Skin tone discrimination in number of red cards given to professional soccer players: Is implicit or explicit skin tone bias the explanation?

Authors: Rickard Carlsson[1]

Affiliations

[1]Department of psychology, Linnaeus University

Abstract

A dataset on the number of red cards professional soccer players have received throughout their careers was analyzed in order to test for whether dark skinned players are discriminated against. We further examined whether this discrimination effect is stronger for referees that come from countries with high levels of explicit (self-reported) or implicit (Implicit Association Test) skin tone bias. Using a clustered robust binomial regression, adjusted for several potentially cofounding variables (e.g., player position), we find an unexplained effect of dark skinned players receiving more red cards than light skinned players, suggesting that referees discriminate based on skin tone. However, we find no indication that the level of implicit or explicit skin tone bias in the referee's home country moderates this discrimination.

One Sentence Summary

Using a clustered robust binomial regression adjusted for several potentially cofounding variables, we find that dark skinned players receive more red cards, but that this is not related to the average levels of implicit or explicit skin bias in the referee's home country.

Results

Initial Approach

The two ratings of skin tone had good internal consistency (alpha = 0.96) so we averaged them into a single index. 21407 observations had missing values on this crucial skin tone variable and were thus omitted from subsequent analysis. Of the remaining 124,621 observations 123,056 resulted in zero red cards, 1541 in one red card, and 24 in two red cards. Of course, having received a red card in 47 games is quite different from having received a red card based on a single played game. The obvious thing to do here is to disaggregate the dyads into games. However, this was not straightforward, since the correlation with other variables also aggregated (e.g., number of yellow-cards, number of goals) is unknown on the level of games. Instead, we divided each game-specific outcome (red cards, yellow cards, victories, defeats, goals, ties) by the number of games played, thus rendering the variables the likelihood of the event occurring during a game. Although this approach resulted in fewer observations than if the data had been disaggregated, the multiple observations should instead add to the reliability of the estimate. Further, because there was very little variation beyond 0 or 1 red card, we decided on treating the dependent variable as a dummy and choose to analyze the results by means of binary logistic regression, treating any non-zero probability as having received (at least) one red card.

Regarding research question 1, we note that the 124, 621 dyads come from 1572 unique players and 3147 unique referees, meaning that we cannot assume the observations to be independent of each other, making the standard errors potentially downward biased. To correct for this, we used a two-way clustering approach suggested by Peterson (2009). We then analyzed the data in STATA 12.1 using an add-on that allowed for two-way clustering (logit2, Petersen, 2014). In other words, we analyzed the data by means of logistic regression with two-way clustering on referee and player, with the dependent variable being whether a single encounter between a player and a referee resulted in a red-card or not. We first examined the unadjusted effect, and then added controls for potentially confounding variables player height, weight and position.

The unadjusted odds ratio for one standard deviation darker skin tone is 1.07, 95 % CI [1.01,1.13]. Hence, darker skin tone is associated with a higher likelihood of receiving a red card. This is a weak, but detectable effect of discrimination against dark skinned players and it cannot be explained by difference between dark and light skinned players in terms of height, weight or position (e.g., forward or goalkeeper) they play. Indeed, the adjusted effect is virtually identical: OR = 1.07, 95 % CI [1.02, 1.14].

Turning to research question 2, we note that because the data on implicit and explicit measures are not directly measured for each referee, but rather based on the overall values of the referees home country, the player dyads are nested within the referee's home country (N = 155). Thus, we conducted a logistic regression clustering on referee’s home country (logit, STATA 12.1). As in the previous analysis, we started without controls. The first model had the following predictors (all standardized): skin tone rating, implicit bias, explicit bias, interaction between skin tone and implicit bias, and interaction between skin tone and explicit bias. The interaction effects served as our answer to Q2. If implicit and explicit biases matter, then there should be a positive interaction effect in that the tendency to give more red cards to dark skinned players should be stronger for those with high levels of implicit and or explicit bias.

First, we note that the effect of skin tone remains very similar in size: OR = 1.08, 95 % CI [0.98, 1.18]. The main effects of both the implicit (OR = 1.12, 95 % CI [.91, 1.39] and the explicit measure (OR =0.87, 95 % CI [0.72, 1.05] have CIs that are too wide to draw any conclusions. Further, these main effects have no direct theoretical interpretation. Of more relevance are the interaction effects that should be positive if skin tone discrimination is moderated by implicit or explicit bias. The interactions are close to zero for the implicit measure, OR = .97, 95 % CI [0.90, 1.05], and only slightly positive for the explicit measure, OR = 1.04. 95 % CI [.97, 1.11]. These interaction effect remain very similar after adding the controls (player position, height and weight), with OR = 0.97, 95 % CI [.90, 1.05] for the implicit measure and OR = 1.05, 95 % CI [0.98, 1.12] for the explicit measure.

Final Approach

The two ratings of skin tone had good internal consistency (alpha = 0.96) so we averaged them into a single index. 21407 observations had missing values on this crucial skin tone variable and were thus omitted from subsequent analysis, leaving 124, 621 observations. Each observation comprised a summary of the number of games (1 - 47, M = 2.99, SD = 3.58) played and the number of red cards received (0 - 2, M = .013, SD = .11) across these games for the player referee dyad. Of course, having received a red card in 47 games is quite different from having received a red card in a single game. We decided to view our dependent variable as the probability of receiving a red card during a single game, and thus used binomial logistic regression for all our analyses (the binreg command in STATA 12.1) that takes into account the number of games played when predicting whether a red card has been received or not.

