Letters to the editor

5 Homan S. A histopathological and immunohistological study of the layers of the lesion in the pericardial sinuses. *Radiographic assessment of the knee joint in OA*.

Sir: The article of Cooper *et al*., on reproducibility of radiological assessment of the knee joint in osteoarthritis, implies that statistics can be considered as one of the various forms of lies, the failure to achieve statistical significance may be misleading. The blinded nature of the analysis may not overcome the apparent fundamental statistical premise violation, related to non-random selection of who reviewed which films. β Error (related to limited sample size) and phenomenon infrequency (making use of x² inappropriate) further compound interpretation.

The reproducibility is a measure of the ability in order of appearance and of clustering of findings and use of the peculiar ill-defined phrase ‘abnormality of bony contour’ raises an intriguing possibility, as only osteophytes have actually been demonstrated as a reliable radiological sign of osteoarthritis. Do the different appearance order and clustering indicate that perhaps we have been grouping phenomenon with osteoarthritis that are at best tangential? Are we dealing with the traditional overlapping circles of phenomenon which so punctuate the epidemiological chapters in McCarty’s text? (Cooper C, Cushnaghan J, Kirwan J R, Dieppe P A, McAlindon T, McCro F. Radiographic assessment of the knee joint in osteoarthritis. *Ann Rheum Dis* 1992; 51: 80–2.)


AUTHORS’ REPLY: We were interested to read Dr Rothschild’s comments on our study of radiological assessment of the knee joint in osteoarthritis. We agree that the reproducibility of these assessments is a subject integral to radiological research and practice. Although we cannot be expected to comment on the American College of Rheumatology study which is still in press, we are uncertain that the reproducibility with which osteophytes was assigned in our study was markedly different from that reported using a similar scale by Altman *et al*. The lowest intraclass correlation between observers for spurs in that study was 0·56, as compared with a x² statistic of 0·58 in our study. Indeed, we emphasised in our paper that assessment of joint space narrowing and osteophyte in the tibia/femoral compartment performed rather better than did sclerosis and cyst formation. Dr Rothschild is correct that one of the reasons why reproducibility in our study might have been generally worse than in others relates to the observers chosen for the study. One of our specific aims, however, was to examine the assessment of these features in routine practice rather than to ascertain the optimum performances which could be attained by highly trained observers. Indeed, we highlighted the potential role of training to reduce observer variation as an area for further research. We were unable to understand Dr Rothschild’s statistical comments. If he is lamenting the widespread use of statistical inference in biomedical research, space does not permit us to do more than profoundly disagree with him. The study was designed so that all the observers viewed all the radiographs unavailable to their own or colleagues’ previous assessments. The statistical analysis of observer variation is controversial, but we chose to use a x² statistic (rather than the x² statistic as your correspondent implies). We took pains to emphasise the pitfalls of this analytical technique, but ultimately decided that a summary index of the concordance for various measurements justified its use. It remains one of the most widely used tools to assess the reproducibility of categorical scales. (Cooper C, Cushnaghan J, Kirwan J R, Dieppe P A, McAlindon T, McCro F. Radiographic assessment of the knee joint in osteoarthritis. *Ann Rheum Dis* 1992; 51: 80–2.)
