A comparison of the measurement properties and estimation of minimal important differences of the EQ-5D and SF-6D utility measures in patients with systemic sclerosis

L. Kwakkenbos¹,²,³, J. Fransen⁴, M.C. Vonk⁴, E.S. Becker⁵, M. Jeurissen¹, F.H.J. van den Hoogen¹, C.H.M. van den Ende¹

ABSTRACT
Objective. To compare measurement properties of the EQ-5D and SF-6D utility measures, to assess the association and agreement between these measures and to estimate minimal important differences (MID) in patients with systemic sclerosis (SSc).

Methods. Both measures were assessed twice in an observational prospective design over a 12-month period (n=211). Spearman’s rank correlation between the EQ-5D and SF-6D was calculated at baseline. Agreement was assessed using Lin’s concordance coefficient (LCC) and a Bland-Altman plot. MIDs were estimated using three anchors: the global rating of change item (SF-36) and changes on the Health Assessment Questionnaire-Disability Index (HAQ-DI) of ≥0.14 and ≥0.22.

Results. At baseline, the mean EQ-5D and SF-6D scores were 0.64 (SD=0.25) and 0.65 (SD=0.11), respectively. The correlation between EQ-D and SF-6D scores was r=0.74. Agreement was moderate (LCC=0.49), and the Bland-Altman plot showed a mean difference of 0.003 but wide limits of agreement (-0.38 to 0.39) and a structural bias for lower scores. The mean MID estimate for the EQ-5D was 0.08 for the improved subgroup, and -0.13 for the deteriorated subgroup. For the SF-6D, the MID estimate was 0.05 for the improved and -0.04 for the deteriorated subgroup.

Conclusion. Although there was a marked correlation between the measures, the moderate agreement implies that EQ-5D and SF-6D scores cannot be used interchangeably. The MID estimates we provided can be used to calculate sample sizes for clinical trials involving SSc patients, and in interpreting the relevance and importance of treatment effects.

Introduction
Self-administered utility measures are increasingly used in economic evaluations of treatments and policy-making (e.g. 1, 2). Utility measures like the EQ-5D (EuroQol 5D) (3, 4) and SF-6D (5) (derived from the Medical Outcomes Trust Short Form-36; SF-36) (6) are designed to assess value of health in a single summary measure, with a value of 0 for death and 1 for perfect health. Utility measures cover different domains of health-related quality of life that might be influenced by (chronic) diseases, including pain, physical limitations and mental health.

The EQ-5D and SF-6D are frequently used and considered as well-established measures of utility across a diversity of chronic diseases, including rheumatic diseases (7, 8). Although both utility measures seem promising and are often used interchangeably, it is not clear whether this is justified given the differences in scoring and reporting of low agreement between the two measures, especially for the lower and higher ranges of utility scores (9–11). Possible EQ-5D scores range from -.59 to 1.00 (negative scores reflect patient belief that health status is valued worse than death), while the SF-6D ranges from 0.29 to 1.00.

Because of the low agreement between measures, differences in calculations of cost-effectiveness outcomes have been reported (12). This decreases the comparability of (cost-) effectiveness studies tremendously and potentially leads to different resource allocation decisions. In previous studies, levels of agreement between the EQ-5D and SF-6D differed wildly across diseases and patient samples (11). Thus, the level of agreement should preferably be examined in a patient sample in which the
instruments will be applied as an outcome measure. Improvement of health related quality of life is one of the most important goals of pharmacological as well as non-pharmacological interventions in many chronic diseases, particularly when there is no cure, as in systemic sclerosis (SSc, scleroderma). SSc is a rare disease of connective tissue, characterised by abnormalities of the vascular and immune system (13). Fibrosis leads to thickening of the skin and might lead to significant organ dysfunction in the internal organs as well. Usually, two clinical subtypes are recognised based on the extent of skin involvement: limited SSc, with skin involvement distal to the elbows and knees, with or without face involvement and diffuse SSc, with skin involvement proximal to the elbows and knees and the trunk (14). SSc has an impact on physical as well as mental health related quality of life (15). Currently, there is no treatment available to cure SSc, but an increasing number of interventions are being developed and tested (e.g. 16-18). To improve health care, accurate measurement of outcomes is crucial. Studies assessing the psychometric properties of patient-reported outcome measures usually include properties such as construct validity, discriminative ability and responsiveness. In addition, in the evaluation and interpretation of the effectiveness of treatments, it is important to determine the clinical significance of change scores, and which changes in outcomes could be interpreted as meaningful improvements to patients and clinicians. The minimal important difference (MID) could be defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s (health care) management” (19). For a particular instrument, MIDs may vary by population and context, for instance the baseline from which the patient starts, and whether they are improving or deteriorating (20). Thus, MIDs for utility measures, notably the EQ-5D and SF-6D, should be estimated for the population in which the measures are used, for instance, among patients with SSc.

