The Minimally Important Difference for the Fatigue Visual Analog Scale in Patients with Rheumatoid Arthritis Followed in an Academic Clinical Practice

DINESHKHANNA, JANET E. POPE, PUJA P. KHANNA, MICHELLE MALONEY, NOOSHIN SAMEDI, DEBBIENORRIE, GILLIANOUIMET, and ROND. HAYS

ABSTRACT. Objective. To estimate the minimally important difference (MID) for a fatigue visual analog scale (VAS) using patient-reported anchors (fatigue, pain, and overall health).

Methods. Patients with rheumatoid arthritis (RA; n = 307) had 2 clinic visits at a median of 5.9 months apart. They completed a fatigue VAS (0–10 scale) and the retrospective anchor items, “How would you describe your overall fatigue/pain/overall health since the last visit?” with response options: Much worsened, Somewhat worsened, Same, Somewhat better, or Much better. The fatigue anchor was used for primary analysis and the pain/overall health anchors for sensitivity analyses. The minimally changed group was defined by those reporting they were somewhat better or somewhat worsened.

Results. The mean [standard deviation (SD)] age was 59.4 (13.2) years, disease duration was 14.1 (11.5) years, and 83% of patients were women. The baseline mean (SD) Health Assessment Questionnaire–Disability Index score was 0.84 (0.75). The baseline fatigue VAS score was 4.2 (2.9) and at followup was 4.3 (2.8) [mean change of –0.07 (2.5); p = not significant]. The fatigue change score (0–10 scale) for Somewhat better and Somewhat worsened for the fatigue anchor averaged –1.12 and 1.26, respectively. Using the pain anchor, the fatigue change score for Somewhat better and Somewhat worsened averaged –0.87 and 1.13; and using the global anchor, the fatigue change score for Somewhat better and Somewhat worsened averaged –0.82 and 1.17, respectively. Effect size estimates using 3 anchors were small for the Somewhat better (range 0.27–0.39) and Somewhat worsened (0.40–0.44) groups, but larger than for the no-change group (0.03–0.08).

Conclusion. The MID for fatigue VAS is between –0.82 for –1.12 for improvement and is 1.13 to 1.26 for worsening on a 0–10 scale in a large RA clinical practice, and is similar to that seen in RA clinical trials. This information can aid in interpreting fatigue VAS in day-to-day care in clinical practice. (First Release Nov 1 2008; J Rheumatol 2008;35:2339–43; doi:10.3899/jrheum.080375)

Key Indexing Terms:
- MINIMAL CLINICALLYIMPORTANT DIFFERENCES
- RHEUMATOID ARTHRITIS
- FATIGUE
- RELIABLE CHANGE INDEX
- VISUAL ANALOG SCALE
- CLINICAL PRACTICE

Fatigue is very common in patients with rheumatoid arthritis (RA) and is associated with poor day-to-day functioning and overall sense of well-being. Fatigue is related to the RA disease activity and other non-RA factors (poor sleep, stress, etc). The prevalence of fatigue in RA, defined as a rating of ≥ 2 on a 0 to 10 visual analog scale (VAS), is...
Fatigue remains an important concern for patients with RA and Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) recommends inclusion of fatigue as a patient-centered outcome in RA studies.

Recent clinical trials have used fatigue as an outcome measure and fatigue VAS has been found to be sensitive to change in RA-related clinical outcomes. In addition, Wolfe recently compared the fatigue VAS scale to multi-item fatigue instruments in a large observational study, and showed that VAS scale and multi-item scales had comparable responsiveness statistics and similar correlation coefficients with clinical outcomes in RA. As future clinical trials and observational studies in RA are likely to include a measure of fatigue, it is important to estimate the minimal clinically important difference (MCID) or minimally important difference (MID) — the smallest improvement in score that patients perceive as beneficial and that may lead to a change in the patient’s management of fatigue. MID can help clinicians understand whether fatigue score differences between 2 treatment groups are meaningful and if changes within 1 group over time are clinically meaningful. The MID estimates were assessed using an anchor-based approach. An “anchor” is a clinically relevant indicator or pointer to which a patient-reported outcome change can be tied. These measures are of clinical relevance and can be “subjective,” such as self-reports of change, or “objective,” such as clinical indicators of response to treatment (joint count or disease severity). Subjective anchors rely on an individual’s (subject’s or his/her physician’s) assessment of the patient’s condition. A global rating of change is a well accepted subjective anchor in patient-reported outcome research. It is a retrospective assessment of change where a person thinks back to a previous timepoint and states whether he/she has experienced change in a domain of health from that timepoint to present.

