Editorial

Imaging Damage: Scoring Versus Measuring

Few things stimulate advances in medicine so well as the ability to measure. Take, for example, the developments in osteoporosis since bone mineral density became widely available and then contrast our struggle with the role of "stress" in disease and disability. So imagine in the future receiving an imaging report on your rheumatoid patient like the report in Table 1.

Perhaps the data in it will have been drawn from a variety of imaging techniques: radiography, magnetic resonance imaging (MRI), ultrasound — the output of each integrated and refashioned with some nifty computer software. Periarticular bone density may well be an additional part of the assessment.

It goes almost without saying that measurements must accurately reflect the true anatomy or physiology, and be reliably reproducible. If they fail these essential tests they are a liability. The practical requirement is that they are readily obtainable within a clinical context.

But if the data in the Table passed all these tests and could be generated as easily from the hand as the foot, knee, hip or spine, and as well from the patient with inflammatory as erosive or degenerative disease, what insights and advances they should provoke. Individual patient management could be better fine tuned to limit damage. New therapies and treatment strategies for inflammatory arthritis and osteoarthritis (OA) could be evaluated with greater speed and accuracy. The possible dissociation in rheumatoid arthritis (RA) between rates of bone erosion and cartilage damage, with its important pathogenic and treatment implications, could be more readily analyzed. The question of the relationship between structure and function in the rheumatic diseases could be better studied. Perhaps our concept of the natural history of OA would be revised. And so on.

It seems easily arguable then, that it would be a major advance to be able to measure accurately, reliably, and easily the physical characteristics of the articular and periarticular tissues that suffer damage in the rheumatic diseases. But we have a way to go.

For RA, scores of radiographic images remain the standard method by which to capture damage and its progression. There have been many attempts to devise an optimal scoring system[1], but the Larsen[2] and Sharp[3] methods and their several modifications[4-6] are used most often. The OMERACT Imaging Group members have worked hard to characterize and reduce the problems in these scoring systems, but many are inherent: a score depends on both the scoring system and the scorer, and scorers make choices — about the order in which they read radiographs, about categorical assignment of borderline lesions. Floor or ceiling effects limit the representation of disease at the either end of the spectrum and "ordinal level scores" (grades of difference that may be of quite unequal degree) often masquerade as an interval level measurement (i.e., numbers with arithmetic significance, where 2 + 2 = 4)[7]. Inter and intrareader variability is a major potential difficulty, but can be improved by training and calibration and recent efforts allow measurement error to be taken into account[8].

MRI offers the prospect of 3 dimensional representation of RA lesions and is a particularly attractive and sensitive imaging method for articular disease. But even with this technique, the best we can currently do to capture RA damage is provided by another scoring system, the OMERACT Toulouse score, which is supported, if not validated, by reliability data from multicenter studies[9]. It is notable that while this scoring system can provide an assessment of erosions, bone lesions, and synovitis, it is insufficiently sensitive to image reliably the degree of cartilage damage in the small joints of the hands and wrist[10].

Measurement of radiographic damage in OA has a much longer history[11] probably because, with positioning standardized, direct measurement of joint space width in the hip and knee is not too arduous a task[12]. For RA, the prospect of similarly measuring multiple bone lesions and joint spaces with a ruler, in so many small joints, is dauntingly tedious, although Buckland-Wright and colleagues directly measured both erosion areas and joint space width in a small

See: Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis, *page 1825*

Table 1. A hypothetical imaging report.

PT: Ms B.T. d.o.b. 25.4.88 Date of Study: 16.9.20xx	Right			Left		
	Carpus	MCP	PIP	Carpus	MCP	PIP
Joint cartilage volume						
Vol μl	3997	1176	256	3479	1298	224
Δ vol μd*	–147	–75	–27	–156	0	0
T score†	–1.77	–1.5	–1.4	–2.1	–2.4	–1.9
Eroded bone volume						
Vol μl	112	65	24	85	17	7
Δ vol μl*	–14	–5	–8	–11	–1	–1.5
Joint capsule tension						
Ten. st. Pa	1.1×10^5	0.9×10^5	1.3×10^5	1.1×10^5	0.9×10^5	1.3×10^5
Ten. st* ΔPa	$0.1 \times 10^{5-}$	$0.01 \times 10^{5-}$	0	$0.1 \times 10^{5-}$	$0.02 \times 10^{5-}$	$0.001 \times 10^{5-}$

* Change since the last study on: 6.3.20xx.
† T score based on normative data from age, sex, habitus matched controls.
(NB: The values in the table make no claim at all on reality.)

study on patients with RA using microfocal radiographs[13]. For a more general move from scores to measurement, it was necessary to wait for an automated method, a fusion of computer and image.

An early application of image analysis for measurement of RA changes in the small joints of the hands was undertaken by Gaydecki, *et al*[14]. In 1995, James, Heald, and colleagues reported on the use of computerized image analysis to measure the radiographic joint space in the metacarpophalangeal and proximal interphalangeal joints in patients with RA. They found the computerized measurement to be accurate and reliable, but considered it compared poorly with Sharp scores for joint space since measured joint space width had to decrease considerably [especially in the metacarpalphalangeal (MCP) joints] before the score increased noticeably, "highlighting the non-linear, variable nature of the subjective scoring process"[15]. This was not quite the experience of Sharp and his colleagues[16], who used a different computer program to measure radiographic joint space width and to estimate erosion volumes in the hands of patients with RA under treatment with either gold therapy or placebo. The computerized system again proved reproducible; moreover, for joint space width, the system was consistent with the scoring method. For estimated erosion volume, the computerized measurements showed a greater difference between the gold and placebo treated patients, probably indicating the greater responsiveness of measurement over scoring.

