

The Minimally Important Difference in Clinical Practice for Patient-centered Outcomes Including Health Assessment Questionnaire, Fatigue, Pain, Sleep, Global Visual Analog Scale, and SF-36 in Scleroderma

SUNEET SEKHON, JANET POPE, Canadian Scleroderma Research Group, and MURRAY BARON

ABSTRACT. Objective. We studied a large clinical practice and multicenter database to estimate the minimally important difference (MID) in systemic sclerosis (SSc) using global rating of change anchors for the Health Assessment Questionnaire-Disability Index (HAQ-DI) and visual analog scale (VAS) in pain, fatigue, sleep, global status, and the Medical Outcomes Study Short-Form 36 (SF-36) in clinical practice.

Methods. Longitudinal data were collected from a scleroderma clinic on patients with scleroderma ($n = 109$) who had completed the HAQ-DI and pain/fatigue/sleep/global status VAS (0 to 100 mm) questionnaires at 2 consecutive visits, and rated their change in overall status since the last visit as much better, better, same, worse, or much worse. Data were extracted from the Canadian Scleroderma Research Group (CSRG) database ($n = 341$) for 2 consecutive annual visits where the patients had completed HAQ-DI and SF-36, and the SF-36 "change in health" item.

Results. For the single site, the mean baseline HAQ-DI was 0.895 and 0.911 at followup, with a mean change of 0.016. The MID estimates for improvement and worsening respectively were -0.0125 (0.125, 75th percentile)/ 0.042 (0.217, 75th percentile) for HAQ-DI, $-8.00/3.61$ for pain, $-10.00/3.79$ (25.32) for fatigue, $-18.50/5.92$ for sleep, and $-6.70/4.05$ for global VAS. In the CSRG, baseline scores were 0.787 for HAQ-DI, 37.20 for the Physical Component Summary (PCS) of SF-36, and 48.57 for the Mental Component Summary (MCS). The MID estimates for improvement and worsening were -0.037 (0.250, 75th percentile)/ 0.140 (0.375, 75th percentile) for HAQ-DI, $2.18/-1.74$ for PCS, and $1.33/-2.61$ for MCS.

Conclusion. This study provides MID estimates in SSc from 2 large databases for commonly used patient-reported outcomes in a clinical practice setting, which could differ from MID in trials. (First Release Jan 15 2010; J Rheumatol 2010;37:591-8; doi:10.3899/jrheum.090375)

Key Indexing Terms:

SCLERODERMA SYSTEMIC SCLEROSIS MINIMAL IMPORTANT DIFFERENCE
MINIMAL CLINICALLY IMPORTANT DIFFERENCE PATIENT REPORTED OUTCOMES

Scleroderma is a rare connective tissue disease characterized by collagen overproduction, as well as vascular and immunological abnormalities^{1,2}. The collagen overproduction leads to thickening or hardening of the skin, and in systemic cases can also involve the internal organs³. Systemic

sclerosis (SSc) involves both skin and potentially internal organs, with Raynaud's phenomenon commonly being present as well^{1,3}. Internal organ manifestations of SSc include damage to the lungs, the gastrointestinal system, the heart and kidneys, with pulmonary complications being the leading cause of death from scleroderma³. Depending on the extent of skin involvement, SSc can be further subdivided into limited and diffuse⁴. The diffuse form is more severe, with proximal and distal skin involvement and increased risk of internal organ manifestations, while the limited form has skin involvement distal to knees and elbows; the face and neck can be involved in both forms^{1,3,4}.

Because of the multisystem expression of SSc, it can greatly affect a patient's functioning and quality of life. Patient-centered outcomes are important in trials in scleroderma as well as clinical practice. The Health Assessment Questionnaire Disability Index (HAQ-DI) is one of the most commonly used measures of function in scleroderma⁵ and

From the Schulich School of Medicine and Dentistry, University of Western Ontario (UWO), St. Joseph's Health Care, London, Ontario; and SMBD-Jewish General Hospital, and McGill University, Montreal, Quebec, Canada.

The CSRG was funded in part by the Canadian Institutes of Health Research, the Scleroderma Society of Canada, The Ontario Scleroderma Society, and unrestricted grants from Actelion and Pfizer.

S. Sekhon, Medical Student, Schulich School of Medicine and Dentistry, UWO; J. Pope, MD, MPH, FRCPC, Professor of Medicine, Schulich School of Medicine and Dentistry, UWO, St. Joseph's Health Care; M. Baron, SMBD-Jewish General Hospital and McGill University.

