Poisson regression analysis indicates that skin tone is likely to be related to the probability of receiving a red card

Authors: R. Clay[1*]

Affiliations

[1]College of Staten Island, CUNY

*Correspondence to: Russ.Clay@csi.cuny.edu

Abstract

Hierarchical Poisson regression was used to investigate the presence of skin tone bias and referee prejudice associated with red cards that players received across the soccer games represented in the data set. The analysis indicated that after controlling for differences in red cards associated with physical characteristics of the player (height, weight, and age), the number of goals that the player had scored, the country of the league that they played in, and the position that they played and after adjusting for variability in the number of games included in the data set for each player, skin tone bias still existed such that players rated as having darker skin received more red cards on average than players rated as having lighter skin. Although the number of red cards received also demonstrated a relation to the average level of prejudice in the referee's home country across the player-referee interactions (independent of the player's skin tone), there was no evidence to indicate that referee prejudice interacted with player's skin tone to influence the number of red cards received (No skin tone X prejudice interactions were observed).

One Sentence Summary

Evidence from Poisson regression analysis indicates that darker skin tone soccer players receive more red cards relative to lighter skin tone players, but it does not appear that average prejudice levels in the home country of the referee play a role in this bias.

Results

Initial Approach

Hierarchical logistic regression was originally pursued as a strategy for both research questions. This was because a large number of players in the data set had never received a red card and very few had received multiple red cards. As such, dichotomization of the red card variable seemed an appropriate approach to investigate whether the likelihood that a player had ever received a red card was related to the player's skin tone rating. However, based on initial feedback from the Crowdstorming participants, this strategy was altered and Poisson regression (using the count of the number of red cards as the outcome variable) was used instead. Poisson regression was considered a superior analysis strategy as it allowed for the retention of information regarding variability in the overall number of red cards received, which would have been lost if the outcome variable was dichotomized.

Technical Limitations

It should be noted that in the analysis of the second research question (exploring the role of prejudice in the referee's home country on skin-tone bias in the distribution of red cards), technical limitations prohibited the pursuit of what was perceived to be the optimal research strategy. More specifically, mixed effects analysis with player and referee assigned as level 2 variables was the preferred approach for this question. This would have been an ideal strategy for addressing the issue of non-independence in the data set. However, technical limitations prevented the use of this analysis strategy: In short, the available processing power was limited and on multiple attempts, the analysis could not be completed before the computer overheated and shut itself down. As such, the strategy outlined below for the investigation of the second research question represents what is believed to be a viable, albeit suboptimal analysis strategy.

Final Approach

For all research questions, Poisson regression with games entered as an exposure (offset) variable was used to investigate the associations between the predictor variables of interest and the total number of red cards received by players in the dataset. Only direct red cards, and not yellow-reds, were used for the outcome variable. This decision was made because it was felt that red cards represent more of a direct, subjective evaluation by the referee that a player should be removed from the game immediately, and therefore are likely to tap into referee biases than yellow-reds. The overall analytical approach allowed for the investigation of the effect of each predictor variable on the observed number of red cards for each player. Additionally, by including the number of games as an exposure variable, the outcome variable represented a rate (number of red cards per game) rather than a count, which helped to address the vast differences between players in opportunities to receive a red card (number of games played).

All analysis was completed using the R Statistical Software Package (version 3.1.0) and was conducted following the methodology for Poisson regression provided by UCLA (2014), the Pennsylvania State University Department of Statistics (2014), and additional guidance for investigating the assumptions of regression were found in Field, Miles, and Field (2012).

It should also be noted that though Poisson regression was determined to be the most appropriate analysis technique given the research question, the researcher lacked any familiarity with the approach prior to beginning this project. Care was taken to follow reliable resources for the approach, citations are included to acknowledge these sources, and the analysis steps are explained in detail. As the nature of the current crowdstorming project precludes the same level of peer review of the analysis that would occur in a typical academic publication, the reader should take care to scrutinize the steps of the approach closely before drawing strong conclusions from the results of the analysis.

