

EXTENDED REPORT

Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study

D Khanna, D E Furst, R D Hays, G S Park, W K Wong, J R Seibold, M D Mayes, B White, F F Wigley, M Weisman, W Barr, L Moreland, T A Medsger Jr, V D Steen, R W Martin, D Collier, A Weinstein, E V Lally, J Varga, S R Weiner, B Andrews, M Abeles, P J Clements



Ann Rheum Dis 2006;**65**:1325–1329. doi: 10.1136/ard.2005.050187

See end of article for authors' affiliations

Correspondence to: D Khanna, Division of Immunology, Department of Medicine, University of Cincinnati, ML 0563, Cincinnati, OH 45267-0563, USA; dinesh.khanna@uc.edu

Accepted 5 March 2006
Published Online First
15 March 2006

Objective: To estimate minimally important differences (MIDs) in scores for the modified Rodnan Skin Score (mRSS) and Health Assessment Questionnaire—Disability Index (HAQ-DI) in a clinical trial on diffuse systemic sclerosis (SSc).

Participants and methods: 134 people participated in a 2-year, double-blind, randomised clinical trial comparing efficacy of low-dose and high-dose D-penicillamine in diffuse SSc. At 6, 12, 18 and 24 months, the investigator was asked to rate the change in the patient's health since entering the study: markedly worsened, moderately worsened, slightly worsened, unchanged, slightly improved, moderately improved or markedly improved. Patients who were rated as slightly improved were defined as the minimally changed subgroup and compared with patients rated as moderately or markedly improved.

Results: The MID estimates for the mRSS improvement ranged from 3.2 to 5.3 (0.40–0.66 effect size) and for the HAQ-DI from 0.10 to 0.14 (0.15–0.21 effect size). Patients who were rated to improve more than slightly were found to improve by 6.9–14.2 (0.86–1.77 effect size) on the mRSS and 0.21–0.55 (0.32–0.83 effect size) on the HAQ-DI score.

Conclusion: MID estimates are provided for improvement in the mRSS and HAQ-DI scores, which can help in interpreting clinical trials on patients with SSc and be used for sample size calculation for future clinical trials on diffuse SSc.

Chronic diseases often have a relapsing and remitting course, with substantial effect on functioning and well-being or health-related quality of life (HRQOL). Systemic sclerosis (SSc) or scleroderma, a multisystem autoimmune disorder with no effective treatment or cure, is an example of a disease in which patients must cope with pain, disfigurement, disability and feelings of helplessness, each of which can affect their HRQOL and outlook on the future.¹ The ultimate goal of healthcare is to improve, restore or preserve HRQOL.² An important progress in the research on HRQOL is the concept of minimally important difference (MID), defined as “the smallest difference in score of a HRQOL [or measure of interest] instrument that patients perceive as beneficial and that would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management”.³ For the clinicians, it has been defined as “the smallest effect size that would lead them to recommend a therapy to their patients”.⁴ In short, “MID is the smallest difference in a score that is considered to be worthwhile or important”.⁵

Estimation of the MID requires identification of an external anchor that provides an indication of underlying change—a clinically relevant indicator or pointer to which a change in HRQOL can be tied.⁶ An anchor is of clinical relevance and can be “subjective”, such as self-reports of change as assessed by the patient or the investigator, or “objective”, such as clinical indicators of response to treatment (skin thickness or disease severity). Self-reports of change require the respondents to evaluate whether they have experienced change in a domain of health from a previous time point to the present.⁷

In this report, we estimate MIDs for the modified Rodnan Skin Score (mRSS) and Health Assessment Questionnaire—

Disability Index (HAQ-DI) in a 2-year, double-blind, randomised clinical trial comparing efficacy of low-dose and high-dose D-penicillamine (D-Pen) in patients with diffuse SSc.^{8–10} The investigator's assessment of the patient's disease was used as the anchor at time points of 6, 12, 18 and 24 months.

