Joint space narrowing, cartilage and physical function: are we deceived by measurements and distributions?

Robert Landewé,[1,2] Désirée van der Heijde[3]

Irreversible physical function loss—that part of physical function loss that remains in the absence of clinically perceptible disease activity—was elegantly conceptualised by Aletaha et al[1] a few years ago. They introduced the idea of the irreversible health assessment questionnaire (HAQ) score, which is the residual HAQ score if a patient is in clinical remission. While one may argue the construct validity of an irreversible HAQ score (eg, the absence of clinical disease activity does not necessarily imply the absence of joint inflammation), the model is valuable because it allows the investigator to disentangle the contribution of signs and symptoms and that of structural damage on physical function.

From many studies performed during the past 10 years we have learnt that structural joint damage independently of disease activity contributes to explaining physical function (eg, Welsing et al).[2] Therefore, in the absence of disease activity, it is structural joint damage that explains irreversible physical function loss.

Aletaha et al[1] have taken up his hypothesis as a starting point, and have tried to unravel the relationship between the two components of radiographic damage—erosions and joint space narrowing (JSN)—and irreversible physical function loss. The analysis they have done, which is presented in this issue of the journal, is in all aspects provocative and challenging, and we truly commend the authors for taking up this exercise (see page 733). Strong elements of their study are the access to databases of large clinical trials with inherently high-quality data and completeness, as well as a hypothesis-driven approach. It is a broadly endorsed idea among clinicians that JSN matters more than erosions. So, the main conclusion of the analysis by Aletaha et al,[1] namely that JSN more than erosions impacts irreversible physical function loss, will be easily accepted by the clinical readership.

The analyses and interpretations the authors have presented have a number of peculiarities that may warrant a closer look in order to judge the results more effectively. First, we will argue the assumed relationship between JSN assessed on radiographs and cartilage loss. Then, we will discuss causality, metric properties of the HAQ score and problems of skewed data distributions.

JSN VERSUS CARTILAGE LOSS

It is attractive to assume that JSN on radiographs means loss of cartilage. Formally, this association has not been proved, as far as we know. Aletaha et al[1] have analysed trials that have been scored by either the Sharp method or by the van der Heijde modification of the Sharp method. The former method excludes joints with (sub)luxation for scoring, the latter method assigns scores of 3 to joints with either severe JSN or subluxation and 4 to joints with either absence of joint space or luxation. So part of the higher JSN scores is not caused by cartilage loss, but is a result of soft tissue damage, and the authors herewith disregard the fact that (sub)luxation may importantly contribute to irreversible function loss.

Another issue of potential importance with regard to the interpretation of the authors' results is that erosions and JSN are measured congruently in the small joints of the hands and feet, whereas JSN and erosions are measured separately in many joints of the wrist. We know that the wrist disproportionally determines the total JSN score, while scoring erosions in the wrist is inherently difficult. It is not clear how these scoring characteristics have impacted irreversible function loss: mainly by wrist involvement (because of its predominant contribution to the JSN score) or by involvement of the small joints of the hands and feet (by their more similar contribution to the erosion and JSN score).

CAUSALITY

Correlation does not imply causation. The authors describe an association of JSN and irreversible HAQ score, and claim that JSN more than erosions contributes to irreversible function loss. An implication of this claim could be that treatment should preferably be aimed at preserving JSN rather than preventing erosions, but that part of the analysis is missing. What we should investigate, and the authors agree in this, is how a change in JSN over time is associated with a change in the HAQ score, and if so whether a change in the JSN score matters more than a change in the erosion score. Until we see such data, the results of this study can still be explained as if patients with the worst HAQ score also have the worst radiographic scores, both as an independent result of the underlying disease process. In such a scenario there is a close relationship between the HAQ score and the Sharp score, but the causal chain ('does JSN lead to impaired physical function') is not clear.

MEASUREMENT INSTRUMENTS

The authors use the HAQ score as a continuous measure. Rasch analysis has shown that the HAQ score performs like an ordinal measure with intervals of 0.125,[3] not necessarily as a continuous measure. Although we usually handle the HAQ score satisfactorily as a continuous measure in clinical trials if we compare group means, it is questionable to what extent the subtle relationships between radiographic scores and HAQ scores are hampered by the ordinal character of the HAQ, and by floor effects and ceiling effects inherent in this measure.[4][5]

RADIOGRAPHIC DATA

Distribution histograms of radiographic scores of patients with rheumatoid arthritis have a peculiar shape that is best described as 'positively skewed'. Typically, up to 50% of patients have relatively low scores, whereas a minority of patients, often less than 10%, may have

[1]Maastricht University Medical Center, Maastricht, The Netherlands
[2]CAPHRI School for Public Health and Primary Care, Maastricht, The Netherlands
[3]Leiden University Medical Center, Leiden, The Netherlands

Correspondence to Professor Robert Landewé, Maastricht University Medical Center, PO Box 5800, Maastricht 6202 AZ, The Netherlands; r.landewe@mumc.nl

scores that go as high as 100 units for both erosions and JSN separately. As a consequence, this minority grossly influences the mean sample score, while the median score, a better statistic for this purpose, is far lower than the mean.

