# A mixed-effects model of race and player penalization in sports

**Authors:** Eli Awtrey[1]

### Affiliations

[1]University of Washington*

*Correspondence to: eawtrey@uw.edu

### Abstract

In order to answer the question of racial bias of referees in sports, a mixed-effects model was used to test whether or not darker player skin tones had a systematic impact on their likelihood of penalization. Standard regression procedures (e.g. ordinary least squares) will likely provide incorrect coefficients due to the highly non-normal nature of the dependent variable Thus, a zero-inflated Poisson model was used to determine the impact of skin color on receiving red cards in soccer. Analysis shows slim effects ($p < .10$) of an interaction between skin color and implicit bias in the referee's country of origin on red cards issuance, but no direct effect of skin color itself.

### One Sentence Summary

There is a relationship ($p < .10$) between player skin color, implicit racial biases of a referee's home, and red card issuance in European football.

# Results

## Initial Approach: Analytical method

The initial approach to this analysis is focused on dealing with substantial non-normality in the dependent variable. A standard OLS approach assumes a normal distribution for the DV, so that is not an option. A standard approach with count data is to employ some version of a Poisson distribution. Since these data are relatively rare events, there is a much higher number of zeroes than a standard Poisson distribution can account for. There are two methods to deal with a "high" zero distribution in count data--a hurdle distribution and a zero-inflated distribution. Both of these are mixture distributions, in which one distribution is chosen for the point mass at zero and another for the non-zero counts. A hurdle distribution assumes there is some sort of "threshold" that must be passed, after which counts are distributed in a particular manner. A zero-inflated distribution (used in this analysis) makes no modeling assumptions about the distribution for the zeroes. Of important note in both of these approaches is that coefficients are produced for both parts of the distribution, so we can answer both "what creates non-zero counts of red cards" and "what creates more red cards in one dyad than another". Unfortunately, this dual-coefficient approach makes interpretation somewhat more difficult. For the sake of this question, regression coefficients will be from the former example, the binomial zero-inflated portion of the model (see Table 1). Cases without skin color ratings were removed from the dataset.

## Initial Approach: Variables

*Dependent variable:* Number of red cards given in the referee-player dyad was the outcome of interest.

*Independent variables:* The primary predictor of interest is skin color of the focal player. This was created by averaging between the two raters and was approached as a continuous variable. Secondary predictors include mean implicit and explicit bias in the referee’s country of origin.

*Controls:* To reduce noise and confounds, several controls were included as well. In order to get an idea of the physical stature of the player (under the naïve assumption that bigger players would seem more “physical”, and thus warranting more penalization), player weight was included. Player position is included to address differing propensities in penalties. (The role of a defensive player is to stop the opponent’s progress—thus creating a higher likelihood of penalties—whereas the offensive player’s role is to elude the opponent.) Number of games were included in the regression as a control, since different dyads have different levels of exposure. Note that there is theoretical ambiguity regarding the effect of increased games. Certainly, more games provide more opportunity for red cards, but increased exposure to a player should cause a referee to rely less on surface-level diversity (Harrison, Price, Gavin, & Florey, 2002).

## Initial Approach: Results

Two models, the first with only main effects and the second including both main effects and interactions between the primary predictor (skin color) and the two secondary predictors (implicit and explicit bias in the referee’s country of origin), were used for interpreting results.

The first model tests the first hypothesis and does not support the claim that skin color plays a role in the quantity of red cards a player receives (-0.030, *ns*). The second model partially supports the interactions of the second hypothesis, with a somewhat significant implicit bias interaction (11.763, $p < .10$) but no significance for the explicit bias interaction (-1.431, *ns*)

### Final Approach: Analytical method

Input on the initial approach highlighted the need to address the nesting in the data by referee. Several mixed-effects models were attempted (both Poisson and binomial), but all of them failed to converge. As a conservative alternative, dummy codes for referee were used in the ZIP regression from the initial round. This account for systematic fixed effects based on referee, but is less efficient than a true mixed-effects model (requires more degrees of freedom).