For all research questions, data was aggregated at the level of the individual player (N = 2053) prior to removing cases associated with missing data. The effects of interest are presented as odds ratios with confidence intervals to indicate the percentage increase in the rate of red cards per game associated with the relevant predictors. Considerations and results specific to each research question are presented in the following subsections.

Transformations. Reliability analysis indicated high reliability across the two skin tone ratings of each player ($\alpha = 0.96$) and as such, the ratings were averaged to create an overall index of player's skin tone, (M = 0.29, SD = 0.29). Player position was transformed into five dichotomous variables as follows: The variable 'Forward' was coded 1 for players whose position was 'Left Winger' 'Right Winger', or 'Center Forward' (n = 349), and 0 otherwise. The variable 'Midfielder' was coded 1 for players whose position was 'Right Midfielder', 'Left Midfielder', 'Center Midfielder', 'Attacking Midfielder', or 'Defensive Midfielder' (n = 598), and 0 otherwise. The variable 'Fullback' was coded 1 for players whose position was 'Right Fullback', 'Left Fullback', or 'Center Back' (n = 543), and 0 otherwise, and the variable 'Goalkeeper' was coded 1 for players whose position was 'Goalkeeper' (n = 196) and 0 otherwise. In addition, a variable 'Other' was created which was coded 1 for any player whose position was listed as 'NA' (n = 367) and 0 otherwise. This transformation was made so that the variables used to control for player position would be more manageable, and because meaningful differences in the propensity for receiving red cards were expected between the four primary categories of positions (excluding the 'Other' category), but were not expected to differ due to the side of the field that the player tended to play on. Dichotomous variables were also created using dummy coding to represent the four league countries that were present in the database: England (*n* = 564), France (*n* = 533), Germany (*n* = 489), and Spain (*n* = 467). Finally, player's age was calculated as the number of years from their birth date to an arbitrarily selected date (January 1, 2013).

Additional transformations were made to the implicit and explicit prejudice variables for investigation of the second research question (exploring the relevance of measures of referee prejudice in red card bias), and are explained in detail in a later subsection.

Research Question 1 Analysis and Results. Poisson regression was used to investigate the relation between a player's skin tone and the number of red cards they had received across the observations reported in the dataset. Prior to building the model, cases that contained missing values for any of the outcome, predictor, control, or exposure variables were dropped (*n* = 489) and the main analysis was conducted using the remaining sample of unique players (*N* = 1,564). The total number of red cards received (*M* = 1.01, *SD* = 1.34) was entered as the outcome variable, and the total number of games played (*M* = 237.73, *SD* = 141.99) was entered as an exposure variable (log transformed). Unstandardized regression coefficients along with calculations of robust standard errors, and odds ratio with associated 95% confidence intervals for each predictor variable can be found in Table 1. Robust standard errors were calculated using the formulas provided by UCLA (2014)

An initial model containing only control variables as predictors was evaluated first. This model included continuous variables representing players age (*M* = 26.77, *SD* = 4.43), height in cm (*M* = 182.04, *SD* = 6.74), weight in kg (*M* = 76.03, *SD* = 7.11), and number of goals scored (*M* = 28.05, *SD* = 41.72). These continuous variables were centered so that the resulting intercept could be interpreted as the rate of red cards per game associated with the 'average' player in the data set. The dichotomous dummy-coded variables representing France, Spain, and Germany as well as the Forward, Midfielder, Fullback, and Goalkeeper positions were included as control variables as well. Thus, the resulting model intercept also represented players from England (arbitrarily selected as the reference category) whose position was not available in the dataset (selected as a reference category because this likely represents a mix of players from all positions). Data associated with the number of victories, ties, and defeats that each player experienced were not included as control variables as these were all highly related to the number of games played. This base model was significantly better than the null model at predicting variability in the outcome variable, $\Delta\chi^2(11) = 157.42$, $p < .001$. Additionally, the rate of red cards received per game was related to the league country of the player and the position that they played. In comparison to someone playing in England, a player of average age, height, weight, and goal scoring ability playing in France had a 46% higher rate of red cards per game, *B* = 0.38, *OR* = 1.46, *95% CI* = [1.21, 1.76], and the same player playing in Spain had a 48% higher rate of red cards per game, *B* = 0.40, *OR* = 1.48, *95% CI* = [1.27, 1.73], whereas playing in Germany demonstrated little difference in the rate of red cards per game compared to England *B* = -0.15, *OR* = 0.86, *95% CI* = [0.73, 1.02]. In regards to a player's position, the average forward only received 66% of the red cards, *B* = -0.42, *OR* = 0.66, *95% CI* = [0.51, 0.85], and the average midfielder only 72% of the red cards, *B* = -0.33, *OR* = 0.72, *95% CI* = [0.57, 0.90], as players in the 'Other' position category received. Fullbacks did not demonstrate meaningful differences from the reference category, *B* = 0.10, *OR* = 1.11, *95% CI* = [0.60, 1.06]. Finally, a player's age, height, weight, and the number of goals they scored did not relate meaningfully to the number of red cards received.

