Logistic Regression Reveals Referees' Fairness on Darker Skin Players

Authors: S. S. Yoon[1], N. M. Fong[1]

Affiliations

[1]Department of Marketing and Supply Chain Management, Temple University.

*Correspondence to: Sangsuk Yoon (ssyoon@temple.edu)

Abstract

We evaluated two research questions regarding soccer referees' red card decision biases. First, using logistic regression we examined whether players with darker skin color received more red cards than the ones with lighter skin color. We looked at data from four major soccer leagues in Europe. Our results showed that skin color does not influence the likelihood of receiving red cards. Second, we evaluated whether implicit/explicit cultural preference on lighter skin color of referees' country influences referees' tendency to give more red cards to darker skin colored players than to lighter skin colored players. Our results showed that neither implicit nor explicit attitude toward lighter skin color influence referees' tendency to give red cards to darker skin color players.

One Sentence Summary

Our logistic regression results showed that the players' skin colors, and the explicit and implicit attitudes held by the referee's country of origin do not influence the distribution of red cards.

Results

Please write a brief manuscript-style results section in American Psychological Association (APA) format. Please begin by providing some details on your statistical approach and included and excluded variables. You can use text that you supplied already when submitting your analytical approach. If you made changes to your statistical approach and results as a result of the feedback round, please describe both your initial and final approach in the same detail. You may include other subheadings if you see fit.

Initial Approach

Transformations:
First, we encoded several nominal string variables into numeric variables. For example, "club", "leaguecontry", and "position" were encoded into numeric values (i.e., dummy variable) to calculate them as fixed effects in the logistic regression model. Next, we created a new binary dependent variable, "red". Since we are interested in whether or not referees gave red cards, we created an independent variable including 0 and 1. Zero indicated that the player did not receive any red card, while 1 indicated that the player received red cards either from a direct red card or from two yellows. Indeed, for the Research Question 2, we created interaction terms by multiplying skin color rating and two cultural preference scores (i.g., mean IAT and explicit bias score) on lighter skin color.

Exclusions:
First of all, we excluded responses which have at least one data point missing for any of the variables. As a result, a total 22,313 responses were exluded. Next, to rule out rater effect for the categorization of skin color, we excluded the responses which were rated differently by the two raters. A total 28,812 responses were exluded.

Technical Explanations:
Since the key research questions are whether skin color or cultural preference of skin color influence the referees' distribution of red cards, we used logistic regression to evaluate both studies. To run logistic regression, we created a new dependent variable which we coded as 0 for those who did not receive any red card from the matched referee and coded as 1 for those who received any red card either directly or from the accumulation of two yellows. For binary dependent variable, logistic regression is one way to analyze it, so our analysis was mainly based on logistic regression with robust standard error using Stata 12.

Research Question 1

We ran logistic regression to investigate the the first research question, whether or not skin color influences referees to give red cards more to darker skin players than to lighter skin players. To rule out other possible explanations, we added height, weight, and the number of games as covariate variables in the model. Also, we added club, leaguecountry, position, and refcountry variables as fixed effect in the regression model, since we are interested in the association between skin color and the likelihood of receiving red cards after controlling those covariate variables and regardless of those fixed effects in the Research Question 1.

A total 91,933 responses were used in the analysis. In initial stage, our result supported the claim that darker skin colored players receive more red cards than lighter skin color players. The result showed that skin color positively influenced the odds of receving red cards (odds ration: 1.05, $p < .05$, 95% CI: 1.01, 1.10), after controlling weight, height, and the number of games that players played. Also, when we looked at the marginal effect, when one unit of skin color rating increases, the probability of getting red cards increases by 0.12% ($p = .10$, 95% CI: 0.0003, 0.0022).

Research Question 2

We also ran logistic regression for the second research question, whether or not the cultural preference of lighter skin color influences the tendency to give red cards more to darker players than to the lighter color players. So, we took a look at the interaction effects between skin color rating and either mean IAT score or mean explicit bias score on the probability of receiving a red card. To rule out possible explanations, we also included the same covariates and fixed effects we used in the Research Question 1, except that in the Research Question 2, we did not include "refcountry" as a fixed effect, because we are interested in cultural effect of the referee's country.

