Internet hand x-rays: A comparison of joint space narrowing and erosion scores (Sharp/Genant) of plain versus digitized x-rays in rheumatoid arthritis patients

Hector O Arbillaga, Gregory P Montgomery, Luis P Cabarrus, Margaret M Watson, Liam Martin and Steven M Edworthy *

Address: Department of Medicine, University of Calgary, Calgary, Alberta, Canada

E-mail: Hector O Arbillaga - hectorarbillaga@hotmail.com; Gregory P Montgomery - gpj_montgomery@hotmail.com; Luis P Cabarrus - lupero@hotmail.com; Margaret M Watson - pmwatson@telusplanet.net; Liam Martin - lmartin@ucalgary.ca; Steven M Edworthy* - sedworth@ucalgary.ca

*Corresponding author

Background: The use of radiographic images as a means of assessing the progression of rheumatoid arthritis (RA) in individual patients and clinical studies has been standard practice for several decades. Underlying this practice is the belief that radiographic changes are the consequence of inflammatory...
The aim of this article is to report the reliability of Internet hand x-rays for scoring erosions and joint space narrowing in patients with Rheumatoid Arthritis.

Methods

Radiograph Selection

Plain x-ray films of the hands of 36 RA patients were selected as a representative sub-set of a group of 235 patients who had taken part in a clinical study involving an examination of hand function. Sample characteristics are summarized in Table 1.

Scanning and posting of radiographs on the internet web site

The 36 plain film x-rays were scanned using a Scanmaster DX with a digitizing area from $\frac{1}{2}^\prime \times \frac{1}{2}^\prime$ to 14" $\times$ 17" and a maximum film size of 15" $\times$ 18". The resolution was 1K, 2K, 4K, and 8K on standard film sizes, and up to 9 lp/mm for custom resolution. The grayscale resolution was 12-bits (4096 gray levels), and the detector for the scanner was a high definition CCD. The interface was SCSI-2.

After scanning, the digitized x-rays were converted to a grayscale mode that discarded color information. This file format permitted rotation and cropping of the image, and enabled necessary alterations in contrast and brightness. The file was then saved in TIFF format (i.e., large files, 12 bits, grayscale). The digitized images were post-processed using Adobe Photoshop, and were subsequently converted to JPEG format (i.e., small files, 8 bits, RGB). The images were ultimately posted to a secure Internet web site in JPEG format (see Figure 1).

Scoring of erosions and joint space narrowing

Erosion scores (ERO) were recorded based on the Sharp [8][4,9] scoring method with Genant [6] modification. This method of scoring (see Figure 2 and Figure 3) uses an 8-point scale with 0.5 increments (each increment denoted by a "\*" sign). The descriptors for the numerical values on the scale were as follows: 0 = normal, 0+ = questionable or subtle change, 1 = mild, 1+ = mild, worse, 2 = moderate, 2+ = moderate, worse, 3 = severe, and 3+ = severe worse. Fourteen sites (4 quadrants per site) were selected for scoring on each hand. These sites were the interphalangeal (IP) joint of digit I, the proximal interphalangeal (PIP) joint of digits II-V, the metacarpophalangeal (MCP) joint of digits I-V, the carpometacarpal (CMC) joint of digit I, the scaphoid, the distal radius, and the distal ulna (see Figure 2 and Figure 3). The four scores, one for each quadrant, were summed to produce a total score for each joint. The ERO scores were then transformed mathematically to produce a Genant score, based on a 4-point scale. The transformed data corresponded to one of the following designations: 0 = normal, 1 = one well-defined erosion, 2 = two or three well-defined erosions, and 3 = two or more well-defined erosions plus an erosion score of at least 0.5 on any combination of two of the four quadrants (i.e. 1&2, 3&4, 1&3, 2&4). The possible range of scores for ERO in the combination of both hands using

Table 1: Basic characteristics of the patient sample.

