Half standard deviation estimate of the minimally important difference in HRQOL scores?

Sepideh S Farivar, Honghu Liu and Ron D Hays†

In addition to statistical significance, it is important to evaluate the magnitude of differences in health-related quality of life over time. Interest in establishing the minimal difference that is clinically important or the minimally important difference has burgeoned over the last few years. This review summarizes some of the leading approaches to estimating the minimally important difference, offers caveats on the minimally important difference estimation based on existing literature and provides recommendations for future work. The authors recommend using multiple anchors to estimate the minimally important difference, using only anchors that correspond to minimal change in health-related quality of life, reporting information about the variation around the estimates, and providing bounded estimates to reflect the uncertainty.

Expert Rev. Pharmacoeconomics Outcomes Res. 4(5), 515–523 (2004)

CONTENTS

†Author for correspondence
UCLA Department of Medicine, Division of General Internal Medicine & Health Services Research, 911 Broxton Plaza, Los Angeles, CA 90095 1736, USA
Tel.: +1 310 794 2294
Fax: +1 310 794 0732
hays@rand.org

KEYWORDS:
clinically important differences, effect size, health-related quality of life, minimally detectable differences, minimally important differences, patient reported outcomes, self-reported health

Health-related quality of life (HRQOL) measures assess functioning and well-being, such as physical functioning, role functioning, social functioning, anxiety, depression, positive well-being, pain and general health perceptions. HRQOL measures are essential ingredients in the assessment of outcomes of healthcare. Accompanying the increased emphasis on HRQOL is a growing recognition of the need to provide objective and unbiased methods to help with interpretation of score differences [1,2].

The significance of difference in HRQOL change between two or more groups is typically used to evaluate the effect of alternative interventions. For example, patients may be randomized to control versus treatment groups and the treatment group produce significantly greater improvement in HRQOL over time. It could then be concluded that the treatment had a significantly positive effect on HRQOL. However, it is also important to consider the magnitude of the difference between the groups. With a large enough sample size, even tiny differences could differ significantly. It is possible that a difference could be so small that it would be considered trivial and unimportant even if statistically significant. Interest in establishing the minimal difference that is clinically important or the minimally important difference (MID) has burgeoned and a review of studies pertinent to MID estimation was recently published [3]. Norman and colleagues concluded that the MID across 33 published studies was approximately 0.50 of a standard deviation (SD) 'for all but six studies, the MID estimates were close to one half of a SD (mean = 0.495; SD = 0.155)' [3]. This review summarizes current approaches and assesses the state of the science of estimating the MID.

Estimating the minimally important difference

Responsiveness refers to the ability of a measure to reflect underlying change [4]. HRQOL change can be compared with change in clinical status, intervening health events, interventions of known or expected efficacy and retrospective reports of change by patients or providers. Responsiveness to change is most frequently evaluated using the effect size (ES), standardized response mean or the responsiveness statistic. For all of these indices, the numerator is the mean change and the denominators are the SD at baseline (ES), the SD of change for the sample (standardized response

 ISSN 1473-7167

mean) and the SD of change for people who are deemed to have not changed according to an external standard (responsiveness statistic).

Evaluating the MID is a special case of examining responsiveness to change that is limited to the subgroup of people who are deemed to have had minimal change. Hence, a fundamental aspect of estimating the MID is to define the subgroup of people who have changed by a minimal amount. This is done using external information or anchors. These anchors include clinical parameters, retrospective measures of change (self or physician reported), or knowledge about the course of health over time. For example, people who have changed by a minimal amount might be identified by asking study participants at follow-up to report how much they changed since the baseline of a study using a multiple categorical response scale, such as:

- Got a lot better
- Got a little better
- Stayed the same
- Got a little worse
- Got a lot worse

People who reported either getting a little better or a little worse would constitute the minimal change subgroup according to the anchor. The average change in HRQOL reported by this subgroup of people (the change in HRQOL for the group who said they had gotten a little worse would be multiplied by negative one to account for the directional difference) would typically be used to index the MID as perceived by the patient. This variant of estimating the MID has been referred to by some as the minimally detectable difference.

The same principles apply to whatever anchor is used to establish the subgroup of people upon which the MID calculation is based. For a clinical parameter it is necessary to establish the amount of change on the anchor that is a reasonable indicator of minimal. Hence, estimating the MID requires agreement on what constitutes a minimal change in the anchor. Kosinski and colleagues defined minimal improvement on their clinical measures as a 1 to 20% improvement in measures of joint swelling and tenderness in their study of 693 patients with rheumatoid arthritis [5]. Although this may be a reasonable threshold, other investigators might argue for something different.