Address correspondence to Dr. J. Pope, Division of Rheumatology, St. Joseph's Health Care, 268 Grosvenor Street, London, Ontario N6A 4V2, Canada. E-mail: janet.pope@sjhc.london.on.ca

Accepted for publication September 14, 2009.

has been validated for use in scleroderma patients⁶⁻⁸. It is a musculoskeletal-targeted self-report tool that assesses the functional status for performing the activities of daily living and is scored from 0 (no disability) to 3 (severe disability), representing the averaging of the worst score in 8 domains of daily functioning⁹. Visual analog scales (VAS) that assess disease activity in various domains have been demonstrated as being useful in SSc^{6,8}. In our study, VAS rated from 0 (none/no problem) to 100 mm (very severe/major problem) were used to assess pain, fatigue, sleep, and global status. The Medical Outcomes Study Short-Form 36 survey (SF-36) is another widely used general self-report tool for assessing health-related quality of life (HRQOL) that has been demonstrated as valid. It is composed of 36 items in 8 domains and can be summarized as the Physical Component Summary (PCS; 0–100) and Mental Component Summary (MCS; 0–100) scores¹⁰⁻¹³. Since these patient-centered outcomes are frequently used in assessing scleroderma patients, it is important to determine how much of a change in score is perceived as improved or worsened by the patient in clinical practice. The minimally important difference (MID) or minimal clinically important difference is the smallest change in a score that is meaningful for the patient and that would lead the physician to consider a change in management¹⁴. The MID is also useful in determining sample sizes for future studies, to see if differences between 2 treatment groups are clinically relevant¹⁵. It can help a healthcare provider by determining whether a treatment has been successful or whether a patient's state of health has changed.

The MID estimates are assessed using an anchor-based approach¹⁶. An “anchor” is a clinically relevant indicator or pointer to which an HRQOL change can be tied^{16,17}. These measures are of clinical relevance and can be “subjective,” such as self-reports of change, or “objective,” such as clinical indicators of response to treatment (a change in erythrocyte sedimentation rate, or a 6-minute walk distance). Subjective anchors rely on an individual's assessment of their disease. A global rating of change is a well accepted subjective anchor in HRQOL research¹⁸. It is an assessment of change in which a person thinks back to a previous time and states whether there has been a change in a domain of health from then until the present¹⁹.

Other studies have been conducted to estimate the MID for HRQOL measures in SSc^{13,20,21}. MID estimates have been calculated for the SF-6D (a selection of SF-36 items) and the European Quality of Life (EQ-5D) index, which are both preference-based health measures, using data from 8 longitudinal studies and the SF-36 “change in health” item as an anchor in addition to distribution-based methods. The MID of 0.041 was reported for the SF-6D (scored 0.29 to 1.00) and 0.074 for the EQ-5D (scored –0.59 to 1.00). Another study estimated the improvement MID for the SF-6D using data from 2 studies and also used the SF-36 “change in health” item as well as the HAQ-DI and skin

score as anchors¹³. Khanna, *et al*²¹ used data from a randomized controlled trial to estimate MID for the HAQ-DI and modified Rodnan Skin Score in patients with diffuse SSc using an investigator health-change rating as an anchor. The estimated MID for the HAQ-DI was 0.10 to 0.14 for improvement.

We aimed to determine the MID for various patient-reported outcomes (such as HAQ, pain, SF-36, and fatigue) in a clinical practice setting using patient information from 2 databases, anchored at different times of followup. It is possible that the MID results could be different in a clinical practice setting compared to a clinical trial setting and could vary over different times of followup. The variety of patient-reported outcomes studied as well as the clinic-based characteristic of the data in our study can provide additional information when interpreting scores in randomized controlled trials (RCT) and in clinical practice.

MATERIALS AND METHODS

Single-site study. Multiple data are collected routinely on patients seen at St. Joseph's Hospital Rheumatology Clinic, which is affiliated with the University of Western Ontario and serves a referral region of about 1 million. Ethics approval was obtained from the University of Western Ontario Ethics Committee. There were 148 patients who had been diagnosed with SSc by expert opinion. Most of the patients met the American College of Rheumatology (ACR) criteria²² and/or had CREST syndrome criteria (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasias), if ACR criteria were not met⁴ (because 12% of patients with SSc in the limited subset may not meet ACR criteria²³), and had at least 2 consecutive visits (mean interval 7.6 months for the single-site and 12 months for the multisite study but as long as 18 months), in order to limit recall bias for inclusion into the study. The patients were seen by 1 rheumatologist serially. Thirty-nine were excluded because of incomplete data, lack of consecutive visits, or too long an interval between visits, leaving a final sample of 109 patients. We did not perform a sample size calculation, but used instead an available sample as we did not know *a priori* the distribution of the data of interest. Data were extracted from medical charts by a trained data extractor and entered into a database (Excel and SPSS). At each visit, patients completed the HAQ-DI (0–3) and VAS for pain, fatigue, sleep, and global status, which ranged from 0 (none) to 100 mm (very severe). Patients also completed a 5-point Likert scale of change that asked, “How would you describe your overall status since the last visit?” on a scale labeled much better, better, the same, worse, much worse. This “change in health” question (on the second visit) was used as the anchor to estimate the MID. Patients who reported themselves to be better or worse on a followup visit were defined as the “minimally changed” subgroups. For each measurement, change scores were calculated as the difference in scores between 2 consecutive visits, with a negative HAQ-DI or VAS change signifying an improvement. The mean change scores for the “minimally changed” groups were used to estimate the MID. For HAQ-DI the MID was also calculated as the 75th percentile of change scores of “better” and “worse” subgroups, which is a method to estimate the MID described by Tubach, *et al*²⁴. We did not plan to do this *a priori*, but when we found that the MID for HAQ-DI were very small (below the instrument measurement detection), we calculated the MID for HAQ-DI by a 75th percentile method to determine if it gave potentially meaningful changes. All differences were visit 2 minus visit 1.

