# Health-Related Quality of Life in Systemic Sclerosis: A Systematic Review

Marie Hudson, MD MPH, Brett Thombs, PhD, Russell Steele, PhD, Pantelis Panopalis, MD, Evan Newton, Canadian Scleroderma Research Group (CSRG)*, Murray Baron, MD

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

**Author institutional affiliations**: SMBD-Jewish General Hospital and McGill University, Montreal, Canada

**Funding**: This study was funded in part by the Canadian Institutes of Health Research, the Scleroderma Society of Canada and educational grants from Actelion Pharmaceuticals and Pfizer Inc. Dr. Hudson is a New Investigator funded by the Canadian Institutes of Health Research.

**Correspondence and request for reprints**: Dr Marie Hudson, SMBD-Jewish General Hospital, Room A-216, 3755 Cote Ste Catherine Road, Montreal, Quebec, H3T 1E2, tel. 514-340-8222 x. 8231, fax 514-340-7906, e-mail marie.hudson@mcgill.ca

**Word count**: 3795

**Abstract**

**Background:** A number of studies (all N < 200) have assessed health-related quality of life (HRQoL) in systemic sclerosis (SSc), but no systematic review of the effect of SSc on HRQoL has been done. The objective of this study was to systematically review the literature on HRQoL in SSc measured using the Medical Outcomes Trust Short Form 36 (SF 36).

**Methods:** A comprehensive search was conducted in August 2007 using Medline, CINAHL and EMBASE to identify original research studies reporting SF 36 scores of SSc patients. Selected studies were reviewed and characteristics of the study samples and SF 36 data were extracted. Bayesian meta-analysis and meta-regression were performed to obtain pooled estimates of SF 36 physical and mental component summary scores (PCS and MCS, respectively) for all patients as well as by limited and diffuse status.

**Results:** Twelve datasets with a total of 1,127 SSc patients were included in the systematic review. HRQoL was impaired in SSc, with pooled SF 36 PCS scores being more than one standard deviation below the general population (38.3, 95% credible interval (CI): 35.2, 41.5) and pooled MCS scores almost ½ of a standard deviation below the general population (46.6, 95% CI: 44.2, 49.1). SF 36 PCS scores were 3.5 points (95% CI -1.0, 8.0) lower in patients with diffuse compared to limited disease.

**Conclusions:** This study provides robust evidence of the presence and magnitude of impairment in HRQoL in SSc. Although the impairment appears greater in physical health, mental health impairment is also reported.

## Introduction

A focus of medical research has traditionally been measurement of mortality and morbidity. As chronic diseases have become more prevalent[1], researchers have begun to realize that these are not sufficient to capture the experience of disease. Patient-reported outcomes, including measurement of health-related quality of life (HRQoL), have emerged as important outcomes of interest. In addition, information regarding HRQoL serves a number of other purposes. First, *in clinical trials*, treatment efficacy and/or improved survival need to be balanced against adverse effects and impaired HRQoL. Second, HRQoL data can be used by *health care policy-makers* to identify needs and allocate resources for patients with various diseases. Finally, *in the clinical setting*, HRQoL data can allow busy clinicians to monitor their patients' status and make treatment decisions.

Systemic sclerosis (SSc) is a multisystem disorder characterized by a disturbance in fibroblast function, microvascular disease and immune system activation, culminating in fibrosis of skin and internal organs[2]. Although it is a heterogeneous disorder, two common clinical subsets are recognized in terms of skin involvement, *limited* (skin involvement distal to the elbows and knees) or *diffuse* (skin involvement proximal to the elbows and knees in addition to the trunk)[3]. SSc is associated with significant morbidity, including disfiguring skin thickening, finger ulcers, joint contractures, pulmonary hypertension, interstitial lung disease, chronic diarrhea and renal failure. Functional disability is considerable[4] and rates of clinically significant depressive symptoms are high even compared to other medical patient groups[5]. The disease thus encompasses broad multi-dimensional issues including biological, psychological and social processes. Thus,

it would not be surprising that health-related quality of life (HRQoL) should be impaired. However, to date, there has been relatively little work on HRQoL in SSc, and experts have recommended additional research in this area[6].

Given the paucity of data, this systematic review of the literature was carried out to gain greater insight into the HRQoL of patients with SSc. Specifically, we had two objectives: (a) Primary objective - to determine to what extent HRQoL is impaired in SSc; and (b) Secondary objective – to determine whether there are differences in HRQoL between patients with limited and diffuse SSc? The Medical Outcomes Trust Short Form 36 (SF 36)[7] is a widely used generic measure of HRQoL. Thus, to maximize the comparability of the studies selected in this systematic review, we decided to limit the review to those studies using the SF 36 as the main outcome measure of HRQoL.

## Methods

*Methodology of the systematic review* We performed this systematic review of the literature according to guidelines proposed by Stroup et al. for the reporting of meta-analysis of observational studies in epidemiology[8].

*SF 36* The SF 36 is composed of 36 questions that can be grouped into eight domains: physical functioning, role physical, bodily pain, general health, vitality, social function, role emotional and mental health. Each of the domains can be scored separately and have a range of 0-100, with 0 indicating worst and 100 best HRQoL. The scores of the domains can also be combined into two summary scores: the Physical Component Summary (PCS) and the Mental Component Summary (MCS) scores. The summary scores are standardized to responses from the US general population, for which the mean score is 50 and the standard deviation is 10.

