“There is definite racial bias in which soccer players are sent off, but its locus is unclear”

Tom Stafford*[1]†, Mathew H. Evans[2], Tim J. Heaton[3], Colin Bannard[4]

[1]Department of Psychology, University of Sheffield.

[2]Faculty of Life Sciences, University of Manchester.

[3]School of Mathematics and Statistics, University of Sheffield.

[4]Department of Linguistics, University of Texas at Austin.

*Correspondence to: t.stafford@shef.ac.uk

†TS was supported in this work by a Leverhulme Trust Research Project Grant

Abstract

Data from 335537 soccer player-referee interactions were analyzed, with the aim of discovering if player skintone and racial bias measures for referees’ countries affected sending-off decisions. Our final model was arrived at after a process of data visualization, model comparison and validation using frequentist and Bayesian methods for calculating a mixed model logistic regression. Player skintone (as well as playing position and league) significantly altered the odds a player-referee interaction would result in a sending-off, but this did not interact with either implicit or explicit bias measures of the referee’s country. Although we can observe that dark skinned players are more likely to be sent off, we have no evidence to suggest racism on the part of the referees. Inferring the mechanism by which player skintone comes to be associated with a higher likelihood of being sent off – and whether this mechanism involves racial bias on the part of any individuals, and which individuals these might be – is not possible with this dataset. An analysis which did not rely on league as a variable replicated these essential results

One Sentence Summary

Darker skinned players are more likely to be sent off the soccer pitch, but – since this is not predicted by measures of implicit or explicit bias associated with the country of the referee - the locus of this bias remains unclear.

Results

We used mixed model logistic regression, both frequentist and Bayesian. Our dependent variable was the outcome of each player-referee interaction; i.e. whether a player was sent off or not for each game. Independent variables were the individual player and referee (as random effects) and player position and skintone, game league and the interaction between skintone and referee country implicit bias and the interaction between skintone and the referee country explicit bias (as fixed effects). This model was arrived at after processes of both data visualization and model comparison. When it became apparent that the dataset included players' entire playing history, we opted to filter out data from referees who definitely were not in the selected premier leagues.

Initial Approach

Data visualization: Using ipython notebooks we created a series of visualisations, some of which were interactive. These allowed us to appreciate the scope, distribution and potential interaction of the variables in the dataset.

We noted that ratings of skintone could be more reliable. The ratings are fairly different at the light end of the spectrum. The two raters disagree on 28742 or 19% of the time, and looking at the histograms of their responses most of these are between the first two categories. These first two categories account for ~ 70% of both rater's classifications, so biases/inconsistencies/uncertainty in this part of the dataset could have a large effect on the rest of the analysis. There could be many reasons for this, but one obvious way of dealing with it would be to use N>2 raters.

Data recoding: This was done in python using the pandas library.

A variable 'skintone' was constructed by averaging rater1 and rater2. In the final analysis this was considered as a continuous variable and was centred around the mean. We also centred the implicit and explicit bias scores.

A variable 'allreds' was constructed by adding yellowReds and redCards

A variable 'allredsStrict' was created based on redCards only.

The data was recoded from player-ref dyads per row into single games per row, with allreds and allredsStrict becoming binary variables (and, obviously, the sum of these variables for each player-ref dyad remaining the same as before recoding)

We dropped all rows where there were missing values in any of:
1) the position of the player
2) the skin tone of the player

3) the bias scores of the country from which the referee came
4) the league in which the player was active

This was done on the presumption that such data were missing at random

Analysis:

Our analytic strategy employed mixed model logistic regression - both frequentist and Bayesian. Our analysis was performed in R using the lme4 package (frequentist) developed by Bates et al.; and the MCMCglmm package (Bayesian).

Logistic regression predicts the outcome on a binary variable - in this case whether a game resulted in a red card for that player. We suppose that the probability p_{ij} of any individual i obtaining a red card in a game j depends upon a number of predictor variables. These predictor variables can either be 'fixed effects' x_{ij} (which are typically of most interest and would stay the same if the experiment were to be hypothetically repeated e.g. skintone of the player) and 'random effects' (which are usually not of specific interest and may well differ in any new experiment e.g. referee identifier). In this data study we might expect there to be some referees who have a greater propensity to award cards in any game and similarly some who award fewer. Which specific referees these are is not our primary question of interest and may be different in a new set of games. We would therefore include a 'random effect' to incorporate this random variation in referee strictness. Random effects are required in order to be able to generalise the results of the analysis outside of the population of study.

Specifically we model the probability of a red card to player i in game j to be

$$\log p_{ij}/(1-p_{ij}) = \sum \beta^T_{ij} x_{ij} + w_{ij}$$

where x_{ij} are the fixed effects and w_{ij} the random effect.

The analysis was run using R, using the lme4 (for frequentist analysis) and MCMCglmm (for Bayesian) packages

Model selection was performed via Akaike's Information Criterion (AIC). Multiple models were considered before selecting that with the lowest AIC as our final model.

The player and referee were included as random effects. The player random effect aims to account for variation amongst players in propensity to be booked (independent of skin tone) and the referee random effect aims to account for variation in strictness amongst referees (again independent of player skin tone)

As fixed effects we included the covariates (as factors) of position played and league. This aimed to take account of the potentially unbalanced nature of the dataset i.e. that the four different leagues may have different numbers of cards awarded and that the skin tone of players in different leagues could be unbalanced. Similarly we felt that different positions are more/less likely to be red carded and certain positions may be more associated with particular skin tone.