The SF-6D has been shown to be a valid measure of utility in a sample of patients with diffuse SSc (21), whereas the EQ-5D has not yet been tested in SSc. Therefore, the purpose of the present study was to compare measurement properties of the EQ-5D and SF-6D utility measures, to assess the association and agreement between these measures and to estimate minimal important differences (MID), effect sizes (ES) and standardised response means (SRM) in SSc patients. Patients with limited as well as diffuse disease were included in our sample.

Materials and methods

Design

Measurement properties of the EQ-5D and SF-6D were assessed using prospective data collection in the cohort study “Psychological factors in scleroderma” in patients with SSc. Patients completed both questionnaires at inclusion (baseline) and after 12 months follow-up. Details of this study were described elsewhere (22). The study was approved by the local medical ethics board (CMO Arnhem-Nijmegen 2008/109).

Patients and procedures

Patients with SSc of the departments of Rheumatology of the Sint Maarten-skliniek or Radboud University Nijmegen Medical Center, both in Nijmegen, the Netherlands, were included between June 2008 and August 2009. All patients had a diagnosis of SSc according to the preliminary ACR classification criteria (23). Exclusion criteria for participation in the cohort were a life expectancy of less than a year, acute serious complications, severe psychiatric comorbidity (e.g. severe substance abuse, psychosis or dementia), other serious comorbidities (e.g. cancer) and insufficient competence in the Dutch language. For the current analyses, patients were included if they completed either the EQ-5D or SF-6D at baseline. For 4 out of the 215 patients who completed the baseline measures, both the EQ-5D and SF-6D could not be calculated due to missing values (n=211). In addition, 4 patients had missing values for the EQ-5D and 12 patients for the SF-6D. Of the 211 patients who were included at baseline, 163 (77.3%) also completed the follow-up measure. Drop-out reasons were: death (n=8), illness (n=3), logistic reasons (n=11), unknown (n=26). Additionally, 9 patients had missing values preventing the calculation of both the EQ-5D and SF-6D scores at follow-up (n=154). Of the patients with follow-up, 6 patients had no EQ-5D score and 7 patients had no SF-6D score.

Assessments

i. Demographics and disease characteristics

Age, sex, marital status and current employment status were assessed by questionnaire. In addition, at baseline the attending rheumatologist assessed disease duration (time since onset of first non-Raynaud symptom), SSc limited or diffuse subtype (14), and the modified Rodnan skin score (mRSS), which is a rating of skin involvement ranging from 0 (no involvement) to 3 (severe thickening) in 17 body areas (total score range 0–51) (24). The SF-36, EQ-5D, and Scleroderma modified disability index of the Health Assessment Questionnaire (SHAQ) were self-assessed by the patient at home using paper versions.

ii. EQ-5D (EuroQol 5D)

The EQ-5D (3, 4) is a 5-item standardised health-related quality of life questionnaire, measuring 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The items are rated from 1 (no problems) to 3 (extreme problems). The EQ-5D ranges from -0.59 to 1.00, with 1.00 indicating full health and 0 representing death. Since negative EQ-5D scores are possible, these indicate health status valued worse than death.

iii. Medical Outcomes Trust Short Form-36 (SF-36)