PARTICIPANTS AND METHODS

Participants

The D-Pen study has been detailed elsewhere.⁸ Briefly, this was a 2-year, multicentre US study that recruited and randomised 134 people with diffuse SSc (≤ 18 months) to receive 750–1000 mg/day of D-Pen versus 125 mg of D-Pen every other day. Diffuse SSc was characterised by the presence of skin thickening, proximal as well as distal to the elbows and knees, with or without involvement of the face.¹¹ No statistically significant differences were observed between treatment groups at baseline and at 2 years in the three primary outcomes in this trial: change in the skin score, incidence of renal crisis and survival. Therefore, for the present analysis, the data from the two groups were pooled. The skin score was assessed using the mRSS (referred to as skin score from this point) and is a clinical measure of the extent and severity of skin thickening.^{12–13} In addition to serving as the primary basis for disease classification, skin thickening in SSc is viewed as a clinical surrogate of disease progression, at least in early disease (usually considered to be < 3 years from the onset of the first typical sign or symptom of SSc, not including Raynaud's phenomenon).¹⁴ Skin

Abbreviations: D-Pen, D-penicillamine; HAQ-DI, Health Assessment Questionnaire—Disability Index; HRQOL, health-related quality of life; MID, minimally important difference; mRSS, modified Rodnan Skin Score; SSc, systemic sclerosis

thickening is assessed on 17 body areas: fingers, hands, forearms, arms, feet, legs and thighs (bilaterally), and the face, chest and abdomen (singly). Each area is scored from 0 to 3, where 0 is, normal skin and 3 is, severe thickening (range 0 (no thickening) to 51 (severe thickening) in all 17 areas).

The HAQ-DI is a self-administered arthritis-targeted measure intended for assessing activities of daily living in arthritis.¹⁵ It assesses a patient's level of functional ability and includes questions on fine movements of the upper extremity, locomotor activities of the lower extremity, and activities that are associated with both upper and lower extremities. In the original HAQ-DI, an additional grade of difficulty was added in patients using assistive or adaptive devices (such as a cane or walker). In our study, the patient responses were not modified for patient use of assistive or adaptive devices.^{9–10} The standard HAQ-DI is scored from 0 (no disability) to 3 (severe disability), representing the averaging of the worst score in each of eight domains of daily function.

At 6, 12, 18 and 24 months, the investigator was asked "Since entering the study, the patient is: markedly worsened, moderately worsened, slightly worsened, unchanged, slightly improved, moderately improved, markedly improved." In the question, there is no direct reference to the patients' health or their disease. Patients who improved slightly were defined as the minimally changed subgroup. The changes in the skin and HAQ-DI scores for the group that "slightly improved" at 6, 12, 18 and 24 months were determined in order to estimate the MID. This was compared to change scores for the group that moderately or markedly improved.

Only improvement is considered in this article because the number of patients who slightly worsened was too small to make a statistically or clinically robust conclusion ($n = 5-6$ at different time points) and the data are not presented.

Statistical analysis

The normality of change in skin score and HAQ-DI score for patients who slightly improved (ie, $\text{skin score}_{\text{month } x} - \text{skin score}_{\text{month } 0}$, where $x = 6, 12, 18$ or 24 months) was assessed using the Shapiro-Wilk test. The data were normally distributed for the change in skin score and HAQ-DI score except for 18-month change in HAQ-DI score; therefore, the MIDs are presented as mean (standard deviation (SD)). As pointed by Hays and colleagues,¹⁶

... [an] anchor should have a non-trivial association with change in HRQOL measure [or measure of interest, eg, skin score here]. If the correlation between the anchor and HRQOL [or measure of interest] change is zero, the anchor is not useful for establishing MID (pp 64–5).

