Interpreting measurements of physical function in clinical trials

Michael M Ward

Ann Rheum Dis 2007;**66**(Suppl III):iii32–iii34. doi: 10.1136/ard.2007.079806

Correspondence to:
Dr Michael Ward, NIAMS/
NIH, Building 10 CRC, Room
4-1339, 10 Center Drive,
MSC 1468, Bethesda, MD
20892, USA;
wardm1@mail.nih.gov

Accepted 25 July 2007

Improving physical functioning is one of the major goals of anti-rheumatic treatment. However, functional limitations can have several different causes, which may differ in their capacity to respond to a given treatment. Functional limitations due to pain or other acute symptoms or signs may be readily reversible with efficacious treatment, while those due to chronic structural changes may be relatively irreversible in the short term. Because measures of physical function characterise the degree of limitation without regard to cause, patients with the same apparent degree of functional limitation may differ greatly in their ability to demonstrate response to treatment. Structural damage accumulates over the course of disease, so measures of functional limitations tend to be less responsive among patients with more longstanding disease. This decreased responsiveness leads to a decreased ability to discriminate between treatments in patients with more longstanding arthritis. In addition, the criteria for minimal clinically important improvement may be underestimated when patients with irreversible functional limitations are included as test subjects, because judgments of improvement may be associated with smaller measured changes in physical functioning. The interpretation of measurements of physical function in clinical trials should consider the composition of the study sample, with attention to the stage of disease and the heterogeneity in disease duration or structural damage among subjects.

Physical function is a major component of health status and health-related quality of life, and is affected in virtually all musculoskeletal conditions. Because of its central role, physical function is included in the core set of measures to be assessed in clinical trials for many rheumatic diseases.^{1–7} A number of well-designed and tested measures of physical function, most based on patient report of the degree of difficulty encountered in attempting everyday tasks, have been included as endpoints in clinical trials and have provided valuable information on the effects of treatment. However, the nature of physical functioning makes the interpretation of the changes in physical function measures more complicated than that of other trial endpoints, and more intricate than it may seem at first glance.

In patients with musculoskeletal diseases, limitations in physical functioning can have several different causes.^{8–11} Limitations in some patients may be primarily due to symptoms of pain, stiffness, fatigue or to acute joint swelling. In the absence of these symptoms or signs, the patient would have no functional limitations. Conversely, limitations in other patients may be primarily due to joint deformity, weakness, deconditioning, reflecting cumulative musculoskeletal damage, or to comorbidities. In these patients, functional limitations would persist in the absence of symptoms, and would not be expected to improve in the short term with anti-rheumatic treatment. In other patients, functional limitations due to acute symptoms may be superimposed on limitations from more chronic causes.

Measures of physical function denote the degree of functional limitation (or the state of the patient), regardless of cause. The same score in different patients may represent different proportions of limitations due to acute, reversible causes and limitations due to chronic, less readily reversible causes. It is possible to demonstrate the composition of physical function measures by reversible and irreversible limitations.¹² In trials of anti-rheumatic treatments directed at improving symptoms or reducing inflammation, the greater the contribution of irreversible limitations to the patient's score, the less likely the patient's functional measure will be able to change and demonstrate improvement. In this situation, the irreversible

limitations provide a new floor for the functional measure. Similarly, the greater the number of patients in a trial who have irreversible functional limitations, the less responsive the functional measure will be. Conversely, functional measures will have more capability of registering improvement with treatment when all (or most) of the patient's functional limitations are due to acute symptoms, and when most of the patients in the trial have functional limitations solely due to reversible causes.

