Optimisation of rheumatic disease assessments in clinical trials, clinical care, and long-term databases

R.B.M. Landewé[1], D. van der Heijde[2]

[1]*Department of Clinical Rheumatology and Immunology, Academic Medical Center, Amsterdam, & Atrium Medical Center Heerlen;*
[2]*Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands.*

Robert B.M. Landewé, MD, PhD
Désirée van der Heijde, MD, PhD

Please address correspondence to:
Robert B.M. Landewé, MD, PhD
Academic Medical Center/
University of Amsterdam,
Department of Clinical Immunology & Rheumatology,
Meibergdreef 9,
P.O. Box 22660,
1100 DD Amsterdam, the Netherlands.
E-mail: landewe@rlandewe.nl

Received on September 14, 2014; accepted in revised form on September 16, 2014.

Clin Exp Rheumatol 2014; 32 (Suppl. 85): S2-S6.

Key words: clinical trials, disease assessment, databases, optimisation

Competing interests: none declared.

ABSTRACT

The assessment of disease in rheumatological diseases is rather complicated, because it may involves different contexts (clinical practice, clinical trials, observational studies, registries, etc) as well as different domains (disease activity, physical function, radiographic damage, quality of life, etc). Furthermore, available tools can be comprehensive but also rather condense, may be patient-oriented or rather physician-oriented, and so on.

In this article all these levels that may matter in case of a choice of disease assessment tool are discussed, ariiving at a conclusion that choosing the appropriate tool for the assessment of disease is not 'cookbook medicine'.

Introduction

Progress in rheumatology has been phenomenal during the last 3 decades. When asked for an appropriate explanation many will respond that the advent of biological disease-modifying anti-rheumatic treatments (bDMARDs) has fostered this.

It should not be forgotten, though, that such developments have only been possible since the clinical scientific community in rheumatology has provided the appropriate measurement tools that allowed credibly demonstrating that new treatments are superior to placebo, and to standard-of-care treatment. The best example is rheumatoid arthritis (RA) where the availability of a core outcome set (1) and the American College of Rheumatology (ACR) response criteria (2) has been particularly instrumental in the registration of the new and highly efficacious biological treatments. Comparable examples are present in the field of psoriatic arthritis (PsA)(3) and axial spondyloarthritis (axSpA)/ankylosing spondylitis (AS)(4,5).

Several questions are relevant for the discussion of which measures should be used to *assess the disease* in the chronic inflammatory diseases. These questions will be dealt with point-by-point.

Furthermore, *assessing the disease* is quite a vague and generic formulation that incorporates many domains of interest: Disease activity is an important domain in chronic inflammatory diseases, and many instruments have been developed to measure disease activity. But damage is another important domain, as is physical function and quality of life. These domains, and their relative importance, will be discussed separately in the text below.

Is the assessment of disease context-specific?

This question refers to the different contexts in which disease is actually assessed. We as rheumatologists are used to hear about new treatments when randomised controlled trials (RCT) investigating the efficacy and short-term safety of new treatments for regulatory purposes are presented at medical conferences or in medical literature. Somewhat further along the line, when new treatments have been approved, but experience needs to be built up, observational studies of all types (*e.g.* registries) are often initiated that may assess multiple facets of the disease under study (since a good gold standard assessment does not exist). And then, there is an increasing need for rheumatologists to document how they (and their patients) perform in daily clinical practice, and how they, as care-providers, can improve that performance (*quality of care*). For that purpose, they must record what they do in clinical practice. That means they must assess patients regularly and document the results of these assessments and thus of their interventions. Alternatively, they ask nurses or dedicated assessors to collect the measures for them. Of interest, RA is among the diseases with most contribution of information obtained from patient history and physical examination to clinical decision making (6).

In the absence of a true *gold standard* outcome assessment, it is arguable, if not very unlikely, that the same disease assessments can be applied to all these different contexts that all have very different goals: The objective of an RCT is to investigate (drug) efficacy; the objective of observational research more often is to investigate aspects of benefit-risk, in other words effectiveness; the main objective of disease assessment in all-day clinical practice is to provide *benchmark data* to be used for improvement of daily care; or to provide group-data as a reflection of practice performance. While some of the disease assessments can be used appropriately in all contexts, most of them have characteristics that make them more or less suitable for a specific context. Therefore, in the following paragraphs they will be discussed in these contexts.