The skin tone variable was added in a second step of the model after accounting for the effect of all of the control variables (The skin tone variable was not centered, as its values ranged from 0 to 1. As such, the intercept of the model with skin tone included as a predictor represents the rate of red cards per game by players rated as having the lightest possible skin tone, and the odds ratio of the skin tone variable represents the increased rate of red cards per game for players rated as having the darkest possible skin tone relative to those rated as having the lightest). The addition of skin tone significantly improved the fit of the model in comparison to the model with only control variables, $\Delta\chi^2(1) = 9.08$, $p = .003$. Skin tone demonstrated a meaningful relation with the number of red cards received, $B = 0.28$, $OR = 1.32$, *95% CI* = [1.06, 1.63], indicating that players with the darkest possible skin tone ratings received 32% more red cards per game relative to the players with the lightest possible skin-tone ratings, though the confidence interval demonstrated considerable variability around the estimate of this effect. Investigation of outliers, influential cases, residuals, and multicolinearity (as recommended by Field et. al., 2012) did not reveal any violations of model assumptions, though statistics did indicate that the model left a significant proportion of the variability in the number of red cards received per game unexplained (residual deviance = 2007.23, df=1551, $p < .001$). This indicates that it is likely that other variables exist which are meaningfully related to the number of red cards that a player receives.

To summarize, evidence stemming from the investigation of the data indicated a bias in the number of red cards a player received associated with skin tone. Across the data set, players varied considerably with respect to the number of opportunities they had to receive a red card (i.e., the number of games played). However, after including the number of games played as an exposure (offset) variable in the analysis, and after controlling for differences associated with league country and the player's position, skin tone was still associated with meaningful differences in the number of red cards received such that darker skin players received more red cards per game relative to lighter skinned players.

Research Question 2 Analysis and Results. The second research question investigated the role of prejudice in the referee's home country in the bias associated with the relation between a player's skin tone and their propensity for receiving red cards. As stated previously, technical difficulties prevented the use of mixed effects modeling to investigate this research question, which would have been the preferred method of analysis. Instead, the data was once again aggregated at the level of each unique player in the data set, and Poisson regression was used to investigate factors associated with the number of red cards that players received. Referee prejudice was represented by historical average scores of implicit and explicit prejudice taken from the home country of the referee. Prior to aggregating the data, the implicit and explicit prejudice scores were multiplied by the number of games represented by the player – referee dyad. These prejudice scores were then summed when the data was aggregated to the level of the player, and the summed scores were divided by the total number of games the player had played which resulted in a variable that represented the average prejudice level of the referee's home country that a player encountered over the course of all of their games that were included in the data set. These scores for implicit ($M = 0.35$, $SD = 0.02$) and explicit ($M = 0.44$, $SD = 0.11$) prejudice had small standard deviations which made interpretation of the effect size measures difficult in the subsequent analysis (e.g. an odds ratio representing the difference in the predicted number of red cards associated with a 1 unit change in implicit prejudice represents a change that is 50 times greater than one would expect). As such, the index for implicit prejudice was multiplied by a factor of 100 and the index for explicit prejudice was multiplied by a factor of 10. Thus, the equations for the prejudice indices for each player are as follows (meanIAT = average implicit prejudice score from a referee's home country; meanExp = average explicit prejudice score from a referee's home country):