RESPONSIVENESS TO TREATMENT

To examine the association between the presence of irreversible functional limitations and the responsiveness of measures of physical function, we used rheumatoid arthritis (RA) as a model and the Health Assessment Questionnaire Disability Index (HAQ)¹³ as the measure of physical function. Because structural damage accumulates in patients with RA, the duration of RA can be used as a measure of the likelihood of irreversible functional impairments. In an analysis of individual patient data from recent RA clinical trials, all of which enrolled patients with active RA, we selected a subgroup of patients who entered clinical remission during the trial.¹² These patients all had a large improvement in RA activity during the trial. We assessed corresponding improvements in HAQ as a function of the duration of RA. The HAQ was less responsive among patients with more longstanding RA than among patients with early RA, as would be predicted if irreversible functional limitations comprised a larger proportion of their total functional impairment (table 1). The HAQ in remission was also much higher among patients with more longstanding RA, demonstrating the higher floor of the measure among these patients. Results were similar when analyses were repeated using radiographic damage scores as the measure of irreversible functional impairment.

These findings are supported by data from an ongoing observational study of treatment responses in 156 patients with RA (table 2). Patients with early RA and late RA had similar

Abbreviations: DAS28, Disease Activity Score-28; HAQ, Health Assessment Questionnaire Disability Index; RA, rheumatoid arthritis

Table 1 Change in the Health Assessment Questionnaire Disability Index (HAQ) among patients in rheumatoid arthritis clinical trials who entered remission during the trial, by duration of rheumatoid arthritis

	Duration of rheumatoid arthritis (years)			
	<2	2–4.9	5–9.9	≥10
Mean baseline HAQ	1.31	1.22	1.27	1.28
Mean HAQ in remission	0.36	0.44	0.51	0.73

HAQ scores can range from 0 to 3.0, with higher scores indicating more functional limitations.

Disease Activity Score-28 (DAS28) values at entry to the study, and similar improvements in RA activity with treatment. However, the HAQ was less responsive among patients with late RA than those with early RA. The effect size, a measure of responsiveness that represents the change in HAQ divided by the SD of the HAQ at study entry, was 0.50 among patients with less than 10 years of RA, 0.28 in patients with 10–19.9 years of RA, and 0.18 in patients with 20 or more years of RA. These results demonstrate that the responsiveness of the HAQ varies with the duration of RA.

This observation appears to be generalisable, as it can be demonstrated using different analytical approaches. For example, in a pooled analysis of published RA clinical trials, the responsiveness of the HAQ was lower in trials that enrolled patients with higher mean durations of RA.¹⁴

DISCRIMINATION BETWEEN TREATMENTS

One potential consequence of decreased responsiveness is a decrease in the ability of a measure to demonstrate differences between treatments. By definition, responsive measures are those that register large changes with effective treatment, and can discriminate easily between effective and ineffective treatments (including placebo). Less responsive measures demonstrate less change with effective treatment. Therefore, even with effective treatments, the changes of less responsive measures may overlap, or be indistinguishable from, the changes seen with ineffective treatment or with placebo.

Because functional measures are less responsive in late RA than in early RA, one would predict that the ability to discriminate between treatments would be more difficult in patients with late RA. To test this hypothesis, we performed a pooled analysis of RA clinical trial results in which we compared improvements in HAQ scores between conventional disease-modifying medications, biological medications and

placebo.¹⁵ For 37 trials with 87 active treatment arms, we computed effect sizes for the HAQ, and modelled the association between the effect size and the mean duration of RA among trial participants for each of the three classes of treatments. Biological treatments had large effect sizes in early RA, but the effect size progressively decreased among trials of patients with RA of longer duration, so that the effect of biological treatments on the HAQ could not be distinguished statistically from placebo in trials of patients with an average RA duration of 12 years or longer. Similarly, the discrimination between biological medications and conventional disease-modifying medications was appreciable in early RA but diminished as the duration of RA increased. The decreased ability to differentiate between classes of medications in improvement in physical functioning resulted directly from the decreased responsiveness of the HAQ in later RA, which in turn was due to the irreversible limitations that contributed to physical functioning in patients with late RA.