To assess the usefulness of an anchor, change in the anchor and change in the HRQOL scores should have a correlation coefficient of at least ≥ 0.37 .¹⁶ The correlation coefficient of ≥ 0.37 corresponds to an effect size of 0.80 (large effect as proposed by Cohen¹⁷). This was assessed using the Spearman correlation coefficient (as change in anchor is an ordinal variable) between the investigator global assessment (the anchor) and change in skin and HAQ-DI scores. MID was estimated by examining change in skin score and HAQ-DI score in patients who had slightly improved. These estimates were compared with those of patients who were rated to improve more than slightly. Responsiveness to change was evaluated using the effect size.¹⁸ Effect size is the mean change in the skin score or HAQ-DI from the baseline to month x (where $x = 6, 12, 18$ or 24 months) divided by the SD at baseline (8.0 for the skin score and 0.66 for HAQ-DI).

Cohen's rule of thumb for interpreting effect size is that a value of 0.20–0.49 represents a small change, 0.50–0.79 a medium change and ≥ 0.80 a large change.^{17–21}

As an exploratory analysis, the effect of baseline HAQ-DI and skin scores was assessed on the MID estimates.⁶ People with different baseline HAQ-DI or skin scores may require different amounts of improvement to consider a change to represent an MID. We divided the baseline HAQ-DI and skin scores into three groups: between 0 and 25th centile, group 1; between 26th and 74th centile, group 2; and ≥ 75 th centile, group 3. We estimated MID for 6-month change score; the numbers of participants were too small at other time points to estimate changes for the three groups.

RESULTS

Table 1 shows the baseline characteristics of the 134 patients. The mean (SD) values were 43.7 (12.4) years for age, 1.0 (0.7) for HAQ-DI and 21 (8) for skin score. At baseline, the 66 patients who completed the study had a statistically significant ($p < 0.05$) higher haematocrit than non-completers ($n = 68$); other baseline variables were similar between the two groups. MID analysis is based on completers at 6 ($n = 104$), 12 ($n = 97$), 18 ($n = 72$) and 24 months ($n = 68$; table 2).

Table 3 shows the Spearman's correlation coefficient between the anchor (investigator assessment of change) and change in the HAQ-DI and skin score in the slightly improved group from baseline at four time points. The correlation coefficient was > 0.37 at all four time points for both measures, except change in the HAQ-DI score at month 6 ($r = 0.35$). This satisfies the requirement that the correlation be ≥ 0.37 .

Table 2 shows the mean (SD) MID scores for improvement in the skin score and HAQ-DI. MID estimate for the skin score ranged from 3.2 to 5.3 (0.40–0.66 effect size), except at month 12 (MID = 1.9). HAQ-DI score ranged from 0.10 to 0.14 (0.15–0.21 effect size), except 0.03 at month 12. Patients who were rated to improve more than slightly were found to improve by 6.9–14.2 (0.86–1.77 effect size) on the skin score and 0.21–0.55 (0.32–0.83 effect size) on the HAQ-DI. Table 2 also provides mean (SD) scores of patients who had not changed according to the doctors (< 0.20 effect size at all four time points for skin score and HAQ-DI). MID estimates of those who were rated to improve more than slightly were consistently higher and in the expected direction than those of the unchanged group.

Table 4 shows the mean (SD) MID scores at 6 months for improvement in the skin and HAQ-DI stratified by severity of skin involvement at baseline. Although exploratory, the data suggest that patients with very high baseline scores require a larger improvement to be characterised by their doctors as minimally improved as exemplified by the skin score (MID for the minimally improved group for low-score group, -2.6 ; medium-score group, -1.8 ; and high-score group, -6.1) and HAQ-DI (MID for the minimally improved group for low score group, $+0.01$; medium score group, -0.11 ; and high score group, -0.21).