A total of 94,424 responses were used in the analysis. Our results did not show any evidence supporting the argument that cultural preference of lighter skin color of a referee's country influences the tendency to give more red cards to darker skin players. As seen in the Table 1, the interaction between skin color rating and mean IAT (odds ratio: 1.6712, $p = .23$, 95% CI: 0.7255, 3.8498), the interaction between skin color rating and explicit bias score (odds ratio: 1.1121, $p = .09$, 95% CI: 0.9839, 1.2569), and the three-way interaction among skin color rating, IAT, and explicit bias score (odds ratio: 1.2377, p = .90, 95% CI: 0.0418, 36.6581) were not statistically significant.

Final Approach

There were six main points from other researchers in the feedback phase. First, since inter-rater reliability is high enough ($\alpha = .96$), using average score between two raters' skin color ratings is more reasonable. So, in the final analysis, we used the averaged score between two raters' skin color ratings. Second, since the number of games matters (e.g., receiving one red card from one game and receiving one red card from 30 games are different), weighting the dependent variable or DV ("redcards") for the number of games is required. We ran logistic regression by weighting "redcards" for the number of games in final analysis. Third, since receiving red cards directly and receiving red cards from two yellow cards are different, we coded dependent variable differently compared to initial approach. We created a new variable "red" by coding "red" as zero if their "redcards" response is non-zero and by coding "red" as 1 if their "redcards' response is zero. Fourth, since referees had participated in multiple games, in the final analysis, we used clustered standard error for "refnum" variable to consider referee's tendency to give a red. Five, "goals" variable can be an exogenous variable (e.g., outcome of receiving red card in the game), so we dropped this "goals" variable from covariate variables.

Last, since the data did not provide enough information about players' switching in or out of a league, we dropped league and club variables in our final analysis.

Transformations:
As we stated in the initial approach, we encoded several nominal string variables into numeric variables (e.g., "club", "leaguecontry", "position", "alpha_3") to put them as fixed effects in the logistic regression model. Next, we created a new binary dependent variable. As we described above, zero indicates no red card, while 1 indicates red cards only from non-zero "redcards" response. Again, we coded "red" as zero for those who recevied red cards from two yellow cards and not from a direct red card. Also, in final analysis, we did not exclude inconsistent skin rating resopnses, but we created a new variable "skin" which is the averaged skin rating scores of the two raters. Last, same as initial approach, we created interaction terms by multiplying skin color rating and two cultural preference scores (i.e., mean IAT and explicit bias score) on lighter skin color for the Research Question 2.

Exclusions:
In final analysis, we excluded responses which have at least one missing data point for any of the variables. As a result, a total 22,313 responses were exluded.

Technical Explanations:
As in the initial approach, our anlaysis was mainly based on logistic regression. The main differences from our initial approach were that: (1) we used clusterred standard error for "refnum" since referees have multiple observations and those observations are not independent, (2) we weighted the data for the number of games since receving a red card from a single game and from multiple games are different, and (3) since "goals" seems to be an outcome of receiving red cards, we dropped this variable in the model.

Research Question 1
We ran logistic regression with clusturred standard errors for the "refnum" variable to investigate Research Question 1. Also, we added height, weight, the number of games, "position", and "refcon" variables as covariates in the logistic regression model because we are interested in looking at the association between skin color and the probability to get red cards after considering those covariate variables for the Research Question 1.

A total 362,706 responses were used in the analysis. The logistic regression result showed that the main effect of skin color was not statistically significant. (odds ration: 1.2030, p = .118, 95% CI: 0.9545, 1.5163). From our result, we can conclude that there is no evidence that skin color influences referees to give more red cards to darker skin color players than to lighter skin color players.

Research Question 2
We also ran logistic regression for Research Question 2. The main aim is to investigate the effect of implicit/explicit cultural lighter skin color preference of referees' country origin on the referees' tendency to give red cards more to darker players than to the lighter color players. In the initial approach, we examined the interaction effects between skin color rating and either mean IAT score or mean explicit bias score on the probability of getting red card.

To rule out other possible explanations, we used the same clusterred standard errors, weighting, covariates, and fixed effects as in the Research Question 1.