### Final Approach: Variables

The same variables are used in both approaches, although the race color variable was rescaled to a 0-1 continuous variable.

### Final approach: Results

Results are consistent with the first analysis. Model 3 indicates no support for the direct effect of skin color on red carding (-0.119, *ns*). Model 4 continues to offer mixed support for the country interactions, with a somewhat significant implicit bias interaction (45.090, $p < .10$) and a non-significant explicit interaction (-5.334, *ns*). Another way of interpreting these results, given the significance of the IAT variable in Model 3, is that player race has no impact on red card issuance but implicit bias does.

## Conclusion

Based on this analysis, it appears that player skin color is not categorically associated with higher penalization in European football, except when the referee is from a country with higher implicit racial biases. Interestingly, a more standard OLS approach attempted by the author would indicate a large and significant effect of player skin color on red card receipt. This reveals the sensitivity of our results to appropriate analytical assumptions.

## Tables

**Table 1: Binomial zero-inflated coefficients (logit)**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | p-value | Estimate | p-value | Estimate | p-value | Estimate | p-value |
| Intercept | -2.208 | 0.336 | 6.075 | 0.202 | -2.253 | 0.319 | 2.357 | 0.470 |
| Dark skin* | -0.030 | 0.678 | -3.506 | 0.053 . | -0.119 | 0.684 | -13.520 | 0.063 . |
| Weight | -0.004 | 0.766 | -0.003 | 0.845 | -0.004 | 0.780 | -0.002 | 0.864 |
| Games | -0.636 | < .001 *** | -0.631 | < .001 *** | -0.653 | < .001 *** | -0.644 | < .001 *** |
| Mean IAT | 14.437 | 0.036 * | -14.084 | 0.375 | 14.941 | 0.025 * | -1.165 | 0.912 |
| Mean Exp. | 0.292 | 0.769 | 3.828 | 0.099 . | 0.294 | 0.755 | 2.288 | 0.131 |
| Skin : IAT | | | 11.763 | 0.055 . | | | 45.090 | 0.067 . |
| Skin : Exp | | | -1.431 | 0.110 | | | -5.334 | 0.137 |
| Controls | *Position* | *Position* | *Position* | *Position* | *Position, referee* | *Position, referee* | *Position, referee* | *Position, referee* |

* Note: Scaling of Dark Skin was 1-5 in Models 1 & 2, and 0-1 in Models 3 & 4.

**Table 2: Poisson count model coefficients (log)**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | p-value | Estimate | p-value | Estimate | p-value | Estimate | p-value |
| Intercept | -8.220 | < .001*** | -4.563 | 0.119 | -8.305 | < .001*** | -6.256 | 0.002 ** |
| Dark skin* | 0.073 | 0.034* | -1.385 | 0.178 | 0.293 | 0.032 * | -5.146 | 0.207 |
| Weight | -0.002 | 0.717 | -0.002 | 0.798 | -0.002 | 0.699 | -0.002 | 0.782 |
| Games | 0.054 | < .001*** | 0.054 | <.001*** | 0.055 | < .001 *** | .0543 | < .001 *** |
| Mean IAT | 11.778 | 0.008** | -1.360 | 0.892 | 12.110 | 0.006 ** | 4.649 | 0.497 |
| Mean Exp. | 0.204 | 0.766 | 2.167 | 0.167 | 0.193 | 0.762 | 1.322 | 0.198 |
| Skin : IAT | | | 5.170 | 0.149 | | | 19.270 | 0.176 |
| Skin : Exp | | | -0.770 | 0.175 | | | -2.860 | 0.205 |
| Controls | *Position* | *Position* | *Position* | *Position* | *Position, referee* | *Position, referee* | *Position, referee* | *Position, referee* |

* Note: Scaling of Dark Skin was 1-5 in Models 1 & 2, and 0-1 in Models 3 & 4.

## References and Notes

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. 2002. Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning. ***Academy of Management Journal***, 45(5): 1029–1045.