DISCUSSION

MID estimates of HRQOL measures have propelled the development of a new drug in the treatment of different arthritides and successful design of controlled studies to improve HRQOL.²² Rapid advances in the field of immunology have spawned novel treatments targeting the cytokines associated with the pathogenesis of SSc. In addition, MID estimates provide a benchmark for future design of clinical trials on patients with SSc by helping researchers and clinicians understand whether differences in HRQOL scores between two treatment groups are meaningful or whether

Table 1 Baseline characteristics of all patients, classified by completion status of the 2-year study

Characteristic	All (n = 134)	Completers (n = 68)	Non-completers (n = 66)
Demographic			
Age (years)	43.7 (12.4)	44.2 (13.1)	43.1 (11.7)
Women (%)	77.6	75.0	80.3
History			
Duration of SSc before entry (days)	290 (125)	285 (129)	295 (122)
Physical examination			
mRSS, range 0–51	21 (8)	20 (8)	22 (7)
HAQ-DI score, range 0–3	1.0 (0.7)	1.0 (0.6)	1.1 (0.7)
Tender joint count, range 0–8	1.5 (2.4)	1.5 (2.2)	1.5 (2.5)
Swollen joint count, range 0–8	0.9 (1.6)	0.8 (1.3)	1.0 (1.8)
Laboratory results			
Haematocrit (%)	39.0 (3.8)	39.8 (3.3)*	38.2 (4.1)
ESR (mm in 1 st h)	24 (17)	21 (14)	27 (20)
Creatine kinase (% upper limit of normal)	81.7 (124.1)	83.3 (156.0)	80.1 (81.1)
DL _{CO} (% of predicted)	75.3 (18.4)	75.7 (17.7)	74.9 (19.2)
FVC (% of predicted)	84 (17)	85 (19)	83 (15)
Serum creatinine (mg/dl)	0.89 (0.19)	0.91 (0.19)**	0.85 (0.19)

DL_{CO}, diffusion capacity; ESR, erythrocyte sedimentation rate; FVC, forced vital capacity; HAQ-DI, Health Assessment Questionnaire—Disability Index; mRSS, modified Rodnan Skin Score; SSc, systemic sclerosis. Values are mean (SD).

Student's *t* tests between completers and non-completers: **p*<0.05, ***p*<0.01.

changes within one group over time are meaningful.⁶ MIDs are also useful for determining sample size for future studies.²³

We used data from the D-Pen study to derive the MIDs for skin scores and HAQ-DI. We used investigator global assessment of change as our anchor. In this study, the correlation coefficient between the anchor and change in skin score and HAQ-DI was ≥ 0.37 . Investigator global assessment of change could capture changes in skin score and HAQ-DI.

Although a previous study²¹ and consensus statement²⁴ by the SSc experts suggest that a change of $\geq 30\%$ is clinically meaningful, to our knowledge, no study has estimated MID using a data-driven approach. Our MID estimates of skin score were 3.2–5.3 (or 15–25% of the baseline skin score of 21), which are smaller than the proposed consensus improvement of $\geq 30\%$. Also, the range of MID estimates of skin score (3.2–5.3) is greater than the reported intraobserver skin score variability of 2.45,²⁵ making these data-driven estimates statistically credible as well.

MID for improvement in HAQ-DI of patients with rheumatoid arthritis has been estimated to be 0.22²⁶ and our MID estimate is lower (0.10–0.14) than that reported in patients with rheumatoid arthritis. The difference might be explained by the fact that MID in this study is based on the doctor's assessment rather than the patient's assessment of the disease course. Also, MID estimates are reported as a range rather than a point estimate, as MID may change with different anchors.¹⁶

The effect size was consistently lower for HAQ-DI than for skin score—negligible to small for the slightly improved group, small to medium (except month 18, which was large) for the moderately and markedly improved groups. This is compared with a small to medium effect size for the slightly improved groups and a large effect size for the other two groups.