Should measures of change or measures of status be used?

For an appropriate answer to this question again the context matters. In a clinical trial, outcome measures should have sensitivity-to-change (since it only makes sense to apply measures that will rapidly change after the start of effective treatments) and discriminatory ability (to find out if the drug to be tested is better than placebo or a comparator drug). This context is relatively easy: Core outcome sets and response measures are available for all inflammatory rheumatic diseases, and many successful trials testify of their usefulness. Response measures have a statistical advantage over state measures, since they efface the unwarranted variability at baseline that usually exists in groups of patients. Well known examples of successful response measures are the ACR response measures for RA (2) and the ASAS-response criteria for AS (6a).

Many have propagated the use of the core-outcome measures in a broader context, eg in observational studies or even in clinical practice. Unfortunately, response measures fall short here, since they typically are relative measures: they express improvement relative to baseline values. In a short-term RCT this will suffice, since all patients will have high disease activity at baseline (they have been selected for that), there is always a comparator group, and follow up is only short. But if one plans to monitor and analyse the course of a patient over many years, a high disease activity baseline value does not make a lot of sense anymore: It overestimates efficacy of a particular treatment because disease activity at a certain point in time is always expressed relative to baseline, when disease activity was high anyway (7). Here we need *disease activity state-measures*: A state measure assesses if a patient 'does well'. It is a reflection of actual disease activity at a certain point in time. For RA we have the disease activity score (DAS) (8) and derived measures (DAS28 (9), SDAI (10) and CDAI (11)) and also the patient-reported RAPID3 (12). For axial SpA/AS we have the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (13) and the Ankylosing Spondylitis Disease Activity Score (ASDAS) (14), and for PsA a few state measures, such as the PASDAS, have been proposed (15). These arguments of response *versus* state measures pertain to all other assessments of disease; in the context of RCTs change measures (either a change score in a continuous measures such as DAS or a response index) usually have a statistical advantage. In the observational study context, or in daily clinical practice, where the physician is interested in 'how a patient is actually doing', state measures are preferred.

Should patient-reported or physician-assessed measures be used to assess disease*?*

This is an interesting discussion, which is not always easy. For some measures of disease the answer to this question is obvious. Radiographic damage is an important outcome in RA, PsA and AS, and entirely observer-dependent. Typically, radiographic progression is assessed by trained experts on x-rays using validated scoring methods. Modified Sharp score is used for RA (16) and for PsA (17); modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) is the method of choice to assess radiographic progression in axial-SpA/AS (18). It should be clear that these measures require specific training, are time-consuming and are particularly designed for application in clinical trials and observational studies. For clinical practice simplified scoring methods have been proposed but not (yet) widely implemented (*e.g.* Simple erosion and narrowing score (SENS) (19)). Since patients obviously cannot influence the results of these assessments, they are considered more *objective* (albeit measurement variation by observers cannot be ignored).

The discussion becomes less trivial if one looks at disease activity, quality of life (QoL) and physical function. Generic instruments have been developed to measure QoL, and include measures such as Short-Form 36 and Euroqol. These are measures frequently used in cost-effectiveness analysis that find their application in RCTs and observational research. For clinical practice they do not serve a meaningful purpose. In axial SpA/AS a brand-new disease-specific health-related quality of life index directly derived from the WHO classification of functioning and health has been proposed recently: The ASAS-Health-Index (20) which has been reviewed in this supplement. In fact this fully patient-reported 17-item index has excellent metric properties, and can be applied in trials, observational studies and clinical care, but experience is currently lacking. It replaces the Ankylosing Spondylitis QoL (ASQoL), which is a disease specific measure that cannot freely be used (copyrights).

