Minimum Important Differences for Scales Assessing Symptom Severity and Quality of Life in Patients With Fecal Incontinence

John Eric Jelovsek, MD, MMEd, Zhen Chen, PhD,† Alayne D. Markland, DO,‡ Linda Brubaker, MD,§ Keisha Y. Dyer, MD,|| Susie Meikle, MD, MPH,¶ David D. Rahn, MD,# Nazeema Y. Siddiqui, MD,** Ashok Tuteja, MD,†† and Matthew D. Barber, MD, MHS**

Objectives: The objective of this study was to estimate the minimum important difference (MID) for the Fecal Incontinence Severity Index (FISI), the Colorectal-Anal Distress Inventory (CRADI) scale of the Pelvic Floor Distress Inventory, the Colorectal-Anal Impact Questionnaire (CRAIQ) scale of the Pelvic Floor Impact Questionnaire, and the Modified Manchester Health Questionnaire (MMHQ).

Methods: We calculated the MIDs using anchor-based and distribution-based approaches from a multicenter prospective cohort study investigating adaptive behaviors among women receiving nonsurgical and surgical management for fecal incontinence (FI). Patient responses were primarily anchored using a Global Impression of Change scale. The MID was defined as the difference in mean change from baseline between those who indicated they were "a little better" and those who reported "no change" on the Global Impression of Change scale 3 months after treatment. The effect size and SE of measurement were the distribution methods used.

Results: The mean changes (SD) in FISI, CRADI, CRAIQ, and MMHQ scores from baseline to 3 months after treatment were −8.8 (12.0), −52.7 (70.0), −60.6 (90.0), and −12.6 (19.2), respectively. The anchor-based MID estimates suggested by an improvement from no change to a little better were −3.6, −11.4 and −4.7, −18.1 and −8.0, and −3.2 for the FISI, CRADI (long and short version), CRAIQ (long and short version), and MMHQ, respectively. These data were supported by 2 distribution-based estimates.

Conclusions: The MID values for the FISI are −4, CRADI (full version, −11; short version, −5), CRAIQ (full version, −18; short version, −8), and MMHQ −3. Statistically significant improvements that meet these thresholds are likely to be clinically important.

Key Words: fecal incontinence, health-related quality of life, minimum important difference, questionnaire, responsiveness, patient-reported outcomes

(*Female Pelvic Med Reconstr Surg* 2014;20: 342–348)

Patient perception of symptom improvement is central to clinical care and important in the treatment of fecal incontinence (FI), a disorder that is known to degrade quality of life for affected women. Patient-physician conversation usually suffices for routine clinical care; however, research requires the use of valid, reliable, and responsive instruments that measure health-related quality of life (HRQOL) and symptom severity. Multiple instruments are available to measure severity and HRQOL for patients with FI.[1] The St Mark's Score[2] and the FI Severity Index (FISI)[3] are 2 psychometrically robust instruments for assessing symptom severity. Other patient-reported outcome measures that can be used for patients with FI include the Modified Manchester Health Questionnaire (MMHQ),[4] the Pelvic Floor Distress Inventory (PFDI), and the Pelvic Floor Impact Questionnaire (PFIQ). The PFDI and PFIQ are comprehensive questionnaires for women with pelvic floor disorders. The Colorectal-Anal Distress Inventory (CRADI) and the Colorectal-Anal Impact Questionnaire (CRAIQ) are colorectal scales within the PFDI and PFIQ that measure symptom bother and life impact of lower gastrointestinal tract complaints, respectively.[5] Despite progress in assessing HRQOL, use of these validated instruments is limited without an understanding of whether changes in an instrument's score are associated with clinically meaningful improvements by patients. The smallest change in score that is associated with a clinically meaningful improvement is called the minimum clinically important difference (MID).[6] A scale's MID is essential knowledge for appropriate interpretation of within-group changes or between-group differences in patient-centered outcomes research.

The objective of this study was to estimate the MID using both anchor-based and distribution-based approaches for the FISI, CRADI, CRAIQ, and the MMHQ in women receiving nonsurgical or surgical management of FI.

MATERIALS AND METHODS

This was an ancillary analysis of the Adaptive Behaviors among women with Bowel Incontinence (ABBI) study, a multicenter prospective cohort study, conducted by the Pelvic Floor Disorders Network, a clinical trials network funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, consisting of 7 clinical sites and a data coordinating center. The ABBI study investigated adaptive behaviors among women receiving nonsurgical or surgical management for FI per usual care practice, including behavioral

From the *Cleveland Clinic, Cleveland, OH; †University of Michigan; Ann Arbor, MI; ‡Department of Veteran Affairs and University of Alabama at Birmingham, Birmingham, AL; §Loyola University Chicago, Chicago, IL; ||University of California San Diego, San Diego, CA; ¶Eunice Kennedy Shriver National Institutes of Health, Bethesda, MD; #University of Texas Southwestern, Dallas, TX; **Duke University, Durham, NC; and ††Department of Veteran Affairs and University of Utah, Salt Lake City, UT.