| Number of Patients (%) |
|------------------------|
| Age, years | 54 (30–84) |
| Time from onset of symptoms, years | 13 (1.5–58) |
| Time from diagnosis, years | 12 (1.1–58) |
| Education Level, grade | 12 (9–17) |
| Female | 33 (91) |
| Male | 3 (9) |
| Rheumatoid Factor, positive | 28 (78) |
| RA Diagnosis | 36 (100) |

The Internet has revolutionized the computer and communications world in an unprecedented manner. At once it possesses global broadcasting capabilities, acts as a mechanism for information dissemination and is a medium for collaboration and interaction between individuals without regard for geographic location [7]. New approaches wedding technological advancements in digitizing methods and Internet communication will likely permit clinicians to participate in outcome studies involving radiological progression, particularly if digital radiographs posted on the Internet can be shown to be reliable representations of plain film x-rays.

The permanent nature of radiographs facilitates simultaneous comparison of images taken over an elapsed time period. Also, multiple readers can interpret the same set of images, allowing for a greater degree of reliability [1]. The objectivity of scoring can be heightened through masking names and dates on films, and by randomizing the sequence in which films are viewed [1]. The advancement in x-ray technology, coupled with a greater emphasis on the use of standardized patient positioning and radiographic techniques, has further enhanced the reproducibility of scoring [6].

The aim of this article is to report the reliability of Internet hand x-rays for scoring erosions and joint space narrowing in patients with Rheumatoid Arthritis.
the Genant score was 0–84. The erosion score was then normalized to a scale ranging from 0–100 in accordance with the Sharp-Genant procedures [6].

Joint space narrowing (JSN) scores were based upon the Sharp [8][4,9] with Genant [6] modification method. In this case a 9-point scale with 0.5 increments was used (see Figure 2 and Figure 4). The corresponding qualitative designations for this scale were: 0 = normal, 0+ = questionable or subtle change, 1 = mild (focal), 1+ = mild, worse, 2 = moderate (loss of less than 50% of joint space), 2+ = moderate, worse, 3 = severe (loss of more than 50% of the joint space), 3+ = severe worse, and 4 = ankylosis or dislocation. Thirteen sites (only 1 score per site) were selected for recording JSN scores: the interphalangeal (IP) joint of digit I, the proximal interphalangeal (PIP) joint of digits II-V, the metacarpophalangeal (MCP) joint of digits I-V, the combination of the carpometacarpal joints of digits III-V, the combination of the capitate-scaphoid-lunate, and the radiocarpal joint (see Figure 4). In contrast to the erosion data, the scores for JSN were not transformed from the Sharp with Genant modification to the Genant scoring system as the quadrant system described above was used only for ERO, while a single score per site was recorded for JSN. The range of possible scores for JSN in both hands was 0–104. This score was then normalized to a scale ranging from 0–100, providing a weighting equal to ERO score [6].

After summing the ERO and JSN scores for both hands, a composite score was created by adding the adjusted ERO and JSN scores together. The minimum value that could be obtained was 0, while the maximum value was 200 (i.e. 100+100).

Statistical methods
The statistical analyses were completed using descriptive statistics (SPSS 10.0) and Lin concordance functions [10] (STATA Rel. 7.0). The Lin co-efficient unites measures of accuracy (i.e. nearness of the data’s reduced major axis to the line of perfect concordance) and precision (i.e. tightness of the data about its reduced major axis) to determine whether observed data significantly diverge from the line of perfect concordance, which occurs at 45 degrees. The value of Lin’s co-efficient increases in relation to the accuracy and precision of the observed data. The Bland-Altman limits of agreement procedure uses data-scale assessment in analyzing both the accuracy (i.e. bias) and the amount of variation or precision between any two measured values when the range of data is sufficiently limited. This graphical representation approach is complimentary to the relationship-scale approach of Lin.

