Assessment of Radiographs in Longitudinal Observational Studies

DÉSIRÉE M.F.M. VAN DER HEIJDE

ABSTRACT. Radiographs are important to assess structural damage in longitudinal studies. This article describes several issues on the selection of films, frequency of followup, scoring of radiographs, and presentation of results, especially in the context of longitudinal studies. (J Rheumatol 2004;31 Suppl 69:46–47)

Key Indexing Terms:
RADIOGRAPHS LONGITUDINAL OBSERVATIONAL
RHEUMATOID ARTHRITIS OUTCOME STRUCTURAL DAMAGE

Structural damage is an important outcome for assessment in longitudinal observational studies. Films provide a permanent record necessary for serial evaluation, show cumulative damage as a history of joint pathology, and are cheap, well standardized and widely available[1]. Moreover, many joints are assessed on films of hands and feet. Films can be used to define disease severity at a single time point or, more interestingly, over time, as a result of natural history; films can also be used to assess effectiveness of therapy.

HANDS AND FEET

If films are being used as an outcome measure to assess severity of structural damage, films of hands and feet are sufficient. For purposes of addressing specific research questions, however, films of large joints and the cervical spine could be added. It has been proven that damage in small joints of hands and feet is a good indication of the overall damage in all joints[2]. Hand films give only part of the picture, especially in early disease[3]. How frequently films should be taken depends on the research question. There is little value in taking films more than once yearly if the general aim is to assess structural damage. If there are specific issues to be addressed, films taken every 6 months can be useful.

It is important to obtain complete followup data on all patients; also there should be little variation in the time intervals between films. On a group level the progression of radiographic damage is fairly linear[4,5]. However, in individual patients there might be considerable variation[6]. Therefore differences in followup between films among patients cannot easily be corrected for.

From the University Hospital Maastricht, Maastricht, The Netherlands, and Limburg University, Diepenbeek, Belgium.

D.M.F.M. van der Heijde, MD, PhD, Professor of Rheumatology.

Address reprint requests to Dr. D. van der Heijde, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands.

SCORING FILMS

Scoring films can best be done in batches, i.e., series of films of the same patients scored at the same time. There is still debate on whether films should be scored with known sequence or completely randomized for time. Without any information on time sequence, films are assessed with a minimum of (expectation) bias. However, this approach also introduces extra measurement error, which can result in the loss of the signal: present progression can be lost in measurement error[7]. Recently, it was shown that clinical experts agree most on scores obtained by reading with known sequence[8]. This applies to films taken one year apart. If the time between films is greater, the influence of knowing the order of the films might be smaller. So far there is no consensus on this question. Scoring in sequence gives the highest likelihood of detecting change if present, but might result in the overestimation of progression. In contrast, scoring without information on chronology gives the most conservative estimate of progression, but might result in missing real progression. Choices need to be based on the preference for the specific research setting.

OBSERVERS

Preferably, films should be scored by 2 observers and the average score used for analyses. An average score is the best way to handle random measurement error. Using a consensus score runs the risk of score bias in the direction of one of the observers, often the most experienced, senior, or dominant person. Especially for large longterm observational studies with many films, it may not be feasible to have 2 readers. In this situation, the scores of one reader can be used. However, a subset of the films should be scored by a second observer to assess interobserver agreement, and thereby the generalizability of the results. It is important that the observers be well trained, which reduces the measurement error greatly[9].

SCORING METHODS

The most widely used scoring systems are the Sharp and the

Larsen method with several modifications[10]. The Sharp method is a detailed scoring system for erosions and joint space narrowing separately[11,12]. This can be used for both hand and foot films. The Larsen is a more global grading system of the entire joint[13-15]. The grades are mainly based on erosions. The Larsen method uses reference films and can be used for films of hands and feet, but also for large joints. Both methods are valid, feasible, and reproducible. The (modified) Sharp method is more sensitive than the Larsen. However, this is at the cost of more time needed to score the films[16]. Again, which method is best depends on the research question, the availability of trained readers, and the length of time for study.

PRESENTATION OF RESULTS

Another essential aspect is the presentation of the results. Data can be presented on a patient level, e.g., the number of patients with erosions; on a joint level, e.g., number of damaged joints; or based on abnormalities, e.g., number of erosions. For all these results absolute and progression scores can be calculated. Usually this is done on a group level, e.g., the mean or median increase in erosions with a measure for variability. However, such an approach gives no information on how many patients showed progression, although such information might be significant. To be able to define how many patients progressed, a cutoff value needs to be determined.

What we would like to know is the minimal clinically important difference: the minimum progression that makes a difference in outcome. This is not known for radiographic progression. One way to define a cutoff value is to calculate the smallest detectable difference (SDD)[17,18]. This is the smallest progression that can be detected apart from measurement error. If 2 observers are being used, the SDD should be based on the scores of the 2 readers. If only one observer reads the films, the SDD can be based on a sample of films scored twice by the observer.

It has been proven that clinical experts judge progression in the magnitude of the SDD as a clinically meaningful change[8]. Even progression smaller than this was judged clinically meaningful, especially for the Larsen method. So by applying the SDD, patients with a progression score higher than the SDD have a clinically meaningful change, but patients might be missed who also have a clinically meaningful change. Progressors and nonprogressors can be used for further analyses, such as the evaluation of prognostic factors for radiographic progression. Recommendations for reporting of radiographic data were made at a recent roundtable, and it would be desirable to comply with the recommendations so that all studies present the same minimum set of data to ensure comparability of results[19].

REFERENCES

1. Dawes PT. Radiological assessment of outcome in rheumatoid arthritis. Br J Rheumatol 1988;27 Suppl 1:21-36.
2. Drossaers-Bakker K, Kroon H, Zwinderman A, Breedveld F, Hazes J. Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function. Rheumatology 2000;39:998-1003.
3. Hulsmans HM, Jacobs JW, van der Heijde DM, van Albada Kuipers GA, Schenk Y, Bijlsma JW. The course of radiologic damage during the first six years of rheumatoid arthritis. Arthritis Rheum 2000;43:1927-40.
4. Scott D, Pugner K, Kaarela K, et al. The link between joint damage and disability in rheumatoid arthritis. Rheumatology 2000;39:122-32.
5. Wolfe F, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. Arthritis Rheum 1998;41:1571-82.
6. Plant MJ, Jones PW, Saklatvala J, Ollier WER, Dawes PT. Patterns of radiological progression in early rheumatoid arthritis: Results of an 8 year prospective study. J Rheumatol 1998;25:417-26.
7. van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology Oxford 1999;38:1213-20.
8. Bruynesteyn K, van der Heijde D, Boers M, et al. Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts. J Rheumatol 2001;28:904-10.
9. Fries JF, Bloch DA, Sharp JT, et al. Assessment of radiologic progression in rheumatoid arthritis. A randomized, controlled trial. Arthritis Rheum 1986;29:1-9.
10. van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. Baillieres Clin Rheumatol 1996;10:435-53.
11. Sharp JT, Young DY, Bluhm GB, et al. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? Arthritis Rheum 1985;28:1326-35.
12. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
13. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn Stockh 1977;18:481-91.
14. 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.
15. Scott D, Houssien D, Laasonen L. Proposed modification to Larsen's scoring method for hand and wrist radiographs. Br J Rheumatol 1995;34:56.
16. Wassenberg S, Herborn G, Larsen A, et al. Reliability, precision and time expense of four different radiographic scoring methods [abstract]. Arthritis Rheum 1998;41 Suppl:S50.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
18. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
19. van der Heijde D, Simon L, Smolen JS, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis? Guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.