Reprints: John Eric Jelovsek, MD, MMEd, Cleveland Clinic, 9500 Euclid Ave, Desk A81, Cleveland, OH 44195. E-mail: jelovsj@ccf.org.

The authors have declared they have no conflicts of interest.

Supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the NIH Office of Research on Women's Health (U01 HD41249, U10 HD41250, U10 HD41261, U10 HD41267, U10 HD54136, U10 HD54214, U10 HD54215, and U10 HD54241). Dr Tuteja (Consultant: Salix Pharmaceuticals, Guidepoint and GIG Consultancy; Forest Ironwood, Proctor & Gamble Pharmaceuticals) and Dr. Siddiqui (Grants: Medtronic, Astellas).

This trial is registered at clinicaltrials.gov as #NCT00729144.

Presented at the 2010 International Continence Society/International Urogynecology Association Joint Meeting Toronto, Canada.

All authors (Jelovsek, Chen, Markland, Brubaker, Dyer, Meikle, Rahn, Siddiqui, Tuteja, and Barber) substantially contributed to the following: conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

DOI: 10.1097/SPV.0000000000000078

techniques, pelvic floor muscle exercises, medications, surgery, or a combination of these treatments. Women were eligible if they had a primary complaint of FI consisting of liquid stool, solid stool, or mucus occurring at least monthly for 3 consecutive months and were planning to have treatment of FI. Each clinical site and the data coordinating center received institutional review board approval, and all subjects provided written informed consent.

Subjects completed several questionnaires including patient characteristics, previous treatments of FI, the Bristol Stool Form Scale to assess stool consistency, and several self-administered questionnaires to assess bowel symptom severity and HRQOL including the FISI, CRADI, CRAIQ, and MMHQ at baseline. Three and 12 months after the initiation of treatment, a Global Impression of Change scale was completed, in addition to repeating the 4 FI scales. Only those subjects who received FI treatment and completed at least 1 of the 4 scales at both the baseline and 3-month follow-up evaluations were included in this analysis.

The MMHQ is a validated questionnaire for HRQOL in women with FI.[4] It measures 8 domains (31 questions) including general health, incontinence impact, role, physical function, social function, personal function, emotional problems, sleep, energy, and the severity, which is assessed using the FISI. Total scores and scores for specific domains are measured on a 5-point scale ranging from 1 indicating "never" to 5 indicating "always." Total scores and domain scores both range from 0 to 100. The MMHQ was chosen for HRQOL measurement over the fecal incontinence quality of life scale based on a previous validation study among women with pelvic floor disorders.[4]

As part of the US English modification of the MMHQ, the FISI is contained within the MMHQ. The FISI is composed of 4 questions on incontinence of gas, mucus, liquid stool, and solid stool.[3] There are 2 highly correlated subscales that have been developed for the FISI, 1 from the patient's perspective of bother and 1 from the doctor's perspective of bother. For this study, all questionnaires were administered by telephone via a central interviewing center. Patients are asked to indicate the frequency of symptoms experienced using the following scale: 2 or more times a day, once a day, 2 or more times a week, once a week, 1 to 3 times per month, or never. Responses are weighted on a 1 to 20 severity scale, and a total FISI score is calculated from the patient's and surgeon's perspectives. Although the FISI score has a weighted scoring scale and has been validated, it lacks responsiveness data limiting its use in interventional studies.

The PFDI and the PFIQ,[5] each with 3 scales (urinary, colorectal, and pelvic organ prolapse), are condition-specific quality of life instruments developed to broadly assess pelvic floor disorders in women. They are available in long-form and short-form versions and have each been shown to be psychometrically valid, reliable, and responsive to change.[5,7–9] These instruments are based on the structure and content of 2 widely used condition-specific quality of life questionnaires for women with lower urinary tract dysfunction, the Urogenital Distress Inventory and the Incontinence Impact Questionnaire, originally described by Shumaker et al.[10] The content to measure bowel dysfunction was modified from these original urinary incontinence questionnaires, which are called the CRADI and the CRAIQ. The 4 validated FI measures have ranges as follows: FISI, 4 to 59; CRADI, 0 to 400; CRAIQ, 0 to 400; and MMHQ, 0 to 100. In all measures, a higher score indicates worse symptom bother or greater impact of symptoms on daily functioning.