Results
General X-ray and clinical results
Standard reading of the plain x-rays showed that these films represented a wide range of radiological progression. Erosions ranged from 0 to 93 with average 45.5(SD 29.7). Joint space narrowing ranged from 30 to 92, with average 65.0 (SD 17.0). Global scores ranged from 30 to 183.5, with average 110.5(SD 43.6). This was consistent with the clinical characteristics of the patient population who had average duration from onset of symptoms of 13 years (range 1.5 to 58 years), and who were seropositive in 78% of cases. Eleven patients reported little or no disability, 13 required aids to daily living, and 12 were severely limited in their functional capacity, but were still ambulatory.

Repeated assessment of plain X-rays
To assess intra-rater reliability, the reader, in blinded and randomized fashion, re-scored a randomly selected set of films. Fifteen of 36 plain film x-rays were reassessed using a standard light box, without the aid of a magnifying lens or bright light. The Bland-Altman graphic demonstrated a slight divergence from perfect concordance (accuracy)
and high precision (i.e. tightness of the data about the major axis represented in the Lin concordance model as $r = 0.916$). The divergence from perfect agreement can be summarized numerically by the average difference pretest to post-test. This was $-13.6$ (95% CI $-4.4$ to $-22.8$), indicating that the second reading had higher scores on average for the global measure, with the preponderance of differences occurring on films with higher global scores.

**Intra-method reliability: plain and internet X-rays**

The tightness of the data to the reduced major axis for the global plain and Internet x-rays was 0.879 and accuracy was represented by an $r = 0.874$, to produce a concordance of 0.769. This is considered "substantial" concordance. This global score was a composite of both erosions and joint space narrowing. Erosion scores were very accurate, with almost complete overlay of the reduced major axis on the line of concordance (bias factor = 0.997; see Figure 5). However, joint space narrowing diverged considerably from the line of concordance (bias factor = 0.495; see Figure 6). These findings are reflected in the erosion concordance score, rho_c = 0.887 (see Figure 5) and the joint space narrowing concordance score, rho_c = 0.365 (see Figure 6).

The Bland Altman graphic (see Figure 7) demonstrates that the Global Scores were higher on average for the Internet scoring approach. Differences ranged from -20 to +60 with an average of 19.0 (95% CI 11.0 to 25.4). The greatest contribution to this variance came from the joint space narrowing readings (see Figure 8), as opposed to the erosion scores (see Figure 9). The joint space narrowing range was -1.0 to 43.0, with average 15.0 (95% CI 11.0 to 19.0).

**Discussion**

This study examines the reliability of deriving ERO and JSN scores from rheumatoid hand x-rays posted on a secure Internet web site. Although several previous studies have investigated the efficacy of digital x-rays to interpret the musculoskeletal system [11–13] and others compare digital and plain film x-rays for assessing rheumatoid arthritis [6,14], there has been limited investigation into the reliability of assessing Internet x-ray images of rheumatoid patients.

The anatomic locations selected for scoring the plain and Internet x-rays using the Sharp with Genant modification method were identical to those sites used by Genant et al [6]. The selection of these sites was based on the relative ease of reading and the general frequency of involvement in ERO and JSN in the hand and wrist [6]. The PIP, MCP, and carpal sites selected by Sharp represent the most active areas of hand involvement related to synovial inflammation of rheumatoid arthritis [15].

The literature suggests that the correlation coefficient alone cannot be used to determine an agreement between two quantitative variables [16–18]. Pearson’s correlation coefficient has the capacity to measure the strength of association between two variables, but it does not provide information about the concordance of two variables [17,19–22]. Having acknowledged the inherent limitations of Pearson’s co-efficient, it must be stressed that an expectation for two independent readers, or the same
reader on multiple trials, to replicate the same ERO and JSN scores on a joint by joint basis is too rigorous [19]. Indeed, the literature supports the notion that it is not necessary for the same absolute radiologic scores to be recorded when application of the scoring-system is carried out by different readers [1], or when the scoring-system is applied multiple times by the same reader [23]. Rather, it is the association of readings on a global (composite) level that is important for determining agreement with respect to overall disease severity. Concordance [10] as described by Lin, et al is a more appropriate method of examining both the precision and accuracy of intra-rater and inter-method reliability.