Physical function assessment usually is disease-specific and usually entirely patient-reported. Performance-based measures, in which the patient is asked to perform a task with given difficulty and is observed, are intuitively attractive and methodologically reliable, but likely too time consuming (21). Since there may be a lot of discrepancy between patient-reported and actually measured physical function, these measures could be interesting for application in clinical trials and observational research. They may be too arduous for daily clinical practice

The most widely applied functional assessment tool is the Health Assessment Questionnaire (HAQ) (22) (and its mod-

ifications (such as the M-HAQ)(23) and the methodologically improved MD-HAQ (23a)). The HAQ (or modifications thereof) are widely used in clinical trials and observational studies with RA patients and PsA patients, in order to get an impression of physical functioning. It works well at the group level of clinical trials and some observational studies. According to principles of measurement theory, the HAQ was considered less suitable to follow up physical function in individual patients in clinical practice, and an improved 10-item version (HAQ-II) was constructed based on an item-bank and Rasch analysis (24). Nevertheless MDHAQ has broadly entered clinical practice as the most attractive measure for functioning in RA and PsA. This illustrates that *clinical practice* is not necessarily following best theoretical considerations but may have its own dynamics. Here the Patient Reported Outcomes Measurement Information System (PROMIS), part of the US National Institute of Health (NIH) Roadmap initiative, should be mentioned as an upcoming initiative with great impact on clinical measurement. PROMIS is essentially an item bank containing optimal items to be used for the development of patient-reported functional (and other) indices (25).

In axial-SpA/AS the Bath Ankylosing Spondylitis Functional Index (BASFI) (26) has been most widely implemented, but copyright-claims issued for the *Bath measures* for research may increasingly jeopardise its feasibility. Currently an appropriate user-free physical function instrument is lacking.

An appropriate assessment of the use of patient-reported *versus* physician-based instruments is most difficult -but also most important- for the domain of disease activity.

As said above, clinical trials with short duration tend to exploit change-based disease activity measures with mixed composition: The ACR-(2) and EULAR-(27) response criteria, indices for measuring efficacy in clinical trials, are mixtures of patient reported measures (*e.g.* VAS pain; tender joints, VAS patient global) and physician-assessed measures (*e.g.* swollen joint count, erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)). The advantage of such measures is that they incorporate the patient's and the physician's perspective (which may markedly differ). The disadvantage is that you may find several types of response: One dominated by changes in patient reported outcome (more *subjective*); one dominated by changes in physician assessments (more *objective*). Likely a compromise best reflects the truth: Ideally, changes in PRO's go along to some extent with changes in swollen joints and acute phase reactants, at least in the same direction. It is therefore that combined indices of patient-reported and physician-assessed measures are most often recommended. The combined indices have so-called predictive validity: RA-disease activity measured by DAS is associated with radiographic progression over time (28); and AS-disease activity as measured by ASDAS is associated with syndemophyte formation over time (29).

A word of caution regarding the fully or partly patient-reported indices such as RAPID3 and DAS for RA, and BASDAI and ASDAS for AS/SpA is appropriate here. These indices are propagated because of their patient-centeredness and feasibility. Authors claim that they correlate well enough with *objective measures* (*e.g.* acute phase reactants), so that they may also be used in daily clinical care. The problem here is that such fully or partly patient reported indices may show a reasonably good correlation at the group level, but may show an important discrepancy between 'subjective' and 'objective' signs of the disease in the individual patient. This does not invalidate such measures as monitoring instruments over time in order to trace a change in the clinical situation in a patient, and similar considerations pertain to any index that includes a patient measure. The treating physician must be aware of this possible issue with regard to clinical decision making. A good example is the patient with high pain scores due to 'central sensitisation' (also known as chronic pain syndrome) that is treated with a TNFi biological. Acute phase reactants and swollen joint count may dramatically improve, which improves the prognosis of that patient, but that improvement is not reflected by improvements in pain-related patient-reported measures.While treatment effect in such a patient could be judged as satisfactorily when one looks at individual measures and long-term outcome, the patient may not agree with the physician in this regard and consider him not improved.

Should measures be more comprehensive or rather condensed?

This is a question that frequently comes up when discussing which disease activity measures should be applied. Also here, the context is extremely relevant. Assuming that one has chosen to use an index in stead of a separate measure as key, and that one has decided to opt for a 'mixed' rather than for a fully-PRO measure, especially in the context of RA, there is still a lot to choose. Most of the commonly used status measures are DAS-based. The original DAS included a comprehensive 44-joint swollen joint count, an ESR (not a CRP), as well as a Ritchie articular index with 4 categories per joint. While this DAS version was the only version with a data-driven design and has been best validated against all kinds of external standards (including radiographic progression), it was broadly considered too arduous to apply in all-day clinical practice. The 28-joint version of the DAS (DAS28), but also the SDAI and CDAI, were all modifications of the original DAS with the sole purpose to increase feasibility. While we appreciate the importance of the feasibility-argument when it comes to implementation of a measure in all-day practice, it cannot be ignored that these modifications are methodologically weaker than the original DAS (7, 30, 31). Not only did they lose discriminatory capacity, but they have also lost content validity ('truth'): The condensed 28-joint versions do not include all clinically relevant joints anymore. Whether or not *extent* is a theme to discuss for clinical application is questionable: *Measuring something inappropriately* is likely preferable to *measuring nothing at all*. The choice for a comprehensive *versus* a condensed measure is an individual choice based on a balance between feasibility and best quality. As