A variety of methods have been used to determine MID, and no single method has proven to be superior.[6] These methods can be separated into 2 different categories—anchor-based methods and distribution-based methods. Anchor-based methods determine clinically meaningful changes in relation to an external indicator, an independent "anchor" that has clinical relevance to the outcome; in contrast, distribution-based methods evaluate the meaningfulness of changes based on the statistical characteristics of a population and relate clinical significance to statistical parameters of the obtained sample, such as sample variability (eg, baseline variation for effect size) or reliability (eg, Cronbach α coefficient used in the SE of measurement [SEM]).[6,11–13]

In this study, we used both anchor-based and distribution-based approaches and then considered all estimates to determine the MID for each of the FI measures. For the anchor-based approach, we used the Global Impression of Change scale as the primary anchor.[13] When calculating an MID, it is important to identify the subset of people who have experienced minimal change.[13] Patients who report getting "a little better" constitute the minimal change subgroup.[13] For this study, the smallest change on the Global Impression of Change scale was defined as the difference in mean change in scores (from baseline to 3 months) between those who indicated that they were a little better and those who reported "no change" on the Global Impression of Change scale.[13] To assess the usefulness of the anchor, the correlation between the Global Impression of Change at 3 months and the change from baseline to 3 months in each of the FI measures was calculated. If the correlation coefficient was significantly greater than 0.3, a cutoff value for a moderate association, the anchor-based analysis was performed.[13] In addition, some investigators have recommended to estimate the MID based on several anchor-based methods, with relevant clinical or patient-based indicators.[13] Therefore, we also estimated the MID of the CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ using the results of the FISI as an anchor, with the MID defined as an improvement from no change to 1 level of improvement in the FISI. The FISI was added as secondary anchor, in addition to the Global Impression of Change, to provide a second external patient-reported clinical anchor to estimate the MID of the remaining instruments.

For the distribution-based approaches, we determined the change from baseline to 3 months in scale scores that corresponded to an effect size of 0.5 baseline SD (1/2 SD), which is considered a conservative estimate of MID for patient-reported questionnaires, and 1 SEM, also shown to be a valid estimate of MID.[6,11,12,14] The effect size was based on the baseline variation of the sample representing a standardized measure of change over time.[6,11,12] The SEM based on the measurement precision of the instrument was calculated by the product of the baseline SD and the square root of 1 minus its reliability coefficient, estimated using Cronbach α coefficient in this analysis (1 SEM = baseline SD $\times$ square root [1 $-$ Cronbach α at baseline]).[6,11,12,14]

The 95% confidence interval (CI) was also estimated for each approach. The 95% CI for the anchor-based MID was based on the standard 95% CI for the difference between 2 population means; for the distribution-based 0.5 SD, it was the standard 95% CI for the population variance.[15] The 95% CI for distribution-based 1 SEM was constructed using the bootstrap resampling approach.[16] For each FI measure, the anchor-based and the distribution-based MID estimates along with the 95% CI's were compared based on triangulation, a well-recognized integrated system that makes use of multiple methods to hopefully converge on a small range of values (or 1 single value).[6,13] The recommended MIDs were selected for the FISI, CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ. Analyses were performed in SAS 9.2 for Windows

TABLE 1. Baseline Demographics by Subjects Included and Excluded from Analysis

Characteristic	Analyzed n = 83, n (%)	Not Analyzed n = 50, n (%)	*P**
Age, mean (SD), y	58.9 (13.4)	54.9 (15.0)	0.12
Parity, median (range)	2.0 (9.0)	3.0 (8.0)	0.2
BMI, mean (SD), kg/m^2	28.7 (7.0)	29.2 (6.6)	0.7
Race			0.2
White	71 (85.5%)	43 (86.0%)	
Black/African American	11 (13.3%)	4 (8.0%)	
Asian	1 (1.2%)	1 (2.0%)	
Other		2 (4.0%)	
Highest education, %			0.6
Lower than high school	7 (8.5%)	2 (6.3%)	
High school	22 (26.8%)	9 (28.1%)	
Some college	25 (30.5%)	12 (37.5%)	
College graduate	17 (20.7%)	3 (9.4%)	
Graduate or professional degree	11 (13.4%)	6 (18.8%)	
Health insurance, %			0.15
Private insurance	28 (33.7%)	23 (46.0%)	
Health maintenance organization	10 (12.0%)	2 (4.0%)	
Medicaid	1 (1.2%)	4 (8.0%)	
Medicare	23 (27.7%)	9 (18.0%)	
Self-pay (without insurance)	2 (2.4%)	1 (2.0%)	
Other	18 (21.7%)	11 (22.0%)	
Do not know/refused	1 (1.2%)		
Tobacco			0.8
Current smoker	9 (11.0%)	3 (8.3%)	
Never smoked	46 (56.1%)	21 (58.3%)	
Quit <6 mo ago	2 (2.4%)	2 (5.6%)	
Quit ≥6 mo ago	25 (30.5%)	10 (27.8%)	
Diabetes			1.0
Yes	11 (13.6%)	4 (11.1%)	
No	70 (86.4%)	32 (88.9%)	
Connective tissue disease			1.0
Yes	2 (2.5%)	1 (2.9%)	
No	79 (97.5%)	33 (97.1%)	
Typical bowel movement (Bristol Stool Scale scores)			0.3
Type 1	12 (14.5%)	1 (2.8%)	
Type 2	7 (8.4%)	3 (8.3%)	
Type 3	11 (13.3%)	5 (13.9%)	
Type 4	22 (26.5%)	17 (47.2%)	
Type 5	11 (13.3%)	4 (11.1%)	
Type 6	13 (15.7%)	4 (11.1%)	
Type 7	7 (8.4%)	2 (5.6%)	
Previous physical therapy for FI			0.5
Yes	7 (8.4%)	5 (13.9%)	
No	76 (91.6%)	31 (86.1%)	
Previous biofeedback for FI			0.4
Yes	7 (8.4%)	1 (2.9%)	
No	76 (91.6%)	34 (97.1%)	
Previous surgery for FI			1.0
Yes	4 (4.9%)	1 (2.8%)	
No	78 (95.1%)	35 (97.2%)	
Previous surgery on rectal/anal area			0.4
Yes	13 (15.9%)	3 (8.3%)	
No	69 (84.1%)	33 (91.7%)	