When the total ERO scores were compared as an indication of agreement of disease severity, the concordance was highly significant (0.887). However, interpretation of the Bland Altman graphic (see Figure 7) suggests a trend for patients with low to moderate levels of disease involvement to have Internet x-rays scored higher than corresponding plain x-rays. It has been noted (personal communication with Dr. J. Sharp) that there is a tendency

Figure 5
Lin’s concordance for Internet and plain x-ray erosion scores.

Figure 6
Lin’s concordance for Internet and plain x-ray joint space narrowing scores.

Figure 7
Bland-Altman data-scale assessment of the degree of agreement between plain and Internet global scores.

Figure 8
Bland-Altman data-scale assessment of the degree of agreement between plain and Internet joint space narrowing scores.
to give higher scores for individuals with low levels of disease involvement, perhaps related to "over-reading" of x-rays with early disease progression. However, this phenomenon would occur in both plain and Internet readings. The Internet version may provide the reader with a basis for giving slightly higher scores at the low end of the scoring scale. For example, a joint that received a Sharp with Genant Modification JSN score of 0 (i.e., normal) when viewed on a plain film x-ray, without the use of a magnifying lens or bright light, could quite easily receive a score of 1 (i.e., mild change) when viewed on the Internet x-ray, which may provide clearer contrast on a monitor with super VGA capacity. This phenomenon may also be accentuated by a "ceiling effect" [24,25]. Given the greater deterioration that characterizes higher ERO and JSN scores, the fine details detectable on Internet images are not likely to push the score higher than it would have otherwise been. Nevertheless, statistical anomalies must be considered as well; the sample size may have been too small to accurately reflect the true reliability, such that chance variation could also be an appropriate interpretation.

The potential for alteration of the image during image conversion and compression is a concern associated with posting radiograph images to a web site. In particular, the JPEG lossy compression algorithm consists of an image simplification stage that removes image complexity with some loss of fidelity, followed by a compression step, but in the case of images it is generally not critical to restore all of the data upon decompression [26]. Recent technological advances in the area of image compression have led to the development of wavelet compression. With this technology an algorithm is used that converts the image to a mathematical expression and subsequently allows analysis of the image as a whole [27]. Thus, an optimum compression ratio can be reached without compromising image quality. However, this technique was not used in the study. Nor were the JPEG formats compared to TIFF formats. These two points may be worthwhile areas of future research for the technical advancement of radiographic interpretation on the Internet.

Patient security and confidentiality represent another set of concerns connected with the posting and transmission of x-rays via the Internet [28,29]. Depending on the risks associated with the system and the resources required for minimizing the potential risks an appropriate balance needs to be sought between privacy and connectivity [30,31]. Internet technology can help to ensure security and confidentiality when transmitting patient identifiers through the use of strong encryption methods, such as is now used for credit card purchases on the Internet [31,32]. Security also could be guaranteed through public-key algorithms known only to appropriate individuals [32].

**Conclusions**

The present study suggests that the Internet represents a valid medium for displaying and scoring hand x-rays of patients with RA. This finding was based on the use of standard JPEG radiograph images. The development of wavelet compression technology, where none of the image information is discarded during the compression process [26], will further enhance the quality of x-ray images available for posting on the Internet.

