

“Poisson regression does not suggest influence of perceived skin tone on likelihood of receiving red cards”

Authors: S. Gordon-McKeon*

Affiliations

*Correspondence to: shaunagm@gmail.com

Abstract

In this analysis, Poisson regression was used to explore the influence of skin tone and referee bias on the number of red cards given to players. Number of games was used as an exposure variable, with goals, position, yellow cards, league of player, and country-of-origin-wide bias measures, meanIAT (implicit) and meanExp (explicit) included in the model. The incidence rate ratio calculated, 2.9321, is interpretable as the number of red cards received by players with skin tone rated darkest for each 1 red card received by players with skin tone rated lightest. However, this effect is not significant (95% conf: [0.1085, 78.6596], $p = 0.5266$). Smaller effects were found when the role of bias was explored, with a non-significant negative impact of meanIAT (IRR: 0.0153, 95% conf: [.0000, 466.2143], $p = 0.4355$) and a non-significant positive impact of meanExp (IRR: 1.7724, 95% conf: [0.3820, 8.0215], $p = 0.4661$). These results do not support the hypothesis that skin tone influences the receipt of red cards.

One Sentence Summary

A Poisson regression analysis suggested a higher rate of red cards for players with skin-tones rated as darker, however neither these results nor the weaker effects of implicit and explicit bias were significant.

Results

Initial Approach

Model Specification

Poisson regression was chosen for this analysis due to the count format of the dependent variable, 'redCards'. Poisson regressions are especially suited for rare events, and the distribution of 'redCards' was strongly right-skewed, with the modal count, 0, occurring 98.7% of the time. The distribution's mean and variance were similar (0.01275 and 0.01297, respectively), another characteristic of a Poisson distribution.

A number of independent variables were specified, including skin-tone rating, goals scored, games played, the number of yellow cards given to the player, and the mean explicit bias and mean implicit bias of the referee's country of origin. In the regression for question 1, the regression coefficient for skin-tone ratings was the result of interest, and an interaction term was created for each other variable by multiplying it by the 'rating' variable. In the regression for question 2a, the regression coefficient for 'meanIAT' was the result of interest, and an interaction term was created for each other variable by multiplying it by the 'meanIAT' variable. In the regression for question 2b, the regression coefficient for 'meanExp' was the result of interest, and an interaction term was created for each other variable by multiplying it by the 'meanExp' variable.

The data was kept in player-referee dyad format. Of the 146028 dyads in the original dataset, 21407 were missing ratings from both raters and were excluded from all analyses. For

questions 2a and 2b 'meanIAT' and 'meanExp' data were missing for 153 dyads, which were also excluded.

Data Transformations

The two ratings, 'rater1' and 'rater2', were averaged into a single variable, 'rating'. Spearman's rank correlation coefficient showed that the raters were 85.8% in agreement, so this was considered sufficient. (Spearman's was chosen over Pearson's due to the non-normal distribution of ratings.) For question 2, 'meanIAT' and 'meanExp' ratings were multiplied by 100 to ease interpretation of effect sizes.

Results

Question 1: A regression coefficient of .4975 was found for skin-tone rating. This result can be exponentiated into an incidence rate ratio of 1.6446. This incidence rate ratio can be interpreted as "for each increase of 1 in the skin-tone rating variable, a player is 1.6446 times more likely to receive a red card". However, this finding was not significant (95% conf: [-0.112, 1.107], $p = 0.110$.)

Question 2: A regression coefficient of .0609 was found for meanIAT. This result can be exponentiated into an incidence rate ratio of 1.0628. This more modest effect was closer to significance (95% conf: [-0.004, 0.126], $p = 0.066$).

Question 3: A regression coefficient of .0323 was found for meanExp. This result can be exponentiated into an incidence rate ratio of 1.0328. This effect was significant (95% conf:

[0.009, 0.056], $p = 0.008$).

Final Approach

Model Specification

A Poisson regression was again performed. In addition to the independent variables specified in the initial approach (skin-tone rating, goals scored, games played, yellow cards, and mean explicit and implicit bias) a variable for player position was included, as well as a variable for the league of the player. Feedback from other groups led to the consideration of a hierarchical model using club and league of the player, but this data was deemed incomplete as it only covered the first club and league of a player's career, while 'redCards', as well as several other key variables, contained data from throughout each player's career. Unfortunately, this leaves the hierarchical nature of the dataset mostly unaccounted for. 'club' was not included in the model as the 93 categories ran up against performance limitations.

In response to feedback, the variable 'games' games was specified as an exposure, or offset, variable. This was deemed a more effective way to account for the fact that player-referee dyads contained unequal numbers of games. Additionally, the separate regressions performed to answer questions 2a and 2b were dropped. Instead, the research question was answered by inspecting the 'meanIAT*rating' and 'meanExp*rating' interaction terms, respectively. This was done in order to measure the influence of bias on any tendency of skin-tone to affect red card frequency, as the previous measure had looked at the effect of bias on red card frequency directly.

The data was kept in player-referee dyad format. Of the 146028 dyads in the original dataset, 21407 were missing skin-tone ratings from both raters and were excluded from all analyses. 153 were missing 'meanIAT' and 'meanExp' ratings, and 17,726 were missing position data. These were all excluded, leaving 116014 dyads.

Data Transformations

As with the initial approach, the mean of 'rater1' and 'rater2' was taken to create a variable 'rating'. At the suggestion of the project organizers, the skin-tone ratings were rescaled from 1-5 to 0-1.

Results

Question 1: A regression coefficient of 1.0757 was found for skin-tone rating. This result can be exponentiated into an incidence rate ratio of 2.9321. This incidence rate ratio can be interpreted as “a player with the darkest skin tone is 2.9321 times as likely to receive a red card than a player with the lightest skin tone”. However, this finding is of little value, given the wide confidence intervals (95% conf: [0.1085, 78.6596]) and high p-value ($p = 0.5266$).

Question 2: A regression coefficient of -4.1810 was found for the interaction term of rating and meanIAT. This result can be exponentiated into an incidence rate ratio of 0.0153. This finding was also not significant (95% conf: [.0000, 466.2143], $p = 0.4355$).

Question 3: A regression coefficient of 0.5723 was found for the interaction term of rating and meanExp. This result can be exponentiated into an incidence rate ratio of 1.7724.

This effect was not significant (95% conf: [0.3820, 8.0215], $p = 0.4661$).

Conclusion

The Poisson regression analysis performed did not find a clear influence of skin tone rating on red cards received. The incidence rate ratio calculated, 2.9321, is interpretable as the number of red cards received by players with skin tone rated darkest for each 1 red card received by players with skin tone rated lightest. However, this effect is not significant (95% conf: [0.1085, 78.6596], $p = 0.5266$). Smaller effects were found when the impact of bias was explored. There was a negative impact of meanIAT on the influence of skin-tone in the form of an IRR of 0.0153, but again the effect was not significant (95% conf: [.0000, 466.2143], $p = 0.4355$). Finally, there was a positive impact of meanExp on the influence of skin-tone in the form of an IRR of 1.7724, also not significant (95% conf: [0.3820, 8.0215], $p = 0.4661$). These wide confidence intervals and large significance values are likely due to the poor fit of the model. The results do not allow us to infer an influence of skin-tone on the rate of red cards given to players. This does not allow us to say that an effect does not exist: a different methodology, or a different dataset (perhaps one with IAT or Exp measures for individual referees, or more comprehensive team and country data allowing hierarchical modeling) might find better evidence.

Data and Output

<https://osf.io/janbc/>