CSRG patients. The second set of data (as described²⁵) was from the Canadian Scleroderma Research Group (CSRG) registry, a multicenter database of annual visits collecting skin scores, SF-36, and data on healthcare utilization completed by physicians and patients for 749 patients, of

whom 341 had 2 consecutive visits and complete data. All patients had a diagnosis of SSc (ACR criteria or diagnosis by an expert), and annual visits were a mean of 12.5 months apart. Extracted data included HAQ-DI, the SF-36 “change in health” item, and the SF-36 PCS and MCS. A negative change signified an improvement for all the scores except PCS and MCS, for which the opposite was true. The anchor question was a 5-point Likert scale based on the SF-36 “change in health” item that asked, “Compared to one year ago, how would you rate your health in general now?” on a scale labeled much better now than 1 year ago, somewhat better now than 1 year ago, about the same as 1 year ago, somewhat worse than 1 year ago, much worse now than 1 year ago. Those patients who reported their health to be somewhat better or somewhat worse were defined as the “minimally changed” subgroups. The MID was estimated as the mean change in score for the minimally changed subgroups, while for HAQ-DI the 75th percentile of change for “better” and “worse” subgroups was also presented. It should be noted that the HAQ-DI used in this study for both groups was not a modified HAQ scale targeted specifically toward scleroderma (the SHAQ or Scleroderma HAQ is the HAQ-DI and VAS of various organ involvement or symptoms⁷).

Statistical analysis. The MID was estimated as the mean change in score for the “minimally changed” subgroups (improvement or worsening), as well as the 75th percentile of change score for HAQ-DI. Descriptive statistics were performed and 95% confidence intervals (CI) were calculated on change scores for the different subgroups. The mean change and SD in each group were calculated. In order to calculate the MID, there should be a significant association between the anchor and a change in score¹⁵. It is recommended that the anchor and change in the HRQOL score should have a correlation coefficient of > 0.371 to represent a “large effect” by Cohen’s rules of thumb^{15,26}. We assessed the association between anchor and change in scores using the Spearman correlation coefficient. The MID was estimated by examining changes in scales of interest in the patients who were slightly improved and slightly worsened in health compared to the previous visit (single-site) or 1 year ago (for CSRG). These estimates were compared to those who improved or worsened more than slightly.

As an exploratory analysis for the CSRG patients, results were stratified by limited versus diffuse and early (< 5 years) versus late disease duration for each subgroup. The data were analyzed using SPSS software. $P < 0.05$ was considered significant. Results were presented as mean (SD) unless otherwise specified.

RESULTS

Single-site patients. The 109 single-site chart patients had a mean age of 56.89 (SD 11.67) years, disease duration of 9.18 (6.60) years, and 84.4% were women (Table 1). The

mean interval between consecutive visits was 7.53 (3.33) months. The mean baseline HAQ-DI was 0.895 (0.672) and at followup was 0.911 (0.654). Patients were, on average, stable between visits with a negligible mean HAQ-DI change of 0.016 (0.277). The VAS scores presented are in millimeters, with each scale ranging from 0 (none) to 100 mm (very severe). The mean baseline pain VAS score was 41.18 (25.81) and at followup 41.39 (25.60), with a mean change of 0.21 (21.93); baseline fatigue VAS score was 46.28 (27.33) and at followup 49.29 (29.09), with a mean change of 3.02 (23.77); and initial sleep score was 41.82 (30.93) and at followup 40.50 (30.80), with little change between visits [−1.31 (26.34)]. The mean baseline global VAS score was 41.34 (25.96) and then 39.53 (25.49), with a change of −1.43 (22.35).

For change in the overall health status question, 57 (52.3%) reported they were the same, 38 (34.9%) reported they were worse, and 3 (2.8%) said they were much worse. In contrast, 10 (9.2%) reported they were better, and only 1 (0.9%) patient was much better. Characteristics of the single-site cohort are seen in Table 1.

