQoL scales. Since treatment has the potential to affect outcomes, we ran separate analysis for the two treatments (radical prostatectomy and external beam radiation therapy) and found comparable results. Additionally, results using 24-month follow-up were comparable as well.

Table 7: Minimal Important Difference and Percentage of Patients Returning to Baseline Estimates

	Distribution Method		*Anchor Method*		
Scale	*Mean of Distributions*	*Return to Baseline*	*Mean of Anchors*	*Return to Baseline*	*Grand Mean of Distribution and Anchor Methods*
Generic HRQoL–SF 36					
Physical function	8.3	79.85	5.4	79.85	7.0
Role physical	16.1	80.94	12.4	80.94	14.0
Role emotional	15.3	88.48	9.8	88.48	12.0
Vitality	9.2	70.56	9.1	70.56	9.0
Mental health	7.0	81.47	5.0	81.47	6.0
Social function	9.9	75.32	7.9	75.32	9.0
Bodily pain	9.6	65.74	7.85	65.74	9.0
General health	8.7	67.84	6.8	67.84	8.0
Prostate cancer index					
Urinary function	10.0	50.77	6.9	49.49	8.0
Bowel function	6.6	74.23	6.7	74.23	7.0
Sexual function	11.3	34.13	4.6	24.34	8.0
Urinary bother	11.5	63.40	5.6	63.40	9.0
Bowel bother	9.8	80.41	7.0	80.41	8.0
Sexual bother	15.3	43.01	7.1	43.01	11.0

DISCUSSION

Health-related quality of life represents the fundamental value of a treatment, over and above the treatment's effect on measurable objective health parameters in prostate cancer patients. In this two-part study, we reviewed various measures of MID and analyzed MID in the context of prostate cancer care. The MID is customary for interpreting therapeutic changes in both generic and disease-specific patient-reported outcomes. It can also be used to relate changes in health status to the changes in more established clinical measures when possible, or as a bridge to compare outcomes across different studies. Comparative effectiveness must choose a meaningful threshold and provide information such as the proportion of patients achieving a small but important benefit or those experiencing poorer outcomes.

Major findings from our study are as follows: (1) we have derived MID estimates of generic and prostate-specific HRQoL measures in a cohort of prostate cancer patients using longitudinal data; (2) MID estimates derived from distribution-based method and anchor-based method show variation; (3) it is important to use multiple methods of distribution- and anchor-based

analysis to derive at the MID values; (4) the recovery pattern (return to baseline) corresponding to the two methods is comparable (kappa statistics 0.76–1.00); and (5) the method of MID estimation has implications for sample size computation, but not for recovery pattern. It is important to test these findings via multiple studies, as more and more randomized controlled trials emerge to analyze the efficacy and effectiveness of various treatment interventions.

The concept of MID was developed to help providers and patients interpret the change in health status and can play a role in patient-centered care and comparative effectiveness assessment (Gyatt and Schunemann 2007; Revicki et al. 2008). The interpretation of changes in health status measurements for outcome assessments is understudied (Osoba et al. 1998). An important issue in the interpretability of patient-reported outcome is whether one makes inferences with respect to individuals or populations. The population perspective takes into account the degree of importance at a population level, where small differences may be important due to the large number of individuals who are affected. It also considers the degree of change required for making adjustments in health care policies. On the other hand, an individual perspective takes into account the degree of importance at an individual level (but determined in groups of patients), where large difference may be needed to assess the change (de Vet et al. 2010). Another perspective is the degree of change required to stimulate clinicians to consider an intervention.

Recommended approaches to estimate the MID include several anchor-based methods, with relevant clinical or patient-based indicators and various distribution-based methods (Jaeschke et al. 1991; Osoba et al. 1998; Wyrwich, Tirney, and Wolinsky 1999; Patrick and Yen-Pin 2000; Wells et al. 2001; Guyatt et al. 2002; Norman, Sloan, and Wyrwich 2003; Eton et al. 2004; Kulkarni 2006; Lemieux et al. 2007; Barrett, Brown, and Mundt 2008; Chen, Clark, and Talcott 2009). Estimation and interpretation of MID is as daunting as it is vital in forming recommendations for intervention. However, there is some discordance regarding appropriate anchors and their desirable properties. Also, while some studies show the convergence of MID estimates derived from anchor-based and distribution methods (Eton et al. 2004), currently, there are no guidelines for action during nonconvergence of estimates. Many researchers have attempted to test empirically the relationship between distribution-based and anchor-based approaches. Osoba et al. (1998) compared the subjective significance questionnaire with effect size approach. The subjective significance questionnaire is an anchor-based approach and the term "subjective significance" refers to the changes in HRQoL scores that the subjects consider to be important. Using this approach, Osoba et al.