CRITERIA FOR MINIMAL CLINICALLY IMPORTANT IMPROVEMENT

Differences in the responsiveness of measures of physical functioning due to the presence or absence of irreversible limitations can also complicate the establishment of a criterion for clinically important improvement. The minimal clinically important improvement represents the smallest amount of change in a measure that is considered clinically meaningful.

In one approach to establish these criteria, patients with some level of symptom severity or impairment are examined before and after receiving treatment. After treatment they are asked to judge whether their symptom or impairment improved, and to assess the magnitude or value of any improvement. These subjective judgments are then related to the measured changes in symptoms or impairments, and collated across patients to derive the criterion.

Because a measured change is the basis for determining the criterion for important improvement, sensitivity to change is a prerequisite.¹⁶ If a measure was poorly sensitive to change (for example, had only a few possible response categories, such as “good”, “fair” and “poor”), some patients may experience quite a large degree of improvement but not change in the health status measure. In this situation, their subjective impression of an important improvement would be misattributed to a small (or no) change in the health status measure. While measures are tested for responsiveness before attempting to establish criteria for clinically important improvement, responsiveness is most often viewed as a property of the measure itself, rather than as a property that can be influenced by the nature of the subjects in whom it is tested. As demonstrated above, responsiveness of physical function measures can vary due to the presence or absence of irreversible functional limitations.

Lower responsiveness of physical function measures in more longstanding RA would be predicted to result in an underestimation of the minimal clinically important difference, compared to patients with early RA. Because the HAQ is more responsive in patients with earlier RA, relating judgments of improvement to measured changes in the HAQ would not be interfered with by irreversible functional limitations to the same degree as in later RA. To test this hypothesis, we compared estimates of clinically important improvement in HAQ scores between patients with earlier and later RA in the observational study described above. In each group, we computed receiver operating characteristic curves that related different degrees of measured changes to patient judgments of improvement¹⁷ (fig 1). As hypothesised, the measured changes in HAQ judged as “important” by patients were systematically

Table 2 Change in response to treatment in the Disease Activity Score 28 (DAS28) and Health Assessment Questionnaire Disability Index (HAQ) in patients with active rheumatoid arthritis

	Duration of rheumatoid arthritis (years)		
	<10	10–19.9	≥20
Mean baseline DAS28	6.1	6.0	6.1
Mean change in DAS28	–1.2	–0.8	–1.0
Mean baseline HAQ	1.31	1.48	1.40
Mean change in HAQ	–0.36	–0.22	–0.13

Responses were measured after either 1 month (for patients treated with corticosteroids) or 4 months (all others). The DAS28 can range from 0 to 10, with scores above 5.1 considered to represent highly active rheumatoid arthritis. HAQ scores can range from 0 to 3.0, with higher scores indicating more functional limitations.

lower among patients with more than 15 years of RA, compared to those with RA for 15 years or less. For example, a decrease in HAQ of 0.25 had a sensitivity of approximately 0.70 and a specificity of approximately 0.60 for being considered an important improvement by patients with 15 years or less of RA. However, among patients with more longstanding RA, a decrease in HAQ of only 0.125 had a similar sensitivity and specificity for being considered an important change. The accuracies of the receiver operating characteristic curves were similar in the two groups. This difference in criteria for clinically important change in the two groups likely relates to the decreased responsiveness of the HAQ in patients with more longstanding RA. This difference highlights the importance of considering the nature of test subjects when attempting to determine criteria for clinically important improvement for measures of physical function.

CONCLUSIONS

Physical function end points in clinical trials should be interpreted with consideration of the type of patients studied. Studies of homogenous groups of patients without irreversible functional limitations would provide the best opportunity to detect treatment effects. To the extent that patients with irreversible functional limitations are included, the decreased responsiveness of physical function measures will make it more difficult to demonstrate efficacy of a treatment, to discriminate between treatments and to establish accurate criteria for clinically important improvement. Differences between trials in the types of patients studied will also confound comparisons of the effects of different treatments on physical function. The dual nature of physical function measures also means that claims for improvement in functional impairment in short-term trials likely primarily reflect reversible functional impairments due to symptoms, without necessarily indicating structural improvement or the delay of structural damage.