In Table 2 the results for the mean changes in scores and 75th percentiles for HAQ-DI are presented for the same, better, worse, and much worse subgroups, with Spearman correlation coefficients. As there was only 1 patient in the “much better” subgroup, the data are not shown. Spearman correlation coefficients were calculated between the “change in overall status” anchor question and the mean change in each score. The mean changes followed the expected pattern for the same and minimally changed subgroups “better” and “worse” (that is, more change in the “better” or “worse” groups than in the “same” group, except for global VAS), with larger mean changes for patients perceiving improvement in the VAS outcomes but less change for improvement in the HAQ-DI. The HAQ-DI estimated as the 75th percentile of change score was also presented, and the magnitude was greater than the mean changes; however,

Table 1. Characteristics of the single-site SSc patients.

Characteristics	Mean (SD)		
Number analyzed	109		
Sex (% F, M)	84.4, 15.6		
Age, yrs	56.89 (11.67)		
Disease duration, yrs	9.18 (6.60)		
Visit time interval, mo	7.53 (3.33)		
Measurement scale	Baseline Mean (SD)	Followup Mean (SD)	Change Mean (SD)
HAQ-DI	0.895 (0.672)	0.911 (0.654)	0.016 (0.277)
Pain VAS, mm	41.18 (25.81)	41.39 (25.60)	0.21 (21.93)
Fatigue VAS, mm	46.28 (27.33)	49.29 (29.09)	3.02 (23.77)
Sleep VAS, mm	41.82 (30.93)	40.50 (30.80)	−1.31 (26.34)
Global VAS, mm	41.34 (25.96)	39.53 (25.49)	−1.43 (22.35)

SSc: systemic sclerosis; HAQ-DI: Health Assessment Questionnaire-Disability Index; VAS: visual analog scale.

the “same” and “better” subgroups both had the same positive number. Interestingly, the Spearman coefficients for all outcomes were low. The wide CI for some of the values represent uncertainty of the results, which may be due to small sample size in the improved subgroups.

CSRG patients. The 341 CSRG registry patients were 55.07 (SD 12.18) years old with a mean disease duration of 7.99 (7.54) years, and 88.2% were female, 42.59% (n = 138) had diffuse SSc, 57.41% (n = 186) had limited SSc, and the rest (n = 17) were not defined (Table 3). The mean interval between baseline and followup was 12.49 (2.39) months. The mean HAQ-DI was 0.787 (0.683) at baseline and 0.812 (0.699) at followup, with a mean change of 0.025 (0.383). The mean baseline scores for PCS and MCS were 37.20 (11.12) and 48.57 (11.54) and mean followup scores were 37.53 (10.75) and 49.30 (11.42), respectively.

For the “change in health” item anchor question, 51.9% (n = 177) reported they were about the same, 27.6% (n = 94) somewhat worse, and 2.3% (n = 8) much worse. In contrast, 13.8% (n = 47) were somewhat better and 4.4% (n = 15)

were much better. In Table 4 data are presented for the HAQ-DI mean changes in the different subgroups divided by the “change in health” anchor compared to 1 year ago. The expected direction of change was seen for about the same, somewhat better, and somewhat worse subgroups. For the HAQ-DI, 75th percentiles are also presented, which were greater in magnitude than mean changes, but there was a greater positive change for the “slightly improved” compared to the “same” subgroup.

The MID estimates for the SF-36 are presented separately in Table 5, since a positive change represented an improvement. The expected direction of change was seen for the same, better, and worse subgroups, except for the MCS, where the “about the same” subgroup had a greater positive change than the “somewhat better” subgroup.

CSRG patients – exploratory analysis. As an exploratory analysis, the MID was estimated for the CSRG patients stratified by diagnosis of limited or diffuse disease to assess for differences based on type of disease. The mean change MID estimates for the HAQ-DI were greater in magnitude

Table 2. Minimal important difference (MID) estimates for single-site SSc patients.

