Dr. Christiansen, et al, reply

To the Editor:

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

We acknowledge that k 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 k to assess which of several variables similarly assessed on the same patients provide sufficient agreement, i.e., we primarily used k to order lesion types. This implies that the actual value of k is of minor importance and the above limitation does not alter the conclusion of our paper. Moreover, interpretation of k statistics should be made after considering the characteristics of the data. We presented k 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 k 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 k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k 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,” but on the interpretation of the estimated OR of 13.5 with model. However, our interpretation does not rely on this perceived “statistically significant” if a p value is ≤ 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 limitations becomes relevant when comparing results across multiple studies. However, this has resulted in criticism because it has been shown that the PABAK adjustment produces inflated positive k scores in cases of prevalence issues and negative k scores in cases of bias issues, leading to the conclusion that k values should remain unadjusted and be reported alongside the proportional agreement5.