Does a Joint Count Calibration Exercise Make a Difference? Implications for Clinical Trials and Training

To the Editor:

Formal joint counts are an integral part of disease assessment in rheumatology. They form the basis of disease responder indices including Disease Activity Score (DAS), American College of Rheumatology responder criteria, Clinical Disease Activity Index, and Simplified Disease Activity Index. Wide variability among examiners may therefore have a significant effect on outcomes in clinical trials[1]. In an attempt to reduce examiner variability, the European League Against Rheumatism developed standardized joint assessment criteria for the presence or absence of joint swelling and/or tenderness[2]. Formal training may improve the degree of variability between examiners.

A joint count calibration exercise was organized as part of the New Zealand Treat-to-Target initiative. Twenty-eight tender and swollen joint counts as described by Fuchs, *et al*[3] were undertaken on 5 separate patients with rheumatoid arthritis (RA) by examiners working in groups of 4 or 5. Joint counts on the first 2 patients were done by individual examiners without comparing their findings with those of their colleagues. After examining the second patient, a 10-min calibration exercise was undertaken. This consisted of the groups of examiners reexamining their first 2 patients, observing each other's technique, and discussing their findings. Repeated examination occurred until there was consensus between the examiners on the swollen and tender joints for the individual patient. The remaining 3 patients were then examined with joint counts being kept confidential from the other examiners. DAS28-C-reactive protein[4] was calculated based on the joint counts from each examiner.

A total of 23 individuals participated: 12 (52.2%) rheumatologists, 5 (21.7%) trainees, 5 (21.7%) other health professionals, and 1 (4.3%) patient with RA. The mean number of years of experience (excluding the patient) was 13.7 (range 1–41).

The mean tender joint count (TJC), swollen joint count (SJC), and DAS28 for each patient from all rounds are shown in Table 1. There was wide variability among individual examiners as evidenced by a large standard deviation (SD; Figure 1). There was a significant reduction in the variability of SJC ($p = 0.002$) and DAS28 ($p = 0.038$) as evidenced by a decline in the SD between examiners after the calibration exercise (rounds 3–5) as compared to before the calibration exercise (rounds 1–2). There was a trend toward reduction in variability of TJC after the calibration ($p = 0.063$; Figure 1). Despite the reduction in variability, there was still significant variability between examiners in the final round.

Tender and swollen joint counts are regarded as the most effective ways of assessing disease activity by rheumatologists[5]. Current strategies for treatment of RA as well as funding of biological therapies are guided by validated disease activity assessments such as DAS28, which heavily weight TJC and SJC. As demonstrated in this and other studies, there is marked variation in joint counts between examiners. While calibration exercises may reduce interexaminer variability, they do not eliminate it.

Variation in joint scores in clinical trials protocols is minimized by undertaking calibration exercises at clinical trial investigator meetings and by using the same trained observer at each visit[6,7]. However, in day-to-day clinical practice, variation in examination technique has the potential to produce observer-dependent discrepancies in the provision of access to biologic therapies. Resources such as that recently produced by Starz, *et al* outlining the mechanics of joint examination may also assist in reducing variability[8]. Our training and calibration exercise was brief and informal, yet it was followed by a reduction in the SD. Joint count calibration exercises that teach a standardized examination technique could potentially minimize this variation. There is a need for further studies to validate specific joint examination protocols and to determine the value of formal training. A new video demonstrating the standard techniques for examining the joints in the DAS28 has been developed (Internet: http://youtu.be/RnvsbD6NKoc).

Table 1. Mean tender joint count (TJC), swollen joint count (SJC), and 28-joint Disease Activity Score (DAS28) for each of the 5 patients, including data from all 23 examiners.

Patient	TJC	SJC	DAS28
1	19.38	10.86	5.93
2	12.17	8.35	4.58
3	14.21	11.68	5.69
4	19.94	19.98	6.75
5	1.21	3.03	3.17

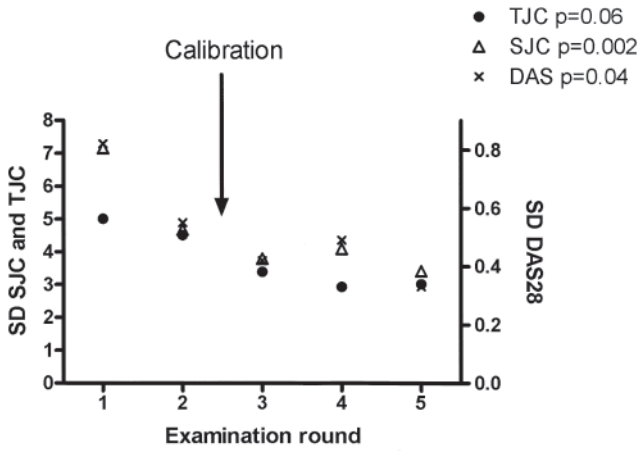

Figure 1. After calibration, variability between examiners declined in the mean tender joint count (TJC), swollen joint count (SJC), and the 28-joint Disease Activity Score (DAS28).

LISA K. STAMP, FRACP, PhD, Associate Professor and Rheumatologist, University of Otago, Christchurch; ANDREW HARRISON, FRACP, PhD, Associate Professor and Rheumatologist, University of Otago, Wellington; CHRISTOPHER FRAMPTON, PhD, Biostatistician, University of Otago, Christchurch; MICHAEL M. CORKILL, FRACP, MBA, Rheumatologist, Waitemata District Health Board, Auckland, New Zealand.

ACKNOWLEDGMENT

Abbott NZ Inc. for supporting the Treat-to-Target NZ meeting.

REFERENCES

1. Pincus T. Limitations of a quantitative swollen and tender joint count to assess and monitor patients with rheumatoid arthritis. Bull NYU Hosp Joint Dis 2008;66:216-23.
2. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. Ann Rheum Dis 2008;67:1360-4.
3. Fuchs H, Brooks R, Callahan L, Pincus T. A simplified twenty-eight joint quantitative articular index in rheumatoid arthritis. Arthritis Rheum 1989;32:531-7.
4. van Gestel AM, Haagsma CJ, van Riel PLCM. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. Arthritis Rheum 1998;41:1845-50.
5. Wolfe F, Pincus T, Thompson AK, Doyle J. The assessment of rheumatoid arthritis and the acceptability of self-report questionnaires in clinical practice. Arthritis Rheum 2003;49:59-63.
6. Bellamy N, Anastassiades T, Buchanan W, Davis P, Lee P, McCain G, et al. Rheumatoid arthritis antirheumatic drug trials. I. Effects of standardization procedures on observer dependent outcome

measures. J Rheumatol 1991;18:1893-900.
7. Scott D, Choy E, Greeves A, Isenberg D, Kassinor D, Rankin E, et al. Standardising joint assessment in rheumatoid arthritis. Clin Rheumatol 1996;15:579-82.
8. Starz T, Moreland L, Levesque M. Quantitative joint assessment to improve rheumatoid arthritis outcomes. J Musculoskel Med 2011;28:79-84.

J Rheumatol 2012;39;4; doi:10.3899/jrheum.111224