A distinction has been drawn between anchor- and distribution-based methods for determining the MID. Distribution-based methods include the ES, standardized response mean and the responsiveness statistic mentioned earlier. The distribution-based indices provide no direct information about the MID. They are simply a way of expressing the observed change in a standardized metric.

Caveats about estimating the minimally important difference

The variety of kinds of anchors and uncertainty in the anchor cutpoint that defines a minimal difference makes it clear that a single estimate is insufficient. Using the retrospective report anchor as an example, the recall item may refer globally to change in health, HRQOL or quality of life. Moreover, the anchor could be worded more specifically, such as physical functioning, pain or getting along with family. The choice of words could lead to variability in the performance of the anchor. Related to this point, is the fact that any specific anchor may be more or less appropriate for different HRQOL domains. For example, an energy/fatigue scale might be expected to change more than a pain scale in response to change in hematocrit [6]. Interpreting change in response to a particular anchor should take into consideration the fact that not all domains should change or change equally in tandem with the anchor. Other factors that can lead to variation in the estimation of the MID include whether the people being evaluated are high or low on the measure at baseline and whether they improve or decline in HRQOL over time [7].

TABLE 1 lists ESs for the SF-36 scales and summary scores for five different anchors used in a clinical trial of 693 people with rheumatoid arthritis [5]. These ES estimates range from 0.04 (joint tenderness anchor for general health perceptions) to 0.83 (self-report anchor for pain); the size of these estimates range from a small to large effect according to Cohen's rules of thumb: 0.20 SD is considered a small effect, 0.50 is a medium effect and 0.80 or above a large effect. The 0.50 SD guideline for a MID from the Norman and colleagues article falls in-between these extremes and is a medium effect [8].

Literature review

This review examines the studies cited by Norman and colleagues and recomputes the MID estimates they reported. The ESs (TABLE 2) that were the basis of the conclusion that 0.50 SD is a universal estimate of the MID were independently estimated. As shown in the second column of TABLE 2, the median (computed due to the skewed distribution of the measures) of the mean of ES estimates reported by Norman and colleagues was 0.48, consistent with their conclusion that the MID is approximately one half of a SD.

The authors estimated the ES reported for different HRQOL scales as well as the average and range of ES within each study and the median of means across all studies. They coded each scale so that a positive ES meant the change on the scale was consistent with the direction of change in the anchor and a negative ES meant the change was inconsistent with the anchor (method 1). Method 1 drives down the estimated MID if the direction of change for a HRQOL measure is in the wrong direction (inconsistent with the anchor). Since some experts may not regard this as an appropriate estimate of the MID, the absolute value of the ES was estimated so that changes that were inconsistent with the anchor increased the magnitude of the MID (method 2). Others would argue that this is not an ideal strategy for estimating the MID. Therefore, the average after recoding any negative ES estimates to zero (method 3) was also estimated. This final approach allows the few estimates that went in the wrong direction to drive the overall MID estimate down but not as extremely as their unadjusted values.

Table 1. Effects sizes for SF-36 changes related to minimal changes in five anchors.

Scale	Self-report	Clinician report	Global report of pain	Joint swelling	Joint tenderness	Average[§]
Physical functioning	0.35	0.33	0.34	0.26	0.32	0.32
Role limitations due to physical health	0.56	0.52	0.29	0.35	0.36	0.42
Bodily pain	0.83	0.70	0.47	0.69	0.42	0.62
General health	0.20	0.12	0.09	0.12	0.04	0.12
Emotional well-being	0.39	0.26	0.25	0.18	0.05	0.23
Role limitations due to emotional problems	0.41	0.28	0.18	0.38	0.26	0.30
Social functioning	0.43	0.34	0.28	0.29	0.38	0.34
Energy/fatigue	0.50	0.47	0.22	0.22	0.35	0.35
Physical component summary	0.49	0.48	0.34	0.29	0.36	0.39
Mental component summary	0.42	0.27	0.19	0.27	0.20	0.27

[§]Median of the means = 0.33.