Patient-rated Overall Status	HAQ-DI Change, mean (SD) [95% CI]	HAQ-DI Change, 75th percentile [†]	Pain VAS ^{††} Change, mean (SD) [95% CI]	Fatigue VAS Change, mean (SD) [95% CI]	Sleep VAS Change, mean (SD) [95% CI]	Global VAS Change, mean (SD) [95% CI]
Better (n = 10)	−0.0125 (0.224) [−0.173 to 0.148]	0.125	−8.00 (32.48) [−31.23 to 15.23]	−10.00 (10.63) [−17.60 to 2.40]	−18.50 (33.91) [−42.76 to 5.76]	−6.70 (26.58) [−25.72 to 12.32]
Same (n = 57)	−0.009 (0.255) [−0.077 to 0.058]	0.125 (0.255)	−0.49 (17.25) [−5.07 to 4.09]	3.25 (21.60) [−2.48 to 8.98]	−3.05 (22.40) [−9.00 to 2.89]	−5.70 (19.48) [−10.87 to −0.53]
Worse (n = 38)	0.042 (0.314) [−0.061 to 0.145]	0.217	3.61 (24.52) [−4.46 to 11.67]	3.79 (25.32) [−4.53 to 12.11]	5.92 (28.83) [−3.55 to 15.40]	4.05 (22.35) [−3.29 to 11.40]
Much worse* (n = 3)	0.327 (0.263) [−0.327 to 0.981]	n is too small	0.67 (34.01) [−83.81 to 85.14]	45.33 (27.93) [−24.06 to 114.73]	3.00 (11.27) [−24.99 to 30.99]	36.33 (15.18) [−1.37 to 74.03]
Spearman Correlation Coefficient**	0.106		0.092	0.211 p < 0.05	0.256 p < 0.01	0.276 p < 0.01

* There was only 1 patient much better (not shown) and 3 patients much worse. ** Correlation is between anchor and mean change in score, significant at the level shown. [†] 75th percentile of HAQ-DI change score was taken for each subgroup except much worse because it contained only 3 patients. ^{††} VAS 0 to 100 mm. HAQ-DI: Health Assessment Questionnaire-Disability Index; VAS: visual analog scale.

Table 3. Characteristics of the CSRG patients.

Characteristics	Mean (SD)		
Number analyzed	341		
Sex (% F, M)	87.7, 11.7 (n = 339, 2 unknown)		
Age, yrs	55.07 (12.18)		
Disease duration, yrs	7.99 (7.54)		
Visit time interval, mo	12.49 (2.36)		
Measurement scale	Baseline Mean (SD)	Followup Mean (SD)	Change Mean (SD)
HAQ-DI	0.787 (0.683)	0.812 (0.699)	0.025 (0.383)
PCS	37.20 (11.12)	37.53 (10.75)	0.28 (7.12)
MCS	48.57 (11.54)	49.30 (11.42)	0.72 (10.18)

CSRG: Canadian Scleroderma Research Group; HAQ-DI: Health Assessment Questionnaire-Disability Index; PCS: Physical Component Summary; MCS: Mental Component Summary.

Table 4. HAQ-DI minimal important difference (MID) estimates for CSRG patients.

Anchor: Change in Health Over Past Year	HAQ-DI Change, mean (SD) [95% CI]	HAQ-DI Change, 75th percentile**
Much better (n = 15)	-0.283 (0.527) [-0.575 to 0.009]	0.000
Somewhat better (n = 47)	-0.037 (0.402) [-0.155 to 0.081]	0.250
About the same (n = 173)	0.0001 (0.343) [-0.050 to 0.053]	0.125
Somewhat worse (n = 93)*	0.140 (0.387) [0.060 to 0.219]	0.375
Much worse (n = 8)	0.156 (0.319) [-0.110 to 0.423]	0.250
Spearman correlation coefficient	0.227, $p < 0.01$	

** 75th percentile of HAQ-DI change score was taken for each subgroup. HAQ-DI: Health Assessment Questionnaire-Disability Index; CSRG: Canadian Scleroderma Research Group.

Table 5. Minimal important difference (MID) estimates for CSRG patients.

Anchor: Change in Health over Past Year	PCS Change, mean (SD) [95% CI]	MCS Change, mean (SD) [95% CI]
Much better (n = 15)	6.97 (7.97) [2.55 to 11.38]	7.37 (11.14) [1.20 to 13.5]
Somewhat better (n = 47)	2.18 (6.58) [0.25 to 4.11]	1.33 (7.78) [-0.95 to 3.61]
About the same (n = 173)	0.20 (6.34) [-0.75 to 1.15]	2.00 (9.31) [0.61 to 3.40]
Somewhat worse (n = 94)	-1.74 (7.72) [-3.32 to -0.16]	-2.61 (10.91) [-4.84 to -0.37]
Much worse (n = 8)	2.18 (8.10) [-4.59 to 8.95]	-3.88 (17.56) [-18.57 to 10.80]
Spearman correlation coefficient	-0.230, $p < 0.01$	-0.203, $p < 0.01$

Positive change signified worsening for all scores except PCS/MCS. CSRG: Canadian Scleroderma Research Group; PCS: Physical Component Summary; MCS: Mental Component Summary.

for the patients with diffuse SSc compared to the patients with limited SSc (Table 6). When calculated using 75th percentile, HAQ-DI estimates were similar for limited and diffuse disease. No clear pattern between limited and diffuse disease was seen for the other HRQOL measures studied

Table 6. CSRG HAQ-DI minimal important difference (MID) estimates for patient-reported outcomes stratified by subset of limited and diffuse SSc.