The SF-36 (6) measures eight domains of health status using 36 items; physical functioning (10 items), role-physical (4 items), bodily pain (2 items),
general health (5 items), vitality (4 items), social functioning (2 items), role-emotional (3 items) and mental health (5 items). In addition, one item assesses global rating of change (GRoC) on a 5-point scale from “Much better now than one year ago” (1) to “Much worse now than one year ago” (5). For the calculation of the SF-6D, 11 items are used covering six domains (5); social functioning, bodily pain, mental health, physical functioning, role-limitation and vitality. The SF-6D ranges from .29 to 1.00, with 1.00 indicating full health.

iv. Scleroderma Health Assessment Questionnaire (SHAQ)
The SHAQ (25) consists of the disability index of the Health Assessment Questionnaire (HAQ) (26) and six Visual Analogue Scales (VAS) measuring perceived severity of pain, digital ulcers, Raynaud’s phenomenon, lung involvement, gastrointestinal problems and patient global assessment (PGA). A VAS was added to measure fatigue, since this is perceived as an important symptom of SSc (27). The HAQ-Disability Index score consists of 20 items, measuring 8 dimensions of functioning (dressing and grooming, ability to get up, eating, walking, personal hygiene, reach, grip strength and activities). The score of each dimension ranges from 0 (best function) to 3 (worst function), and the mean of these scores can be calculated as an indicator of overall physical functioning (HAQ-DI) with higher scores pointing to worse functioning. The HAQ was originally developed for use in Rheumatoid Arthritis (26) but has also demonstrated to be reliable and valid in patients with SSc (25).

Statistical analyses
Descriptive statistics were provided as mean and standard deviation (SD) or median (P25-P75) for continuous variables and percentages for categorical variables. For the EQ-5D and SF-6D, descriptive statistics (mean, median, SD, minimum, maximum and frequencies) were calculated at baseline and follow-up. The COSMIN recommendations (28) were followed in the assessment of measurement properties, except for the recommendation on responsiveness. The COSMIN recommendation includes hypotheses testing and considers traditionally accepted responsiveness measures such as standardised response mean (SRM) and effect size (ES) inappropriate. However, since the SRM and ES are used to compare which outcome measure detects changes over time more accurately, it can be considered appropriate for this purpose, especially in the absence of a gold standard or external criterion (29). All patients had completed the EQ-5D and/or SF-6D measures at baseline. However, some patients had missing data at follow-up. Missing values for the EQ-5D and SF-6D at follow-up were imputed using baseline EQ-5D and SF-6D scores, as well as baseline HAQ-DI score and VAS scores for the different domains included in the SHAQ, and available demographic and disease parameters. The assumption of values missing at random was tested with regression models examining the association of baseline characteristics and missingness of the EQ-5D and SF-6D at follow-up, respectively, as well as t-tests comparing baseline characteristics between patients with and without missing values at follow-up (results not shown). Multiple imputation by chained equations was used to create 20 datasets, using 15 cycles for each dataset (30). Results of these 20 imputed datasets were combined following Rubin’s rules (31).

Floor and ceiling effects were assessed by calculating the percentage of patients scoring the lowest and highest possible score, respectively. The association between the EQ-5D and SF-6D at baseline was assessed using Spearman’s rank correlation coefficient, and agreement between measures was assessed using Lin’s concordance coefficient (32) and a Bland-Altman plot (33). Post-hoc, the agreement between utility measures was further explored for EQ-5D scores <0.45. The construct validity of the EQ-5D and SF-6D was assessed by calculating the correlation of both measures with HAQ-DI scores, VAS scores and skin score (mRSS). A priori, it was hypothesised that both measures would have 1) at least a moderate correlation (r>0.40) with HAQ-DI scores and with PGA, 2) low correlations (r≤0.40) with the 6 VAS scores, and 3) no correlation (r<0.05) with skin score (21). Correlations were interpreted following Franzblau (34): 0.00-0.20 indicating no correlation, 0.21-0.40 indicating a low correlation, 0.41-0.60 indicating a moderate correlation, 0.61-0.80 indicating a marked correlation, and 0.81-1.00 indicating a high correlation. The utility measures were considered valid if ≥75% of our a priori hypotheses were confirmed (35).

