Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis

Patricia Minnock¹, John Kirwan² and Barry Bresnihan¹,³,⁴

Objective. Fatigue is an important symptom in patients with RA. Measurement of fatigue in clinical trials and in clinical practice requires scales that are reproducible, sensitive to change and practical. This study examined the reliability and sensitivity to change of fatigue and its relative independence as an outcome measure in RA.

Methods. Successive patients referred to the rheumatology clinic at St Vincent’s University Hospital and Our Lady’s Hospice were evaluated. Clinical assessments were undertaken at baseline and 3 months after commencing TNF-α blockade. Fatigue was measured using an 11-point numeric rating scale (NRS). Sensitivity to change when compared with current core set outcome measures was determined by calculation of the standardized response mean (SRM). Multiple regression analysis was employed to determine the independent variance of fatigue scores relative to the core set.

Results. Forty-nine patients were evaluated. At baseline, mean (s.d.) fatigue scores were 6.7 ± 2.1. At 3 months, fatigue scores had fallen to 4.3 ± 2.6 (P < 0.001). Test–retest intraclass correlation coefficient for the NRS was 0.79 (P < 0.008). Fatigue was ranked third for relative sensitivity to change as shown by SRM: pain, 1.37; tender joint count (TJC), 1.09; fatigue, 0.92; swollen joint count (SJC), 0.86; HAQ, 0.82; CRP, 0.69; and patient global health (GH), 0.25. The relative independent variance in fatigue of 22% was higher than that of the core set: TJC, 20%; pain, 19%; SJC, 16%; GH, 8%; HAQ, 7%; and CRP, 8%.

Conclusions. This study demonstrates that measures of fatigue are reliable and sensitive to change, and should be considered for inclusion as a core outcome measure in RA.

Key words: Rheumatoid arthritis, Pain, Fatigue.

Introduction

RA is a multisystem immune-mediated chronic disease characterized by both articular and constitutional symptoms [1]. Although RA affects men and women of all ages, it has a female preponderance [2]. Pain and fatigue are prominent symptoms of RA, and both are multidimensional in nature [3–11]. Unlike pain, fatigue is not included in the ACR core set of outcome measurements and improvement criteria [12]. It has been argued that fatigue should be included in the core set of outcome variables in RA clinical trials [9, 13–15].

Proposed candidate instruments suitable for outcome measurement in clinical trials are expected to attain recognized measurement standards of truth, discrimination and feasibility [16]. These properties represent the questions that the instruments must satisfy when utilized: truth captures the issue of face, content, construct and criterion validity; it demands that the measure is truthful, it measures what it claims to measure and that the result is unbiased and relevant. Discrimination captures the issues of reliability and sensitivity to change; it asks if the instrument discriminates between situations that are of interest. These can be states at one time such as classification or prognosis, or states at different times such as capturing change. Feasibility refers to the ease of application of the instruments with regard to constraints such as time, money and interpretability. Feasibility may be the element that determines the choice and even the success of an instrument [16, 17].

Through a process that involved patients and clinicians as research partners, it was agreed that evidence to support the case for inclusion of fatigue, a patient-reported outcome, within the core set measures in RA should be generated [9, 13, 14]. Subsequently, on the basis of such evidence, fatigue received the endorsement of the OMERACT network of international researchers in 2006 [15]. It was agreed and recommended that fatigue should be measured in all future studies of RA whenever possible. Moreover, there was a substantial support of further research into measuring and understanding fatigue in RA [15].

The purpose of this study was to contribute to this data-driven, iterative process, building a body of knowledge in relation to fatigue assessment. The study examined two areas in particular: first, the discriminatory nature of fatigue; and secondly, if the measurement of fatigue provided unique information on outcome in RA in addition to that already captured by the core set variables. The study examined the reliability and sensitivity to change of fatigue in RA as well as its relative independence as an outcome measure in RA.

Methods

Clinical assessments

All patients were attending the rheumatology therapy service at St Vincent’s University Hospital, including Our Lady’s Hospice. The diagnosis of RA was confirmed according to established criteria [18]. All patients who were prescribed anti-TNF-α therapy were evaluated prospectively according to a predetermined protocol, which included the ACR core set of outcome measures [19], and quantitative assessment of fatigue. This survey was conducted as part of an internal audit of biological therapy services and processes. Clinical assessments were undertaken at baseline and 3 months after commencing TNF-α blockade and included six of the internationally agreed core outcome measures, namely: pain, swollen joint count (SJC), tender joint count (TJC), patient global health (GH), HAQ and the acute phase marker CRP [19]. Fatigue severity was quantified using a unidimensional 11-point numeric rating scale (NRS) with anchors of 0 (none) and 10 (a great deal). The words ‘none’ and ‘a great deal’ corresponding to the anchors 0 and 10, respectively, were placed under the NRS.
Fatigue severity units included: none, 0; mild, 1–3; moderate, 4–6; severe, 7–10, giving it ordinal properties of measurement [20].