*Based on χ^2 test for discrete outcomes and 2-sample *t* test for continuous outcomes.

(SAS Inc, Cary, NC), and plots were produced in R 2.7.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Eighty-three (62%) of the 133 women participating in the ABBI study completed at least 1 of the 4 FI scales at both baseline and 3 months after receiving FI treatments per routine clinical care and are the subjects of this analysis. Of these, 100% completed the FISI and CRADI, 92% the CRAIQ, and 80% the MMHQ. Subjects had a mean age (SD) of 59 (13) years, a median (range) parity of 2 (9) deliveries, and most were white (86%). When comparing the ABBI women who were eligible for this analysis (n = 83) with those who were not (n = 50), there were no significant differences in baseline demographics, medical and smoking history, typical stool consistency as characterized by the Bristol Stool Scale, or prior treatments of FI ($P > 0.1$; Table 1).

Multiple different FI therapies were received by participants, with several having more than 1 therapy, including medications (n = 46), pelvic floor exercises (n = 39), surgery (n = 20), biofeedback (n = 10), and "other" (n = 6) per usual care. Overall, significant improvements in FI symptom severity and in HRQOL were observed on all measures 3 months after these interventions compared with baseline ($P < 0.0001$). Mean changes (SD) in FISI, CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ scores were −8.8 (12.0), −52.7 (70.0) and −16.03 (21.5), −60.6 (90.0) and −17.04 (25.6), and −12.6 (19.2), respectively.

Changes from baseline to 3 months after treatment in scale scores are presented in Table 2, stratified by patients' perception of degree of symptom improvement as measured by the Global Impression of Change. The Global Impression of Change and each of the instruments showed correlations above the established threshold of $r = 0.3$. The anchor-based MID (95% CI) estimates suggested by an improvement from no change to a little better for the FISI, CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ are shown in Table 2. These final anchor-based MID estimates were −3.6, −11.4 and −4.7, −18.1 and −8.0, and −3.2 for the FISI, CRADI (long and short version), CRAIQ (long and short version), and MMHQ, respectively (Table 2).

Using a distribution-based method, the MID (95% CI) estimates corresponding to a medium (0.5 SD) effect size and the MID estimates corresponding to 1 SEM are demonstrated in Table 3. Although the MIDs using the 0.5 SD and 1 SEM approaches in Table 3 may seem slightly different for some measures, their respective 95% CIs overlap one another.

The MID estimates for the CRADI, CRAIQ, and MMHQ were also calculated using the degree of improvement in FI severity level as the anchor. The MID (95% CI) estimates suggested by an improvement from no change to 1 level of improvement in the FISI were as follows: −9.11 (−19.7 to 1.5) and −21.6 (−55.0 to 11.9) for CRADI (short and long versions), −5.9 (−20.9 to 9.2) and −11.5 (−59.6 to 36.7) for CRAIQ (short and long versions), and −7.1 (−15.7 to 1.4) for MMHQ (Table 4).

Both the anchor-based and distribution-based estimates for the MIDs of the 4 scales (with 95% CI) are illustrated graphically based on the triangulation methodology in Figure 1. Here, the MID estimate corresponding to 0.5 SD effect size generally seems most conservative, whereas the MID estimate corresponding to 1 SEM closely shadowed the anchor-based estimate. Final MIDs of the scales reflect only the anchor-based estimates using the Global Impression of Change scale.

