

## Reliability of Radiographic Assessment of Sacroiliac Joints in Patients with Suspected Early Spondyloarthritis: Methodological Issue

To the Editor:

I was interested to read the paper by Christiansen, *et al* published in the January 2017 issue of *The Journal of Rheumatology*<sup>1</sup>. The purpose of the authors was to determine the reproducibility of sacroiliac joint (SIJ) radiographs among readers with varying levels of experience, and to identify potential drivers of disagreement in classification among 5 predefined radiographic lesion types<sup>1</sup>. A hundred and four patients with low back pain  $\geq 3$  months who met the Assessment of Spondyloarthritis international Society definition for a positive SIJ magnetic resonance image were recruited. Seven blinded readers (2 musculoskeletal radiologists, 5 rheumatologists) classified pelvic radiographs according to the modified New York criteria (mNY) and recorded presence/absence of 5 lesion types in both SIJ: erosion, sclerosis, ankylosis, joint space widening, and joint space narrowing. Reproducibility of mNY classification among 21 reader pairs was assessed and potential drivers of disagreement were identified among 5 lesion types<sup>1</sup>. Based on their results, mean  $\kappa$  values (percent concordance) were 0.39 (84.1%) for mNY classification over 21 reader pairs, 0.46 (79.8%) between 2 musculoskeletal radiologists, and 0.55 (86.5%) and 0.36 (77.9%) between the most experienced rheumatologist and the 2 radiologists, respectively<sup>1</sup>.

It is crucial to know that using  $\kappa$  is one of the common mistakes in reliability analysis for qualitative outcomes. Two important weaknesses of  $\kappa$  statistics to assess agreement are as follows: (1) it depends on the prevalence in each category, which means it can be possible to have different  $\kappa$  values having the same percentage for both concordant and discordant cells; (2)  $\kappa$  value also depends on the number of categories<sup>2,3,4,5,6,7,8,9,10</sup>. Moreover, in reliability analysis, an individual-based approach should be applied instead of global average. Therefore, using mean  $\kappa$  values is another methodological issue in reliability analysis<sup>2,3,4,5,6,7,8,9,10</sup>. Finally, appropriate correction should be applied in multiple comparison analysis. They concluded that reproducibility of radiographic SIJ classification in a spondyloarthritis (SpA) inception cohort was only fair to moderate among 7 readers with varying levels of experience, questioning the applicability of mNY in early SpA<sup>1</sup>. Such conclusion should be supported by the above-mentioned statistical and methodological issues. Otherwise, misdiagnosis and mismanagement of the patients may occur.

SIAMAK SABOUR, MD, MSc, DSc, PhD, Postdoc, Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, and Department of Clinical Epidemiology, Shahid

Beheshti University of Medical Sciences, Tehran, I.R. Iran. Address correspondence to Dr. S. Sabour, Safety Promotion and Injury Prevention Research Center, Department of Clinical Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran. E-mail: s.sabour@sbmu.ac.ir

## REFERENCES

1. Christiansen AA, Hendricks O, Kuettel D, Hørslev-Petersen K, Jurik AG, Nielsen S, et al. Limited reliability of radiographic assessment of sacroiliac joints in patients with suspected early spondyloarthritis. *J Rheumatol* 2017;44:70-7.
2. Szklo M, Nieto J. *Epidemiology beyond the basics*, 2nd ed. Manhattan: Jones and Bartlett Publisher; 2007.
3. Sabour S. Reproducibility of circulating endothelial cell enumeration and activation: a methodological issue. *Biomark Med* 2016;10:1215.
4. Sabour S. Reliability of lumbar movement dysfunction tests for chronic low back pain patients; methodological concerns to avoid misinterpretation. *Man Ther* 2016;26:e4.
5. Sabour S. Reproducibility of semi-automatic coronary plaque quantification in coronary CT angiography with sub-mSv radiation dose; common mistakes. *J Cardiovasc Comput Tomogr* 2016;10:e21-2.
6. Sabour S. Reliability assurance of EML4-ALK rearrangement detection in non-small cell lung cancer: a methodological and statistical issue. *J Thorac Oncol* 2016;11:e92-3.
7. Sabour S. Adherence to guidelines strongly improves reproducibility of brachial artery flow-mediated dilation. Common mistakes and methodological issue. *Atherosclerosis* 2016; 251:490-1.
8. Sabour S. Reliability of a new modified tear breakup time method: methodological and statistical issues. *Graefes Arch Clin Exp Ophthalmol* 2016;254:595-6.
9. Sabour S. Re: does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder? *Clin Radiol* 2015;70:333-4.
10. Sabour S. The reliability of routine clinical post-processing software in assessing potential diffusion-weighted MRI "biomarkers" in brain metastases, common mistake. *Magn Reson Imaging* 2014;32:1162.

*J Rheumatol* 2017;44:6; doi:10.3899/jrheum.170014