Most of the studies included in the review included more than one HRQOL instrument. While Norman and colleagues captured most of the instruments reported in the cited studies in their analysis, the selection of instruments on which to report average ES was incomplete. For example, the SF-36 was used in seven different studies but Norman and colleagues reported use of the SF-36 in only five of the studies. It is not obvious why the SF-36 was included in the analysis for some studies and excluded for others. Studies in which Norman and colleagues excluded HRQOL instruments are noted below. Jones and Bosh's study of chronic obstructive pulmonary disease (COPD) patients which reported findings for the St George's Hospital Respiratory Questionnaire (SGRQ) and the SF-36 [9], however, Norman and colleagues did not report the SF-36. Bagenstose and Berstein's study of rhinitis patients also included both the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ) and the SF-36, but Norman and colleagues did not report the SF-36 [10]. Bestall's study of COPD patients included the SGRQ, Chronic Respiratory Questionnaire (CRQ) and the Hospital Anxiety and Depression (HAD) [11], but Norman and colleagues did not report the HAD. Miller and colleagues' study of multiple sclerosis patients used the SF-36, SIP and Fatigue Impact Scale (FIS) [12], but Norman and colleagues did not report the FIS. Singh's study used the SGRQ, CRQ, Global Quality of Life Scale (GQOL) [13], People Do and the Breathing Problems Questionnaire (BPQ). Norman and colleagues did not report the GQOL, People DO and the BPQ. Talley's study of heartburn included the Quality of Life in Reflux and Dyspepsia (QOLRAD) and the Gastrointestinal Symptoms Rating Scale (GSRS) [14]. Norman and colleagues did not report the GSRS.

Calculation of the average ES within studies varied. In some studies, the calculation of the mean ES included domain-specific scores and in other studies the mean ES includes the total score only. For example, Cella and colleagues's study regarding lung cancer patients uses the Lung Cancer Subscale (LCS), Functional Assessmend of Cancer Therapy (FACT)-General and the Trial Outcomes Index (TOI) [15]. Norman's mean ES estimate for the FACT includes the total FACT-G score, but there was no reference to the domain-specific ES estimates of physical, social, emotional and functional well-being which should be included in either the average ES for the FACT or as separate domain-specific averages. For this same study, Norman reports the average ES for LCS to be 0.5 with a range of 0.39 to 0.47 (the mean ES should not exceed the upper bound of the range). In another study by Cella and colleagues, the average ES appears to be based on the FACT-G summary scale and not the FACT-G scale scores [16]. On the other hand, in some studies Norman and colleagues appear to include SF-36 scale scores in computing the average ES [5,24,44].

Two studies were included that had not estimated the MID (a study of 605 coronary artery disease/congestive heart failure patients and a study of 417 COPD patients) but referred to the ES that would be needed to demonstrate equivalency to the MID standards previously reported [17,18]. For example, the 0.35 average ES in Norman and colleagues for the CRQ was computed as the ratio of the previously reported MID of 0.5 per item divided by the SDs observed in this sample of 417 COPD patients [17]. One study evaluated 112 people with stable COPD and ES estimates were based on a 6-minute walk test, which is not a measure of HRQOL [19]. The authors excluded these three studies from the calculations. The Santanello and colleagues study [42] included puffs per day of inhaled B-agonist as a dependent variable and this is arguably another estimate that should have been excluded form the pooled estimate.

The median of the mean ES for the studies reported in TABLE 2 was 0.42 for method 1, method 2 and method 3. The SD of the mean ES for each method was 0.31, the coefficient of variation

Table 2. Recalculation of effect size estimates reported in [3].