A total 371,626 responses were used in the analysis. Our results did not show any evidence which supports the argument that cultural preference on lighter color skin in the referees' country influences the tendency to give more red cards to darker skin players. As seen in the Table 2, the interaction between rating and mean IAT (odds ratio: 0.0030, p = .49, 95% CI: 1.74e-10, 51,016.73), the interaction between rating and explicit bias score (odds ratio: 1.5775, p = .89, 95% CI: 0.0023, 1083.99), and the three-way interaction among rating, IAT, and explicit bias score (odds ratio: 1.2026, p = .98, 95% CI: 1.38e-08, 1.05e+08) were not statistically significant.

Conclusion

We investigated two main research questions regarding referees' biases in red card decisions. First, we investigated whether or not referees' red card decisions are influenced by players' skin color using logistic regression. The result showed that players' skin color did not influence referees' red card decisions. Second, we investigated whether or not implicit and explicit cultural preference of referees' country. Our results showed that implicit and explicit cultural preference on light skin color positively influenced the probability of receiving red cards, but these effects were not moderated by players' skin color.

Tables

Table 1. Regression results for the Research Question 2 (initial analysis)

	Model 1	Model 2	Model 3
rating	0.8712	0.9908	1.1032
mean IAT	3.3459		79209.79*
rating × IAT	1.6712		0.7783
mean Explicit		0.8364	51.4335
rating × Explicit		1.1121	0.9987
IAT × Explicit			< 0.0001*
rating × IAT × Explicit			1.2377
Pseudo R-squared	0.0889	0.0885	0.0902
N	94424	94424	94424

- values indicate odds ratio
* p<.05, ** p<.01, *** p<.001

Table 2. Regression results for the Research Question 2 (final analysis)

	Model 1	Model 2	Model 3
rating	5.2783	1.4047	7.5748
mean IAT	16654.2525***		1.1958e+12***
rating × IAT	0.0147		0.0030

mean Explicit		2.6278**	16438.2690***
rating × Explicit		0.7008	1.5775
IAT × Explicit			0.0000***
rating × IAT × Explicit			1.2026
Pseudo R-squared	0.0577	0.0560	0.0599
N	371626	371626	371626

- values indicate odds ratio
* p<.05, ** p<.01, *** p<.001

Data and Output

Please upload to the Open Science Framework (OSF): https://osf.io/ and include both your initial analyses before the feedback round and your final analyses. Links to these files will appear in the published manuscript.

Instruction for Uploading to the Open Science Framework:
In addition to these instructions here is a brief video on the OSF:
https://www.youtube.com/watch?feature=player_embedded&v=c6lCJFSnMcg

1. **Create an account**. Visit the site (www.openciencefraemwork.org). Each contributor to the crowdstorm should create a personal account, by clicking the 'create an account or sign in' button in the top right corner.

2. **Create the project**. One contributor should go to the Dashboard by clicking the link on the top of the page. Create a new project by clicking the 'New Project' button. For the title, write: "Crowdstorming a dataset: Do soccer referees give more red cards to dark skin toned players? Analyses by [Team Member Names]", and then click the 'Create New Project' button.

3. **Add collaborators (Optional)**. The project creator can add collaborators by clicking the 'add' link just below the project title. Type in the name of any other team members of yours (just last name may be enough) and add them. If they have not registered, they will not appear in the search. Do not add them until they are registered. Now all your team members have editing privileges for the project.

4. **Using project space**. The project space includes tags, a wiki, files, and components/nodes. You can use any of these features as they are useful for documenting your research. Nodes operate like folder. These may be most useful to define discrete components of the research process, particularly if they have independent contributor lists, or if you'd like to be able to cite those components independently, or if you'd like to keep some parts of the project private while other parts are public (e.g., a data node that stays private until the article is accepted for publication). Each node has the same features as the project - unique contributor list, tags, wiki, files (a new "project" node creates a project within a project). Project nodes might be useful, for example, if your project consists of multiple studies. Until you click the "make public" button on the top right of any project or node, the project page is private. Only the collaborators can access the materials.

5. **Upload files**. When in the main project, go to the Files tab. Click the Upload button, or simple drag and drop files onto the webpage. The file appears in the upload list. Click the blue Start button to upload the file. If you revise a document and then upload it again with the identical filename, the OSF will retain a version history of the file. You will be able to access any prior version, and when you download it. The date the file was uploaded will be appended automatically to the filename (probably in the same way you manage file edits in your local directory).

6. **Make project public**. Click the ‘Make public’ button on the top right of the project space. If you have multiple components to make public, each one must be made public manually.

References and Notes