*Search strategy and study selection* MEDLINE, EMBASE and CINALH up to August (week 3) 2007 were searched by two investigators (MH and EN) independently using the following search strategies: MEDLINE – ((scleroderma[mh] OR scleroderma[tiab]) AND sf 36[tiab]), and CINALH and EMBASE - (scleroderma and sf 36). The searches were also repeated using systemic sclerosis instead of scleroderma. In addition, reference lists of selected papers and a recent review article[9] were also handsearched by one investigator (MH). Two investigators (MH and EN) independently reviewed the abstracts of each reference identified by the search to identify full-length, published, original research

studies which included SSc patients and which reported SF 36 data. All such studies were selected for full article review. Differences were resolved by consensus.

The same two investigators reviewed the articles identified as potentially eligible based on their abstracts and determined eligibility for study inclusion, again based on consensus. Studies were selected according to the following criteria:

1. the study presented original data;
2. the study included SSc patients;
3. the study reported SF 36 subscale or summary score data;
4. studies of any design without restriction as to language were included;
5. in the case of duplication, with multiple articles publishing data on the same cohort, the most complete dataset or the article whose focus was more specifically on HRQoL in SSc was included;
6. data available only in abstract form was excluded;
7. studies with mixed patient population were included if some data on SSc patients were available, separately.

*Description of studies* Two investigators (MH and EN) independently extracted the data from each selected study using a structured data extraction form. Differences were resolved by consensus. The following information was systematically extracted:

1. Study design (eg. randomized trial, cohort, cross sectional, etc.);
2. Country where study was done;

3. Characteristics of the patients: sample size, criteria used to identify patients with SSc (American College of Rheumatology classification[10] or other classification system); age; percentage of female patients; mean disease duration; percentage of patients with limited and diffuse skin involvement; and
4. SF 36 subscale and summary scores. For clinical trials and cohort studies, SF 36 scores at baseline were recorded.

Authors of individual studies were contacted to obtain complementary data (in particular, SF 36 summary scores) necessary to perform the meta-analysis. Several[11-13] graciously replied and provided the requested data.

*Statistical analysis* Data were extracted and summarized in tabular form. The SF 36 *subscale* scores reported in most studies were not standardized, and were not directly comparable. However, the SF 36 *summary* scores are standardized and can be compared directly. Thus, we performed a meta-analysis with the studies for which we had SF 36 summary scores (nine studies, N 955). We used Bayesian meta-analysis and meta-regression methods to perform our analysis for a number of reasons. First, it was important to be able to select appropriate covariates for the meta-regression and given the small number of studies, Bayesian methods are more reliable for model selection and small sample sizes because they do not depend on large sample properties of the tests[14]. Second, Bayesian methods, particularly when using the WinBUGS/OpenBUGS[15] software, naturally handle problems with missing covariate values. We note that the primary drawback of Bayesian hierarhical modelling is the sensitivity of results to

specification of the prior distributions for the parameters. Therefore, we performed a sensitivity analysis for the most critical choices of prior distribution (Appendix).

We used the Deviance Information Criterion (DIC) to select the covariates for the meta-regression models. The DIC is similar in nature to more commonly used likelihood based model selection criteria for hierarchical models, the AIC and BIC. These three criteria all share a common generic formula by rewarding models that fit the observed data well (as measured by -2 * log-likelihood) and penalizing models that are increasingly complex (where the penalty is some function of the number of model parameters and the number of observations). There are two major practical differences between the DIC and its Frequentist counterparts. First, rather than calculating fit of the model based on the maximized log-likelihood, the model fit component of the DIC is an average of the -2 * log-likelihood over all posterior samples. Secondly, the number of parameters used to penalize for model complexity is estimated from the data, rather than being fixed in advance. This second difference is more important, in that it tries to adjust for the extent to which studies "share" parameters in the model.

All Bayesian parameter estimates and associated credible intervals were obtained using the WinBUGS and OpenBUGS software and the R statistical package[16,17]. All study estimates are based on the runs of three chains with 100,000 samples from each, thinned by a factor of 10 yielding 30,000 iterations for each analysis result presented. The Gelman-Rubin convergence diagnostic[18, 19] in WinBUGS was used to diagnose convergence of the MCMC samplers. Finally, we also analyzed our data with Frequentist methods to allow for comparison in the cases where we have complete data and to assess

more objectively the impact of our Bayesian model assumptions using the MiMa meta-regression package[20].

## Results

*Search results*

The search process identified 22 unique titles (Supplemental data, Figure 1)[11-13, 21-39]. During the title and abstract review process, one study was excluded because it did not report SF 36 data for patients with SSc[29]. Twenty-one articles were selected for full text review. Eight were excluded: five[27, 30-32, 34] because they reported on subsets or duplicated data reported in other included studies[23, 28, 37] with only minor change in the overall sample size and SF 36 scores; one[25] reported data from two other studies, one reported separately and included in the selection[26] and one published only in abstract form[40]; one[35] did not report any SF 36 data; and one[36] reported SF 36 data on 15 patients, of which only four had SSc and their results were not reported separately. One study reported supplementary data[24] to another eligible study[38] and data from the two reports were therefore combined.

*Studies included in the systematic review*

Thus, twelve studies with a total of 1,127 SSc patients were included in the systematic review. Four studies were from the United States[21, 23, 24, 26, 38], four from Italy[13, 33, 37, 39], two from France[22, 28], one from England and Wales[11] and one from Canada[12]. Characteristics of the SSc patients included in the studies are presented in Table 1. SF 36 PCS and MCS scores were availables for nine studies (N 955).

*SF 36 results*

SF 36 data was extracted from the selected studies (Supplemental data, Table 1). The SF 36 PCS scores ranged from 33.4 to 43.8 and the SF 36 MCS scores from 41.0 to 50.7. Bayesian meta-analysis was performed to pool the SF 36 PCS and MCS data from the nine studies (N 955) that reported this data (Figure 1). The resulting posterior mean estimate of the population overall PCS score was 38.3 (95% credible interval: 35.2 to 41.5). Without adjusting for any covariates, we estimated a between-study standard deviation of 4.1 (2.4 to 8.3) (Table 2). Similarly, we obtained an overall estimate of 46.6 (95% credible interval: 44.2 to 49.1) for the MCS score and an unadjusted between-study standard deviation of 3.0 (1.6 to 6.2). The significant heterogeneity between studies, particularly in the PCS score estimate, is visible in the forest plots in Figure 1 and supported by the wide standard deviations mentioned.

