Dr. Christiansen, et al, reply

To the Editor:

We thank Dr. Sabour and Dr. Rothschild for their interest in our manuscript.

We acknowledge that $\kappa$ 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 $\kappa$ to assess which of several variables similarly assessed on the same patients provide sufficient agreement, i.e., we primarily used $\kappa$ to order lesion types. This implies that the actual value of $\kappa$ is of minor importance and the above limitation does not alter the conclusion of our paper. Moreover, interpretation of $\kappa$ statistics should be made after considering the characteristics of the data. We presented $\kappa$ 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 $\kappa$ accordingly, resulting in an adjusted coefficient referred to as PABAK (prevalence-adjusted bias-adjusted kappa). However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive $\kappa$ scores in cases of prevalence issues and negative $\kappa$ scores in cases of bias issues, leading to the conclusion that $\kappa$ values should remain unadjusted and be reported alongside the proportional agreement.

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 $\leq 0.05$ is for the generalized linear mixed 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 Peipert or Sterne and Smith. We believe our conclusions can be drawn from the data provided and do not agree with Dr. Sabour that our analysis, if properly interpreted, may lead to mismanagement and misdiagnosis of patients.

We agree with Dr. Rothschild that recognition of spondyloarthritis remains a complex process of composite deduction based on complementary information obtained from clinical, laboratory, and imaging assessment. 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@gigforeningen.dk

REFERENCES