Implicit prejudice: $(\Sigma(\text{meanIAT} * \text{games played})_{\text{player-referee dyad}}) / \text{total games played} * 100$

Explicit prejudice: $(\Sigma(\text{meanExp} * \text{games played})_{\text{player-referee dyad}}) / \text{total games played} * 10$

Prior to beginning the analysis, the same observations were excluded as in the first research question to arrive at the final sample of unique players (N = 1,564). Additionally, the same base model was used including all of the same control variables that were used in the first research question. Following this, two separate models were developed which investigated the role of implicit and explicit prejudice levels in the home countries of the referees that the players had encountered. In each model, the (centered) main effects of player's skin tone and average prejudice level encountered were entered in the second step of the model, and a skin tone X prejudice interaction term was added in the third and final step. The results of this analysis are presented in Table 2.

In the model investigating the effect of implicit prejudice on skin tone bias, the addition of the main effects significantly improved the fit of the model over the model that included only control variables, $\Delta\chi^2(2) = 17.02$, $p < .001$. Both main effects demonstrated meaningful associations with the number of red cards received. Darker skin tone was associated with a higher number of red cards relative to lighter skin tone, $B = 0.31$, $OR = 1.36$, *95% CI* = [1.09, 1.69], and players encountering higher levels of referee prejudice over the course of their games received more red cards relative to those encountering lower levels of referee prejudice, $B = 0.08$, $OR = 1.08$, *95% CI* = [1.02, 1.15].

Of main interest for the current research question, the skin tone X implicit prejudice interaction term was entered in the final step of the model. The addition of the interaction term did not significantly improve the fit of the model, $\Delta\chi^2(1) = 0.24$, $p = 0.63$, as the interaction term did not demonstrate a meaningful relation to the number of red cards received, $B = -0.03$, $OR = 0.97$, *95% CI* = [0.84, 1.12].

In the model investigating the effect of explicit prejudice on skin tone bias in the awarding of red cards, the addition of the main effects of skin tone and explicit prejudice once again demonstrated significantly improved fit over the model that included only control variables, $\Delta\chi^2(2) = 15.28$, $p < 0.001$ and the main effects of skin tone, $B = 0.30$, $OR = 1.35$, *95% CI* = [1.09, 1.67], and explicit prejudice, $B = 0.10$, $OR = 1.11$, *95% CI* = [1.02, 1.21] demonstrated the same relations to the rate of red cards received.

Once again, of primary interest in the present analysis, the addition of the skin tone X explicit prejudice did not significantly improve the fit of the model, $\Delta\chi^2(1) = 0.39$, $p = 0.53$, and the interaction term did not demonstrate a meaningful relation with the number of red cards received, $B = -0.06$, $OR = 0.94$, *95% CI* = [0.76, 1.17].

Investigation of outliers, influential cases, residuals, and multicolinearity of the predictors did not indicate any violations of model assumptions, though once again, there was a significant amount of residual variability left unexplained (residual deviance = 1999.00, $df = 1549$, $p < .001$).

In sum, the analysis indicated a main effect of both implicit and explicit prejudice such that players encountering referees from countries with higher levels of prejudice (implicit as well as explicit) on average over the course of their games tended to receive more red cards relative players who encountered referees from countries with relatively lower levels of prejudice on average. However, this effect was present independent of the player's skin tone, and there was no evidence of a skin tone X prejudice interaction, regardless of whether prejudice was assessed with an implicit or explicit measure.

Conclusion

The results of the present analysis indicate that skin tone bias in the awarding of red cards existed in the soccer games represented by the current data set. Although other characteristics such as age, height and weight, and the goal scoring ability of the player did not appear to be related to the number of red cards received, skin tone demonstrated a meaningful relation such that darker skin players tended to receive more red cards relative to lighter skin tone players, even after controlling for meaningful differences such as the player's position and the country of the league that they played in and after making an adjustment for the number of games that a player had played.