Norman *et al.* ES (range)	Recalculated method 2: absolute value (range)	Measure	Ref.
0.5 (0.28–1.42)	0.39 (0.01–1.60)	Arthritis Impact Measurement Scales	[33]
0.45 (0.06–0.83)	Same as Norman	Arthritis Impact Measurement Scales	[34]
0.48 (0.38–0.74)	0.48 (0.28–0.74)	Health Assessment Questionnaire	[34]
0.37 (0.27–0.54)	0.37 (0.24–0.59)	Nottingham Health Profile	[34]
0.52 (0.26–0.69)	0.52 (0.26–0.69)	Functional Limitations Profile	[34]
0.48 (0.33–0.83)	0.60 (0.37–0.83)	Health Assessment Questionnaire	[35]
0.22 (–)	0.32 (0.17–0.42)	St George's Hospital Respiratory Questionnaire	§[9]
	0.11 (0.02–0.25)	SF-36	§[9]
0.49 (0.41–0.58)	0.51 (0.35–0.72)	Asthma Quality of Life Questionnaire	[36]
0.65 (–)	0.64 (0.47–0.73)	Chronic Respiratory Questionnaire	[37]
0.5 (0.39–0.47)	0.50 (0.42–0.58)	Lung Cancer Subscale	§[15]
0.55 (0.42–0.58)	0.55 (0.52–0.59)	Trial Outcomes Index	§[15]
0.43 (0.39–0.47)	Same as Norman	Functional Assessment of Cancer Therapy - General total	§[15]
	0.57 (0.30–0.84)	Functional Assessment of Cancer Therapy - General physical well-being	§[15]
	0.15 (0.08–0.23)	Functional Assessment of Cancer Therapy - General SWB	§[15]
	0.24 (0.08–0.41)	Functional Assessment of Cancer Therapy - General emotional well-being	§[15]
	0.42 (0.35–0.48)	Functional Assessment of Cancer Therapy - General functional well-being	§[15]
0.67 (–)	2.31 (1.75–3.33)	Chronic Respiratory Questionnaire	++§[21]
0.53 (–)	0.60 (0.44–0.73)	Asthma Quality of Life Questionnaire	[38]
0.49 (0.07–1.5)	0.35 (0.06–0.76)	St George's Hospital Respiratory Questionnaire	§[23]
0.43 (0.31–0.54)	Same as Norman	6-minute walk	§[19]
0.52 (0.24–0.62)	0.49 (0.24–0.94)	SF-36	§[24]
0.34 (–)	Same as Norman	Global Quality of Life	§[24]
0.92 (–)	0.93 (–)	Carpal Tunnel Syndrome Assessment Questionnaire	§[24]
0.5 (–)	0.72 (0.57–1.00)	Chronic Respiratory Questionnaire	[39]
0.35 (0.14–0.59)	0.31 (0.05–0.46)	+++Incontinence - Quality of Life Instrument	[40]
0.27 (–)	0.26 (–)	++++Incontinence - Quality of Life Instrument	[40]
0.48 (0.32–0.83)	0.38 (0.16–0.54)	Rhinoconjunctivitis Quality of Life Questionnaire	§[10]
	0.31 (0.06–0.68)	SF-36	§[10]
0.54 (0.04–1.07)	0.56 (0.02–1.11)	St George's Hospital Respiratory Questionnaire	[11]
0.55 (–)	0.50 (0.02–1.27)	Chronic Respiratory Questionnaire	[11]
	0.50 (0.00–0.98)	Health Assessment Questionnaire	[11]
0.4 (0.23–0.62)	0.55 (0.23–1.02)	Sickness Impact Profile weighted	[41]
0.3 (–)	0.24 (0.17–0.30)	Puffs/day	§§[42]

Table 2. Recalculation of effect size estimates reported in [3] (cont.).