The discriminative ability among baseline HAQ-DI and PGA categories was assessed for both measures. HAQ-DI scores were classified by convenience in three categories: no-to-mild disability (0.00-1.00), moderate disability (1.01-2.00) and severe disability (2.01-3.00). For PGA, a score of 0.00-3.30 was conveniently categorised as mild disease activity, 33.1-66.0 as moderate disease activity and 66.1-1.00 as severe disease activity. Differences between the HAQ-DI and PGA categories were assessed for the EQ-5D and SF-6D using MANOVA. In case of a significant result, post-hoc analyses were conducted between the three categories. A p-value <0.05 was considered statistically significant.

The Minimal Important Difference (MID) was defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s (health care) management” (19). To calculate MIDs for the EQ-5D and SF-6D, three different anchors were used: 1) the global rating of change item (GRoC) of the SF-36, 2) the HAQ-DI minimal important difference (0.22 points) for patients with rheumatoid arthritis (36), and 3) the upper bound of a published MID of the HAQ-DI for patients with diffuse SSC (0.14 points) (37). For the GRoC, patients who scored 2 (somewhat better) or 4 (somewhat worse) on the item “Compared to a year ago, how would you rate your health now?” of the SF-
36 were considered the minimally improved and deteriorated subgroups, respectively. Patients with a score of 3 (about the same) were considered the unchanged group. For the HAQ-DI groups, minimally improved and deteriorated subgroups were categorised as patients with a decrease or increase of at least 0.22 (or 0.14 points for the third anchor) on the HAQ-DI, respectively. Patients with a change score <0.22 (or <0.14 for the third anchor) were considered the unchanged group.

To assess responsiveness, a distribution-based approach was used. For the abovementioned anchors, responsiveness was calculated using Effect Size (ES = D/SD_{baseline}) (38) and Standardised Response Mean (SRM=D/SD_{difference}) (39), in which D is the difference between baseline and follow-up measures, and SD_{baseline} and SD_{difference} are the standard deviations of the baseline measures and difference scores, respectively. Effect sizes were interpreted following Cohen’s criteria (38): small (0.2), moderate (0.5) or large (0.8). ES and SRMs of the improved and deteriorated groups were compared with those of the unchanged groups. All statistical analyses were conducted using Stata/IC 10.1 software (StataCorp LP, College Station, TX).

Results
Sample characteristics
Demographics and disease characteristics were displayed in Table I. The majority of patients were female and middle-aged. Most patients were living together or married. One-third of the patients were employed at time of the study, and a large minority received higher education. Time since the onset of the first non-Raynaud’s symptom ranged from 2 months to 52 years.

Distribution and agreement
At baseline, the mean EQ-5D score was 0.64 (SD = 0.25, range = 0.18–1.00). The mean SF-6D score was 0.65 (SD = 0.11, range = 0.38–1.00). Mean follow-up scores for the EQ-5D and SF-6D were 0.59 (SD = 0.31) and 0.64 (SD = 0.14), respectively. Neither the EQ-5D nor the SF-6D showed a floor effect (both 0.0%) or a ceiling ef-

| Table I. Baseline patient demographic and disease characteristics. |
|------------------------|------------------------|------------------------|
| Variables              | Values (n=211)          |
| Female (%)             | 143 (67.8)             |
| Mean age, years (SD)   | 56.4 (12.0)            |
| Median time since onset first Non-Raynaud symptom (P25-P75) | 7.4 (3.5–12.3)<p> |
| Higher education (%>12 years) | 86 (41.6)<p> |
| Currently working (%)  | 69 (32.7)              |
| Married or living as married (%) | 159 (75.4)            |
| Limited disease (%)    | 154 (74.4)             |
| Mean modified Rodnan Skin Score (SD; range) | 6.3 (6.0; 0-37) |
| Mean HAQ-DI score (SD; range) | 1.04 (0.73; 0-3.0)<p> |
| Mean EQ-5D score (SD; range) | 0.64 (0.25; -0.18-1.00)<p> |
| Mean SF-6D score (SD; range) | 0.65 (0.11; 0.38-1.00)<p> |