Patient-reported Outcome	Mean MID (SD) in "Somewhat better" Subgroup	Limited	Mean (SD) MID in "Somewhat worse" Subgroup	Mean MID (SD) in "Somewhat better" Subgroup	Diffuse	Mean (SD) MID in "Somewhat worse" Subgroup
		Mean (SD) Change in "Same" Subgroup			Mean (SD) Change in "Same" Subgroup	
HAQ-DI*	-0.022/0.250 (0.366) (n = 24)	-0.025/0.125 (0.292) (n = 100)	0.102/0.375 (0.391) (n = 49)	-0.057/0.250 (0.454) (n = 22)	0.032/0.125 (0.417) (n = 66)	0.207/0.500 (0.333) (n = 35)

* Presented as mean change/75th percentile (SD). CSRG: Canadian Scleroderma Research Group; HAQ-DI: Health Assessment Questionnaire-Disability Index.

(Table 7). Also, mean change in scores was calculated for patients stratified by early and late disease (data not shown), with a disease duration of 5 or fewer years being classified as "early" disease, where 3 out of 4 HRQOL measures that were analyzed showed the MID was greater in patients with early compared to late disease. These analyses were exploratory, so statistical tests were not done.

DISCUSSION

It is useful to determine the MID for various patient-centered outcomes in order to determine how much of a change in score is meaningful to the patient¹⁴⁻¹⁶. The MID can vary based on characteristics of the patient population, baseline scores, the expression of the disease (i.e., whether patients tend to improve or deteriorate), and the choice of anchors used to estimate the MID^{15,27}. Since there is uncertainty in a single MID estimate, it is important to establish estimates for different situations. In our study, the MID for several patient-reported HRQOL scores was estimated for 2 different anchor questions and times and mostly different populations: patients from a single rheumatology clinic and patients who were part of a multicenter database. Pain, fatigue, and sleep problems occurred frequently in the single-site cohort of SSc patients, with high baseline and followup values. Patients in both cohorts had low overall HAQ-DI scores, indicating relatively mild disability. The data were not evenly distributed: patients were skewed toward more being worse than improved at followup, a situation that is not surprising in SSc. The Spearman correlations were not large (in the range of a medium effect). Although it is recommended that correlation should be 0.30-0.35 or > 0.371 to establish an acceptable anchor, our correlations are still significant and were statistically associated with the outcomes of interest^{15,26}.

The MID was estimated for the HAQ-DI and pain, fatigue, sleep, and global VAS for the single-site patients using a retrospective self-report anchor. The VAS MID estimates were different bidirectionally, with the MID for improvement being of a greater magnitude than the MID for worsening. SSc clinic patients may be more able to detect a worsening compared to an improvement for these VAS scores. Correlations were weak for the HAQ-DI and pain VAS, making these MID estimates of uncertain significance,

Table 7. CSRG minimal important difference (MID) estimates for patient-reported outcomes stratified by subset of limited and diffuse SSc.

Patient-reported Outcome	Mean MID (SD) in "Somewhat better" Subgroup	Limited	Mean (SD) MID in "Somewhat worse" Subgroup	Mean MID (SD) in "Somewhat better" Subgroup	Diffuse	Mean (SD) MID in "Somewhat worse" Subgroup
		Mean (SD) Change in "Same" Subgroup			Mean (SD) Change in "Same" Subgroup	
PCS	2.25 (6.97) (n = 24)	0.585 (5.81) (n = 101)	-1.92 (7.27) (n = 50)	2.24 (6.42) (n = 22)	-0.585 (7.00) (n = 65)	-0.953 (8.41) (n = 35)
MCS	0.186 (5.53) (n = 24)	1.67 (8.69) (n = 101)	-2.74 (11.27) (n = 50)	2.61 (9.79) (n = 22)	3.05 (9.83) (n = 65)	-2.66 (10.72) (n = 35)

CSRG: Canadian Scleroderma Research Group; SSc: systemic sclerosis; PCS: Physical Component Summary; MCS: Mental Component Summary.

and they may have little clinical meaning. Other VAS scores had "medium effect" correlations with the anchor according to Cohen's rules of thumb²⁶. We expected from our other publications on MID in rheumatic diseases that fatigue and sleep disturbance may be poorly related to the disease²⁸⁻³¹. Also, although patients may not read the timeframe of the questions, we asked for the change in status "since last visit," but the HAQ asks "over the last week."

For the CSRG patients, MID was estimated based on the SF-36 "change in health" item as the anchor for the HAQ-DI (scored 0-3) as well as the PCS and MCS. As for the single-site patients, the HAQ-DI MID was of a lesser magnitude for improvement compared to worsening. The PCS and MCS did not appear to vary bidirectionally.