long as *assessing 28 joints* is not considered the same as *clinically investigating only 28 joints*, the problem is an academic one.

Another relevant point of concern in this discussion could be that the DAS has been developed in the pre-biologic era in which disease activity in RA patients was on average higher than nowadays. One could argue if the DAS or its modifications (but also ACR response indices, etc.) perform as well as they did before now average disease activity in RA patients is markedly lower, patients start effective treatment far earlier, and respond on average better to treatment than before.

Similar arguments pertain to the situation in PsA, where a scala of measures have been proposed. One of them – PASDAS (15) – is an example of a measure that is fully data-driven, and likely best performing in the context of clinical trials, but far too laborious to apply in daily clinical practice. Besides, multi-domain measures like PASDAS may cover too many distinct topics of the disease inappropriately into one disease activity index (*e.g.* enthesitis, dactylitis and joint inflammation).

Conclusion

It is not easy to recommend a set of disease assessments for chronic inflammatory diseases that may serve as a 'cookbook' for optimal patient care in different settings. Over the years many instruments have been developed and validated for different goals. Some are better, some are worse, but we have tried to make clear here that the optimal instrument to assess disease either does not exist, or is a matter of choice for the physician that wants to measure patients in a particular context. Needless to say that individual considerations, *e.g.* about feasibility *versus* methodological quality, are decisive in this choice process.

References

1. FELSON DT, ANDERSON JJ, BOERS M *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. FELSON DT, ANDERSON JJ, BOERS M *et al.*: American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995; 38: 727-35.
3. GLADMAN DD, MEASE PJ, STRAND V *et al.*: Consensus on a core set of domains for psoriatic arthritis. *J Rheumatol* 2007; 34: 1167-70.
4. VAN DER HEIJDE D, BELLAMY N, CALIN A, DOUGADOS M, KAHN MA, VAN DER LINDEN S: Preliminary core sets for endpoints in ankylosing spondylitis. *J Rheumatol* 1997; 24: 225-9.
5. VAN DER HEIJDE D, CALIN A, DOUGADOS M *et al.*: Selection of instruments in the core set for DC-ART, SMARD, physical therapy and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS working group. *J Rheumatol* 1999; 26: 951-4.
6. CASTREJON I, MCCOLLUM, L, TANRIOVER MD, PINCUS T: Importance of patient history and physical examination in rheumatoid arthritis compared to other chronic diseases: results of a physician survey. *Arthritis Care Res* (Hoboken) 2012; 64: 1250-5.

6a. ANDERSSON JJ, BARON G, VAN DER HEIJDE D, FELSON DT, DOUGADOS M: Ankylosing spondylitis assessment group preliminary definitions of short term improvement in ankylosing spondylitis. *Arthritis Rheum* 2001; 44: 1876-86.