Norman *et al.* ES (range)	Recalculated method 2: absolute value (range)	Measure	Ref.
0.33 (0.25–0.41)	No standard deviation§§	Number of symptoms	§[42]
0.36 (0.34–0.37)	Data not available§§	CHQ	§§[17]
0.35 (0.3–0.38)	Data not available§§	Chronic Respiratory Questionnaire	§[18]
0.45 (–)	0.64 (–)	Chronic Respiratory Questionnaire	[43]
0.45 (0.41–0.49)	0.34 (0.04–0.83)	SF-36	[5]
0.35 (–)	0.28 (0.19–0.35)	Health Assessment Questionnaire	[5]
0.49 (0.09–0.89)	0.42 (0.09–0.89)	SF-36	§[12]
	0.18 (0.01–0.39)	FIS	§[12]
0.81 (-)	0.75 (0.32–1.14)	Sickness Impact Profile weighted	§[12]
0.52 (0.4–0.54)	0.50 (0.46–0.53)	Medical Outcomes Study - HIV	§[20]
0.34 (0.27–0.49)	0.34 (0.27–0.44)	Multidimensional Quality of Life - HIV	§[20]
0.31 (0.05–0.86)	0.31 (0.16–0.47)	SF-36	[44]
0.39 (0.2–0.62)	0.39 (0.19–0.62)	Western Ontario and McMaster Universities Osteoarthritis Instrument	[44]
1.06 (1.03-1.08)	Same as Norman	Short Inflammatory Bowel Disease Questionnaire	§[22]
0.62 (–)	0.15 (0.04–0.26)	SF-36 physical function§§§§	§§§[45]
0.26 (–)	0.26 (0.03–0.39)	St George's Hospital Respiratory Questionnaire	[13]
0.54 (0.39–0.59)	0.68 (0.42-1.09)	Chronic Respiratory Questionnaire	[13]
	0.39 (–)	GQOL	[13]
	0.27 (–)	People Do	[13]
	0.23 (–)	BPQ	[13]
0.51 (0.50–0.55)	0.58 (0.04–1.33)	Quality of Life in Reflux and Dyspepsia	[14]
	0.94(0.64–1.04)	GSRS	[14]
0.55 (0.48–0.62)	(0.23–0.66)+	Lung Cancer Subscale change	[46]
0.59 (0.48–0.67)	Same as Norman	Trial Outcomes Index baseline	[46]
0.57 (0.48–0.72)	Same as Norman	Lung Cancer Subscale baseline	[46]
0.4 (0.38–0.42)	(0.14–0.45)	Trial Outcomes Index change	[46]
0.71 (0.56–1.42)	Data not available§§	DATE	
0.58 (0.33–0.83)	Data not available§§	Pain Visual Analogue Scale	
0.55 (0.4–0.8)	Data not available§§	Functional Assessment of Cancer Therapy - Head and Neck	
0.35 (0.32–0.38)	Data not available§§	Time Trade Off Utility Instrument	
0.51 (0.38–0.73)	Data not available§§	Functional Assessment of Cancer Therapy - General	
0.73 (0.53–1.46)	Data not available§§	Karnofsky	
0.48 (0.05–0.51)	0.49 (0.35–0.62)	Functional Assessment of Cancer Therapy - General	[16]
	0.27 (0.17–0.36)	Functional Assessment of Cancer Therapy - General physical well-being	[16]
	0.31 (0.03–0.59)	Functional Assessment of Cancer Therapy - General SWB	[16]

Table 2. Recalculation of effect size estimates reported in [3] (cont.).

Norman *et al.* ES (range)	Recalculated method 2: absolute value (range)	Measure	Ref.
	0.37 (0.26–0.47)	Functional Assessment of Cancer Therapy - General emotional well-being	[16]
	0.35 (0.20–0.49)	Functional Assessment of Cancer Therapy - General functional well-being	[16]
0.36 (0.06–0.63)	Same as Norman	Western Ontario and McMaster Universities Osteoarthritis Instrument	[47]
0.48 (0.22–1.06)	0.42 (0.11–2.31)	Median (range) of means across studies	

[§]Study may be inappropriate for minimally important differences estimation.
[§§]Unable to use study in calculation of mean and median due to missing information.
[§§§]Baseline standard deviation not available in original study. To calculate ES, standard deviation for females in the US general population was used.
[§§§§]Other SF-36 scales did not show significant differences between groups in change over time.
[+]Only range of ES estimates were reported.
[++]Standard deviation not reported in original study. To calculate ES, the authors used standard deviation from another study [39].
[+++]Norman *et al.* clinically important difference estimate.
[++++]Norman *et al.* minimally important difference estimate.
BPQ: Breathing Problems Questionnaire; CHQ: Chronic Heart Failure Questionnaire; ES: Effect size; GQOL: Global Quality of Life Scale; GSRS: Gastrointestinal Symptoms Rating Scale; FIS: Fatigue Impact Scale; SWB: Social/family well-being.

was 64% and the range was 0.11 to 2.31. (Of course, if one were to drop an extreme value, the SD and range would be reduced.) The bottom quartile includes ES averages of 0.31 or lower. The third column of TABLE 2 provides the method 2 ES estimates.

It is worth noting that many of the studies reviewed by Norman and colleagues had estimates that were based on changes that were not necessarily minimal in size [3]. For example, Badia and colleagues looked at change in the Medical Outcomes Study-HIV and Multidimensional Quality of Life-HIV in 296 people with AIDS who either got better or much better [20]. Several studies had changes reported for interventions of unknown magnitude. Bagenstose and Bernstein estimated change in the RQLQ for 19 new rhinitis patients before and after seeing an allergist who prescribed a new medication regimen [10]. Goldstein and colleagues reported mean differences in the CRQ between 45 treatment and 44 control patients in an evaluation of respiratory rehabilitation for COPD patients [21]. Jones and Bosh compared changes between baseline and 16 weeks later for COPD patients treated with salmeterol (Serevent®, GlaxoSmithKline) [9]. Jowett and colleagues examined 123 ulcerative colitis patients using cross-sectional comparisons of mild versus moderate and mild versus remission subgroups according to the SEO index of disease activity [22]. Miller and colleagues looked at differences between adjacent Expanded Disability Status Scale categories among 300 multiple sclerosis patients [12]. Osman and colleagues studied 266 COPD patients, comparing those with [23]:

- No further COPD admission and survived versus further COPD admission or died
- No nebulizer versus nebulizer
- No domiciliary O_2 versus O_2

Bessette and colleagues had pre–post comparisons of the subgroup of 40 carpal tunnel patients who were somewhat satisfied with surgery [24]. Cella and colleagues used change on the Eastern Cooperative Group Performance Status Self-Report among 41 lung cancer patients [15]. Cella and colleagues used global ratings of change to categorize people into what the authors labeled sizably worse, minimally worse, no change, minimally better and sizably better in another study of lung cancer patients [16]. Although the spirit of this approach is appropriate, some may not agree with the particular collapsing that was done to translate the 15-point change scale into the five categories noted above. Due to the unknown magnitude of true change in each of these studies, it is likely that these provide biased estimates of the MID. One could argue that all ten of these studies should be excluded, but this would reduce the number of included studies by a third.

Preference instruments

MID estimates for preference instruments have been closer to 0.30 SD than to the 0.50 SD threshold suggested in the Norman and colleagues review [3]. The SD of the Health Utilities Index was 0.13 in the 1994–1995 Canadian National Population Health Survey [25]. Drummond reported that differences as little as 0.01 on the Health Utilities Index could be meaningful and important, and differences of 0.03 (ES = 0.23) are definitely important [26]. The SD of the Quality of Well-Being Scale (QWB) is approximately 0.10 [27]. Kaplan and colleagues suggested that the MID for the QWB is 0.03 units or 0.30 ES [28]. Similarly, a study of 134 subjects with schizophrenia or schizoaffective disorder reported an ES of 0.36 for change in the QWB between those improving and not improving according to the the Positive and Negative Syndrome Scale [29]. Seven studies estimated the MID for the SF-6D ranging from 0.01 to 0.05 and the weighted mean

estimate was 0.03 (95% confidence interval [CI]; 0.03–0.04) [30]. The SDs of change were approximately 0.10, indicating an ES of approximately 0.30. The corresponding standardized response means ranged from 0.11 to 0.48, with a mean of 0.30.

Expert opinion & five-year view

We recommend that in the future, effort should be directed at providing the information about the distribution of and reasonable bounds around the MID rather than forcing the MID to be a single value. The inherent uncertainty in estimating the MID indicates the multifaceted attributes and the complexity in measuring the MID. It points out the importance of including multiple anchors and the value in reporting a range and bounded estimate rather than a point estimate. The ES estimates reported in TABLE 2 are quite variable and the median estimate of 0.42 conveys limited information about any particular study or situation. The bounds can be estimated using range, interquartile range (IQR) and CIs. Range and IQR have the advantage that they are robust to possibly asymmetric distributions of MID estimates. CIs can be estimated through large sample theory with the assumption of asymptotically normal distribution of MID estimates. The bootstrap method, for example, could be applied to estimate the distribution of MID estimates including the range, IQR and the standard errors of the measures [31,32]. The sampling process should be stages before obtaining parameter estimates to preserve the sampling variation of the raw data.

In addition, evaluating studies that have anchors that represent an unknown quantity of change is difficult and problematic. Since the size of the observed HRQOL difference should match the true underlying change, anchors that do not represent minimal change are inappropriate for estimating the MID.

In summary, future research should:

- Use multiple anchors to estimate the MID
- Only use anchors that correspond to minimal change in HRQOL
- Report bounded estimates of uncertainty and information about the variation associated with the estimate

Further evaluation of MID estimates for preference measures and exploration as to why they tend to be closer to Cohen's small ES guideline than to the half-standard deviation guideline suggested by Norman and colleagues is also warranted [3]. Finally, it is important to note that the examination of the MID in health services research has focused on group level comparisons. In contrast, parallel work in psychology has emphasized differences for individual patients that are clinically significant [48]. The size of difference that is important for individual patient change exceeds the size for group differences due to the larger error assciated with individual measurement.