MID estimates in the other groups should be larger and in the expected direction compared with the no change group because “if it turns out that the change for the no change

Table 2 Minimal important difference estimates for the skin score and Health Assessment Questionnaire—Disability Index

Months	Mean (SD) skin score				Mean (SD) HAQ-DI			
	No change group	MID in the slightly improved group	Change in the moderately improved group	Change in the markedly improved group	No change group	MID in the slightly improved group	Change in the moderately improved group	Change in the markedly improved group
6	1.5 (5.2) (n=20)	-3.2 (4.3) (n=32)	-6.9 (2.9) (n=10)	*	0.09 (0.4) (n=20)	-0.10 (0.4) (n=31)	-0.30 (0.3) (n=10)	*
Effect size	0.19	0.40	0.86		0.13	0.15	0.45	
12	0.44 (5.6) (n=16)	-1.5 (6.7) (n=31)	-8.5 (5.1) (n=24)	-12.5 (6.3) (n=6)	0.11 (0.4) (n=16)	0.03 (0.5) (n=30)	-0.21 (0.43) (n=24)	-0.35 (0.42) (n=6)
Effect size	0.06	0.19	1.06	1.56	0.17	0.05	0.32	0.53
18	-0.7 (8.3) (n=11)	-5.3 (6.6) (n=20)	-7.2 (8.4) (n=21)	-14.2 (5.1) (n=12)	0.01 (0.7) (n=11)	-0.10 (0.5) (n=20)	-0.09 (-0.55) (n=21)	-0.55 (0.32) (n=12)
Effect size	0.09	0.66	0.90	1.77	0.01	0.15	0.13	0.83
24	1.1 (9.7) (n=11)	-3.9 (6.9) (n=10)	-8.9 (6.2) (n=26)	-12.6 (5.9) (n=16)	0.03 (0.7) (n=11)	-0.14 (0.6) (n=10)	-0.24 (0.48) (n=26)	-0.50 (0.27) (n=15)
Effect size	0.14	0.48	1.11	1.58	0.05	0.21	0.36	0.76

HAQ-DI, Health Assessment Questionnaire—Disability Index; MID, minimally important difference.

Negative score indicates improvement of the HAQ-DI and skin score.

Effect size = mean difference from the baseline to month *x*, where *x* = 6, 12, and 24 months/SD at baseline.

*Only 2 patients in this group.

Table 3 Spearman’s correlation coefficient between the anchor and change in the Health Assessment Questionnaire—Disability Index and skin scores

Change (slightly improved) in physician global assessment from baseline to month	Change in HAQ-DI		Change in mRSS	
6	-0.35 (n = 104)		-0.73 (n = 106)	
12	-0.42 (n = 97)		-0.54 (n = 97)	
18	-0.41 (n = 72)		-0.62 (n = 72)	
24	-0.59 (n = 68)		-0.69 (n = 68)	

HAQ-DI, Health Assessment Questionnaire—Disability Index; mRSS, modified Rodnan Skin Score.

group is similar to that of the minimally changed group, then the MID estimate is suspect” Hayes *et al*¹⁶(p 65). MID estimates for skin score and HAQ-DI were consistently greater and in the expected direction with the no change group with exception at month 12. A review of the 12-month data did not find any non-random pattern of missingness, and we excluded these values from our proposed estimates.

As previously noted,⁶ MID estimates may depend on baseline scores. This trend was seen in our analysis (table 4), where people with higher baseline scores (defined as baseline score ≥75th centile) required a larger change in their skin score and HAQ-DI for them to be considered as minimally improved by their doctors.

Our study has certain limitations. Firstly, as the global ratings of change asked the doctor to remember how his or her patient’s health was a year ago, retrospective self-reports are subject to recall bias. The HRQOL experts suggest choosing multiple and different kinds of anchors owing to inherent uncertainty associated with MID estimates,¹⁶ and future clinical trials should incorporate prospective or concurrent anchors.⁷ Secondly, MID estimates for the HAQ-DI were based on doctor’s assessment rather than patient’s assessment; patient’s assessment was not considered in this study. This assumes that the doctor is able to assess SSc-related disability as reflected by the HAQ-DI. Future studies should incorporate the patient’s assessment to confirm our findings. Thirdly, the trial cohort is not reflective of a general SSc population as the patients had early diffuse SSc.