The data for the MID of HAQ-DI were not identical between the 2 databases. When calculated using the mean changes, the no-change group had similar insignificant mean changes in HAQ-DI, and the CI of the MID overlapped. It was found that the magnitude of the MID was about 3-fold higher for CSRG patients compared to single-site patients for both improvement and worsening using the first approach for MID calculation. This could be due to differences in time of followup, the anchor questions used, and the smaller sample size in the single-site study. However, it may be that with a weak correlation, the MID for HAQ-DI is unreliable. A similar pattern was seen in both patient groups, with the magnitude of the MID being lower for improvement compared to worsening. We also chose to calculate the HAQ-MID using 75th percentile of change score as described by Tubach, *et al*²⁴, because of the low MID found using mean changes (changes that were below the measurements of the scale, which can change in general by increments of 0.125). When calculated using the 75th percentile, the HAQ MID was higher in magnitude, and it remained that the estimates were lower for the single-site patients compared to CSRG and also lower for improvement compared to worsening. Results of MID using patient anchors for patient-reported outcomes may be both different in clinical practice than trials and difficult to interpret because of the wide variability of the data in each category of minimally changed patients.

The exploratory analysis looked at how MID was affect-

ed by limited versus diffuse SSc and early versus late SSc in the CSRG patients. It was seen that the MID estimate for the HAQ-DI was of a greater magnitude for diffuse compared to limited disease. This may be explained by the generally increased severity of disease in patients with diffuse SSc, resulting in a higher baseline HAQ score, and thus a greater change is accepted before it is reported on the anchor question. Looking at disease duration, for all measures except HAQ-DI (for which the pattern was unclear), it was found that early disease had a greater MID estimate compared to late disease. In other words, patients with late disease noticed a change that was on average smaller than early disease. This may be due to longer years with the disease leading to more stable self-report.

Some have used multiple anchors to estimate the MID, and the MID estimates could be different if other anchors such as laboratory tests or physician ratings were used^{15,16,32}. The study by Khanna, *et al* based on the D-penicillamine study found a HAQ-DI MID estimate of 0.10-0.14 for improvement. The anchor question was a 7-point scale that asked the physician to assess the patient's status at 6-month intervals during the 2-year study²¹. This estimate is greater in magnitude compared to the estimates found in our study; however, the 75th percentile estimates fall above or within the range of 0.10-0.14. The difference in MID may partly reflect the patient-centered versus physician-centered anchor, or the use of only early diffuse SSc in the RCT.

Although our data sets were relatively large with respect to MID studies, our study has some limitations. The standard deviations and CI were wide, so estimates are less certain. The low correlations were an unexpected result as both the anchor and outcomes were patient-centered. The lower than expected correlations may partly be due to the fact that the anchors asked about "overall status" or "general health," which may not necessarily be reflected by HAQ-DI or VAS in pain, fatigue, or sleep. It is important to realize that the correlations were statistically significantly related. Another limitation is the potential for recall bias, as the anchor questions led to a recall period of up to 18 months. In addition, the HAQ-DI and VAS, for which anchors were used to estimate MID, asked patients to recall over the past week. The

varying time window could affect the results, especially since 1 anchor was “since last visit” and in the CSRG group the health transition was “since the last year.” However, in RCT, the MID of HAQ is developed comparing an effect often from the beginning of the trial to the end, so this methodology is often used despite the HAQ asking about the last week. The retrospective self-report feature of the anchors leaves them open to recall bias, but this would be true of any patient-centered MID study³³. The anchors asked patients to remember the change since their last visit (single site) or since 1 year (CSRG). Walters and Brazier found correlations were moderate between a retrospective anchor and SF-6D (an HRQOL score) at followup and considerably lower at baseline³⁴. Thus, the answer to such an anchor question is more strongly influenced by the current situation, and this may have affected the correlations in our study. It was assumed that missing HRQOL information was missing at random, and differences in baseline scores may have affected the MID estimates. The MID in clinical practice may be different from what could be observed in trials as the expectations of the patients in clinical practice may be different (many worsening, but also a large proportion were stable), and in clinical trials, there is a population of patients meeting inclusion criteria, often with early disease or active disease. The study by Khanna, *et al*, which used clinical trial patients, found an improvement MID that was different from the estimates in our study and the population was also a subset of all SSc (early diffuse SSc)²¹. Perhaps in mostly prevalent SSc cases such as those we studied, the HAQ does not improve because of lack of reversibility (particularly as HAQ is weighted toward hand function and many had fixed contractures). Thus it is likely that the HAQ does not have reversibility in prevalent disease, so the HAQ may not operate well for improvements over 6 to 12 months once the disease is longstanding.