(1998) supported Cohen's guidelines that small, moderate, and large effect sizes coincided with the same magnitude of change as indicated by the subjective significance questionnaire ratings (Cohen 1988). Studies have demonstrated that effect sizes and minimum clinically important differences provide equivalent information (Juniper et al. 1994; Norman et al. 2001).

Our study has several strengths as well as limitations. To our knowledge, this is the first study reporting MID for popularly used patient-reported outcome measures of generic and prostate-specific HRQoL among newly diagnosed prostate cancer patients. While our assessment of MID has the potential to facilitate outcomes research, it needs further validation using other anchors. Another limitation is that anchors were chosen retrospectively. The results may be stronger in case of prospective anchors. We have used the patient-reported physical signs and symptoms of "more tired or worn out than usual" as a longitudinal anchor. Although this anchor shows good correlation with the generic and prostate-specific HRQoL subscales, there is a potential for better longitudinal anchors. These issues need further evaluation. Also, in the case of longitudinal anchors, some of the confidence intervals are large, indicating the sampling variation of our study cohort. This demands further refinement of the MID values using larger cohorts. Another limitation is the generalizability of our findings as the participants were recruited from an urban academic hospital and a VA hospital. Finally, our study included only African American and Caucasian patients due to nonavailability of other racial and ethnic groups at this health care system.

The patient-reported outcomes measures can help clinicians gain insight into a patient's perspective of his/her care, treatment, and clinical decision making. We hope that further studies as well as clinicians' frequent use of HRQoL measures will ultimately lead to the sort of clear interpretability of HRQoL results that is common for most clinical parameters for prostate cancer care. There exists limited information regarding established MIDs for many generic and disease-specific patient-reported outcome measures. Thus, there is an urgent need of assessments of MIDs across generic and disease-specific instruments for various disease and population settings. In conclusion, we were able to derive the first estimates of the MID for the generic and prostate-specific HRQoL measures in a longitudinal study of prostate cancer patients. Our MID estimates have significant implications for sample size computation when planning for comparative effectiveness studies or clinical trials to analyze meaningful changes over time. Additional studies are needed to refine our estimates and further advance the understanding and utility of patients reported outcomes to enhance prostate

cancer care. Our study provides a template for researchers attempting to derive MID estimates across various diseases domain using prospective cohort design studies.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Partially Supported by the DOD Prostate Cancer Research Program DAMD17-02-1-0126, W81XWH-04-1-0257 and Linda and Laddie Montague Research Fund.

Disclosures: None.

Disclaimers: None.

REFERENCES

Allison, P. D. 1999. *Logistic Regression Using the SAS System: Theory and Application.* Cary, NC: SAS Institute Inc.

Anastasi, A., and S. Urbina. 1997. *Psychological Testing.* Upper Saddle River, NJ: Prentice-Hall.

Barnett, P. G. 1999. "Review of Methods to Determine VA-Health Care Costs." *Medical Care* 37 (4): AS9–17.

Barrett, B., R. Brown, and M. Mundt. 2008. "Comparison of Anchor-Based and Distributional Approaches in Estimating Important Difference in Common Cold." *Quality of Life Research* 17: 75–85.

Barrett, B., D. Brown, M. Mundt, and R. Brown. 2005. "Sufficiently Important Difference: Expanding the Framework of Clinical Significance." *Medical Decision Making* 25: 250–61.

Beaton, D. E., M. Boers, and G. A. Wells. 2002. "Many Faces of the Minimal Clinically Important Difference (MCID): A Literature Review and Directions for Future Research." *Current Opinion in Rheumatology* 14: 109–14.

Bonniaud, V., D. Bryant, B. Parratte, and G. Guyatt. 2008. "Qualiveen, a Urinary-Disorder Specific Instrument: 0.5 Correspons to the Minimal Important Difference." *Journal of Clinical Epidemiology* 61: 505–10.

Cella, D., D. T. Eton, D. L. Fairclough, P. Bonomi, A. E. Heyes, C Silberman, M. K. Wolf, D. H. Johnson. 2002a. "What Is a Clinically Meaningful Change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592." *Journal of Clinical Epidemiology* 55: 285–95.