**Statistics**

The computer software statistical package SPSS 12.0 for windows [21] was used to analyse the data. Descriptive statistics employed included summary statistics and frequency distributions. Two separate tests of reliability were conducted. First, the NRS was subjected to test–retest analysis using the intraclass correlation coefficient (ICC) [22]. Secondly, the repeatability or strength of agreement between repeated measurements of the NRS was examined using the statistical method recommended by Bland and Altman [22]. Similarly, the responsiveness or sensitivity of fatigue measurement to change and its comparison with that of the core set outcome measures was demonstrated using two separate tests, a paired samples t-test and the standardized response mean (SRM) [23]. The formula for the paired samples t-test is the change score on measure divided by the S.E. of the difference. The SRM is the change score on measure divided by the s.d. of the change score. This gives a standard unit of measurement of the ‘effect size’, which is the before and after changes in the respective variables [24]. It is suggested that standardized effect sizes of 0.2–0.5 should be regarded as small, 0.5–0.8 as moderate and those >0.8 as large [25].

Multiple regression analysis was used to examine the relationship between fatigue and the core set variables. Backward deletion technique [26] was used to determine the most statistically meaningful variables in the prediction of fatigue, and to calculate the size of the independent contribution of fatigue to the assessment of RA. The change in the values at 3 months of the six core set variables and fatigue was the data utilized. Fatigue as the dependent variable was regressed against the six variables of the core set to calculate the explained variance ($R^2$, the square of the correlation coefficient, known as coefficient of determination). This gave a measure of variation in fatigue that could be explained by the variation in all other outcomes. Next, subtraction of $R^2$ from 1 gave the unexplained variance or independent contribution made by fatigue. This was repeated for each of the variables, by regressing them against all the other outcomes (including fatigue). The unexplained variance multiplied by 100 and divided by the sum of all the unexplained variances gave the relative unexplained variance of each of the seven measures of outcome.

**Results**

**Clinical and demographic details**

Forty-nine successive patients were evaluated. Twenty-nine (59%) patients were females (mean age (s.d.) (range) was 53 ± 11.17 (22–73) years, mean disease duration was 11 ± 7.2 (0–31) years] 41 (84%) were RF-positive. Mean baseline clinical measures included: pain visual analogue scale (VAS), 5.79 mm, TJC, 11.71; SJC, 13.46; GH, 5.84; HAQ, 1.37; CRP, 34 (0–4) mg/l; and fatigue 6.65.

**Reliability**

The ICC gives a measure of consistency or agreement of values within cases. In a test–retest analysis of the fatigue NRS on a random sample of 12 patients, the ICC was 0.79 ($P < 0.008$). The Bland and Altman plot demonstrated small differences on repeated measurement and no bias in the distribution [mean (s.d.) (range) 5.01 ± 2.87 (0–8)] (Fig. 1).

**Sensitivity to change**

At baseline, mean (s.d.) fatigue score was 6.7 ± 2.1. At 3 months, fatigue score had fallen to 4.3 ± 2.6 ($P < 0.001$) (Fig. 2). Sensitivity to change of fatigue was further examined comparing it with the sensitivity to change of the core set. Fatigue was ranked third for relative sensitivity to change as shown by SRM: pain, 1.37; TJC, 1.09; fatigue, 0.92; SJC, 0.86; HAQ, 0.82; CRP, 0.69; and GH, 0.25 (Fig. 3).
Independent contribution of fatigue to the assessment of RA

Multiple regression analysis was used [1] to determine the explained variation in fatigue, and [2] to compare the independent contributions of fatigue and the six core set measures to the assessment of RA as described in ‘Methods’ section. The final backward stepwise model obtained in this way included GH and TJC as shown in Table 1. The independent variables GH and TJC explained 34% of the variance in fatigue, making a significant unique contribution to fatigue prediction.

Table 2 shows the relative independent contributions of the core set variables and fatigue to the assessment of RA. Fatigue has the highest value (22%), similar to that of tender TJC (20%) and pain (19%), whereas HAQ has the lowest (7%).

Discussion

This study provides evidence that fatigue is a valid outcome measure in RA. The reliability and sensitivity to change of fatigue in patients with RA commencing TNF-α blockade was demonstrated. It was also demonstrated that measurement of fatigue provides additional information on outcome. The amount of information provided by fatigue was proportionate to that of the core set variables, particularly TJC, pain and SJC. Moreover, measuring fatigue provided additional information that is essential to the understanding of disease outcome from the patients’ perspective.