A double extraction procedure was not performed and this may add to the degree of error in the results. Some patients in the single-site cohort were also part of the CSRG database, but the overlap of patient outcomes was only for the HAQ-DI, which was measured at different times and with different anchor questions. Data were at 1 year for CSRG and about half a year for the single-site study. Other outcomes did not overlap as they were not identical between the 2 groups. The MID estimates may vary by limited and diffuse SSc and disease duration, but the analyses were exploratory only. MID can vary depending on the baseline score. For instance, the MID in a group with lower ratings may be different from those with high ratings^{28,29}.

Overall, our study provides MID estimates for clinic-based SSc patients for various HRQOL scores, data that will be useful for patient care and interpretation of clinical trials.

ACKNOWLEDGMENT

The Canadian Scleroderma Research Group (CSRG) investigators: J. Pope, London, Ontario; M. Baron, M. Hudson, J-P. Mathieu, S. Ligier, Montreal,

Quebec; J. Markland, Saskatoon, Saskatchewan; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; N. Khalidi, E. Kaminska, Hamilton, Ontario; P. Docherty, Moncton, New Brunswick; S. LeClercq, M. Abu-Hakima, Calgary, Alberta; A. Masetto, Sherbrooke, Quebec; C. Douglas Smith, Ottawa, Ontario; E. Sutton, Halifax, Nova Scotia; M. Fritzler, Advanced Diagnostics Laboratory, Calgary, Alberta.

REFERENCES

1. Silver RM. Clinical aspects of systemic sclerosis (scleroderma). *Ann Rheum Dis* 1991;50:854-61.
2. Denton CP, Black CM, Korn JH, de Crombrughe B. Systemic sclerosis: current pathogenetic concepts and future prospects for targeted therapy. *Lancet* 1996;347:1453-8.
3. Clements PJ. Systemic sclerosis (scleroderma) and related disorders: Clinical aspects. *Baillieres Best Pract Res Clin Rheumatol* 2000;14:1-16.
4. LeRoy EC, Black CM, Fleischmajer R, Jablonska S, Krief T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): Classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202-5.
5. Khanna D. Health-related quality of life: a primer with a focus on scleroderma. *Scleroderma Care Res* 2006;3:3-13.
6. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. *Arthritis Rheum* 2002;46:2410-20.
7. Poole JL, Williams CA, Bloch DA, Hollak B, Spitz P. Concurrent validity of the Health Assessment Questionnaire disability index in scleroderma. *Arthritis Care Res* 1995;8:189-93.
8. Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997;40:1984-91.
9. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
10. Ware JE Jr, Kosinski M, Keller S. SF-36 physical and mental health summary scales: a user's manual. Boston: New England Medical Center Health Institute; 1994.
11. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
12. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247-63.
13. Khanna D, Furst DE, Wong WK, Tsevat J, Clements PJ, Park GS, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res* 2007;16:1083-92.
14. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
15. Hays RD, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005;2:63-7.
16. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
17. Khanna D, Tsevat J. Health-related quality of life — an introduction. *Am J Manag Care* 2007;13 Suppl 9:S218-23.
18. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol* 2002;55:900-8.
19. Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers

- (FACT-BRM) instrument. *Value Health* 2005;8:117-27.
20. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-32.
 21. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. *Ann Rheum Dis* 2006;65:1325-9.
 22. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980;23:581-90.
 23. Winterbauer RH. Multiple telangiectasia, Raynaud's phenomenon, sclerodactyly, and subcutaneous calcinosis: A syndrome mimicking hereditary hemorrhagic telangiectasia. *Bull Johns Hopkins Hosp* 1964;114:361-83.
 24. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29-33.
 25. Thombs B, Hudson M, Taillefer S, Baron M, Canadian Scleroderma Research Group. Prevalence and clinical correlates of symptoms of depression in patients with systemic sclerosis. *Arthritis Rheum* 2008;59:504-9.
 26. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
 27. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research: how meaningful is it? *Pharmacoeconomics* 2000;18:419-23.
 28. Pope J, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the Health Assessment Questionnaire in rheumatoid arthritis is smaller than in randomized controlled trials. *J Rheumatol* 2009;36:254-9.
 29. Khanna D, Pope J, Khanna PP, Maloney M, Samadi N, Norrie D, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008;35:2339-43.
 30. Wheaton L, Pope J. The minimally important difference for patient centred outcomes in ankylosing spondylitis, including pain, fatigue, sleep and the Health Assessment Questionnaire [abstract]. *J Rheumatol* 2009;36:2579.
 31. Colangelo K, Pope J, Peschken C. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. *J Rheumatol* 2009;36:2231-7.
 32. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.
 33. Schwartz N, Sudman S. Autobiographical memory and the validity of retrospective reports. New York: Springer-Verlag; 1994.
 34. 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* 2003;1:1-8.