Regarding research question 1, we note that the 124, 621 dyads come from 1572 unique players and 3147 unique referees, meaning that we cannot assume the observations to be independent of each other, making the standard errors potentially downward biased. To correct for this, we used a two-way robust clustering approach suggested by Petersen (2009). Because STATA is only able to handle one-dimensional clustering out of the box we used a STATA add-on for logistic regression (logit2.ado) by Peterson (2014) that we could make some changes to for it in order to work with binomial regression.

Having explained our dependent variable, we now turn to the predictors. For Q1, the variable in focus is the skin tone rating (0 - 1, $M = .28$, $SD = .29$). The data material also includes several other variables that may potentially confound the relationship between skin tone rating and number of red cards. Referees may give more red cards to younger, taller and heavier players. Further, playing certain positions may make one more vulnerable for red cards (e.g., defender vs. goal keeper). Another possible confound is if certain leagues give less red cards and have less black players playing in them. Thus the final model included the predictor rating (continuous 0 - 1), as well age, weight and height as continuous controls (all standardized), and position and league as dummy coded controls.

We find that having the darkest skin tone (1) compared to the lightest (0) increases the probability of receiving a red card in a played game: OR = 1.28, 95 % CI [1.04, 1.57). This is a weak, but detectable effect of discrimination against dark skinned players. The unadjusted effect (without the controls) is highly similar: OR = 1.34, 95 % CI [1.10, 1.62].

Turning to research question 2, we note that because the data on implicit and explicit measures are not directly measured for each referee, but rather based on the overall values of the referees home country, the player dyads are nested within the referee's home country (N = 155). Hence, this time we could use the built-in STATA command binreg clustering on referee home country only.

Explicit or implicit skin tone bias should not have a direct relationship with number of red cards. Any such relationship may indicate an effect of referee's home country unrelated to skin tone bias. Specifically, countries with higher levels of skin tone bias may come from soccer cultures that generally give more red cards. However, this is beyond the scope of the present analysis, and as the focus of the analysis is the interaction effect between skin tone discrimination and implicit and explicit skin tone bias. This interaction should have the specific pattern that the skin tone discrimination should be stronger for referees from countries that have stronger implicit and/or explicit skin tone bias in general.

We first examined whether the implicit bias, as measured by the Implicit Association Test, would moderate the skin tone discrimination effect. We included the same controls as in the previous analysis, but added the IAT-score (standardized) and the interaction between IAT and skin tone rating. The skin tone discrimination remained very similar in this model: OR = 1.29, 95 % CI [1.06, 1.57]. The effect of IAT was OR = 1.09, [.98, 1.22]. The odds ratio for the interaction is: 1.00, 95 % CI [.86, 1.17]. In other words, there is no change in the skin tone discrimination effect, but the confidence interval is too wide to be informative. This conclusion remains even when removing the controls: OR = .92, [.76, 1.12]. Curiously, there is in this unadjusted model some indication that the IAT has an effect on the probability of receiving a red card: OR = 1.19, 95 % CI [1.01, 1.39]. A possible explanation for this is that referees from certain home countries are stricter with red cards than others, regardless of the player's skin tone.

Moving on to the explicit measure of skin tone bias in a referees home country, we again find very similar levels of skin tone discrimination: OR = 1.30, 95 % CI [1.07, 1.58]. We find no detectable effect of explicit bias: OR = 1.04, 95% CI [.92, 1.17], nor any interaction with it and skin tone discrimination: OR = 1.08, [.91, 1.27]. Notice that this interaction is stronger than that of the IAT, but the CI is to wide to draw any conclusions. Further, removing the controls, however, makes this interaction disappear entirely: OR = .98, [.80, 1.20] but makes the main effect slightly more prevalent: OR = 1.13, 95 CI [.94, 1.35]. However, as with the IAT, an effect of explicit bias without an interaction effect is most likely due to some confound.

Conclusion

We found a small positive association between dark skin tone of professional soccer players and the probability of receiving a red card. The effect remained robust when controlling for player age, weight, height, position and league. A reasonable interpretation for this unexplained difference in probability of receiving a red card is that referees discriminate soccer players based on skin tone. However, as in all correlation research on discrimination, it is possible that the difference actually lies in unobserved differences between light skinned and dark skinned players. Although this alternative explanation seems unlikely since the discrimination effect was highly insensitive to the included controls, future research based on experimental data will be useful in providing a more definitive answer.

Another goal of the present analysis was to examine whether the levels of explicit bias or implicit bias in the referees home country moderated the discrimination. The results do not offer any evidence for this. It is worth noting that the discrimination effect is so weak that it is unlikely candidate to be moderated in the first place. Further, examining the bias on the level of the referee's home country lacks precision. Hence, the present research is rather informative as to whether implicit or explicit skin tone bias is behind the discrimination effect.

References and Notes

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review Of Financial Studies, 22(1*), 435-480.

Petersen, M. A. (2014). *Programming advice.* http://www.kellogg.northwestern.edu/faculty/petersen/htm/papers/se/se_programming.htm