Cella, D., D. T. Eton, L. Jin-Shei, A. H. Peterman, and D. E. Merkel. 2002b. "Combining Anchor and Distribution-Based Methods to Derive Clinically Important Differences on the Functional Assessment of Cancer Therapy (FACT) Anemia and Fatigue Scales." *Journal of Pain and Symptom Management* 24 (6): 547–61.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. Mackenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Disease* 40: 373–83.

Chen, R. C., J. A. Clark, and J. A. Talcott. 2009. "Individualizing Quality-of-Life Outcomes Reporting: How Localized Prostate Cancer Treatments Affect Patients with Different Levels of Baseline Urinary, Bowel, and Sexual Function." *Journal of Clinical Oncology* 27: 1–13.

Coeytaux, R. R., J. S. Kaufman, R. Chao, J. D. Mann, R. F. Devellis, 2006. "Four Methods of Estimating the Minimal Important Difference Score Were Compared to Establish a Clinically Significant Change in Headache Impact Test." *Journal of Clinical Epidemiology* 59: 374–80.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associated.

Crosby, R. D., R. L. Kolotkin, and R. Williams. 2003. "Defining Clinically Meaningful Change in Health Related Quality of Life." *Journal of Clinical Epidemiology* 56: 395–407.

———. 2004. "An Integrated Method to Determine Meaningful Changes in Health Related Quality of Life." *Journal of Clinical Epidemiology* 57: 1153–60.

Dixon, T., L. L. Y. Lim, and N. B. Oldridge. 2002. "The MacNew Heart Disease Health-Related Quality of Life Instrument: Reference Data for Users." *Quality of Life Research* 11: 173–83.

Engels, J. M., and P. Diehr. 2003. "Imputation of Missing Longitudinal Data: A Comparison of Methods." *Journal of Clinical Epidemiology* 56: 968–76.

Eton, D. T., D. Cella, K. J. Yost, S. E. Yount, A. H. Peterman, D. S. Neuberg, G. W. Sledge, and W. C. Wood. 2004. "A Combination of Distribution-and Anchor Based Approaches Determined Minimally Important Differences (MIDs) for Four Endpoints in a Breast Cancer Scale." *Journal of Clinical Epidemiology* 57: 898–910.

Guyatt, G. H., D. Osoba, A. W. Wu, K. W. Wyrwich, G. R. Norman and the Clinical Significance Consensus Meeting Group. 2002. "Methods to Explain the Clinical Significance of Health Status Measures." *Mayo Clinic Proceedings* 77: 371–83.

Gyatt, G., and H. Schunemann. 2007. "How Can Quality of Life Researchers Make Their Work More Useful to Health Workers and Their Patients?" *Quality of Life Research* 16: 1097–105.

Hayas, L. C., J. M. Quintana, J. A. Padierna, A. Bilbao, P. Munoz, and E. F. Cook. 2007. "Health-Related Quality of Life for Eating Disorders Questionnaire Version-2 Was Responsive 1-Year after Initial Assessment." *Journal of Clinical Epidemiology* 60: 825–33.

Jaeschke, R., G. H. Guyatt, J. Keller, and J. Singer. 1991. "Interpreting Changes in Quality-of-Life Score in N or 1 Randomized Trials." *Controlled Clinical Trials* 12: 226S–33S.

Jeschke, R., J. Singer, and G. H. Guyatt. 1989. "Measurement of Health Status-Ascertaining the Minimal Clinically Important Difference." *Controlled Clinical Trials* 10: 407–15.

Jordan, K., K. M. Dunn, M. Lewis, and P. Croft. 2006. "A Minimal Clinically Important Difference Was Derived for the Roland-Morris Disability Questionnaire for Low Back Pain." *Journal of Clinical Epidemiology* 59: 45–51.

Juniper, E. F., G. H. Guyatt, A. Willan, and L. E. Griffith. 1994. "Determining a Minimal Important Change in Disease-Specific Quality of Life Questionnaire." *Journal of Clinical Epidemiology* 47 (1): 81–7.

Kawata, A. K., D. A. Revicki, R. Thakkar, P. Jiang, S. Krause, M. H. Davidson, H. A. Punzi, and R. J. Padley. 2009. "Flushing Assessment Tool (FAST): Psychometric Properties of a New Measure Assessing Flushing Symptoms and Clinical Impact of Niacin Therapy." *Clinical Drug Investigation* 29 (4): 215–29.

Kelleher, C. J., A. M. Pleil, P. R. Reese, S. M. Burgess, and P. H. Brodish, 2004. "How Much Is Enough and Who Says So? The Case of The King's Health Questionnaire and Overactive Bladder." *BJOG: An International Journal of Obstetrics and Gynecology* 111: 605–12.