We prospectively studied a large clinical practice at a university hospital to determine the MID estimates for fatigue VAS (0–10 scale; for improvement and worsening) using 3 patient-reported anchors — fatigue, pain, and overall health. Previous studies have shown an inherent uncertainty around the MID estimates and experts have recommended using several anchors. Our primary anchor was the fatigue scale, but we also performed our analyses using overall and pain anchors, as a previous study has shown an association between pain, overall health, and fatigue.

MATERIALS AND METHODS

Multiple data are collected routinely on patients seen at St. Joseph’s Hospital Rheumatology Clinic, which is affiliated with the University of Western Ontario and serves a referral region of approximately 1 million. The data are from patients (n = 307) with RA who had at least 2 consecutive visits within 12 months (median duration 5.9 mo) with their rheumatologists. As patients with low disease activity may not have a frequent follow-up with rheumatologists, we chose the 12-month duration so we can include all eligible subjects. Data were extracted from medical charts by trained data-extraction persons and entered into a database. Patients completed the fatigue VAS (0–10), which stated, “How much problem has unusual fatigue or tiredness been for you over the past week?” and was anchored from 0 (fatigue is no problem) to 10 (fatigue is a major problem). In addition, patients completed the Health Assessment Questionnaire-Disability Index (HAQ-DI) and a retrospective anchor item at visit 2, “How would you describe your overall fatigue since the last visit?” with response options: Much worsened, Somewhat worsened, Same, Somewhat better, or Much better. Patients who reported Somewhat better or Somewhat worsened at visit 2 were defined as the minimally changed subgroup. The changes in the fatigue VAS scores (time2 – time1) for the group that reported Somewhat better and Somewhat worsened were determined in order to estimate the MID. This was compared to change scores for the group that reported Same, Much better, and Much worsened.

The normality of fatigue scores was assessed using the Shapiro-Wilk test. The test was nonsignificant (p = 0.06) and we report the MID estimates as mean and 95% confidence intervals (95% CI).

The fatigue change scores were calculated (fatigue at time2 – fatigue at time1). To assess the usefulness of an anchor, experts recommend reporting correlation between the anchor and changed score; for example, a correlation of zero will make the anchor useless, so a correlation coefficient of 0.30–0.35 has been suggested (for details see Hays, Revicki, et al). We assessed the association between anchor and fatigue change score using the Spearman correlation coefficient (as the anchor is an ordinal variable). The MID was estimated by examining change in the fatigue VAS scores in subjects who were slightly better and slightly worsened. These estimates were compared to those who improved or worsened more than slightly. Responsiveness to change was evaluated using the effect size (ES). ES is ratio of observed change to a measure of variance (also known as signal to noise). For ES, the numerator is the mean change in the fatigue VAS from the baseline to followup and the denominator is the standard deviation of fatigue VAS at baseline (SD = 2.87). Cohen’s rule of thumb for interpreting ES is that a value of 0.20–0.49 represents a small change, 0.50–0.79 a medium change, and 0.80 or greater a large change and, in general, an ES of 0.2 to 0.5 is usually considered relevant for MID.

We also estimated MID scores using pain and overall health assessments as anchors. The patients were asked, “How would you describe your pain/overall health since the last visit?” on a scale labeled: Much worsened, Somewhat worsened, Same, Somewhat better, Much better.

In addition, the effect of baseline fatigue VAS score on the MID estimates was estimated. In other words, people with different baseline fatigue scores may require different amounts of improvement or worsening to consider a change to represent an MID. We divided the baseline fatigue VAS scores into 2 groups at the median score — the Less severe group (fatigue VAS ≤ 4.5) and the More severe group (fatigue VAS > 4.5). We also assessed the significance of change at the individual level. The reliable change index (RCI) is a measure to assess the magnitude of change score necessary to be considered statistically reliable and not due to random measurement error, and is a function of the SD and reliability coefficient of an instrument. When RCI is > 1.96, it is unlikely that the posttest score is due to random measurement error and the change is reliable (p < 0.05). RCI is calculated by the following formula:

\[
\text{RCI} = \frac{\text{(individual fatigue score} - \text{individual fatigue score}_1)}{\text{SD}_{\text{baseline}} \times (2 \times [1 - \text{reliability}])^{1/2}}
\]

