Dark-skinned players are more likely to be rejected from a game in world's most popular sport

Seth M. Spain[1]*, Kristin Lee Sotak[1]

[1]Center for Leadership Studies and School of Management, Binghamton University, State University of New York, PO Box 6000, Binghamton, NY 13902-6000, USA.

*sspain@binghamton.edu

Abstract

Previous research in the NBA has shown a same-race bias in fouls, where white referees were more likely to call fouls on black players and black referees were more likely to call fouls on white players. Awareness brought to this issue, however, helped eliminate this bias. The purpose of the current research is to investigate similar relationships in soccer. We used a multilevel logistic regression model to analyze data on male soccer players in top European divisions. Results showed similar racial bias as in the NBA, where dark skinned players were more likely to receive red cards than light skinned players. Moreover, referees from countries with higher mean implicit association test scores were more likely to give red cards, but not necessarily to those with darker skin. We hope these results help bring awareness and change to the racial bias that exists in the world's most popular sport.

One Sentence Summary

Results show that darker skinned players are more likely to receive a red card, and referees from countries with higher mean implicit association test score are more likely to give red cards; however, they do not seem to be particularly more likely to punish darker toned players than other referees, on average.

Results

Initial Approach

Initially, we started with the simplest model and worked from there (Gelman & Shalizi, 2013). Due to the fact that the data were count data, not normally distributed, and that the dependent variable technically had no upper limit, we started with a Poisson model (Gelman & Hill, 2007). From here, we incorporated the data design and nesting structure into our model - the data were at the player-referee dyad, where multiple players were nested under the same referees, which violated the assumption of independence (Bliese & Hanges, 2004). Moreover, there were multiple player-referee dyads for each player, which also violated the assumption of independence. Failing to model non-independent data can influence standard errors, which can influence inferences and increase Type I errors, so we therefore incorporated a cross-classified model (Gooty & Yammarino, 2011) into our analysis. That is, neither player nor referee is strictly nested within the other classifying variable. We modeled the data using a generalized linear mixed effect model with a logarithmic link function, the canonical link for a Poisson-distributed random variable. The varying intercepts on the referees and the players were modeled as a function of the referee's country's mean Implicit Association Test (mean_IAT) score, and the players skin color, as the average of two independent raters (mean_skinCol) and the player's position, respectively.

In our model, we allowed the intercepts for both the player and the referee in each dyad to vary, and used country skin-tone prejudices of referees as a predictor of the s in level 2 of our model. Level 1 variables included skin-tone and control variables. To build off of existing literature, we incorporated similar control variables as Price & Wolfers (2010) and that also made theoretical sense in this research setting (position, goals). We used the glmer function (see attached R code) in the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) in R (R Development Core Team, 2012). Together, our model is as follows:

Level-1: $Y_{ijk} = \alpha_{0j} + \beta_{0k} + \beta_1(\text{goals}_{ijk})$

Level-2: $\alpha_{0j} = a_{00} + a_{01}(\text{mean\_skinCol}_j) + a_{02}(\text{position}_j) + u_{aj}$

$\beta_{0k} = b_{00} + b_{01}(\text{mean\_IAT}_k) + u_{bk}$

Where, Y_{ijk} is the dependent variable (number of red cards in a player-referee dyad), α_{0j} is the level-1 varying intercept for the player, β_{0k} is the level-1 varying intercept for referees, β_1 is the non-varying slope coefficient for goals scored by the player when observed by the referee. In the second level, a_{00} is the mean player intercept, and a_{01} and a_{02} are the coefficients on the player-level predictors of the varying intercepts. In the second equation for level-2, b_{00} is the mean intercept for referees, and b_{01} represents the coefficient on the referee-level predictor of number of red cards, respectively. Last, u_{aj} and u_{bk} represent the residual terms for each level. There is no residual term explicitly modeled for the level-1 equation, because the Poisson distribution's variance is equal to its mean, and therefore the variability is implicit in the mean function above.

Final Approach

We used similar analyses in our final approach as in our original. However, we changed from a Poisson model to a multilevel (hierarchical) logistic model (Gelman & Hill, 2007), as it made more sense to us to model the likelihood of getting a red card rather than modeling the number of red cards players receive. Considering the nested structure of the data, we continued with the multilevel model to account for the nonindependence of cases. Level 1 variables included skin-tone and player position. We let the intercept and slope on skin tone vary by referee, where country skin-tone prejudice of referees (as indexed by the referee's country's mean implicit association test, IAT) was a level 2 predictor for the intercept and slope. and an interaction term (skin color by referee's country skin-tone prejudice). In alternative models, we included height, weight, and goals as additional controls; however, they were not significant and they did not influence point estimates of variables of interest (player skin tone, country skin-tone prejudice, and the interaction between these two variables). We also ran a model with another variable that measures a referee's country skin-tone prejudice, mean explicit white preference scores. Again, results did not differ from those with mean IAT scores. We used the glmer function (see attached R code) in the lme4 package (Bates, et al., 2013) in R (R Development Core Team, 2012). A second analysis followed the same structure, but used the referee's country's mean explicit white preferences as a level-2 predictor.