We performed meta-regression analyses for the nine studies included in the meta-analyses. The goal of the meta-regression was to use selected population characteristics to try to explain (and therefore reduce) the observed heterogeneity between studies. We looked at various models using age, percentage of diffuse patients, percentage of female subjects and duration of disease among study participants as covariates in the regression models. Table 2 shows the DIC values and the estimated standard deviations of the SF 36 PCS and MCS scores for the unadjusted baseline models and for models adjusting for the selected covariates. For purposes of interpretation, lower DIC values indicate better model fit. For the SF 36 PCS scores, we found that age (DIC 37.0) or percentage of diffuse patients (DIC 37.4) yielded slightly better fitting models than the baseline unadjusted model (DIC 37.7). Of note, however, was that the effect of age and percentage

of patients with diffuse disease went in opposite directions, with increasing age being associated with better and diffuse disease worse SF 36 PCS scores. The model containing both covariates actually performed worse, due to the correlation between age and percentage of patients with diffuse disease. Percentage of female subjects and disease duration did not have any effect, either by themselves or in addition to the other covariates. None of the covariates helped to explain the observed heterogeneity in the SF 36 MCS scores.

Forest plots for the pooled study estimates adjusting for percentage of diffuse patients are also presented in Figure 1. Although the pooled unadjusted and adjusted estimates appear similar, the heterogeneity in the forest plots is reduced in the adjusted models for the SF 36 PCS. In addition, the standard deviations in the adjusted models for the SF 36 PCS are almost half of those in the unadjusted models (Table 2). Thus, adjusting for percentage of patients with diffuse disease considerably decreases the heterogeneity among studies included in the meta-analysis for the SF 36 PCS. Similar findings were obtained when adjusting for age (data not shown).

*Comparison of Limited and Diffuse Subsets*

Eight studies (N 797) reported SF 36 scores by extent of skin involvement (Supplemental data, Table 2). SF 36 PCS scores ranged from 36.8 to 43.8 in patients with limited disease, compared to 32.4 to 43.7 in patients with diffuse disease. SF 36 MCS scores ranged from 40.0 to 54.1 in patients with limited disease, compared to 40.0 to 50.6 in patients with diffuse disease. Two studies (N 157)[12, 28] reported significantly *worse PCS scores in patients with diffuse compared to limited disease* and one[33] (N 24) reported

significantly *worse MCS in patients with limited compared to diffuse disease*. Of the eight studies, two contained *only* subjects of one disease type (one study with only diffuse patients[24] and the other with only limited patients[11]) which were excluded from the meta-analysis looking at *differences* between subsets. Thus, the results in this section are derived from the data of six studies (N 414, Table 2 and Figure 2).

Since there were few studies and most had small sample sizes, we pooled the results using meta-analytic methods rather than using meta-regression. The bottom section of Table 2 shows the results of two different models for this meta-analysis. The fixed effect model assumes that there is no between-study heterogeneity in SF 36 difference between limited and diffuse patients. The random effects model assumes that there exists between-study heterogeneity in the differences between limited and diffuse patients. The DIC values for the random effects models for both the SF 36 PCS and MCS were the lowest, indicating better fit than the fixed effect models. Using the random effects model, we estimated that patients with limited disease had an SF 36 PCS score that was 3.5 (-1.0 to 8.0) higher than patients with diffuse disease. On the other hand, limited and diffuse patients did not seem to differ in their SF 36 MCS scores (the estimates for the difference in scores between the two groups were close to 0 in both models). The forest plots in Figure 2 show the results for the random effects model.

Of note, we also found considerable heterogeneity between studies included in this meta-analysis (bottom section, Table 2). However, since we did not have covariate information on the individual subsets (limited or diffuse) for many of the studies, we decided not to perform a meta-regression on these data. Instead, we examined how the inference about the difference in HRQoL between limited and diffuse patients would

change under different values of true between-study heterogeneity (as measured by the standard deviation of the true study differences). Figure 3 shows that for PCS, the amount of between-study heterogeneity does not affect the estimate of the difference, but does affect the perceived likelihood of the difference being greater than zero. For the MCS, there is no amount of between-study heterogeneity that would lead to the conclusion that there was a difference between limited and diffuse patients' mental HRQoL.

*Sensitivity Analyses and Computational Details*

We conducted a sensitivity analyses to assess the robustness of our results to the specification of our models. First, we re-fit the meta-analyses of the preceding sections to the 8 studies with completely observed data using Frequentist random effects meta-analysis methods in the R statistical package via the meta library and the MiMa meta-regression software. We found that the Frequentist approach yielded very similar parameter estimates to what we obtained with the Bayesian models. In the one situation where they differed, the Frequentist meta-analyses showed a more statistically significant difference between diffuse and limited patients ($p = 0.047$) than the Bayesian meta-analysis of the differences. Similarly, we tested the robustness of our Bayesian model prior specification by obtaining results for the four prior specifications detailed in the Appendix, in particular the important prior distribution on the random effects variance. We did not see any substantial difference in the parameter estimates themselves (Figure 4), although in the case of the differences one interval contained 0 (for Prior d) whereas the three others did not.