DISCUSSION

Patient-reported outcomes represent 1 of the most important measures for evaluating the effectiveness of treatments of FI. However, these outcome measures must have evidence of reliability and validity to be useful in assisting patients' and clinicians' understanding of the impact of treatment. Responsiveness is an aspect of construct validity and is determined by evaluating the relationship between changes in clinical and patient-based end points and changes in outcome scores over time using observational studies or clinical trials.[13] Responsiveness is demonstrated by showing that scale's scores are sensitive to actual changes in health status as demonstrated in Table 2. Although responsiveness is a key component to establishing a scale's construct validity, it is also important to determine the MID to assist in interpreting statistically significant results in clinical trials. Thus, responsiveness represents the instrument's ability to detect changes, whereas the MID denotes the smallest score or change in score that would likely be important from the patient's or clinician's perspective. In the

TABLE 2. Changes in FISI, CRADI, CRAIQ, and MMHQ Stratified by Global Impression of Change With Anchor-Based MIDs

		FISI Patient Weights		FISI Surgeon Weights		CRADI	
Global Impression of Change	n	3 Months Baseline Mean (SD)	Effect Size	3 Months Baseline Mean (SD)	Effect Size	3 Months Baseline Mean (SD)	Effect Size
Very much better	19	−18.79 (12.05)	−1.41	−20.16 (11.95)	−1.65	−93.77 (65.21)	−1.37
Much better	22	−11.23 (8.55)	−0.95	−11.77 (9.60)	−1.03	−76.62 (88.67)	−0.86
A little better	26	−4.73 (8.31)	−0.41	−4.50 (7.53)	−0.41	−27.84 (33.26)	−0.36
No change	12	−1.17 (12.76)	−0.09	−0.92 (13.90)	−0.07	−16.41 (60.97)	−0.2
A little worse	2	−7.50 (2.12)	−0.82	−6.50 (2.12)	−0.61	−8.57 (9.09)	−0.39
Very much worse	2	13.00 (8.49)	0.63	10.00 (7.07)	0.57	17.02 (36.53)	0.18
MID*	**38**	**−3.56**		**−3.58**		**−11.43**	
95% CI		−10.56 to 3.43		−12.75 to 5.58		−51.66 to 28.80	

*MID for improvement, difference in a little better and no change category mean changes from baseline. Effect size = (3 months mean − baseline mean)/baseline SD.

NA indicates not available.

TABLE 3. Changes in Scores and Distribution-Based MIDs for FISI, CRADI, CRAIQ, and MMHQ

	N	Baseline Mean (SD)	3 Months Mean (SD)	3 Months Baseline Mean (SD)	*P**	MID: 0.5 SD† (95% CI)	MID: 1 SEM‡ (95% CI)
FISI patient weights	83	26.7 (12.1)	17.9 (12.4)	−8.8 (12.0)	<0.0001	−6.1 (−7.1 to −5.3)	−7.7 (−8.6 to −6.9)
CRADI	83	133.5 (80.9)	80.9 (68.1)	−52.7 (70.0)	<0.0001	−40.4 (−47.7 to −35.1)	−33.3 (−36.8 to −30.0)
CRADI short form	83	42.1 (22.6)	26.0 (21.6)	−16.0 (21.5)	<0.0001	−11.3 (−13.4 to −9.8)	−12.7 (−13.7 to −11.9)
CRAIQ	76	150.0 (118.4)	89.4 (103.3)	−60.6 (90.0)	<0.0001	−59.2 (−70.5 to −51.1)	−19.9 (−25.6 to −14.0)
CRAIQ short form	76	43.4 (31.8)	26.4 (28.1)	−17.0 (25.6)	<0.0001	−15.9 (−18.8 to −13.8)	−9.2 (−10.3 to −8.3)
MMHQ	66	37.7 (24.7)	25.1 (23.1)	−12.6 (19.2)	<0.0001	−12.3 (−14.9 to −10.5)	−6.0 (−7.1 to −5.0)

*P values associated with the paired t test.

†0.5 SD = 0.5 × baseline SD.

‡1 SEM = baseline SD × square root (1 − Cronbach α at baseline).

gastrointestinal literature, MIDs have been determined for the Nepean Dyspepsia Index and the St Mark's (Vaizey) Score, outcome measures for dyspepsia and FI, respectively.[17,18] This study has determined that reasonable estimates of MID are 4, 11 and 5, 18 and 8, and 3 points for the FISI, CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ, respectively. Statistically significant improvements on these scales that meet these thresholds are likely to be clinically important and would assist the patient and physician in making treatment decisions.[12,19]

Because no one method for determining MID has been shown to be superior, we used a combination of methods including 2 anchor-based approaches and 2 distribution-based approaches to determine the MID for each scale. The application of multiple methods to determine the MID in a specific patient population almost always results in a range of values for the MID as seen in this study. This integrated approach using multiple methods is referred to as triangulation (ie, examining multiple values from different approaches and converging on a single MID value or narrow range of values).[13] Once the different MID estimates are determined, they may be graphed[13,20] to depict the ranges of estimates with a single MID value (or narrow range of MID values) determined for each scale with the anchor-based estimates assigned the most weight.[13]

