Multilevel logistic regression does not show support for referee bias driven by explicit or implicit attitudes

Authors: Š. Bahník[1]*

Affiliations

[1]University of Würzburg, Department of Psychology.

*Correspondence to: bahniks@seznam.cz

Abstract

While dark skin toned players were more likely to get a red card, it cannot be concluded from the data that this is a proof of referee bias. An influence of explicit or implicit attitudes on size of the skin tone effect would be indicative of such bias, but neither explicit nor implicit attitudes influenced the effect. However, the data were not well suited for finding influence of attitudes, so they do not provide good evidence for absence of referee bias as well.

One Sentence Summary

Dark skin toned players were more likely to get a red card, but the effect of skin tone did not seem to be dependent on explicit or implicit attitudes.

Results

All analyses were conducted using multilevel modelling (Bates, D., Maechler, M., & Bolker, B., 2012; Gelman & Hill, 2007). Before analysis, the data were transformed to a format where a single match was taken as one observation. The format was obtained by decomposing each row from the original data to multiple rows where the number of rows obtained was equal to the number of games for a given player-referee dyad. Therefore, each game for a given player-referee dyad served as one observation. A red card variable was set to 1 in *x* of these games, where *x* was the number of red cards for a given player-referee dyad. For the remaining games, the variable was set to 0.

The initial model included weight, height, skin tone, age, position and average number of goals in a game as predictors. Players and referees were included in the model as random factors. Positions were recoded such that they did not take into account whether a player plays on a right or left side (e.g. right winger and left winger were treated as the same position). The skin tone predictor was the average of ratings of the two raters. For answering question 2, mean implicit or explicit bias for referee country as well as an interaction between the mean bias and skin tone were included in the analysis. Note that mean implicit and explicit bias were standardized for countries included in analysis (see below for description of exclusion criteria). However, means and standard deviations of these predictors across observations were $M = 0.38$, $SD = 0.27$ for implicit bias and $M = 0.24$, $SD = 0.30$ for explicit bias. Reported odds ratios therefore correspond to increase of one standard deviation of bias between countries and to increase of about 3.5 standard deviations of bias between observations.

From each model estimation, all observations that did not contain full data for all predictors were excluded. Furthermore, in answering question 2, observations for referees from countries that did not have enough data for estimation of a given attitude measure were excluded. In particular, countries were excluded if they had standard errors of an attitude measure within the country higher than ⅕ of standard deviation of mean values between countries. The ⅕ of standard deviation threshold was chosen arbitrarily and meant that referees from countries that had a measure of attitudes computed from less than approximately 25 people were removed from analysis. This exclusion was done so that unreliable values of attitude measures do not influence the results.

Initial Approach

The initial approach used both straight red cards and red cards obtained after two yellow cards in a summary dependent variable. Based on Akaike information criterion, age and average number of goals were discarded from the initial full model. The remaining predictors were used for all further analyses. The initial approach used multilevel linear regression. The initial approach led to similar results as the final approach and the latter was better suited for the present data, so only its results are described in more detail.

Final Approach

The final approach was the same as initial approach, but included only straight red cards as a dependent variable. Furthermore, multilevel logistic regression was used since it is better suited for binary dependent variable. Otherwise all models were kept same as in the initial approach.

The results for question 1 revealed a significant effect of skin tone, $z = 2.88$, $p = .004$, $OR = 1.38$, 95% CI = [1.11, 1.72] showing that players with darker skin tone were more likely to get a red card.

Analysis for question 2a showed that skin tone and implicit bias ratings were both associated positively with probability of getting a red card, $z = 1.68$, $p = .09$, $OR = 1.35$, 95% CI = [0.95, 1.93] for skin tone, and $z = 1.87$, $p = .06$, $OR = 1.40$, 95% CI = [0.98, 1.98] for implicit bias ratings. Most importantly, the interaction between the two terms was not significant and did not indicate that the effect of skin tone on probability of getting a red card is dependent on implicit bias, $z = 0.28$, $p = .78$, *ratio of OR* = 1.11, 95% CI = [0.54, 2.27].

A similar pattern of results was observed for explicit bias ratings. Both skin tone and explicit bias ratings were positively associated with probability of getting a red card, $z = 2.49$, $p = .01$, $OR = 1.39$, 95% CI = [1.07, 1.81] for skin tone, and $z = 1.75$, $p = .08$, $OR = 1.28$, 95% CI = [0.97, 1.69] for explicit bias ratings. However, the interaction was again not significant, $z = -0.02$, $p = .99$, *ratio of OR* = 0.99, 95% CI = [0.56, 1.77].

Conclusion

The results showed that dark skin tone players were more likely to get a red card. However, it is not clear whether this effect is driven by dark skin tone itself or by some other characteristic associated with it. It is for example possible that players with darker skin tone differ in their playing style or physical characteristics. In fact, such a finding was obtained by Price and Wolfers (2010) for basketball players. While some of the possible confounding variables were included in modelling, many other could not have been accounted for. For this reason, it is important to look for other support for possible bias. If it was shown that the effect of skin tone on probability of getting a red card is higher for referees from countries with more negative racial attitudes, it could be taken as indication of racial bias. However, no such effect was found in the present analysis. It is possible to see that the analysis does not yield precise estimates for the effect of attitudes on the effect of skin tone on probability of getting a red card. Therefore, a possibility of bias cannot be dismissed. Future study with information about attitudes or skin tone of referees might be better suited for answering this important question.

Data and Output

https://osf.io/25gf9/

References and Notes

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. The Quarterly Journal of Economics, 125, 1859-1887.