Acknowledgements

Sepideh S Farivar was supported by the National Research Services Award (T32-HS00046) from the Agency for Healthcare Research Quality. Ron Hays was supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities, (P20-MD00148-01) and the UCLA Center for Health Improvement in Minority Elders/Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging, (AG-02-004).

Key issues

- Multiple anchors should be used to estimate minimally important differences of health-related quality of life scores.
- Only anchors that correspond to minimal change should be used.
- Information about the variation around the estimates should be reported.
- Bounded estimates should be provided to reflect the uncertainty.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J. Clin. Epi.* 56, 395–407 (2003).
•• Provides a comprehensive overview of the concept of clinically meaningful differences in health-related quality of life.

2 Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin. Proc.* 77, 371–383 (2002).
•• Reviews in detail the various methods that can be used to estimate clinically significant differences.

3 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med. Care* 41, 582–592 (2003).
•• Examines the existing literature and concludes that the minimally important difference was close to a half standard deviation.

4 Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual. Life Res.* 1, 73–75 (1992).

5 Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining the minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum.* 43, 1478–1487 (2000).
•• Uses five different anchors to estimate the minimally important differences for the SF-36 in a clinical trial of people with rheumatoid arthritis.

6 Beusterien KM, Nissenson AR, Port FK, Kelly M, Steinwald B, Ware JE. The effects of recombinant human erythropoietin on functional health and well being in chronic dialysis patients. *J. Am. Soc. Neph.* 7, 763–773 (1996).

7 Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research: how meaningful is it? *PharmacoEconomics* 18, 419–423 (2000).

8 Cohen J. A power primer. *Psychol. Bull.* 112, 155–159 (1992).

9 Jones PW, Bosh TK. Quality of life changes in COPD patients treated with salmeterol. *Am. J. Resp. Crit. Care Med*. 155, 1283–1289 (1997).

10 Bagenstose SE, Bernstein JA. Treatment of chronic rhinitis by an allergy specialist improves quality of life outcomes. *Ann. Allergy Asthma Immunol*. 83, 524–528 (1999).

11 Bestall JC, Paul EA, Garrod R, Garnham R, Joes PW, Wedzicha JA. Usefulness of the medical research council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 54, 581–586 (1999).

12 Miller DM, Rudick RA, Cutter G, Baier M, Fischer JS. Clinical significance of the multiple sclerosis functional composite: relationship to patient-reported quality of life. *Arch. Neurol.* 57, 1319–1324 (2000).

13 Singh SJ, Sodergren SC, Hyland ME, Williams J, Morgan MDL. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. *Resp. Med*. 95, 71–77(2001).

14 Talley NJ, Fullerton S, Junghard O, Wiklund I. Quality of life in patients with endoscopy-negative heartburn: reliability and sensitivity of disease-specific instruments. *Am. J. Gastroenterol*. 96, 1998–2004 (2001).

15 Cella DF, Bonomi AE, Lloyd SR, Tulsky DS, Kaplan E, Bonomi P. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. *Lung Cancer* 12, 199–220 (1995).

16 Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual. Life Res.* 11, 207–221 (2002).

17 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J. Clin. Epi*. 52, 861–873 (1999).

18 Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med. Care* 37, 469–478 (1999).

19 Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the 6-minute walk test in chronic lung disease patients. *Am. J. Resp. Crit. Care Med*. 155, 1278–1282 (1997).

20 Badia, X, Podzamczer D, Casado A, López-Lavid C, Garcia M, Spanish MOS-HIV and MQOL-HIV Validation Group. Evaluating changes in health status in HIV-infected patients: Medical Outcomes Study-HIV and Multidimensional Quality of Life-HIV quality of life questionnaires. *AIDS* 14, 1439–1447 (2000).

21 Goldstein RS, Gort EH, Stubbing D, Avendano MA, Guyatt GH. Randomised controlled trial of respiratory rehabilitation. *Lancet* 344, 1394–1397 (1994).

22 Jowett SL, Seal CJ, Barton R, Welfare MR. The short inflammatory bowel disease questionnaire is reliable and responsive to clinically important change in ulcerative colitis. *Am. J. Gastroenterol*. 96, 2921–2928 (2001).

23 Osman LM, Godden DJ, Friend JAR, Legge JS, Douglas JG. Quality of life and hospital readmission in patients with chronic obstructive pulmonary disease. *Thorax* 52, 67–71 (1997).

24 Bessette L, Sangha O, Kuntz K *et al*. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Med. Care*. 36, 491–502 (1998).