As an increasing number of clinical trials emphasize the radiographic progression of the disease as a primary outcome measure [5,33–35], the capacity to employ a reliable method of scoring radiographs is placed at a premium. X-rays posted on the Internet have the potential to enhance reliability of scoring protocols used in clinical trials through heightened standardization and through the facilitation of information sharing and education. The ease of transmission of images and information over the Internet essentially renders geographic obstacles irrelevant, thereby fostering collaboration. Web sites can be set up with an atlas of x-ray images permitting clinicians around the globe to hone their scoring abilities in a tutorial like fashion. Similarly, the capacity to clearly display and disseminate standard scoring protocol via the Internet has the potential to enhance the consistency of x-ray scoring on a broad scale. Such advantages and opportunities are not available with the use of traditional plain film x-rays. We conclude that coupling Internet technology with standard approaches to radiographic analysis is a neces-
sary step toward the realization of a new spectrum of potential benefits.

Competing interests
None declared.

Author's Contributions
Author 1 (HA) performed the scoring of x-rays, and assisted in data management and data analysis. Author 2 (GM) drafted the manuscript, participated in data management and statistical analysis. Author 3 (LC) digitized x-rays and set up Internet web site from which x-rays were assessed. Author 4 (MW) participated in data collection and data management. Author 5 (LM) provided periodic review and design input for the study. Author 6 (SE) conceived of the study, participated in its design, and coordinated the study throughout.

Acknowledgements
The authors would like to thank Dr. J.B. Houpt, of Toronto, Ontario for his observations regarding the reduced quality of plain x-ray reading environments, in the absence of “bright lighting” and use of a magnifying lens.

References
1. Sharp JT: Radiologic assessment as an outcome measure in rheumatoid arthritis. [Review] [59 refs]. Arthritis & Rheumatism 1989, 32:221-9.

2. Gravallese EM, Goldring SR: Cellular mechanisms and the role of cytokines in bone erosions in rheumatoid arthritis. [Review] [92 refs]. Arthritis & Rheumatism 2000, 43:2143-51.

3. Feldmann M, Maini RN: The role of cytokines in the pathogenesis of rheumatoid arthritis. [Review] [40 refs]. Rheumatology (Oxford) 1999, 38 Suppl 2:1-7.

4. Sharp JT: An overview of radiographic analysis of joint damage in rheumatoid arthritis and its use in metaanalysis. [Review] [41 refs]. J Rheumatol 2000, 27:254-60.

5. Kremer JM: Rational use of new and existing disease-modifying agents in rheumatoid arthritis. [Review] [87 refs]. Ann Intern Med 2001, 134:695-706.

6. Genant HK, Jiang Y, Peterfy C, Lu Y, Redei J, Countryman PJ: Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs [see comments]. Arthritis & Rheumatism 1998, 41:1583-90.

7. Pullen M: Introducing the Internet. [Review] [25 refs]. B M J 1995, 311:1422-4.

8. 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 & Rheumatism 1971, 14:706-20.

9. Sharp JT, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, Genant HK, Gofton JP, Goodman N, Larson S: Reproducibility of multiple-observer scoring of radiographic abnormalities in the hands and wrists of patients with rheumatoid arthritis. Arthritis & Rheumatism 1985, 28:16-24.

10. Lin L: A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989, 45:253-268.

11. Richmond BJ, Powers C, Piraino DW, Freed H, Meziane MA, Hale JC, Schluchter MD, Schils J, Gragg LA: Diagnostic efficacy of digitized images vs plain films: a study of the joints of the fingers. AJR American Journal of Roentgenology. 1992, 158:1071-80.

12. Buckwalter KA, Braunstein EM: Digital skeletal radiography. [Review] [62 refs]. AJR American Journal of Roentgenology. 1992, 158:437-41.

13. Bramble JM, Murphy MD: Comparison of digital and conventional musculoskeletal radiography: observer performance study. [Letter; comment]. Radiology 1990, 177:587-9.

14. Jonssson A, Borg A, Hannesson P, Hertlin K, Jonsson K, Sloth M, Peterson H: Film-screen vs. digital radiography in rheumatoid arthritis of the hand. An ROC analysis. Acta Radiologica 1994, 35:311-8.

15. Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, Genant HK, Gofton JP, Goodman N: How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? Arthritis & Rheumatism 1985, 28:1326-35.

16. Rickman A, Ebbe B, Trampisch Hj: How to evaluate measuring methods in the case of non-defined external validity. J Rheumatol 1995, 22:1998-2000.

17. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986, 1:307-10.

18. O’Sullivan MM, Lewis PA, Newcombe RG, Broderick NJ, Robinson DA, Coles EC, Jessop JD: Precision of Larsen grading of radiographs in assessing progression of rheumatoid arthritis in individual patients. Ann Rheum Dis 1990, 49:286-9.

19. Genant HK, Gofton JP, Goodman N, Larsen A: A concordance correlation coefficient to evaluate reproducibility: comment on the article by Genant et al [letter; comment]. Arthritis & Rheumatism 1999, 42:1556-7.

20. Lee J, Koh D, Ong CN: Statistical evaluation of agreement between two methods for measuring a quantitative variable. Computers in Biology & Medicine 1989, 21:11-23.

21. Muller R, Burtner P: A critical discussion of intraclass correlation coefficients. [see comments]. Statistics in Medicine 1994, 13:2465-76.

22. Meeuw AA, Pui M, Cockshott WP, Buchanan WW: Observer differences in detecting erosions in radiographs of rheumatoid arthritis. A comparison of posteroanterior, N&Kx00F8rgaard and Brewerton views. J Rheumatol 1983, 10:216-21.

23. Wassenberg S, Rau R: Problems in evaluating radiographic findings in rheumatoid arthritis using different methods of radiographic scoring: examples of difficult cases and a study design to develop an improved scoring method. J Rheumatol 1995, 22:1990-7.

24. Larsen A, Thoen J: Hand radiography of 200 patients with rheumatoid arthritis repeated after an interval of one year. Scand J Rheumatol 1987, 16:395-401.

25. Copeland L, Kay R: Data and image compression. Computerworld 2000, 34:

26. Hill J: JPEG for the millennium. Presentations 2000, 14:17-8.

27. Menduno M: Prognosis: wired. Why Internet technology is the next medical breakthrough. Hospitals & Health Networks 1932, 72:28-30.

28. Winkler MA, Flanagan A, Chi-Lum B, White J, Andrews K, Kennett RL, DeAngelis CD, Musacchio RA: Guidelines for medical and health information sites on the internet: principles governing AMA web sites. American Medical Association. JAMA 2000, 283:1600-6.

29. Jadad AR: Promoting partnerships: challenges for the internet age. [see comments]. [Review] [42 refs]. B M J 1999, 319:761-4.

30. Edworthy SM: World wide web: opportunities, challenges, and threats. Lupus 1999, 8:596-605.

31. Parente ST: Beyond the hype: a taxonomy of e-health business models. Health Affairs 2000, 19:89-102.

32. Lipsky PE, van der Heide DM, St Clair EW, Forst DE, Breedveld FC, Kalden JR, Smolen JS, Weisman M, Emery P, Feldmann M, Harrigan GB, Maini RN: Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis, Concomitant Therapy SG: Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Inhibitor Treatment Study Group. [see comments]. N Engl J Med 2000, 343:1594-602.

33. Sharp JT, Strand V, Leung H, Hurley F, Low-Efriedrich I: Treatment with leflunomide slows radiographic progression of rheumatoid arthritis: results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group. [erratum appears in Arthritis Rheum 2000 Jun;43(6):1345]. Arthritis & Rheumatism 2000, 43:495-505.

34. Strand V, Cohen S, Schift M, Weaver A, Fleischmann R, Cannon G, Fox R, Moreland L, Olsen N, Forst D, Caldwell J, Kaine J, Sharp J, Hur-
ley F, Loew-Friedrich I: Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Leflunomide Rheumatoid Arthritis Investigators Group. *Arch Intern Med* 1999, 159:2542-50

**Pre-publication history**
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2474/3/13/prepub