Comment [SMS1]: Note, 11/18/2015. By this we mean the outcome was of the form c(redCards, nGames) so that we were modeling the number of redCards the particular player received from the particular referee in the total number of games the player played under the referee. This seemed more natural to us than using games as a control variable or offset, since players are ejected upon receiving a red card. We noted this in the survey-based report, but it is not clear here.

Results are shown in Tables 1 and 2. Parameter estimates are reported as odds ratio, and confidence intervals for the parameters of interest are given. Results using meanIAT (Table 1) show that higher player skin color (darker skin-tone) increases the likelihood of receiving a red card (Skin Col = 1.42, 95% CI [1.19, 1.71]). Moreover, a referee from a country with a higher mean IAT score is also more likely to give a red card (MeanIAT = 1.17, 95% CI [1.04, 1.31]). However, the interaction between player skin color and referees' country mean IAT was not found to be influential (Skin Col x meanIAT = 0.96, 95% CI [0.77, 1.18]).

Results using mean explicit white preference scores are similar to those using meanIAT (Table 2). First, darker skinned players are more likely to receive a red card (Skin Col = 1.43, 95% CI [1.19, 1.72]). Additionally, the interaction between player skin color and a referee's country prejudice was also not found to be influential (Skin Col x mean Explicit = 1.03, 95% CI [0.84, 1.27]). Unlike results using meanIAT, the main effect of mean explicit prejudice was not found to be influential on likelihood of giving a red card card (mean Exlicit = 1.10, 95% CI [0.99, 1.23]), but it was close to being influential, and should therefore be interpreted with caution.

Note that results from our initial approach were similar in that soccer referees were found to be more likely to give red cards to dark skin toned players than to light skin toned players (skin color odds ratio, e^{a01} = .077, p < .007). Moreover, country skin-tone prejudice for a referee did not appear to influence how likely a referee is to give a red card (meanIAT odds ratio, e^{b01}= 1.540, p = .240).

Conclusion

Results from both our initial and final analyses converge on one particular finding: darker skinned players are more likely to receive a red card than lighter skinned players. These results mirror initial findings found in other professional sports leagues, such as basketball and baseball (Price & Wolfers, 2010; Parsons, Sulaeman, Yates, & Hamermesh, 2011). However, awareness brought to these biases in referee decisions was so strong that these biases were eliminated and did not appear in subsequent analyses (Pope, Price, and Wolfers, 2013). Therefore, results from these analyses of soccer referee bias will hopefully influence FIFAs (international government body for soccer) approach to dealing with referee racial bias on the pitch.

Though results show that referees are more likely to give dark skinned players a red card than light skinned players, does the country from which a referee comes influence his decision to give a red card? Results partially support this idea: Referees from countries with higher mean implicit association test score were more likely to give red cards in one of the analyses. *However*, these referees, from such countries, do *not* appear to be more likely to give darker skinned players red cards, compared to other referees (the 95% confidence interval for the interaction included zero). The implication here is that FIFA should *not* use a country from which a referee comes as a selection tool for eliminating referees that are biased. These referees may be more likely to card players, but not necessarily in a biased manner.

TABLE 1
Results from multilevel logistic model, using mean implicit prejudice (meanIAT)

	Estimate	LowerCI	UpperCI*
Skin Col	1.42	1.19	1.71
Mean IAT	1.17	1.04	1.31
Skin Col x meanIAT	0.96	0.77	1.18

Note: A zero regression coefficient on the log-odds scale = 1.00. Therefore, confidence intervals for logistic models can be interpreted similarly as linear regression models, except confidence intervals that include 1 can be considered non-influential (as opposed to zero).

TABLE 2
Results from multilevel logistic model, using mean explicit prejudice (mean Explicit)

	Estimate	LowerCI	UpperCI*
Skin Col	1.43	1.19	1.72
Mean Explicit	1.10	0.99	1.23
Skin Col x mean Explicit	1.03	0.84	1.27

Note: A zero regression coefficient on the log-odds scale = 1.00. Therefore, confidence intervals for logistic models can be interpreted similarly as linear regression models, except confidence intervals that include 1 can be considered non-influential (as opposed to zero).

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. http://CRAN.R-project.org/package=lme4

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods, 7*, 400-417.

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

Gelman, A., & Shalizi, C.R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66*, 8 – 38.

Gooty, J. & Yammarino, F. J. (2011). Dyads in organizational research: Conceptual issues and multilevel issues. *Organizational Research Methods, 14*, 456-483.

Parsons, C., Sulaeman, J., Yates, M., & Hamermesh, D. (2011). Strike three: Discrimniation, incentives, and evaluation. *American Economic Review, 101*, 1410-1435.

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

Pope, D., Price, J., & Wolfers, J. (2013). Awareness reduces racial bias. NBER working paper No. 19765.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/