## Conclusions

In this systematic review of 12 studies with a total of more than 1000 SSc patients, we found significant impairment in the HRQoL of SSc patients. Although the impairment in physical health appeared greater (SF 36 PCS was more than 1 standard deviation below that of the general population), mental health was also impaired (SF 36 MCS was almost one half standard deviation below that of the general population). Moreover, the physical health of patients with diffuse disease was approximately one half standard deviation below that of patients with limited disease, whereas mental health was impaired to the same extent in both subsets of diseases. The minimal important clinical differences (MICD) is an important measure of change in HRQoL[41] and represents the smallest change in the score that patients can perceive. A change in score of 2.5-5.0 has been suggested as representing a MICD and has been previously used in SSc[23]. Thus, although the differences identified in this study were obtained using cross-sectional data, the differences in PCS but not MCS scores between diffuse and limited disease are also likely to be *clinically meaningful*.

The significance of this study is two-fold. Firstly, it provides evidence of the presence and magnitude of impairment in HRQoL in SSc, both in physical and mental health. Indeed, although several small studies had found that SF 36 PCS scores were impaired in SSc, the results were inconsistent (Supplemental data, Tables 1 and 2). Moreover, SF 36 MCS scores were thought to be relatively "preserved" in SSc, leading some to argue that, despite significant impairments in physical health, SSc patients adapt well to their slowly progressing disease[23]. This was incongruent, however, with reports of high rates of depressive symptoms in SSc[5, 42]. Thus, this study provides strong evidence

that HRQoL is considerably impaired both in the physical and mental health domains in SSc. Secondly, policy-makers may be unfamiliar with this rare but devastating disease. These data, which show significant impairment in the physical health of SSc patients *of more than 1 standard deviation compared to the general population*, provide valuable evidence that physicians and patient groups can use to advocate for resources for patients who suffer from this severe and devastating disease.

In the meta-regression analyses, we found that controlling for age or the percentage of patients with diffuse disease, but not both together, decreased the heterogeneity between studies. In addition, we found that the effect of age and percentage of diffuse patients was in opposite directions, with increasing age being associated with better and diffuse disease worse SF 36 PCS scores. Although the finding related to age appears counter-intuitive, we hypothesize that it may be the result of confounding and survival bias, with patients with limited disease having better survival than those with diffuse disease and thus surviving to older ages. Indeed, in our analyses stratified by limited and diffuse disease, we did show that patients with limited disease had better SF 36 PCS scores than those with diffuse disease. However, although data was not available to allow us to demonstrate that the patients with limited disease included in the systematic review were older than those with diffuse disease, there is nevertheless some independent evidence to suggest this. First, patients with diffuse disease are believed to have worse survival that those with limited disease[43, 44]. Second, in the only paper included in the review that reported age separately according to limited or diffuse status, 70% of patients with limited disease were $\geq$ 55 compared to only 49% of those with diffuse disease. Thus, we believe that the model adjusting for limited or diffuse disease

(Figure 1) is the model that provides the best estimate of the pooled SF 36 PCS (38.2, 95% credible interval 36.2 to 40.4) and MCS (46.6, 95%credible interval 43.9 to 49.2) scores in SSc.

A study such as this is not without limitations. A systematic review of published studies is limited by the fact that it excludes unpublished data and this may result in publication bias, whereby studies with negative results may be less likely to have been published and included in the analysis. We attempted to examine this using funnel plots (Funnel Plots, Supplemental data). These showed that only the overall meta-analysis of MCS showed evidence of asymmetry, suggesting that the larger studies tended to have more normal MCS scores. The other three analyses did not show significant asymmetry, although this is difficult to assess fully given the small number of studies. Confounding is also possible, given the lack of individual patient data. Nevertheless, we attempted to control for some of the possible confounding in the meta-regression by adjusting for common confounders including age and gender, albeit at the group level. Unfortunately, we did not have characteristics of patients by subset of disease (limited or diffuse) and could not perform a meta-regression for that part of the analysis. Thus, we acknowledge that confounding remains a possibility in that analysis. Finally, the patient inclusion criteria for each study with regards to disease subset were clearly quite varied, with studies ranging from having only limited patients to only diffuse patients (Table 1). Such heterogeneity in selection could in fact affect the analysis. First, it could cause our estimates to be less precise due to the possibility of estimating different population parameters in each study. Secondly, if the patients selected for the studies were somehow different than the general population of SSc patients, this could also have an impact on

the generalizability of our results. However, we view the heterogeneity of the patient population as a strength, rather than a weakness of our study. We were still able to detect a difference in physical HRQoL, in spite of the very different patient populations and the limited number of studies. In our opinion, it is far more likely that having studies with more homogeneous populations would strengthen our results, rather than reveal a systematic bias.

In conclusion, this study provides robust evidence that HRQoL is considerably impaired in SSc. This finding should now serve as our call to action to identify targets and implement interventions that have the ability to improve the HRQoL of those living with this devastating disease.