Kulkarni, A. V. 2006. "Distribution and Anchor-Based Approaches Provided Different Interpretability Estimates for the Hydrocephalus Outcome Questionnaire." *Journal of Clinical Epidemiology* 59: 176–84.

Lee, B. B., M. T. King, J. M. Simpson, M. J. Haran, M. R. Stockler, O. Marial, and G. Salkeld. 2008. "Validity, Responsiveness, and Minimal Important Difference for the SF-6D Health Utility Scale in a Spinal Cord Injured Population." *Value in Health* 11 (4): 680–8.

Lemieux, J., D. E. Beaton, S. Hogg-Johnson, L. J. Bordeleav, and P. J. Goodwin. 2007. "Three Methods for Minimally Important Difference: No Relationship Was Found with the Net Proportion of Patients Improving." *Journal of Clinical Epidemiology* 60: 448–55.

Litwin, M. S., R. D. Hays, A. Fink, et al. 1998. "The UCLA Prostate Cancer Index: Development, Reliability, and Validity of Health-Related Quality of Life Measure." *Medical Care* 36: 1002–12.

Norman, G. R., J. A. Sloan, and K. W. Wyrwich. 2003. "Interpretation of Changes in Health Related Quality of Life-The Remarkable Universality of Half a Standard Deviation." *Medical Care* 41 (5): 582–92.

Norman, G. R., F. G. Sridhar, G. H. Guyatt, and S. D. Walter. 2001. "Relation of Distribution-and Anchor Based Approaches in Interpretation of Changes in Health Related Quality of Life." *Medical Care* 39 (10): 1039–47.

Oakes, M. 1990. *Statistical Inference.* Chestnut Hill, MA: Epidemiological Resources.

Osoba, D., G. Rodrigues, J. Myles, B. Zee, and J. Pater. 1998. "Interpreting the Significance of Changes in Health Related Quality of Life Score." *Journal of Clinical Oncology* 16: 139–44.

Patrick, D. L., and C. Yen-Pin. 2000. "Measurement of Health Outcomes in Treatment Effectiveness Evaluations." *Medical Care* 38 (suppl II): II-14–15.

Puhan, M. A., M. Frey, S. Buchi, and H. J. Schunemann. 2008. "The Minimal Important Difference of the Hospital Anxiety and Depression Scale in Patients with Chronic Obstructive Pulmonary Disease." *Health and Quality of Life Outcomes* 6 (46): 1–6.

Quek, M. L., and D. Penson. 2005. "Quality of Life in Patients with Localized Prostate Cancer." *Urologic Oncology: Seminars and Original Investigations* 23: 208–15.

Raj, A. A., D. I. Pavord, and S. S. Birring. 2009. "Clinical Cough IV: What Is the Minimal Important Difference for the Leicester Cough Questionnaire?" *Handbook of Experimental Pharmacology* 187: 311–20.

Redelmeier, D. A., G. H. Guyatt, and R. S. Goldstein. 1996. "Assessing the Minimal Important Difference in Symptoms: A Comparison of Two Techniques." *Journal of Clinical Epidemiology* 49 (11): 1215–9.

Revicki, D., R. D. Hayes, D. Cella, and J. Sloan. 2008. "Recommended Methods for Determining Responsiveness and Minimal Important Differences for Patient-Reported Outcomes." *Journal of Clinical Epidemiology* 68: 102–9.

Ringash, J., A. Bezjak, B. O'Sullivan, and D. S. Redelmeier. 2004. "Interpreting Differences in Quality of Life of Life: The FACT-H&N in Laryngeal Cancer Patient." *Quality of Life Research* 13: 725–53.

Ringash, J., B. O'Sullivan, A. Bezjak, and D. A. Redelmeier. 2007. "Interpreting Clinically Significant Changes in Patient-Reported Outcomes." *Cancer* 110: 196–202.

Schmitt, J. S., and R. P. De Fabio. 2004. "Reliable Change and Minimum Important Difference (MID) Proportions Facilitated Group Responsiveness Comparisons Using Individual Threshold Criteria." *Journal of Clinical Epidemiology* 57: 1008–18.

Shikiar, R., G. Harding, M. Leahy, and R. D. Lennox. 2005. "Minimal Important Difference (MID) of the Dermatology Life Quality Index (DLQI): Results from Patients with Chronic Idiopathic Urticaria." *Health and Quality of Life Outcomes* 3 (36): 1–5.