Due to missing values: n=204, n=207, n=210, n=207, n=199

| Table II. Correlations [95% CI] of SF-6D an EQ-5D with physical functioning measures and disease-related visual analogue scales at baseline. |
|------------------------|------------------------|------------------------|
| SF-6D                  | Hypothesis confirmed   | EQ-5D                  | Hypothesis confirmed |
| HAQ-DI                 | -0.63 [-0.71, -0.54]   | Yes                    | -0.63 [-0.71, -0.54]   | Yes |
| PGA                    | -0.58 [-0.66, -0.48]   | Yes                    | -0.62 [-0.70, -0.53]   | Yes |
| Pain VAS               | -0.57 [-0.65, -0.46]   | No                     | -0.55 [-0.64, -0.45]   | No  |
| Ulcer VAS              | -0.36 [-0.48, -0.23]   | Yes                    | -0.30 [-0.42, -0.17]   | Yes |
| Raynaud’s VAS          | -0.32 [-0.44, -0.19]   | Yes                    | -0.28 [-0.40, -0.15]   | Yes |
| Lung VAS               | -0.45 [-0.56, -0.33]   | No                     | -0.38 [-0.49, -0.26]   | Yes |
| GI VAS                 | -0.31 [-0.43, -0.18]   | Yes                    | -0.32 [-0.44, -0.19]   | Yes |
| Fatigue VAS            | -0.63 [-0.71, -0.54]   | Yes                    | -0.55 [-0.64, -0.44]   | Yes |
| Skin score             | -0.15 [-0.29, -0.01]<p> | Yes                    | -0.13 [-0.27, 0.01]<p> | Yes |

HAQ-DI: Health Assessment Questionnaire-Disability Index; PGA: Patient Global Assessment; VAS: Visual Analogue Scale; GI: Gastrointestinal. All p<0.001;*p<0.05;**p<0.06.

Fig. 1. Bland and Altman plot of differences between EQ-5D and SF-6D for patients with systemic sclerosis.
Lower line: Mean difference +1.96 SD (Value = -0.38); Upper line: Mean difference -1.96 SD (Value = 0.39).
Utility measures in systemic sclerosis / L. Kwakkenbos et al.

Effect (7.7% and 0.5%, respectively) at baseline. Spearman’s correlation between both measures at baseline was \( r = 0.74 \) (\( p < 0.01 \)), indicating a marked association between EQ-5D and SF-6D utility scores. Lin’s concordance coefficient for baseline was 0.49 (95% CI = 0.42–0.56), indicating moderate concordance between the EQ-5D and SF-6D. The Bland-Altman plot (Fig. 1) showed a mean difference of 0.003 (95% CI = 0.024–0.030) but wide limits of agreement (-0.38–0.39). In an additional analysis, only EQ-5D scores <0.45 were taken into account. A structural bias was found for the lower scores on the utility measures. In this group (n=24), the mean difference between the EQ-5D and SF-6D measures was -0.44 (SD=0.14; 95% CI=0.38–0.50).

Validity: construct
Correlations of the EQ-5D and SF-6D with physical functioning measures are displayed in Table II. Consistent with our hypotheses, a marked correlation was found for both utility measures and HAQ-DI and PGA scores, and a small (<0.20) correlation was found with skin score. Our hypotheses regarding a low correlation with the six VAS scores were confirmed for ulcers, Raynaud’s phenomenon and gastrointestinal tract. For lung symptoms, a low correlation was found for the EQ-5D, but a moderate correlation with the SF-6D. Furthermore, pain and fatigue both showed a moderate to marked correlation with the EQ-5D and SF-6D. There were no significant differences between the EQ-5D and SF-6D in correlations with HAQ-DI, PGA, the 6 VAS scores and skin score. For the EQ-5D, 77.8% of the \textit{a priori} hypotheses were confirmed, and 66.7% for the SF-6D.