**Table 1 Characteristics of studies included in the systematic review**

| | Design | Country | Sample size | Method of diagnosis | Age, years | % Female | Disease duration, years | % Diffuse |
|---|---|---|---|---|---|---|---|---|
| Khanna, SLS, 2007[23] | RCT | US | 158 | ACR criteria | Mean 48.5 (SD 12.3) | 71 | Mean 3.1 (SD 2.1)* | 59 |
| Sallam, 2007[21] | Cohort | US | 17 | ACR criteria | Mean 57.2 (SD 1.9) | 82 | Not reported | 53 |
| Khanna, Relaxin, 2007[24, 38] | RCT | US | 196 | ACR criteria | Mean 47.2 (SD 10.3) | 85 | Mean 2.2 (SD 1.4)* | 100 |
| Rannou, 2007[22] | Cross sectional | France | 50 | ACR criteria and/or Leroy and Medsger criteria | Mean 54 (SD 12) | 88 | Mean 9.1 (SD 8.8)** | 46 |
| Khanna, HV, 2007[26] | Cross sectional | US | 107 | ACR criteria | 61% were > 55 | 91 | Median 7 (IQR 4-12)** | 45 |
| Georges, 2006[28] | Cross sectional | France | 89 | ACR criteria and/or Leroy and Medsger criteria | Median 51 (range 19-77) | 80 | Median 5 (range 1-34)** | 75 |
| Gliddon, 2006[11] | Cross sectional | England and Wales | 187 | Not reported | Mean 55 (SD 11.8) | 86 | Median 3.9 (IQR 1.2-8.0) | 0 |
| Johnson, 2006[12] | Cross sectional | Canada | 68 | ACR criteria | Mean 48.2 (SD 12.5) | 87 | Mean 7.1 (SD 5.9)** | 51 |
| Milio, 2006[39] | RCT | Italy | 60 | Leroy and Medsger criteria | Mean 39 (SD 21) | 82 | Mean 6 (4)** | Not reported |
| Danieli, 2004[13] | Cross sectional | Italy | 76 | ACR criteria | Median 58 (IQR 48-65) | 92 | Median 8 (IQR 4-13)** | 32 |
| Del Rosso, 2004[33] | Cross sectional | Italy | 24 | Not reported | Mean 53.4 (SD 15.1) | 88 | Mean 8.3 (SD 6.6)* | 38 |
| Cossutta, 2002[37] | Cross sectional | Italy | 95 | Not reported | Median 60 (range39-83) | 97 | Median 6 (range 39-83) | 57 |

SLS – Scleroderma Lung Study; Relaxin – Relaxin Study; HV – Health Values Study; RCT – Randomized Clinical Trial; US – United States; ACR – American College of Rheumatology; SD – standard deviation; IQR – interquartile range

*Since the onset of the first non-Raynaud's disease manifestation.

**Time of onset for measurement of disease duration not reported.

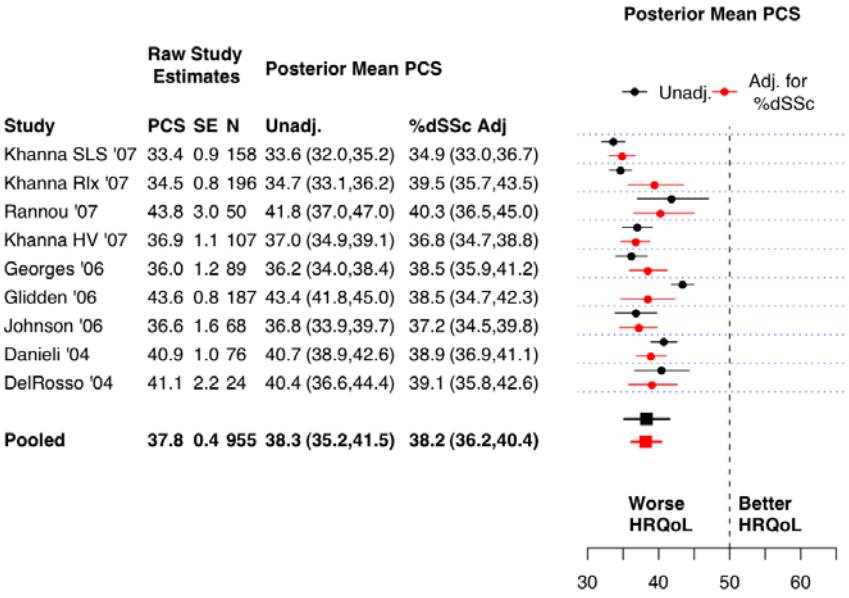
Highlighted studies reported SF 36 summary scores and were included in the meta-analysis. Details can be found in the Supplementary tables.

**Table 2 Meta-analysis and meta-regression models for SF 36 PCS and MCS.**

**Smaller Deviance Information Criterion** (DIC) values indicate better model fit. Between study heterogeneity is assessed using the standard deviation of adjusted study means and the coefficient estimate is the simple meta-regression coefficient for the model of interest. 95% credible intervals for all parameter estimates are contained in ( )'s.

| **Analysis of Overall HRQoL (9 studies, N = 955)** | | | | | | |
|---|---|---|---|---|---|---|
| | **SF 36 PCS** | | | **SF 36 MCS** | | |
| **Covariates** | **DIC** | **Between study heterogeneity** | **Coefficient estimate** | **DIC** | **Between study heterogeneity** | **Coefficient estimate** |
| **Unadjusted** | 37.7 | 4.1 (2.4, 8.3) | ---------- | 37.3 | 3.0 (1.6, 6.2) | ---------- |
| **Age** | 37.0 | 2.3 (0.9, 5.3) | 0.8 (0.2, 1.4) | 40.4 | 3.2 (1.4, 6.8) | -0.0 (-0.7, 0.7) |
| **% Diffuse** | 37.4 | 2.3 (1.0, 5.5) | -0.1 (-0.2,-0.0) | 37.4 | 3.2 (1.7, 6.7) | -0.0 (-0.1, 0.1) |
| **% Female** | 37.6 | 3.8 (1.8, 7.6) | 0.3 (-0.1, 0.8) | 38.9 | 3.2 (0.8, 6.6) | -0.1 (-0.5, 0.4) |
| **Disease duration** | 37.4 | 3.7 (2.1, 7.9) | 0.8 (-0.5, 2.1) | 39.3 | 2.9 (0.5, 6.5) | -0.3 (-1.4, 0.8) |
| **Analysis of HRQoL Difference, Limited – Diffuse (6 studies, N = 414)** | | | | | | |
| **Model** | **DIC** | **Between Study Heterogeneity** | **Difference in SF 36 PCS scores (Limited - Diffuse)** | **DIC** | **Between Study Heterogeneity** | **Difference in SF 36 MCS scores (Limited - Diffuse)** |
| **Fixed Effect** | 37.3 | ---------- | 3.6 (1.3, 5.8) | 39.7 | ---------- | –0.4 (–2.8, 2.0) |
| **Random Effects** | 34.7 | 4.0 (1.1, 8.0) | 3.5 (-1.0, 8.0) | 36.3 | 5.0 (0.2, 9.4) | –0.5 (–5.0, 4.0) |