For the primary anchor-based approach, we used a Global Impression of Change scale, the most commonly used type of anchor for MID determination, which represents the best measure of change from an individual perspective.[13] We chose to use the anchor-based method to determine the MID because it is recommended that the patient's perspective be given the most weight, although the clinician's perspective should also be considered. For the CRADI, CRAIQ, and MMHQ, we also used a second anchor, the FISI, which revealed similar results to those of the Global Impression of Change. When it is difficult or there is uncertainty about which MID may be best, it is recommended to base the final selection of MID values on some systematic review and evaluation process such as the modified Delphi method for achieving consensus, which we did not have to use in this study.[13]

These anchor-based estimates were subsequently supported using 2 distribution-based estimates, effect size (1/2 SD), and 1 SEM. Distribution-based methods are anchor free and convey the notion that an MID can be estimated based on the distribution of observed scores.[13] It is important to distinguish distribution-based from anchor-based methods; distribution-based methods are simply a way of expressing responsiveness in that they describe an observed change using a standardized metric, and they provide some insight into how large or small that change is. Distribution-based methods, unlike anchor-based methods, do not inform us about whether the observed change is important from the patient's or clinician's perspective thus representing an MID. Nevertheless, there are increasing body of evidence and growing consensus that an effect size of 0.5 (or change of 1/2 SD) is a conservative estimate that is likely to be clinically significant across different patient-reported questionnaires and that in the absence of other

CRADI Short Form		CRAIQ		CRAIQ Short Form		MMHQ	
3 Months Baseline Mean (SD)	Effect Size	3 Months Baseline Mean (SD)	Effect Size	3 Months Baseline Mean (SD)	Effect Size	3 Months Baseline Mean (SD)	Effect Size
−32.07 (17.80)	−1.57	−105.90 (96.58); n = 17	−0.96	−32.54 (27.80)	−1.03	−28.19 (24.02); n = 14	−1.05
−23.30 (24.54)	−0.94	−83.16 (100.56); n = 19	−0.63	−18.62 (26.11)	−0.6	−16.99 (21.68); n = 16	−0.63
−7.06 (15.28)	−0.36	−43.58 (65.55); n = 24	−0.41	−13.57 (21.96)	−0.42	−6.57 (9.08); n = 24	−0.3
−2.38 (14.36)	−0.1	−25.49 (81.15)	−0.2	−5.56 (18.03)	−0.17	−3.41 (9.68); n = 9	−0.15
0.00 (4.42)	0	−5.33 (7.54)	−0.05	0.00 (0.00)	0	0.47 (6.70)	0.02
1.56 (2.21)	0.06	67.36 (8.84)	0.36	16.67 (23.57)	0.27	20.52 (NA); n = 1	NA
−4.68		**−18.09**		**−8.01**		**−3.15**	
−15.30 to 5.94		−69.09 to 32.90; n = 36		−22.76 to 6.74		−10.52 to 4.21; n = 33	

TABLE 4. Change in FISI Levels and Change in MMHQ, CRADI, and CRAIQ in All FI Subjects

	CRADI			CRAIQ			MMHQ		
Change in FI Severity Level From Baseline to 3 mo	**n**	**3 Months Baseline Mean (SD)**	**Effect Size**	**n**	**3 Months Baseline Mean (SD)**	**Effect Size**	**n**	**3 Months Baseline Mean (SD)**	**Effect Size**
No change	21	−30.46 (42.25)	−0.35	20	−37.33 (91.68)	−0.34	17	−6.18 (5.77)	−0.22
One-level improvement	31	−52.00 (77.23)	−0.74	28	−48.81 (73.92)	−0.51	24	−13.29 (19.31)	−0.72
Two-level improvement	30	−66.44 (75.57)	−0.79	27	−79.84 (87.33)	−0.65	24	−13.64 (20.19)	−0.55
MID*	52	−21.55		48	−11.48		41	−7.12	
95% CI		(−55.03 to 11.94)			(−59.64 to 36.69)			(−15.68 to 1.44)	

*MID for improvement, difference in "one-level improvement" and "no change category means." Effect size = (3 months mean − baseline mean)/baseline SD.

No change category represents the subjects whose maximum level change in any of FI types (solid, liquid, mucus, and gas) is zero (eg, 6 to 6, 5 to 5…).

One-level improvement category represents the subjects whose maximum level improvement in any of FI types (solid, liquid, mucus, gas) is 1 (eg, 6 to 5, 5 to 4, 4 to 3, 3 to 2, 2 to 1).

Two-level improvement category represents the subjects whose maximum level improvement in any of FI types (solid, liquid, mucus, gas) is 2 (eg, 6 to 4, 5 to 3, 4 to 2, 3 to 1).

information, 1/2 SD or 1 SEM is a reasonable and scientifically supportable estimate of a meaningful effect.[21,22] The MIDs for each scale using the anchor-based approach are, as predicted, less than that for 0.5 SD approach, a conservative estimate for MID, and consistent with the one estimated by using the 1 SEM distribution-based approach. Therefore, the distribution-based methods for determining MID support the MID derived from the anchor-based approach.