7. LANDEWÉ RB, VAN DER HEIJDE D: Follow up studies in rheumatoid arthritis. *Ann Rheum Dis* 2002: 61; 479-81.
8. VAN DER HEIJDE DMFM, VAN 'T HOF M, VAN RIEL PLCM, VAN DE PUTTE LBA: Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* 1993; 20: 579-81.
9. PREVOO MLL, VAN 'T HOF MA, KUPER HH, VAN LEEUWEN MA, VAN DE PUTTE LBA, VAN RIEL PLCM: Modified disease activity scores that include twenty-eight-joint counts. *Arthritis Rheum* 1995; 38: 44-8.
10. SMOLEN JS, BREEDVELD FC, SCHIFF MH *et al.*: A simplified disease activit index for rheumatoid arthritis for use in clinical practice. *Rheumatology* (Oxford) 2003; 42: 244-57.
11. ALETAHA D, NELL VP, STAMM T *et al.*: Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005; 7: R796-806.
12. PINCUS T, YAZICI Y, BERGMAN MJ: RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: Similar results to DAS28 and CDAI in clinical trials and clinical care. *Rheum Dis Clin North Am* 2009; 35: 773-8.
13. GARRETT S, JENKINSON T, KENNEDY LG, WHITELOCK H, GAISFORD, P, CALIN A: A new approach to defining disease status in ankylosing spondylitis: The Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* 1994; 21: 2286-91.
14. LUKAS C, LANDEWÉ R, SIEPER J *et al.*: Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009; 68: 18-24.
15. HELLIWELL PS, FITZGERALD O, FRANSEN J *et al.*: The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). *Ann Rheum Dis* 2013; 72: 986-91.
16. SHARP JT, LIDSKY MD, COLLINS LC, MORELAND J: Methods of scoring the progression of radiological changes in rheumatoid arthritis: correlation of radiological, clinical and laboratory abnormalities. *Arthritis Rheum* 1971; 14: 706-20.
17. RAVINDRAN J, CAVILL C, BALAKRISHNAN C, JONES SM, KORENDOWYCH E, MCHUGH NJ: A modified sharp score demonstrates disease progression in established psoriatic arthritis. *Arthritis Care Res* 2010; 62: 86-91.
18. WANDERS AJB, LANDEWÉ RBM, SPOORENBERG A *et al.*: What is the most appropriate radiologic scoring method for ankylosing spondylitis? *Arthritis Rheum* 2004; 50: 2622-32.
19. VAN DER HEIJDE DMFM, DANKERT T, NIEMAN F, RAU R, BOERS M: Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. *Rheumatology* (Oxford) 1999; 38: 941-7.
20. KILTZ U, VAN DER HEIJDE D, BOONEN A *et al.*: Development of a health index in patients with ankylosing spondylitis (ASAS HI): Final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis*. 2014 Jan 7 [Epub ahead of print].
21. VAN WEELY SFE, VAN DENDEREN JC, STEULTJENS MPM *et al.*: Moving instead of asking? Performance-based tests and BASFI-questionnaire measure different aspects of physical function in ankylosing spondylitis. *Arthritis Res Ther* 2012; 14: R52.
22. FRIES JF, SPITZ P, KRAINES RG, HOLMAN HR: Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980; 23: 137-45.
23. PINCUS T, SUMMEY JA, SORACI SA, WALLSTON KA, HUMMON NP: Assessment of patients' satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. *Arthritis Rheum* 1985; 26: 1346-53.

23a. PINCUS T, SWEARINGEN C, WOLFE F: Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. *Arthritis Rheum* 1999; 42: 2220-30.

24. WOLFE F, MICHAUD K, PINCUS T: Development and validation of the health assessment questionnaire II. *Arthritis Rheum* 2004; 50: 3296-305.
25. FRIES JF, KRISHNAN E, ROSE M, LINGALA B, BRUCE B: Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther* 2011; 13: R147.
26. CALIN A, GARETT S, WHITELOCK H *et al.*: A new approach to defining functional ability in ankylosing spondylitis: The development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* 1994; 21:1281-5
27. VAN GESTEL AM, PREVOO MLL, VAN 'T HOF MA, VAN RIJSWIJK MH, VAN DE PUTTE LBA, VAN RIEL PLCM: Development and validation of the European League against Rheu-

matism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1996; 39: 34-40.

28. WELSING PMJ, LANDEWÉ RBM, VAN RIEL PLCM *et al*.: The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis. *Arthritis Rheum* 2004; 50: 2082-93.
29. RAMIRO S, VAN DER HEIJDE D, VAN TUBERGEN A *et al*.: Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. *Ann Rheum Dis* 2014; 73: 1455-61.
30. LANDEWÉ R, VAN DER HEIJDE D, VAN DER LINDEN SJ, BOERS M: Twenty-eight joint counts invalidate the DAS28 remission definition owing to an omission of the lower extremity joints: A comparison with the original DAS remission. *Ann Rheum Dis* 2006; 65: 637-41.
31. LANDEWÉ R, VAN DER HEIJDE D: The validity of a rheumatoid arthritis medical records-based index of severity compared with the DAS28. *Arthritis Res Ther* 2006; 8: 107.