In conclusion, using a data-driven approach, we provide MID estimates (generated by the patient’s doctor) for improvement in the skin score and HAQ-DI, which can be used in interpreting clinical trials on patients with SSc and which can help in sample size calculation for future clinical trials on patients with diffuse SSc.

Authors’ affiliations

D Khanna, Division of Immunology, Department of Medicine; Institute for the Study of Health, University of Cincinnati, Cincinnati, Ohio, USA; Veterans Affairs Medical Center, Cincinnati

D E Furst, G S Park, S R Weiner, P J Clements, Department of Medicine, Division of Rheumatology, David Geffen School of Medicine, Los Angeles, California, USA

R D Hays, Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine

G S Park, W K Wong, Department of Biostatistics, David Geffen School of Medicine

R D Hays, RAND, Los Angeles

J R Seibold, University of Michigan Scleroderma Program, Ann Arbor, Michigan, USA

M D Mayes, Division of Rheumatology and Clinical Immunogenetics, The University of Texas—Houston Medical School, Houston, Texas, USA

B White*, Department of Rheumatology and Clinical Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

F F Wigley, Rheumatology Division, The Johns Hopkins University, Baltimore, Maryland, USA

M Weisman, Division of Rheumatology, Cedars-Sinai Medical Center, Los Angeles

W Barr, J Varga, Division of Rheumatology, Northwestern University Medical School, Chicago, Illinois, USA

L Moreland, Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama, USA

T A Medsger Jr, Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

V D Steen, A Weinstein, Division of Rheumatology, Georgetown University Medical Center, Washington, DC, USA

R W Martin, Division of Rheumatology, College of Human Medicine, Michigan State University, Grand Rapids, Michigan, USA

D Collier, Division of Rheumatology, University of Colorado Health Sciences Center, Denver, Colorado, USA

E V Lally, Division of Rheumatology, Brown Medical School, Providence, Rhode Island, USA

B Andrews, Division of Rheumatology, University of California, Irvine, California, USA

M Abeles, Division of Rheumatic Diseases, University of Connecticut Health Center, Farmington, Connecticut, USA

*Current address: Amgen, Thousand Oaks, California, USA

Funding: This study was supported in part by grants from the Scleroderma Federation; United Scleroderma Foundation; FDA Orphan Drug Program; Arthritis Foundation; CRC grant numbers M01RR00865 and M01RR00827; WH Conzen Endowment in Clinical Pharmacology of the Schering-Plough Foundation; and bequests from the estates of Winifred Krause, Morris Goldsmith and Takako Ito. DK was supported by the Arthritis and Scleroderma Foundations (Physician Scientist Development Award), the Scleroderma Foundation (New Investigator Award), a National Institutes of Health BIRCWH Award (grant number HD051953) and a grant from the Scleroderma Clinical Trial Consortium. RDH was supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities (P20-MD00148-01), the UCLA

Table 4 Minimal important difference estimates for the skin score and Health Assessment Questionnaire—Disability Index at 6 months stratified by severity at the baseline

		Mean (SD) score at baseline				
		Low		Medium		High
Month	MID in the slightly improved group	No change group	MID in the slightly improved group	No change group	MID in the slightly improved group	No change group
6	-2.6 (3.3) (n = 9)	-1.9 (4.3) (n = 10)	-1.8 (3.8) (n = 15)	1.2 (2.5) (n = 6)	-6.1 (5.3) (n = 8)	-0.75 (9.9) (n = 4)
		Mean (SD) HAQ-DI at baseline				
		Low		Medium		High
Month	MID in the slightly improved group	No change group	MID in the slightly improved group	No change group	MID in the slightly improved group	No change group
6	0.01 (0.35) (n = 10)	0.37 (0.52) (n = 5)	-0.11 (0.38) (n = 14)	0.02 (0.18) (n = 6)	-0.21 (0.42) (n = 7)	-0.03 (0.43) (n = 9)

HAQ-DI, Health Assessment Questionnaire—Disability Index; MID, minimally important difference.