The strengths of this study are the use of multiple approaches to triangulate MID estimates, the use of validated and widely accepted patient-reported FI outcome measures, and the use of a longitudinal prospective cohort of subjects with FI who were treated with the variety of treatment options per usual care. Usual care cohort studies, such as ABBI, are suited for MID determination because mild to moderate treatment effects are more common than those seen in randomized, controlled clinical trials.

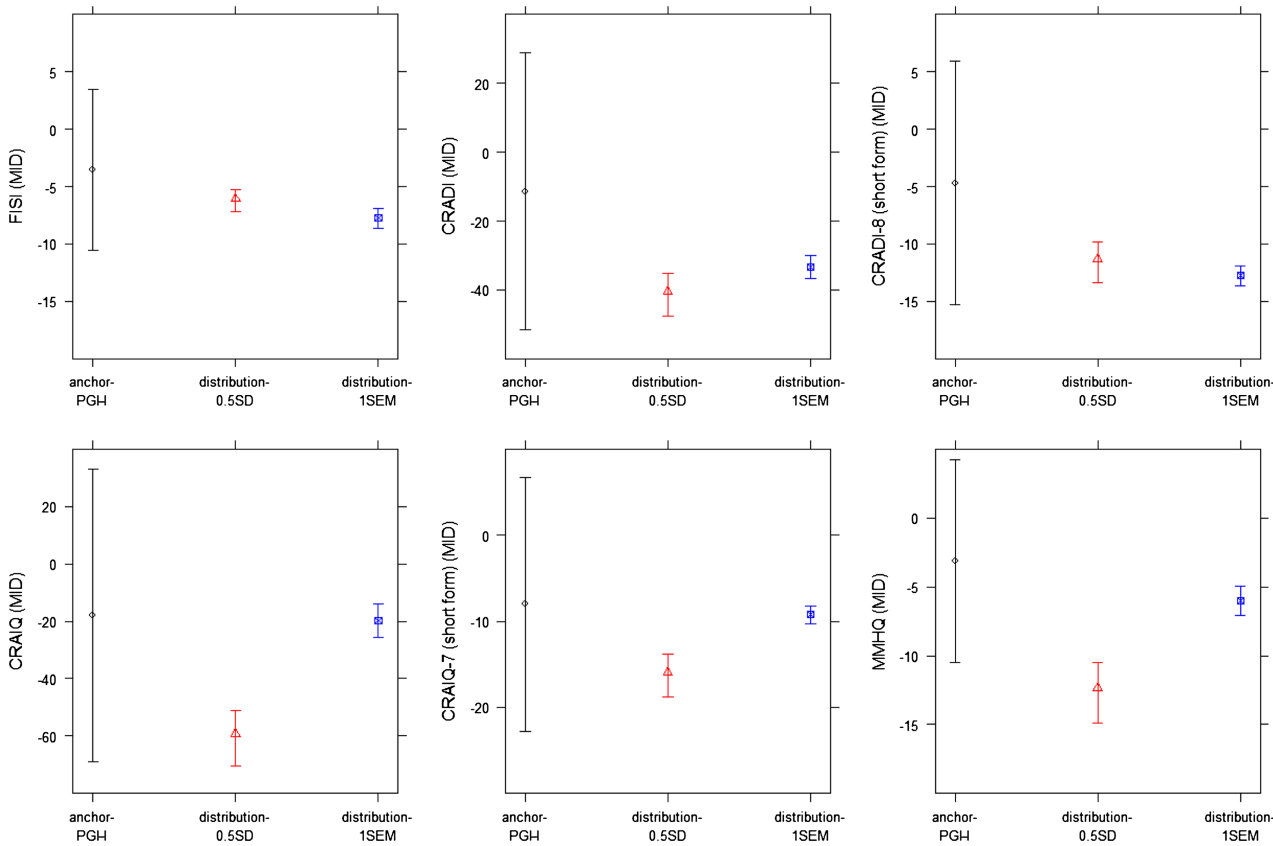

FIGURE 1. Estimates of MIDs for FISI, CRADI, CRAIQ, and MMHQ using anchor-based and distribution-based methods. The MID estimates using 3 different methods are presented for 4 questionnaires with 95% CIs. Anchor–Global Impression of Change, difference in a little better and no change category means from the Global Impression of Change (0.5 SD = 0.5 × baseline SD; 1 SEM = baseline SD × square root [1 − Cronbach α at baseline])