In this issue of *The Journal*, Angwin, James, Heald, and colleagues extend their earlier work, this time examining the effect of variable hand position on computerized measurement of joint space width in healthy men[17]. The paper carefully explores sources of measurement error. The authors calculate an individual cutoff value (ICO) [the same notion as the "smallest detectable difference" (SDD) in the OMERACT studies[18]], which is the mean joint space width (JSW) difference greater than the measurement error of the difference, reflecting true change in JSW in an individual subject. However, measurement error is usually study-specific and the ± 0.05 mm JSW (average of all MCP and proximal interphalangeal joints of both hands) calculated in this group of healthy men cannot be generalized to a disease population that is characterized by a range of joint space width and beset by other problems such as joint malalignment and osteoporotic bone[19].

In addition to the individual cutoff, the authors also calculate a JSW group cutoff value (± 0.01 mm). Whereas the individual cutoff is the 98% confidence interval around the *standard deviation* of the differences between paired measurements, the group cutoff is the 95% confidence interval around the *mean* of the differences between paired measurements, a statistical procedure better known as the paired t test. The concept of a group cutoff is useful and aptly parallels that of the individual cutoff but its correspondence to the paired t test is not transparent in the paper. We believe it should be, because the assessment of *reliability* is already beleaguered by such variable nomenclature. Even the term itself — reliability — is replaced by a variety of other terms that reflect sometimes only subtle differences in context or perspective: precision, reproducibility, repeatability, replicability, stability, consistency, test-retest, agreement and concordance.

Finally, the paper shows well that the advantage of the computer in measurement is not its greater mensural accuracy but its ability to make many more measurements more rapidly than is possible manually. The improved precision is primarily a function of averaging 180 measurements rather than 8.

The pioneers in this field of radiographic and MRI lesion measurement will surely have many followers, but a rate limiting step could be access to critical computer software. Intellectual property needs protection, but open evaluation

and testing require broad access to common acquisition and computer based technologies. In this respect, Sharp, *et al*[16] set a great example: their work on joint space width used a public domain program developed at the US National Institutes of Health for which they list the website address; they conclude their paper with an offer to make their software programs available on application. Angwin, *et al*[17] are similarly generous.

Our fictional imaging report is coming dimly into view and measurement is on its way. It will require a lot of work: deciding exactly what is worthwhile measuring and how many sites are representative; finding the best mix of imaging modalities — radiography, computerized tomography, MRI, ultrasound — to capture different types of pathology; devising the right computer programs to handle and integrate these different imaging data; testing these systems on the highly diverse anatomy and pathology encountered in the field; gathering normative data and then rigorously evaluating the discriminatory and accuracy performance of the measures in different study contexts. A big agenda. It will need much expertise, not less in measurement theory than in imaging and computing.

JOHN EDMONDS, MB, BS, MA, FRACP,
Director of Rheumatology,
St. George Hospital,
Professor of Rheumatology
University of New South Wales;
MARISSA LASSERE,
MB, BS(Hons), Grad Dip Clin Epi, PhD, FRACP, FAFPHM,
Staff Specialist in Rheumatology,
St. George Hospital,
Senior Lecturer in Medicine,
University of New South Wales,
Sydney, Australia 2217.

Address reprint requests to Dr. Edmonds.

REFERENCES

1. van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of the scoring methods, their reliability and applicability. Baillieres Clin Rheumatol 1996;10:435-53.
2. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn 1977;18:481-91.
3. Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971;14:706-20.
4. Scott D, Houssien D, Laasonen L. Proposed modification to Larsen's scoring method for hand and wrist radiographs. Br J Rheumatol 1995;34:56.
5. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999;26:743-5.
6. Rau R, Herborn G. A modified version of Larsen's scoring method to assess radiologic changes in rheumatoid arthritis. J Rheumatol 1995;22:1976-82.
7. van der Heijde D, Boers M, Lassere M. Methodological issues in radiographic scoring methods in rheumatoid arthritis. J Rheumatol 1999;26:726-30.
8. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
9. Ostergaard M, Klarlund M, Lassere M, et al. Inter-reader agreement in the assessment of MRI images of rheumatoid arthritis wrist and finger joints — an international multicenter study. J Rheumatol 2001;28:1143-50.
10. Conaghan P, Edmonds J, Emery P, et al. MRI in rheumatoid arthritis: summary of OMERACT activities, current status and future plans. J Rheumatol 2001;28:1158-61.
11. Ravaud P. Quantitative radiography in osteoarthritis: plain radiographs. Baillieres Clin Rheumatol 1996;10:409-14.
12. Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. J Rheumatol 1996;24:786-91.
13. Buckland-Wright JC, Carmichael I, Walker SR. Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis. Ann Rheum Dis 1986;45:379-83.
14. Gaydecki P, Browne M, Mamtora H, Grennan DM. Measurement of radiographic changes occurring in rheumatoid arthritis by image analysis techniques. Ann Rheum Dis 1987;46:296-301.
15. James MF, Heald G, Shorter JH, Turner RA. Joint space measurement in hand radiographs using computerized image analysis. Arthritis Rheum 1995;38:891-901.
16. Sharp J, Gardner JC, Bennett EM. Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. Arthritis Rheum 2000;43:1378-86.
17. Angwin J, Heald G, Lloyd A, et al. Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis. J Rheumatol 2001;28:1825-36.
18. Lassere M, van der Heijde D, Johnson K, et al. The reliability of measures of disease activity and disease damage in rheumatoid arthritis: Implications for the smallest detectable difference, the minimum clinical important difference and the analysis of treatment effects in randomized controlled trials. J Rheumatol 2001; 28:892-903.
19. Lassere M, van der Heijde D, Johnson K, et al. Robustness and generalizability of smallest detectable difference in radiological progression. J Rheumatol 2001;28:911-3.