Validity: discrimination
Discrimination between the HAQ-DI and PGA categories for both utility measures at baseline are displayed in Figure 2. According to the MANOVA, both the EQ-5D and SF-6D were able to discriminate between HAQ-DI categories (F(2,203) = 41.2, \( p < 0.001 \)) and F(2,195) = 35.9, \( p < 0.001 \), respectively), and between PGA levels (F(2,200) = 53.3, \( p < 0.001 \) and F(2,192) = 32.9, \( p < 0.001 \), respectively). Post-hoc analyses revealed that both measures were able to discriminate between all HAQ-DI and PGA categories.

Minimally important differences and responsiveness
MIDs, SRMs and effect sizes for the three utilised anchors are displayed in Table III. For the minimally improved categories, MIDs were 0.05 for the SF-6D and ranged from 0.05 to 0.10 for the EQ-5D. For the minimally deteriorated categories, MIDs for the SF-6D were -0.03 to -0.04, and ranged from -0.12 to -0.14 for the EQ-5D. The MIDs in the minimally changed groups were of larger magnitude than the unchanged groups for both measured and all three anchors. For the SF-6D, the mean MID estimate was 0.05 for the improved and -0.04 for the deteriorated subgroups. The mean MID estimate for the EQ-5D was 0.08 for the improved subgroups, and -0.13 for the deteriorated subgroups. The SRMs and ES for the both utility measures were small to moderate [38]. For the SF-6D, ES and SRM were ranging from -0.29 to -0.41 for the deteriorated categories and from 0.38 to 0.54 for the improved categories. The SRMs and ES for the EQ-5D ranged from -0.50 to -0.55 for the deteriorated categories and from 0.19 to 0.38 for the improved categories.

Discussion
This was the first study to compare measurement properties of the EQ-5D and SF-6D utility measures, to assess the association and agreement between these measures and to estimate MIDs in a sample of patients with SSc. A marked correlation was found between both utility measures, indicating that higher scores on the EQ-5D are associated with higher scores on the SF-6D and vice versa. However, according to the results of this study, the agreement between the EQ-5D and SF-6D was moderate. Especially for lower scores (including about 10% of the sample), results showed that EQ-5D and SF-6D scores, in particular for patients with worse health status, are not reasonably comparable. Further investigation is warranted to assess whether and to what extent the choice for the EQ-5D versus SF-6D to measure utility might...
have an influence on cost-utility evaluations and treatment decisions. Neither of the measures showed floor- and ceiling effects, the mean scores were comparable for both measures, and the EQ-5D and SF-6D were both able to discriminate between levels of disability and PGA of health. For the EQ-5D, 77.8% of our a priori hypotheses were confirmed, whereas for the SF-6D this percentage was 66.7%. Based on our a priori cut-off (≥75% confirmed), our results support the construct validity of the EQ-5D but not for the SF-6D. In contrast to our a priori hypotheses, a moderate to marked correlation was found with pain and fatigue for both measures. However, both symptoms are experienced frequently by patients with SSC (27), and were found to have at least moderate impact on daily activities in the majority of patients, indicating that pain and fatigue might significantly reduce health related quality of life in patients with SSC. Therefore, it might not be surprising that these symptoms are associated with lower utility scores in SSC, and our hypotheses regarding pain and fatigue might have been suboptimal.

The mean MIDs for the SF-6D reflect a change of 6.1%–7.7% for SF-6D scores, well within the suggested change of 5–10% that is generally considered a reasonable MID (40). The mean MIDs for the EQ-5D were equivalent of a change of 12.5–20.3%, somewhat larger than the suggested cut-offs (40). The ES and SRM for both measures were consistent with analyses showing MID estimates for health related quality of life instruments ranging from 0.30 to 0.50 SD units (41, 21). In addition, the MIDs for the changed groups were at least twice as large as the estimates in the unchanged groups and could therefore be used to calculate sample sizes for clinical trials involving patients with SSC, and in interpreting the relevance and importance of treatment effects (42).

To test stability over time, the SF-6D was found to be excellent. In future studies, assessing both utility measures twice in a narrower time frame contributes to the comparability of the stability over time.