Siegal, R., E. Ward, O. Brawley, and A. Jemal. 2011. "Cancer Statistics, 2011." *CA A Cancer Journal for Clinicians* 61: 212–36.

Swigris, J. J., K. K. Brown, J. Behr, R. H. du Bois, T. E. King, G. Raghu, and F. S. Wamboldt. 2010. "The SF-36 and SGRQ: Validity and First Look at Minimum Important Differences in IPF." *Respiratory Medicine* 104 (2): 296–304.

Taris, T. W. 2000. *A Primer in Longitudinal Data Analysis.* London: Sage Publications.

Turner, D., H. J. Schunemann, L. E. Griffith, D. E. Beaton, A. M. Griffiths, J. N. Critch, and G. H. Guyatt, et al. 2010. "The Minimal Detectable Change Cannot Reliably Replace the Minimal Important Difference." *Journal of Clinical Epidemiology* 63: 28–36.

de Vet, H. C., R. W. Ostelo, C. B. Terwee, N. van der Roer, D. L. Knol, H. Beckerman, M. Boers, L. M. Bouter. 2007. "Minimally Important Change Determined by a Visual Method Integrating an Anchor-Based and a Distribution-Based Approach." *Quality of Life Research* 16: 131–42.

de Vet, H. C. W., B. Terluin, D. L. Knol, L. D. Roorda, L. B. Mokkink, R. W. Ostelo, E. J. Hendriks, L. M. Bouter, and C. B. Terwee. 2010. "Three Ways to Quanitfy Uncertinty in Idividually Applied "Minimallly Important Change" Values." *Journal of Clinical Epidemiology* 63: 37–45.

van Walraven, C., J. L. Mahon, D. Moher, C. Bohn, and A. Laupacis. 1999. "Surveying Physicians to Determine the Minimal Important Difference: Implications for Sample-Size Calculation." *Journal of Clinical Epidemiology* 52 (8): 717–23.

Walters, S. J., and J. E. Brazier. 2005. "Comparison of the Minimally Important Difference for Two Health State Utility Measures: EQ-5D and SF-6D." *Quality of Life Research* 14: 1523–32.

Ware, J. E., and C. D. Sherbourne. 1992. "The MOS 36-Item Short-for Health Survey (SF-36)-Conceptual Framework and Item Selection." *Medical Care* 30: 473–83.

Wells, G., D. Beaton, B. Shea, M. Boers, L. Simon, V. Strand, P. Brooks, and P. Tugwell. 2001. "Minimal Clinically Important Differences: Review of Methods." *Journal of Rheumatol* 28: 406–12.

Witek, T. J. Jr, and D. A. Mahler. 2003. "Meaningful Effect Size and Patterns of Response of the Transition Dyspnea Index." *Journal of Clinical Epidemiology* 56: 248–55.

Wyrwich, K. W. 2004. "Minimal Important Difference Threshold and the Standard Error of Measurement: Is There a Connection?" *Journal of Biopharmaceutical Statistics* 14 (1): 97–110.

Wyrwich, K. W., W. M. Tirney, and F. D. Wolinsky 1999. "Further Evidence Supporting an SEM-Based Criterion for Identifying Meaningful Intra-Individual Changes in Health Related Quality of Life." *Journal of Clinical Epidemiology* 52 (9): 861–73.

Wyrwich, K. W., W. M. Tierney, A. N. Babu, K. Kroenke, and F. D. Wollinsky. 2005. "A Comparison of Clinically Important Differences in Health Related Quality of Life for Patients with Chronic Lung Disease, Asthma, or Heart Disease." *Health Services Research* 40 (2): 577–97.

Yalcin, I., D. L. Patrick, K. Summers, K. Kinchen, and R. C. Bump. 2006. "Minimal Clinically Important Differences in Incontinence Quality-of-Life Scores in Stress Urinary Incontinenece." *Urology* 67: 1304–8.

Yost, K. J., D. Cella, A. Chawla, E. Holmgren, D. T. Eton, J. Z. Ayanian, and D. W. West. 2005. "Minimally Important Differences Were Estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) Instrument Using a Combination of Distribution- and Anchor-Based Approaches." *Journal of Clinical Epidemiology* 58: 1241–51.

Yost, K. J., D. T. Eaton, S. F. Garcia, and D. Cella. 2011. "Minimally Importand Differences Were Estimated for Six Patient-Reported Outcomes Measurement Information System-Cancer Scales in Advanced-Stage Cancer Patients." *Journal of Clinical Epidemiology* 64: 507–16.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.