We did not assess test-retest reliability of our fatigue VAS. However, a reliability (test-retest) coefficient of 0.70 for the fatigue VAS was found in patients with RA. Therefore, we based our primary analysis on the reliability coefficient of 0.70 and performed sensitivity analyses with 0.80 and 0.90. We report the scores that are statistically reliable, and proportion of patients who had individual change greater than the statistically reliable change (improved or worsened) and proportion that had individual changes less than the statistically reliable change.
The data were analyzed using Stata 9.2. p < 0.05 was deemed to indicate statistical significance.

RESULTS

The mean (SD) age of the patients was 59.4 (13.2) years, the mean (SD) disease duration was 14.1 (11.5) years, and 256 (83%) were women; the median (25th, 75th percentile) followup between 2 visits was 5.9 (3.9, 7.6) months. Twenty-eight patients (9%) reported no fatigue at baseline and 5 (2%) reported maximum fatigue at baseline (Figure 1). The majority of patients (92%) were using disease-modifying antirheumatic drug (DMARD) therapy, 16% prednisone, 50% nonsteroidal antiinflammatory drugs (NSAID), and 17% were prescribed biologics. The baseline HAQ-DI was 0.84 (SD 0.75), suggesting mild functional disability. The baseline fatigue VAS was 4.2 (SD 2.9) and at followup was 4.3 (SD 2.8) (mean change of –0.07 ± 2.5; p = nonsignificant).

The Spearman correlation between the patient assessment of fatigue and change in the fatigue VAS was 0.37 (p < 0.001). Of 307 patients, 171 (56%) reported no change at followup visit, 40 (13%) reported being somewhat improved, and 17 (5%) reported being much better. In contrast, 65 (21%) reported being somewhat worsened and 14 (5%) reported being much worsened.

The fatigue change scores for the Somewhat better (n = 40) and Somewhat worsened (n = 65) groups averaged –1.12 and 1.26 (95% CI –1.86 to –0.37 and 0.68 to 1.86), respectively (Table 1). The change in fatigue scores for those who were much better (n = 17) was –1.44 (95% CI –2.78, –0.10) and it was +3.35 (95% CI 1.94, 4.78) for much worse (n = 14). All the change scores were statistically different (p < 0.05) from the score for the no-change group [n = 171; score –0.23 (95% CI –0.57, 0.11)]. ES estimates were small for the Somewhat better (0.39) and Somewhat worse (0.44) groups, but larger than for the no-change group (ES = 0.08).

Table 2 provides the mean MID scores for improvement in the fatigue scores stratified by the “less severe” fatigue (fatigue VAS ≤ 4.5) versus “more severe” (fatigue VAS > 4.5) groups at baseline. The data suggest that patients with “more severe” baseline fatigue scores require a larger change to be minimally improved (MID for the minimally improved group for the low score group = –0.15, high score group = –3.38). In contrast, the change required for the slightly worsened group with “more severe” fatigue (0.21) was smaller than that for the “less severe” (3.08) fatigue group.

**Reliable change index.** The amount of change on the fatigue VAS needed to have the RCI to be 1.96 was 3.47 when the reliability coefficient was 0.70. In other words, a change of score from time1 to time2 (improvement or worsening) of > 3.47 (on a 0–10 scale) was a real change and was not due to measurement error of the instrument. A change < 3.47 may be real, but a random measurement error cannot be ruled out. Out of 307 patients, 26 (8%) had worsening and 30 (10%) had improvement of scores of > 3.47. The RCI was 3.14 and 2.64 when the test-retest reliability coefficients were 0.80 and 0.90, respectively.

DISCUSSION

Fatigue is very common in patients with RA, is likely to be multifactorial, and was recommended as a patient-cen-
tered outcome during the OMERACT 8 meeting. Work by Kirwan and Hewlett and Wolfe and Michaud has shown that fatigue is an important concern for patients with RA.

We found that our baseline mean fatigue score of 4.2 (SD 2.9) is very similar to a mean score of 4.2 (SD 2.8) seen by Wolfe in 7760 patients with RA. Our data add to the literature by comparing various patient-reported outcome anchors and assessing the MID overall in fatigue and in high and low fatigue. Our MID estimates for improvement ranged from –0.82 to –1.12 for improvement and 1.13 to 1.26 for worsening. Previous studies have shown an inherent uncertainty around the MID estimates and experts have recommended using several anchors. Our primary anchor was the fatigue scale, but we also performed our analyses using overall and pain anchors, as previous studies have shown an association between pain, overall health, and fatigue. Our results were very similar using the 3 anchors.

