Instructions for peer evaluation in Stage 7

Thank you very much for reviewing the final report assigned to you. Below you will find a series of guiding points to help your assessment. These points are based on the feedback given to the initial analytical approaches.

Please carefully examine the final report. We would like to know to what extent each point is (still) an issue in the described approach, or whether it has been (fully) addressed. If you need verifying information from the authors, please get in touch.

Please note that the validity of the inclusion of covariates will be assessed separately.

Dependent Variable:

Point 1. In the dataset, that the dependent variable (red cards given) depends on the number of games played. It has been suggested that a remedy is to transform the data (for instance to be dyadic so that each line represents a single referee player interaction). Alternatively, it has been suggested that 'Games' should be used as an offset in a regression (rather than a predictor) so that observations are weighted depending on the number of games in each player/referee dyad.

Point 2. The value of red cards in the dataset is either 0, 1 or 2 and there are many cases in which no red card was given and two red cards was very few. The dependent variable thus cannot be assumed to be linear.

Point 3. Red cards are dependent on the number of games played. If red cards per game was specified as a proportion, this represents a ratio and a linear model would also not be appropriate. Further, transforming red cards into a proportion has limitations in that it equates getting 0 red cards in only 1 or 2 games with a referee and getting 0 red cards in 20 games with the referee.

Point 4. Many players received 0 red cards from a referee. Therefore the dependent variable often takes the value of 0. Was a model chosen that addresses this issue? It has been suggested that a negative binominal regression is more appropriate than a Poisson regression, because of the high number of zeros in the distribution (and the associated low mean and high variance in this variable).

Point 5. YellowRed and redCards are qualitatively different: the yellowRed depends on a previous yellow card and is given to a less serious foul, while the pure red can be given at any moment during the game. Pooling them assume they come from the same or similar processes, which is not the case.

This point is an issue in the described approach
Strongly Disagree - Somewhat Disagree - Neither Agree nor Disagree - Agree Somewhat - Strongly Agree

Model:

Point 1: The dataset is based on repeated observations of referees and players. Many regression analyses such as OLS – classical linear regression models and also standard logistic regression requires each observation to be independent. It is an issue if the analytical technique doesn't treat the data as independent, instead of nested, multi-level, and thus accounting for repeated observations of referees and players.

This point is an issue in the described approach
Strongly Disagree - Somewhat Disagree - Neither Agree nor Disagree - Agree Somewhat - Strongly Agree

Exclusions & Transformations / Missing Data

Point 1. Have cases been unnecessarily been excluded, potentially leading to a loss in information? For instance, dichotomizing skintone (and exluding "neutrals"); excluding cases where the raters disagree; excluding dyads or players for whom no red card was given.

This point is an issue in the described approach
Strongly Disagree - Somewhat Disagree - Neither Agree nor Disagree - Agree Somewhat - Strongly Agree

Overall, how convinced are you that the presented approach successfully addressed most concerns and misspecifications.

Very convinced *Rather unconvinced* *Neither convinced nor unconvinced* *Rather unconvinced* *Very unconvinced*