Negative score indicates improvement in HAQ-DI and skin score.

Low score, between 0 and 25th centile; medium score, between 26th and 74th centile; and high score, ≥75th centile.

Center for Health Improvement in Minority Elders/Resource Centers for Minority Aging Research, National Institutes of Health, and National Institute of Aging (AG-02-004).

Competing interests: None declared.

REFERENCES

- 1 **Khanna D**, Clements PJ, Furst DE, Chon Y, Elashoff R, Roth MD, *et al*. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. *Arthritis Rheum* 2005;**52**:592–600.
- 2 **Osoba D**, King M. Meaningful differences. In: Fayers P, Hays RD, eds. *Assessing quality of life in clinical trials*. Oxford Press: New York, 2005:244–57.
- 3 **Jaeschke R**, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**:407–15.
- 4 **van Walraven C**, Mahon JL, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. *J Clin Epidemiol* 1999;**52**:717–23.
- 5 **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.
- 6 **Crosby RD**, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;**56**:395–407.
- 7 **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.
- 8 **Clements PJ**, Furst DE, Wong WK, Mayes M, White B, Wigley F, *et al*. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. *Arthritis Rheum* 1999;**42**:1194–203.
- 9 **Clements PJ**, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, *et al*. The Disability Index of the Health Assessment Questionnaire is a predictor and correlate of outcome in the high-dose versus low-dose penicillamine in systemic sclerosis trial. *Arthritis Rheum* 2001;**44**:653–61.
- 10 **Clements PJ**, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, *et al*. Correlates of the disability index of the health assessment questionnaire: a measure of functional impairment in systemic sclerosis. *Arthritis Rheum* 1999;**42**:2372–80.
- 11 **Clements PJ**. Systemic sclerosis (scleroderma) and related disorders: clinical aspects. *Baillieres Best Pract Res Clin Rheumatol* 2000;**14**:1–16.
- 12 **Clements P**, Lachenbruch P, Seibold J, White B, Weiner S, Martin R, *et al*. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995;**22**:1281–5.
- 13 **Merkel PA**, Clements PJ, Reveille JD, Suarez-Almazor ME, Valentini G, Furst DE. Current status of outcome measure development for clinical trials in systemic sclerosis. Report from OMERACT 6. *J Rheumatol* 2003;**30**:1630–47.
- 14 **Black CM**. Measurement of skin involvement in scleroderma. *J Rheumatol* 1995;**22**:1217–19.
- 15 **Fries JF**, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.
- 16 **Hays RD**, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD J Chronic Obstructive Pulm Dis* 2005;**2**:63–7.
- 17 **Cohen J**. A power primer. *Psychol Bull* 1992;**112**:155–9.
- 18 **Hays RD**, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992;**1**:73–5.
- 19 **Liang MH**, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;**28**:632–42.
- 20 **Kim S**, Hays RD, Birbeck GL, Vickrey BG. Responsiveness of the quality of life in epilepsy inventory (QOLIE-89) in an antiepileptic drug trial. *Qual Life Res* 2003;**12**:147–55.
- 21 **Khanna D**, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, *et al*. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;**32**:832–40.
- 22 **Hakkinen A**, Sokka T, Kotaniemi A, Hannonen P. A randomized two-year study of the effects of dynamic strength training on muscle strength, disease activity, functional capacity, and bone mineral density in early rheumatoid arthritis. *Arthritis Rheum* 2001;**44**:515–22.
- 23 **Wyrwich KW**, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;**14**:285–95.
- 24 **Seibold JR**, McCloskey DA. Skin involvement as a relevant outcome measure in clinical trials of systemic sclerosis. *Curr Opin Rheumatol* 1997;**9**:571–5.
- 25 **Clements PJ**, Lachenbruch PA, Seibold JR, Zee B, Steen VD, Brennan P, *et al*. Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. *J Rheumatol* 1993;**20**:1892–6.
- 26 **Wells GA**, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;**20**:557–60.