25 van Doorslaer E, Jones AM. Inequalities in self-reported health: validation of a new approach to measurement. *J. Health Econ.* 22, 61–87 (2003).

26 Drummond M. Introducing economic and quality of life measurements into clinical studies. *Ann. Med.* 33(5), 344–349 (2001).

27 Fryback DG, Lawrence WF, Martin PA, Klein R, Klein BEK. Predicting Quality of Well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. *Med. Dec. Making 17* (1), 1–9 (1997).

28 Kaplan RM. The minimally clinically important difference in generic utility-based measures. Prepared for the workshop on minimally clinically important differences in COPD. Bel Harbor, FL, USA. January 11–13 (2004).

29 Pyne JM, Sullivan G, Kaplan R, Williams DK. Comparing the sensitivity of generic effectiveness measures with symptom improvement in persons with schizophrenia. *Med. Care 41* (2), 208–217 (2003).

30 Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual. Life Outcomes* 1(1), 4 (2003).

31 King M, Dobson A. Estimating the responsiveness of an instrument using more than two repeated measures. *Biometrics* 56, 1197–1203 (2000).

32 Efron B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26 (1979).

33 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for changes in health status. *Med. Care* 27, S178–S189 (1989).

34 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Importance of sensitivity to change as a criterion for selecting health status measures. *Qual. Health Car*e 1, 89–93 (1992).

35 Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J. Rheumatol*. 20, 557–566 (1997).

36 Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. *Qual. Life Res*. 5, 35–46 (1996).

37 Wijkstra PJ, Ten Vergert EM, van Altena R *et al*. Long-term benefits of rehabilitation at home on quality of life and exercise tolerance in patients with chronic obstructive pulmonary disease. *Thorax* 50, 824–828 (1995).

38 Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J. Clin. Epidemiol*. 47, 81–87 (1994).

39 Guell R, Casan P, Sangenis M, Morante F, Belda J, Guyatt GH. Quality of life in patients with chronic respiratory disease: the Spanish version of the chronic respiratory questionnaire. *Eur. Respir. J.* 11, 55–60 (1998).

40 Patrick DL, Martin ML, Bushnell DM, Yalcin L, Wagner TH, Buesching DP. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). *Urology* 53, 71–76 (1999).

41 Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the health utilities index mark II. *PharmacoEconomics* 15, 141–155 (1999).

42 Santanello NC, Zhang J, Seidenber B, Reiss TF, Barber BL. What are minimal important changes for asthma measures in a clinical trial? *Eur. Resp. J.* 14, 23–27 (1999).

43 Troosters T, Gosselink R, Decramer M. Short- and long-term effects of outpatient rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. *Am. J. Med.* 109, 207–212 (2000).

44 Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimally clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Care Res.* 45, 384–391 (2001).

45 Segal R, Evans W, Johnson D *et al.* Structured exercise improves physical functioning in women with states I and II breast cancer: results of a randomized controlled trial. *J. Clin. Oncol.* 19, 657–665 (2001).

46 Cella D, Eton DT, Fairclough DL *et al.* What is a clinically meaningful change on the functional assessment of cancer therapy-lung (FACT-L) questionnaire? results from eastern co-operative oncology group (ECOG) study 5592. *J. Clin. Epidemiol.* 37, 469–478 (2002).

47 Angst F, Aeschlimann A, Michel BA *et al.* Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J. Rheumatol.* 29, 131–138 (2002).

48 Wise EA. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. J. Pers. Assess. 82, 50–59 (2004).

Affiliations

- *Sepideh S Farivar*
  UCLA School of Public Health, Department of Health Services, 650 Charles E Young Dr. South, Room 31–269, Los Angeles, CA 90095 1772, USA
  Tel.: +1 949 378 6877
  Fax: +1 310 825 3317
  sfarivar@ucla.edu
- *Honghu Liu*
  Associate Profesor, UCLA Department of Medicine, Division of General Internal Medicine & Health Services Research, 911 Broxton Plaza, Los Angeles, CA 90095 1736, USA
  Tel.: +1 310 794 0700
  Fax: +1 310 794 0732
  hhliu@mednet.ucla.edu
- *Ron D Hays, PhD*
  Professor, UCLA Department of Medicine, Division of General Internal Medicine & Health Services Research, 911 Broxton Plaza, Los Angeles, CA 900951736, USA
  Tel.: +1 310 794 2294
  Fax: +1 310 794 0732
  hays@rand.org