The most significant limitation in this study is the small sample size in the no change and a little better groups (n = 38) that was used for the anchor-based approach resulting in wide CIs around the MID for each scale. This is largely attributable to the relatively small sample sizes within the global rating of change categories of no change and a little better combined with missing data on several of the scales. Similar to virtually every measure used in medicine, all patient-reported outcome measures include some measurement error. One needs confidence that observed changes in scores over time with treatment or disease progression are not primarily attributable to error.[13] Confidence in the MID values of each scale from this study should evolve over time through additional research on different populations and contextual characteristics. As with other aspects of construct validity, responsiveness and the MID value are confirmed based on accumulating evidence from multiple studies, and with additional data, we can be more confident in the MID value.[13] The second limitation is that anchor-based approaches are 1 measure of the patient's perspective (global rating of change), but the study did not include any objective clinical measures of FI frequency or severity such as incontinent episodes recorded on a bowel diary. The use of these types of objective measures of FI might strengthen the validity of our findings, and future studies should incorporate these if possible. In addition, our MID values for the CRADI, CRAQI, and MMHQ when using the Global Impression of Change scale were approximately half the value of those generated using the degree of improvement in FI severity level as the anchor. Possible explanations for this may be that changes in symptom severity scales, such as FISI, have a lower degree of association with changes in quality of life scores when compared with global impression scales or that our sample size is again simply too low to determine if the resulting MIDs are actually different using the 2 types of scales as anchors. Finally, questionnaires for this study were administered by telephone, and MIDs may differ if questionnaires are administered in person.

From this study, we recommend that reasonable estimates of MID are 4, 11 and 5, 18 and 8, and 3 points for the FISI, CRADI (long and short versions), CRAIQ (long and short versions), and MMHQ, respectively. These MID estimates may provide the basis for interpreting clinical trial results and help clinicians and patients understand the effects of treatment and patient functioning in those with FI. These values may also be identified a priori by researchers in sample size planning for clinical trials of FI. When caring for patients with FI in routine clinical care or during observational studies or clinical trials, statistically significant improvements that meet or exceed these MID thresholds should be considered clinically important. However, some patients with changes in scores less than these estimates may perceive clinically important improvements. The values for these MID estimates should be refined with accumulating evidence from future studies.

REFERENCES

1. Rockwood TH, Church JM, Fleshman JW, et al. Fecal Incontinence Quality of Life Scale: quality of life instrument for patients with fecal incontinence. *Dis Colon Rectum* 2000;43(1):9–16.
2. Vaizey CJ, Carapeti E, Cahill JA, et al. Prospective comparison of faecal incontinence grading systems. *Gut* 1999;44(1):77–80.
3. Rockwood TH, Church JM, Fleshman JW, et al. Patient and surgeon ranking of the severity of symptoms associated with fecal incontinence: the fecal incontinence severity index. *Dis Colon Rectum* 1999;42(12):1525–1532.
4. Kwon S, Visco AG, Fitzgerald MP, et al., Pelvic Floor Disorders Network (PFDN). Validity and reliability of the Modified Manchester Health Questionnaire in assessing patients with fecal incontinence. *Dis Colon Rectum* 2005;48(2):323–331.
5. Barber MD, Kuchibhatla MN, Pieper CF, et al. Psychometric evaluation of 2 comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. *Am J Obstet Gynecol* 2001;185(6):1388–1395.
6. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56(5):395–407.
7. Barber MD, Walters MD, Cundiff GW, PESSRI Trial Group. Responsiveness of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ) in women undergoing vaginal surgery and pessary treatment for pelvic organ prolapse. *Am J Obstet Gynecol* 2006;194(5):1492–1498.
8. Wren PA, Janz NK, Brubaker L, et al. Reliability of health-related quality-of-life measures 1 year after surgical procedures for pelvic floor disorders. *Am J Obstet Gynecol* 2005;192(3):780–788.
9. Barber MD, Chen Z, Lukacz E, et al. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). *Neurourol Urodyn* 2011;30(4):541–546.
10. Shumaker SA, Wyman JF, Uebersax JS, et al. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res* 1994;3(5):291–306.
11. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;14(2):285–295.
12. Guyatt GH, Osoba D, Wu AW, et al., Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77(4):371–383.
13. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61(2):102–109.
14. Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. *J Clin Epidemiol* 2008;61(4):350–356.
15. Ott L. *An Introduction to Statistical Methods and Data Analysis*. North Scituate, MA: Duxbury Press, 1977.
16. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
17. Jones M, Talley NJ. Minimum clinically important difference for the Nepean Dyspepsia Index, a validated quality of life scale for functional dyspepsia. *Am J Gastroenterol* 2009;104(6):1483–1488.
18. Bols EM, Hendriks EJ, Deutekom M, et al. Inconclusive psychometric properties of the Vaizey score in fecally incontinent patients: a prospective cohort study. *Neurourol Urodyn* 2010;29(3):370–377.
19. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40(2):171–178.
20. Yost KJ, Cella D, Chawla A, et al. 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. *J Clin Epidemiol* 2005;58(12):1241–1251.
21. 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* 2003;41(5):582–592.
22. Wyrwich KW, Nienaber NA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37(5):469–478.