

Dr. Christiansen, *et al*, reply

To the Editor:

We thank Dr. Sabour¹ and Dr. Rothschild² for their interest in our manuscript³.

We acknowledge that κ statistics depend on the prevalence of the variable under investigation and we have made this transparent. This limitation becomes relevant when comparing results across multiple studies. However, we use κ to assess which of several variables similarly assessed on the same patients provide sufficient agreement, i.e., we primarily used κ to order lesion types. This implies that the actual value of κ is of minor importance and the above limitation does not alter the conclusion of our paper. Moreover, interpretation of κ statistics should be made after considering the characteristics of the data. We presented κ values along with positive and negative percent agreements, thus allowing readers to make a fully informed judgement. Others have suggested to examine the prevalence and bias indexes and to adjust κ accordingly, resulting in an adjusted coefficient referred to as PABAK (prevalence-adjusted bias-adjusted kappa)⁴. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive κ scores in cases of prevalence issues and negative κ scores in cases of bias issues, leading to the conclusion that κ values should remain unadjusted and be reported alongside the proportional agreement⁵.

Our article focuses on providing statistical inference by giving CI for the quantities of interest. The only instance where we claim a result to be “statistically significant” if a p value is ≤ 0.05 is for the generalized linear mixed model. However, our interpretation does not rely on this perceived “statistical significance,” but on the interpretation of the estimated OR of 13.5 with 95% CI ranging from 9.1–20.1, i.e., we prioritize assessment of clinical relevance over statistical significance. Even if we had Bonferroni-corrected the family-wise error rate for performing 500 tests (which is a number of tests far beyond the number of variables looked at in our paper), i.e., compared p values to $0.05 \div 500 = 0.0001$, the conclusion with regard to which variables were “significant” in Table 3 of our article would remain the same.

Regarding concepts and differences between significance tests, CI, and hypothesis tests, we refer to Blume and Peipert⁶ or Sterne and Smith⁷. We believe our conclusions can be drawn from the data provided and do not agree with Dr. Sabour¹ that our analysis, if properly interpreted, may lead to mismanagement and misdiagnosis of patients.

We agree with Dr. Rothschild² that recognition of spondyloarthritis remains a complex process of composite deduction based on complementary information obtained from clinical, laboratory, and imaging assessment⁸. Our study is in support of previous reports that radiography of the sacroiliac

joints has a limited involvement in assessment of patients with back pain clinically suspected to have early spondyloarthritis. Whether early recognition of this multifaceted disorder might be enhanced by expanded clinical evaluation, considering also response to treatment, remains to be shown in the future.

ALICE ASHOURI CHRISTIANSEN, MD, King Christian 10th Hospital for Rheumatic Diseases, Gråsten, and Hospital of Southern Jutland, Jutland, and Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; KASPAR RUFIBACH, PhD, Rufibach rePROstat, Biostatistical Consulting and Training, Basel, Switzerland; ULRICH WEBER, MD, King Christian 10th Hospital for Rheumatic Diseases, Gråsten, and Hospital of Southern Jutland, Jutland, and Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. Dr. Rufibach is founder and owner of Rufibach rePROstat and is an employee of F. Hoffmann-La Roche, Department of Biostatistics, Basel, Switzerland. Address correspondence to Dr. A.A. Christiansen, Department of Research, King Christian 10th Hospital for Rheumatic Diseases, Toldbodgade 3, 6300 Gråsten, Denmark.
E-mail: achristiansen@gigtforeningen.dk

REFERENCES

1. Sabour S. Reliability of radiographic assessment of sacroiliac joints in patients with suspected early spondyloarthritis: methodological issue. *J Rheumatol* 2017;44:957.
2. Rothschild B. Back to basics: clinical versus radiologic recognition of spondyloarthropathy. *J Rheumatol* 2017;44:957-8.
3. 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.
4. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423-9.
5. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol* 2000;53:499-503.
6. Blume J, Peipert JF. What your statistician never told you about p -values. *J Am Assoc Gynecol Laparosc* 2003;10:439-44.
7. Sterne JA, Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ* 2001;322:226-31.
8. Weber U, Jurik AG, Lambert RG, Maksymowych WP. Imaging in spondyloarthritis: controversies in recognition of early disease. *Curr Rheumatol Rep* 2016;18:58.

J Rheumatol 2017;44:6; doi:10.3899/jrheum.170315