**Conclusion**

The present study is the first to compare measurement properties and MIDs of two widely used utility measures, notably the EQ-5D and SF-6D, in patients with SSC. Measurement properties were generally acceptable and similar for both measures. Although there was a marked correlation between the measures, the moderate agreement implies that caution is needed when interpreting and comparing results that are

---

**Table III.** Minimal important differences and responsiveness (SRM and ES) of the EQ-5D and SF-6D measures.

<table>
<thead>
<tr>
<th>Anchor</th>
<th>n²</th>
<th>MID [95%CI]</th>
<th>SRM</th>
<th>ES</th>
<th>MID [95%CI]</th>
<th>SRM</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRoC (somewhat worse)</td>
<td>49</td>
<td>-0.05 [-0.06, 0.00]</td>
<td>-0.34</td>
<td>-0.29</td>
<td>-0.12 [-0.18, -0.05]</td>
<td>-0.53</td>
<td>-0.53</td>
</tr>
<tr>
<td>GRoC (unchanged)</td>
<td>87</td>
<td>-0.01 [-0.03, 0.01]</td>
<td>-0.14</td>
<td>-0.12</td>
<td>-0.01 [-0.05, 0.04]</td>
<td>-0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>GRoC (somewhat better)</td>
<td>24</td>
<td>0.05 [0.00, 0.09]</td>
<td>0.44</td>
<td>0.38</td>
<td>0.05 [0.04, 0.14]</td>
<td>0.26</td>
<td>0.19</td>
</tr>
<tr>
<td>HAQ-DI (≥0.22 worse)</td>
<td>63</td>
<td>-0.04 [-0.07, -0.02]</td>
<td>-0.41</td>
<td>-0.35</td>
<td>-0.14 [-0.20, -0.07]</td>
<td>-0.55</td>
<td>-0.52</td>
</tr>
<tr>
<td>HAQ-DI (unchanged)</td>
<td>95</td>
<td>-0.01 [-0.03, 0.00]</td>
<td>-0.16</td>
<td>-0.11</td>
<td>-0.04 [-0.08, 0.00]</td>
<td>-0.20</td>
<td>-0.15</td>
</tr>
<tr>
<td>HAQ-DI (≥0.22 better)</td>
<td>38</td>
<td>0.05 [0.01, 0.09]</td>
<td>0.45</td>
<td>0.54</td>
<td>0.10 [0.01, 0.18]</td>
<td>0.38</td>
<td>0.29</td>
</tr>
<tr>
<td>HAQ-DI (≥0.14 worse)</td>
<td>68</td>
<td>-0.04 [-0.06, -0.01]</td>
<td>-0.39</td>
<td>-0.33</td>
<td>-0.13 [-0.20, -0.07]</td>
<td>-0.53</td>
<td>-0.50</td>
</tr>
<tr>
<td>HAQ-DI (unchanged)</td>
<td>84</td>
<td>-0.02 [-0.04, 0.00]</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.04 [-0.08, 0.00]</td>
<td>-0.20</td>
<td>-0.15</td>
</tr>
<tr>
<td>HAQ-DI (≥0.14 better)</td>
<td>42</td>
<td>0.05 [0.02, 0.09]</td>
<td>0.45</td>
<td>0.53</td>
<td>0.08 [0.01, 0.16]</td>
<td>0.35</td>
<td>0.27</td>
</tr>
</tbody>
</table>

GRoC: Global rating or change; HAQ-DI: Health Assessment Questionnaire-Disability Index; MID: Minimal important difference; ES: D/SDunits; SRM: D/SDachieve; *Calculations based on minimum number of imputations.
obtained with the EQ-5D and SF-6D. Results from the present study suggest that scores of the EQ-5D and SF-6D are not reasonably comparable, in particular for patients with low utility scores. The MID estimates we provided can be used to calculate sample sizes for clinical trials and guide the interpretation of results, in particular when examining the usefulness of a clinical intervention for SSC patients.

References


35. TERWEE CB, BOT SDM, DE BOER MR et al.: Quality criteria were proposed for measure- ment properties of patient-reported outcome measures. J Clin Epidemiol 2007; 60: 34-42.


38. COHEN J: Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Erlbaum; 1988