This bias was found independent of the average level of prejudice that the player encountered from the referees that officiated their games (as measured by the average level of prejudice in the referee's home country). It should be reiterated that the statistic used to make this inference of referee prejudice in the current analysis was not a direct measurement of the referees themselves, and that a sub-optimal analysis technique had to be used which did not allow for investigation of differences among individual referees. As such, it would be premature to conclude that referee prejudice (implicit or explicit) did not play a role in the observed skin-tone bias in the awarding of red cards. However, the specific variables used for the present analysis did not provide any evidence for the role of prejudice.

Tables

Table 1. *Summary of Model Statistics for Research Question 1*

Predictor	B (SE_{robust})	Odds Ratio	95% Confidence Interval LCL	UCL
Step 1				
Intercept	-5.45 (0.00)	0.00	0.00	0.01
Age	-0.01 (0.01)	0.99	0.98	1.01
Height	0.01 (0.01)	1.01	0.99	1.02
Weight	0.01 (0.01)	1.01	1.00	1.03
Goals	0.00 (0.00)	1.00	1.00	1.00
France	0.38 (0.14)	1.46	1.21	1.76
Germany	-0.15 (0.07)	0.86	0.73	1.02
Spain	0.40 (0.12)	1.48	1.27	1.73
Forward	-0.42 (0.08)	0.66	0.51	0.85
Midfielder	-0.33 (0.08)	0.72	0.57	0.90
Fullback	0.10 (0.12)	1.11	0.90	1.37
Goalkeeper	-0.22 (0.12)	0.80	0.60	1.06
Step 2				
Intercept	-5.54 (0.00)	0.00	0.00	0.00
Skin Tone	0.28 (0.14)	1.32	1.06	1.63

Note. Robust standard errors were calculated following the syntax provided by UCLA (2014)

Table 2. *Summary of Model Statistics for Research Questions 2a & 2b*

Predictor	B (SE_{robust})	Odds Ratio	95% Confidence Interval LCL	UCL
Step 1				
Intercept	-5.45 (0.00)	0.00	0.00	0.01
Age	-0.01 (0.01)	0.99	0.98	1.01
Height	0.01 (0.01)	1.01	0.99	1.02
Weight	0.01 (0.01)	1.01	1.00	1.03
Goals	0.00 (0.00)	1.00	1.00	1.00
France	0.38 (0.14)	1.46	1.21	1.76
Germany	-0.15 (0.07)	0.86	0.73	1.02
Spain	0.40 (0.12)	1.48	1.27	1.73
Forward	-0.42 (0.08)	0.66	0.51	0.85
Midfielder	-0.33 (0.08)	0.72	0.57	0.90
Fullback	0.10 (0.12)	1.11	0.90	1.37
Goalkeeper	-0.22 (0.12)	0.80	0.60	1.06
Step 2				
Intercept	-5.36 (0.00)	0.00	0.00	0.00
Skin Tone (implicit)	0.31 (0.15)	1.36	1.09	1.69
Prejudice (implicit)	0.08 (0.03)	1.08	1.02	1.15
Skin Tone (explicit)	0.30 (0.15)	1.35	1.09	1.67
Prejudice (explicit)	0.10 (0.05)	1.11	1.02	1.21
Step 3				
Intercept	-5.35 (0.00)	0.00	0.00	0.00
Skin Tone x Prejudice (implicit)	-0.03 (0.07)	0.97	0.84	1.12
Skin Tone x Prejudice (explicit)	-0.06 (0.10)	0.94	0.76	1.17

Note. Robust standard errors were calculated following the syntax provided by UCLA (2014)

References and Notes

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using r. SAGE Publications Ltd., London.

The Pennsylvania State University (2014). Lesson 9: Poisson Regression. The Pennsylvania State University Department of Statistics Online Learning Website. From https://onlinecourses.science.psu.edu/stat504/book/export/html/165 (accessed July 11, 2014).

UCLA (2014). R Data Analysis Examples: Poisson Regression. UCLA: Statistical Consulting Group. From http://www.ats.ucla.edu/stat/r/dae/poissonreg.htm (accessed July 10, 2014)