**Figure 1 Crude and Diffuse percentage-adjusted SF 36 PCS and MCS scores in studies selected for the systematic review**



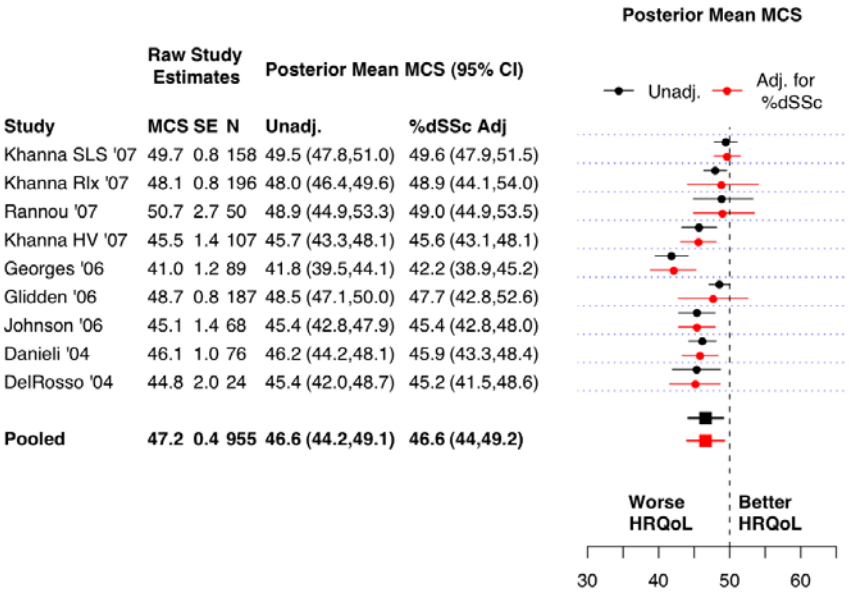
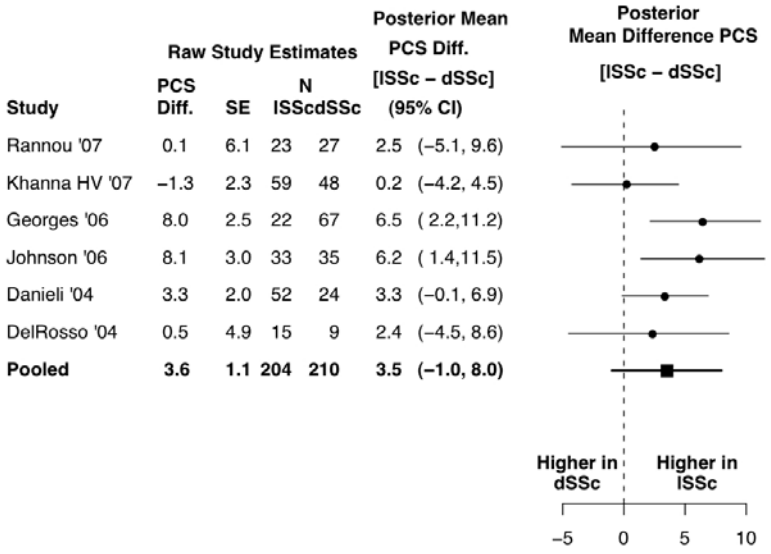



**Figure 2 SF 36 PCS and MCS differences between limited and diffuse patients in studies with data on both limited and diffuse patients**



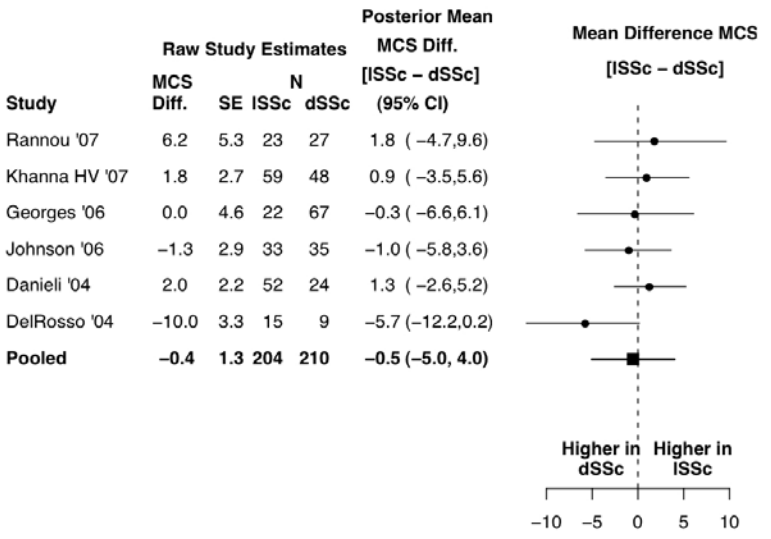



**Figure 3 Sensitivity of estimation of HRQoL difference between limited and diffuse patients to the true heterogeneity in study populations.** The two figures show the sensitivity of the inference about the difference in HRQoL between limited and diffuse patients to the true value of the standard deviation of the study-specific differences. The solid lines represent the estimated SSc population PCS/MCS mean group difference at an assumed value of the true standard deviation of the study differences. The dashed lines represent 95% credible limits for the overall popuation difference if the true study difference standard deviation were known. The figure on the left contains results for the PCS, the figure on the right contains results for the MCS.