The basis of the author's analysis was a subdivision of their sample into tertiles. Such an approach suggests a trend of increasing severity across tertiles, but because of a skewed data distribution such an interpretation may fall short. Making use of the authors' data, we have calculated that the patients with erosion scores in the first two erosion tertiles only accounted for the lower 6% range of all observed erosion scores, while the patients in the third tertile accounted for the remaining (and higher) 94% range (table 1). The same was true for the JSN tertiles.

If one assumes a linear relationship between the JSN score and the irreversible HAQ score, which is what the authors inherently claim, almost all effect of JSN on the irreversible HAQ score should be seen in the third JSN tertile. As it is far more likely that the relationship between JSN and the (irreversible) HAQ score is threshold dependent (ie, you need a certain threshold of JSN before it interferes with functioning) or non-linear, the effect of JSN on the irreversible HAQ score in the third tertile should be even more extreme.

The skewed distribution may also be responsible for spurious effects in the analysis in which erosion tertiles and JSN tertiles are combined. The authors show a positive relationship between JSN tertiles and the irreversible HAQ score across erosion tertiles, whereas they do not find such a relationship between erosion score tertiles and the irreversible HAQ score across JSN tertiles. At first sight, and supported by strong 'visuals', this is convincing evidence, but realise that the first two tertiles for both the erosion score and the JSN score are formed by almost 'clean' patients, whereas the erosion score and the JSN score are usually closely correlated. This implies that the congruent categories (eg, highest tertile for both erosion and JSN) are relatively well filled with patients, but that the incongruent categories (eg, highest tertile for erosion; one of the two lower tertiles for JSN) are filled with far fewer patients. Unfortunately, the authors do not provide numbers per category, so that we cannot weigh this argument appropriately, but it looks as if the authors' claim on the association between JSN tertiles and the irreversible HAQ score relies importantly on the inadvertently low mean irreversible HAQ score in the category of patients with an erosion score in the highest tertile (erosion scores between 11 and 170) and a JSN score in the lowest tertile (JSN scores below 0.5). Such patients must be hard to find in real life, and it is difficult to believe that they do better in terms of the irreversible HAQ score than patients with an erosion score and a JSN score in the lowest tertile.

One may argue that statistical testing and multivariate modelling account for these concerns, and this is partly true, but skewed distributions may behave spuriously when analysed with parametric statistics and models assuming normality and linearity. The most important limitation is overweighing extreme values (such as very high JSN scores): Scatter plots with a cloud of observations in the 'left lower corner' and one extreme observation in the 'right upper corner' produce seemingly meaningful and statistically significant positive correlation coefficients. Linear regression analysis essentially behaves similarly. The authors defend themselves in the text by stipulating the general limit theorem, essentially allowing statistical methodology that assumes normal distributions if sufficiently large sample sizes are used, but the general limit theorem does not account for non-linearity and data transformation is necessary. We would have greater confidence in the results if the analyses show statistically similar results when using transformed data, regardless of the inherent lack of interpretability of such results. We would also welcome a non-linear approach to elucidate these complex relationships further.

So, how should we interpret the data presented by Aletaha et al?[1]

First of all, we praise their provocative efforts to try and shed light on the obscured relationship between components of radiographic damage and their consequences. It is commendable to try and make use of the exquisite databases of pharmaceutical industries and combine them when possible in order to address important clinical questions. We truly believe that there is lot of hidden information in those datasets that is of higher quality than every other imaginable dataset in the field.

Do we believe that JSN may be more important than erosions in explaining the loss of physical function? This may definitely be true. We have tried here to report that we think Aletaha et al[1] have only taken the first step in unravelling these interactions that in the end should lead to refocussed efforts in designing new and even more efficient drugs in rheumatoid arthritis.

We challenge the authors to try and confirm their results and to prove that our methodological concerns are only grumpy futilities.

REFERENCES

1. **Aletaha D,** Smolen J, Ward MM. Measuring function in rheumatoid arthritis: identifying reversible and irreversible components. *Arthritis Rheum* 2006;**54**:2784–92.
2. **Welsing PM,** van Gestel AM, Swinkels HL, *et al.* The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis Rheum* 2001;**44**:2009–17.
3. **Tennant A,** Hillman M, Fear J, *et al.* Are we making the most of the Stanford Health Assessment Questionnaire? *Br J Rheumatol* 1996;**35**:574–8.
4. **Taylor WJ,** McPherson KM. Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. *Arthritis Rheum* 2007;**57**:723–9.
5. **ten Klooster PM,** Taal E, van de Laar MA. Rasch analysis of the Dutch Health Assessment Questionnaire disability index and the Health Assessment Questionnaire II in patients with rheumatoid arthritis. *Arthritis Rheum* 2008;**59**:1721–8.

Table 1 Range and cumulative distribution of erosion scores and JSN scores per tertile of patients

	First	Second	Third
Tertiles of erosion score			
Score range erosions (units (minimum–maximum))	0–2.3	2.5–11	11–170
Percentage range of the maximum observed erosion score (%)	0–1.4	1.4–6.5	6.5–100
Tertiles of JSN score			
Score range JSN (units (minimum–maximum))	0–0.5	0.6–5.4	5.5–121
Percentage range of the maximum observed JSN score (%)	0–0.4	0.4–4.5	4.5–100

The maximum observed scores in the dataset were 169.5 units for the erosion score and 120.5 units for the joint space narrowing (JSN) score.