This interventional study also provides, for the first time, scientific evidence on the reliability of fatigue measurement in clinical practice using a single-item scale. A prior detailed systematic review of scales used to measure fatigue found no data to support the reliability of an NRS in measuring fatigue in RA [27]. For the purpose of this work, reliability was estimated in terms of the test–retest properties of the NRS, and consistency between two measurements was demonstrated [28]. However, because the test–retest was conducted on only 12 cases, the authors recommend repeat evaluation of the test–retest reliability using a larger sample size. Prior reviews of the performance of existing fatigue measurement scales in RA have largely been confined to observational studies [17]. One report compared the discriminatory properties in relation to the performance of various short and long questionnaires for fatigue assessment. It was concluded that the single-item VAS was suitable for use in clinical practice [17]. The VAS and the NRS are both single-item, linear scales with similar measurement properties [20]. They quantify the amount of fatigue experienced over a defined period. This study now provides evidence of the consistency of a single item scale in quantifying the experience of fatigue in RA.

Sensitivity to change is considered as the criterion that determines a good measurement scale [17, 29]. It is defined as the capacity of instruments to measure change statistically. Responsiveness addresses the detection of clinically relevant change [30]. This study calculated two separate change coefficients to determine the sensitivity of the fatigue NRS to therapeutic intervention. First, the responsiveness of the fatigue NRS was demonstrated. A statistically significant difference between fatigue levels at baseline and 3 months after starting TNF-α blockade was shown. Secondly, the magnitude of this change (effect size) when compared with the effect size on the core set measures was demonstrated. Previous studies have examined the sensitivity to change of fatigue in RA measured through single-item scales [17, 27]. Similar to the property of reliability, no data were found to support the sensitivity of NRS in measuring fatigue in a detailed systematic review [27]. Where sensitivity to change of single-item scales was previously reported (the VAS), it included the caveat that the data were from observation studies [17]. Nevertheless, fatigue was measured through a single-item VAS performed as well as more detailed scales with respect to its association with other clinical variables [17]. Through the calculation of the effect size, this study provides further evidence of how fatigue measurement compares with traditional clinical measures, particularly SJC and pain, in standard clinical practice [24]. The strength of the sensitivity data reported here is the one that is derived from an interventional study.

The elucidation of the unique contribution made by the variable fatigue to the assessment of disease outcome in RA added an interesting dimension to the debate on fatigue as an appropriate outcome measure. The rationale behind this approach was to determine whether fatigue is a unique outcome measure in RA; a variable worthy of independent status that contributes unique information to the assessment of the disease. If fatigue is an alternate measure for any of the core set, TJC or pain or both for example, then its measurement would not provide any additional information about outcome in RA. If, however, fatigue not only provides some information that overlaps with other established measures of RA, but also provides additional information that does not overlap, then it is worthy of measurement [15]. This study provides evidence to support the latter viewpoint in several ways. First, taking fatigue as the dependent variable, it was seen that the performance of fatigue was similar to the core set variables. The proportion of both the explained and unexplained variance in the respective variables was similar in size. The proportion of fatigue explained (11%) by the overlap with the core set variables supports the argument that, as a component of RA, fatigue behaves in a similar manner to the core variables. Calculation of the proportion of fatigue unexplained (22%) by the core set, which collectively is representative of the totality of RA, supports the argument that a substantial component of RA remains unexplained. Therefore it can be contended that management strategies for this salient symptom of RA are under investigation in randomized trials, and intervention options are limited in clinical practice.

The feasibility of a measurement instrument is frequently the element that determines initial choice and long-term success of use of an instrument [16]. The NRS is easily understood, quick to complete, simple to score and its usage carried only minimal associated expense as it is easy to replicate. This makes it an attractive instrument of choice for use in routine clinical and research practice. This is an important consideration in relation to instrument choice when encouraging clinicians to measure an important symptom such as fatigue [16, 17]. However, it must be
acknowledged that the one-dimensional measurement property of the NRS is regarded as a weakness in the measurement of a multi-dimensional concept [5, 31]. Whereas single-item scales contribute no qualitative information in relation to the experience and meaning of fatigue, nonetheless, the associated ease of use in practice is a counter strength [17]. This is an important consideration when recommending that the fatigue should be measured whenever possible [15].

This study demonstrates that this simple, single-item measures of fatigue are reliable and sensitive to change in RA, following TNF-α blockade, and feasible for use in clinical practice. The fact that this study is an interventional study contributes to the robustness of these findings. The observations support the suggestion that fatigue is an independent measure of outcome and should be considered for inclusion as a core outcome measure in RA. Moreover, this study provides evidence that fatigue measurement captures information that contributes to disease management in RA.

Rheumatology key messages

- Measures of fatigue in RA are reliable and sensitive to change.
- Fatigue should be included as an outcome measure in studies of RA, both in clinical practice and clinical trials.

Acknowledgements

We wish to thank Prof. Oliver FitzGerald and Prof. Douglas Veale for their valuable contribution to this work.

Disclosure statement: The authors have declared no conflicts of interest.

References

23 Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.