Figure 1 Receiver operating characteristic curves associating measured changes in the Health Assessment Questionnaire Disability Index (HAQ) with patient judgments of whether they had important improvement in physical functioning over the same time. Analyses were performed separately for patients with 15 years of rheumatoid arthritis or less (dashed line), and for patients with more than 15 years of rheumatoid arthritis (solid line). Measured changes in HAQ scores at selected sensitivities and specificities are noted in each group; negative values indicate improvement in HAQ scores.

Although the examples used here involve the HAQ in patients with RA, similar issues apply to other measures of physical function and conditions other than inflammatory arthritis. The central issue relates to the nature of functioning itself. Measures of function, whether of physical function, renal function, pulmonary function or cognitive function, measure the state of an individual without regard to the cause of any dysfunction. Some of the functioning limitations will be reversible, while others will not, and the relative contribution of reversible and irreversible causes will vary among patients and over time. Recognising the different consequences of reversible and irreversible impairments is important for the proper interpretation of any measure of function.

ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health.

Competing interests: None declared.

REFERENCES

- 1 Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;**36**:729–40.
- 2 Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, *et al*. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol Suppl* 1994;**41**:86–9.
- 3 van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan MA. Which domains should be included in a core set for endpoints in ankylosing spondylitis? Introduction to the ankylosing spondylitis module of OMERACT IV. *J Rheumatol* 1999;**26**:945–7.
- 4 Miller FW, Rider LG, Chung YL, Cooper R, Danko K, Farewell V, *et al*. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. *Rheumatology* 2001;**40**:1262–73.
- 5 Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B, Feldman BM, *et al*. Preliminary core set of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. *Rheumatology* 2003;**42**:1452–9.
- 6 Bellamy N, Kirwan J, Altman R, Boers M, Brandt K, Brooks P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;**24**:799–802.
- 7 Stucki G, Cieza A, Geyh S, Battistella L, Lloyd J, Symmons D, *et al*. ICF core sets for rheumatoid arthritis. *J Rehabil Med* 2004;**44**(Suppl):87–93.
- 8 Scott DL, Pugner K, Kaarela K, Doyle DV, Woolf A, Holmes J, *et al*. The links between joint damage and disability in rheumatoid arthritis. *Rheumatology* 2000;**39**:122–32.
- 9 van Leeuwen MA, van der Heijde DM, van Rijnwijk MH, Houtman PM, van Riel PL, van de Putte LB, *et al*. Interrelationship of outcome measures and process variables in early rheumatoid arthritis: a comparison of radiologic damage, physical disability, joint counts, and acute phase reactants. *J Rheumatol* 1994;**21**:425–9.
- 10 Welsing PM, van Gestel AM, Swikels HL, Kiemeny LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis Rheum* 2001;**44**:2009–17.
- 11 Escalante A, del Rincón I. The disablement process in rheumatoid arthritis. *Arthritis Rheum* 2002;**47**:333–42.
- 12 Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis. Identifying reversible and irreversible components. *Arthritis Rheum* 2006;**54**:2784–92.
- 13 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.
- 14 Aletaha D, Ward MM. Duration of rheumatoid arthritis influences the degree of functional improvement in clinical trials. *Ann Rheum Dis* 2006;**65**:227–33.
- 15 Aletaha D, Strand V, Smolen J, Ward MM. Treatment-related improvement in physical function varies with the duration of rheumatoid arthritis. A pooled analysis of clinical trial results. *Ann Rheum Dis*. Published Online First: 20 July 2007. doi: 10.1136/ard.2007.071415.
- 16 Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006;**4**:70.
- 17 Ward MM, Marx AS, Barry NN. Identification of clinically important changes in health status using receiver operating characteristic curves. *J Clin Epidemiol* 2000;**53**:279–84.