Wells, et al assessed MID estimates for the fatigue VAS scale in 2 randomized controlled studies comparing abatacept versus placebo in RA and found similar results; their MID estimates for improvement ranged from 6.7 to 17.0 for 2 studies on a 0–100 scale (or 0.67 to 1.70 on a 0–10 scale). Apart from differences in our study design and that of Wells, et al, we used external anchors to determine our MID estimates, whereas Wells, et al used internal anchors (change in HAQ-DI, pain VAS, and patient global VAS) based on previous studies. They followed their estimates with a Delphi exercise and reached a consensus estimate of 10 (on a 0–100 scale).

Previous literature has shown that an effect size of 0.2 to 0.50 corresponds to MID for a patient-reported outcome measure. The ES of the Somewhat better and Somewhat worse groups were small (0.39 and 0.44, respectively), supporting the MID results. In addition, the MID estimates were in the right direction and of larger magnitude than that for the no-change group.

As noted, the MID estimates may depend on the baseline scores. This trend was seen in our analysis (Table 2), where people with higher baseline scores required a larger change in their fatigue VAS for improvement to be considered as minimally improved. Conversely, patients with lower baseline scores required a larger change in their fatigue VAS for worsening. This may be related to floor and ceiling effect (where people near the bottom of the scale are limited by how much they can improve or near the top of the scale are limited by how much they can worsen) or may represent difference in interpretation of the scale along the continuum.

We introduce the concept of statistically significant change in individual patient-reported outcome scores rather than group scores. RCI indicates whether change for an individual patient is beyond the measurement error of the instrument; RCI has the advantage of yielding a direct test of the significance of individual change. As expected, clinically significant change was larger for individual patients (3.47) compared to group change (–0.82 to –1.12 for improvement and 1.13 to 1.26 for worsening) and RCI decreased as the test-retest reliability of the VAS increased.

Our study is not without limitations. First, the global ratings of change ask people to remember how their health was

| Table 1. Minimal important difference (MID) estimates for the fatigue VAS (0–10). |
|---------------------------------|-----------------|-----------------|------------------|------------------|
| Patient-rated Fatigue           | Mean Fatigue    | 95% CI          | Effect Size      | p                |
|                                 | VAS Change      |                 |                  |                  |
| Much better (n = 17)            | –1.44           | –2.78, –0.10    | –0.50            | 0.04             |
| Somewhat better (n = 40)        | –1.12           | –1.86, –0.37    | –0.39            | 0.03             |
| Same (n = 171)                  | –0.23           | –0.57, 0.11     | –0.08            | —                |
| Somewhat worse (n = 65)         | 1.26            | 0.68, 1.86      | 0.44             | <0.001           |
| Much worse (n = 14)             | 3.35            | 1.94, 4.78      | 1.17             | <0.001           |

Negative scores signify improvement and positive scores signify worsening. VAS: visual analog scale.

| Table 2. Mean MID scores for improvement in the fatigue scores stratified by “less severe” fatigue vs “more severe” fatigue at baseline. |
|---------------------------------|-----------------|-------------------|-----------------|
| Patient-rated Fatigue           | n   | Fatigue VAS ≤ 4.5 mean score | 95% CI | n   | Fatigue VAS > 4.5 mean score | 95% CI |
| Much better                     | 11  | –0.41             | 1.16, 0.34      | 6   | –3.33           | –7.05, 0.39 |
| Somewhat better                 | 28  | –0.15             | –0.86, 0.55     | 12  | –3.38           | –4.50, –2.25 |
| Same                            | 102 | 0.61              | 0.24, 0.98      | 69  | –1.47           | –2.01, –0.94 |
| Somewhat worse                  | 24  | 3.08              | 3.08, 0.45      | 41  | 0.21            | –0.34, 0.76  |
| Much worse                      | 5   | 5.70              | 3.66, 7.74      | 9   | 2.06            | 0.69, 3.42   |

Negative scores signify improvement and positive scores signify worsening. MID: minimal important difference; VAS: visual analog scale.
REFERENCES