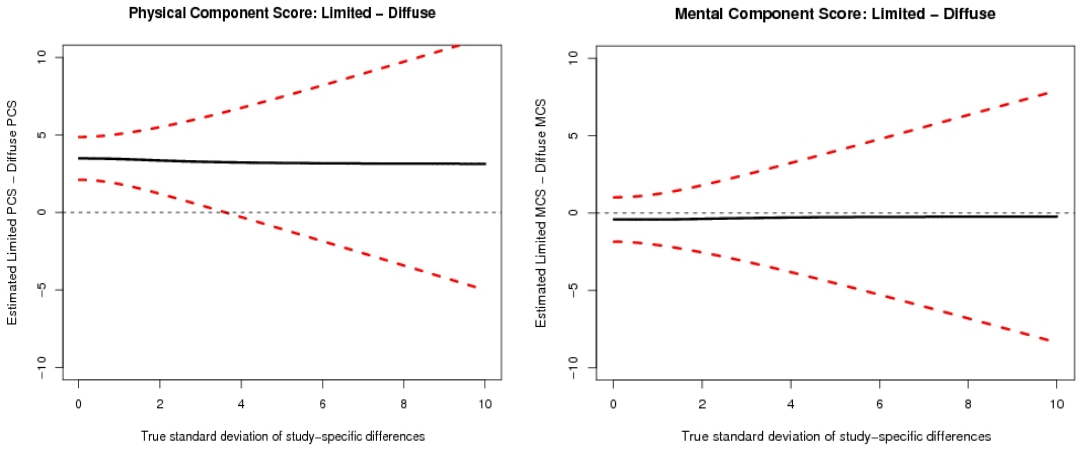
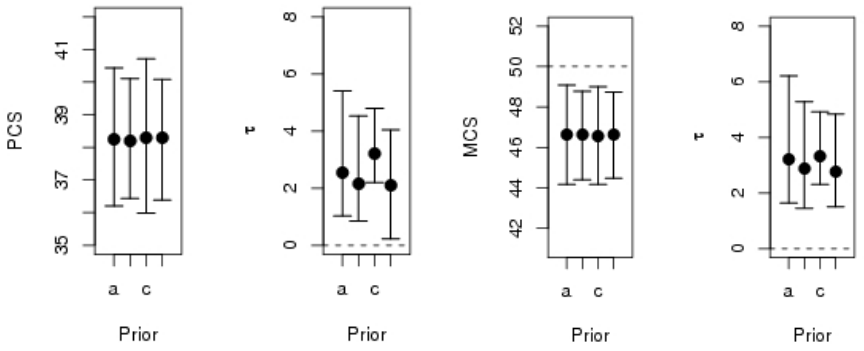
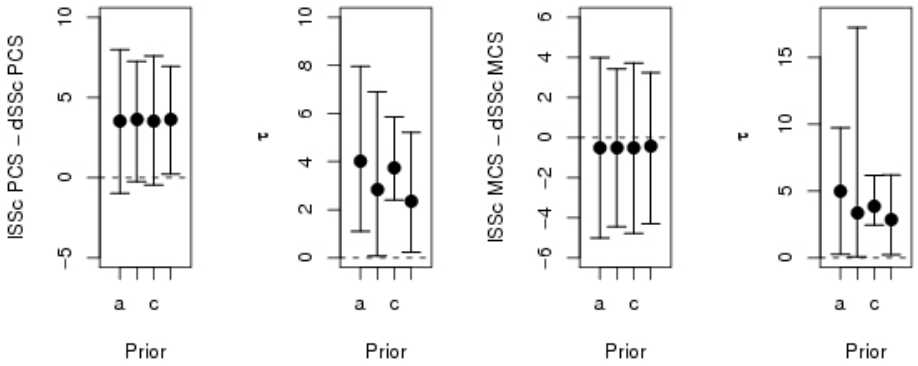


**Figure 4 Sensitivity of results to different choices of prior distribution.** The four graphs below show the sensitivity of the main results to the choice of prior distribution (a,b,c,d defined in the Appendix) for the DIC-best models using posterior means (circles) and the 95% credible interval limits (bars). Figure 4(i) shows estimates of the overall mean PCS and of between-study heterogeneity for the meta-regression model with age as a covariate. Figure 4(ii) shows estimates of the overall mean MCS and of between-study heterogeneity for the meta-analysis model with no covariates. Figures 4(iii) and 4(iv) show the mean difference between limited and diffuse patients and between-study heterogeneity of the differences for PCS and MCS, respectively.

**(i)** **(ii)**




**(iii)** **(iv)**



**References**


1. World Health Organization. Preventing Chronic Diseases: A Vital Investment: WHO Global Report. 2005. http://whqlibdocwhoint/publications/2005/9241563001_engpdf, accessed October 1, 2008.

2. Seibold J. Scleroderma. In: Harris E, Budd, RC, Firestein GS, Genovese, MC, Sergent, JS, Ruddy, S, Sledge, CB, ed. Kelley's Textbook of Rheumatology. 7th ed. Philadelphia: Elsevier; 2005.

3. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15(2):202-5.

4. 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(11):1984-91.

5. Thombs B, Taillefer S, Hudson M, M B. Depression in patients with systemic sclerosis: a systematic review of the evidence. Arthritis Rheum 2007;57(6):1089-97.

6. 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(7):1630-47.

7. 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.

8. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama 2000;283(15):2008-12.

9. Khanna D. Health-Related Quality of Life: A primer with a Focus on Scleroderma. Scleroderma Care and Research 2006;3(2):3-13.

10. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23(5):581-90.

11. Gliddon AE, Dore CJ, Maddison PJ. Influence of clinical features on the health status of patients with limited cutaneous systemic sclerosis. Arthritis Rheum 2006;55(3):473-9.

12. Johnson SR, Gladman DD, Schentag CT, Lee P. Quality of life and functional status in systemic sclerosis compared to other rheumatic diseases. J Rheumatol 2006;33(6):1117-22.

13. Danieli E, Airo P, Bettoni L, et al. Health-related quality of life measured by the Short Form 36 (SF-36) in systemic sclerosis: correlations with indexes of disease activity and severity, disability, and depressive symptoms. Clin Rheumatol 2005;24(1):48-54.

14. Spiegelhalter D, Myles J, Jones D, Abrams K. Bayesian methods in health technology assessment: a review. Health Technology Assessment 2000;4(38).

15. Spiegelhalter DJ, Thomas A, Best NG. WinBUGS version 1.2 User Manual. 1999 MRC Biostatistics Unit.

16. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2007.

17. Schwarzer G. meta: Meta-analysis. 2007, R package version 0.8-2.

18. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. Statistical Science 1992;7:457-511.

19. Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 1997;7:434-55.

20. Viechtbauer W. MiMa: An S-Plus/R function to fit meta-analytic mixed-,random-,and fixed-effects models [Computer software and manual]. Retrieved from http://www.wvbauer.com/.

21. Sallam H, McNearney TA, Doshi D, Chen JD. Transcutaneous electrical nerve stimulation (TENS) improves upper GI symptoms and balances the sympathovagal activity in scleroderma patients. Dig Dis Sci 2007;52(5):1329-37.

22. Rannou F, Poiraudeau S, Berezne A, et al. Assessing Disability and Quality of Life in Systemic Sclerosis: Construct Validities of the Cochin Hand Function Scale, Health Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ, and Medical Outcomes Study 36-Item Short Form Health Survey. Arthritis Rheum 2007;57(1):94-102.

23. Khanna D, Yan X, Tashkin DP, et al. Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. Arthritis Rheum 2007;56(5):1676-84.

24. Khanna D, Park GS, Seibold J. SF-36 Scales in the Relaxin study. Rheumatology (Oxford) 2007;46(4):724.

25. Khanna D, Furst DE, Wong WK, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. Qual Life Res 2007;16(6):1083-92.

26. Khanna D, Ahmed M, Furst DE, et al. Health values of patients with systemic sclerosis. Arthritis Rheum 2007;57(1):86-93.

27. Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354(25):2655-66.

28. Georges C, Chassany O, Toledano C, et al. Impact of pain in health related quality of life of patients with systemic sclerosis. Rheumatology (Oxford) 2006;45(10):1298-302.

29. Sandqvist G, Akesson A, Eklund M. Daily occupations and well-being in women with limited cutaneous systemic sclerosis. Am J Occup Ther 2005;59(4):390-7.

30. Khanna D, Clements PJ, Furst DE, 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(2):592-600.

31. Georges C, Chassany O, Mouthon L, et al. Validation of French version of the Scleroderma Health Assessment Questionnaire (SSc HAQ). Clin Rheumatol 2005;24(1):3-10.

32. Georges C, Chassany O, Mouthon L, et al. [Quality of life assessment with the MOS-SF36 in patients with systemic sclerosis]. Rev Med Interne 2004;25(1):16-21.

33. Del Rosso A, Boldrini M, D'Agostino D, et al. Health-related quality of life in systemic sclerosis as measured by the Short Form 36: relationship with clinical and biologic markers. Arthritis Rheum 2004;51(3):475-81.

34. Belotti Masserini A, Zeni S, Cossutta R, Soldi A, Fantini F. [Cost-of-illness in systemic sclerosis: a retrospective study of an Italian cohort of 106 patients]. Reumatismo 2003;55(4):245-55.

35. Kenefick NJ, Vaizey CJ, Nicholls RJ, Cohen R, Kamm MA. Sacral nerve stimulation for faecal incontinence due to systemic sclerosis. Gut 2002;51(6):881-3.

36. Kenefick NJ, Vaizey CJ, Cohen RC, Nicholls RJ, Kamm MA. Medium-term results of permanent sacral nerve stimulation for faecal incontinence. Br J Surg 2002;89(7):896-901.

37. Cossutta R, Zeni S, Soldi A, Colombelli P, Belotti Masserini A, Fantini F. [Evaluation of quality of life in patients with systemic sclerosis by administering the SF-36 questionnaire]. Reumatismo 2002;54(2):122-7.

38. Khanna D, Furst DE, Clements PJ, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. J Rheumatol 2005;32(5):832-40.

39. Milio G, Corrado E, Genova C, et al. Iloprost treatment in patients with Raynaud's phenomenon secondary to systemic sclerosis and the quality of life: a new therapeutic protocol. Rheumatology (Oxford) 2006;45(8):999-1004.

40. Postlethwaite A, Furst D, Wong W, Clements P. Oral Tolerance (OT). Induction to type I collagen (Cl) significantly reduces the skin score in patients with diffuse systemic sclerosis (SSc) with late-phase disease. Results of a NIAMS/NIAID multicenter phase II placebo-controlled double blind clinical trial. Arthritis and rheumatism (abstract) 2005.

41. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407-15.

42. Thombs BD, Hudson M, Taillefer SS, Baron M. Prevalence and clinical correlates of symptoms of depression in patients with systemic sclerosis. Arthritis Rheum 2008;59(4):504-9.

43. Mayes MD, Lacey JV, Jr., Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48(8):2246-55.

44. Silman AJ. Scleroderma--demographics and survival. J Rheumatol Suppl 1997;48:58-61.