One could potentially consider league as a random (instead of fixed) effect but we felt there may be specific interest in the four major European leagues and not more general. We did also investigate if there was an interaction between league and skin tone (i.e. some leagues were more likely to give red cards to certain skin tones) but found no evidence for this.

Finally as specific interest was in whether the bias score of the referee's home country was significant we also included two interactions: the first between the meanIAT (Implicit Bias) score and skin tone in our model; and the second between the mean (Explicit Bias) score and skin tone. Inclusion of both bias scores simultaneously means that one needs to be careful when interpreting results - comparisons are made conditional on all other variables being equal.

It is not therefore possible to consider the Implicit and Explicit bias effects separately. One can instead compare two countries with the same Implicit bias scores but different Explicit bias scores; or vice versa. We are therefore concerned that results could be misinterpreted due to high correlation between implicit and explicit bias. Generalised statements about a general effect of explicit bias to compare two countries (as if one can ignore also the implicit bias) must be treated with caution. We are hence presuming that in any specific comparison/prediction between countries all the fixed variables are known or at least considered to be equal. All questions were answered using this model.

We would have preferred to have had a single measure of bias but this did not seem possible according to the project brief. In our investigation we fitted separate models including explicit and implicit bias separately (and both together). Explicit bias was seen to be less significant than implicit bias - although there was very little evidence to support either in the model. As a result, if we were to choose our 'best' model we would have dropped explicit bias and calculated the odds ratio simply for implicit bias.

From background knowledge of soccer we reasoned that player identity, league, player position and referee identify would effect likelihood of a red card being awarded.

We tested a number of different models and selected that with the lowest AIC as our final model (conditional on including the terms required to answer the questions posed).

Since the data set was very large we felt it was appropriate to model the parameter estimate as normally distributed and so used +/- 1.96 * standard error to estimate the confidence intervals.

Upon request we repeated our final analysis omitting the 'league' variable

Final Approach

Our final approach was essentially similar to that we initially developed. The major change was that we decided that the issue of the data including players' entire playing histories needed addressing. Although players were selected for inclusion by virtue of playing in the Spanish, German, French or English premier leagues in the 2012-2013 season, the dataset included player games from different leagues in prior years (we deduced this by close inspection of the data for the players of Liverpool FC). This will introduce considerable heterogeneity, since referees without premier league experience will give cards in diverse contexts (both with respect to style of play and with respect to player skintone profiles) which will then contribute to the dataset. We concluded that the intention of the research question is to answer how the effect of skintone affects referee judgments within the selected premier leagues. Consequently, we pruned the data by excluding data from referees who interacted with less than 22 players (and so, by definition, cannot have refereed in the selected premier leagues in the 2012-13 season). This pruning reduces the data set by 16,958 interactions but does not significantly alter the results.

Results from both the frequentists (lme4 package) and Bayesian (MCMCglmm package) agreed. We report the exact results based on the frequentist analysis, since this was the analysis we used for model development. The full list of variables in our final model and full R output are shown in tables 2 and 3. Full specification of the models and data processing methods is provided by the code available via the Open Science Framework project page.

Conclusion

Player skintone significantly altered the odds a player-referee interaction would result in a sending-off, but this did not interact with either implicit or explicit bias measures of the referree's country. There are significant differences between player position and leagues in likelihood of being sent-off (as expected), as well as between the referee country implicit bias measure and sending-offs. This suggests that differences in referee style exist between countries and affect the likelihood of carding (although this doesn't interact with player skintone). These results are summarized in table 1.

Although we can observe that dark skinned players are more likely to be sent off, we have no evidence to suggest racism on the part of the referees. We offer a partial list of possible loci for the effect: referees may be biased, but in ways not predicted by the country-level measures we have; other players may be biased, creating or exacerbating situations where dark skinned players may be sent of; managers or trainers may be biased, encouraging a playing style in dark skinned players which increase their odds of being sent off; spectators may be biased, influencing referee decisions for actions by dark-skinned players; dark skinned players may be more likely to play in ways that get them sent off, for example because they are more likely to emerge from leagues with a 'high-carding' playing style, or because a culture of hostile racism provokes a high-carding playing style. We could go on. Although we are convinced that skintone is associated with differences in referee decisions, the identification of causes for this difference – including whether blame can be placed on any individual's racial bias – is not possible with this dataset.

Tables

Table 1. Odds ratios and 95% confidence intervals for selected predictor variables

Skintone

Variable	Odds ratio (2 d.p.)	95% CI lower	95% CI upper
Skintone *	1.31	1.10	1.56
Implicit Bias *	107.90	1.16	10059.93
Explicit Bias	0.93	0.49	1.77
Skintone x Implicit Bias	0.00	0.00	23.25
Skintone x Explicit Bias	1.84	0.49	6.84

* 95% confidence intervals of estimated odds ratio do not overlap with 1

Table 2. Model variables

	Variable name	Variable description
Dependent Variable	allreds	Is player sent off or not in this game
Fixed Effects	ContSkinTone*ImpBias	Interaction of skintone and referee country implicit bias
	ContSkinTone*ExpBias	Interaction of skintone and referee country explicit bias
	league ‡	Player league
	SpecificPos	Player playing position
Random Effects	Player	Player identity
	Ref	Referee identity

‡ omitted in a final, supplemental analysis (see below)

Table 3. Odds ratios and 95% confidence intervals for all model variables (frequentist model using lme4)

Variable	Estimate	95% CI lower	95% CI upper
ContSkinTone *	1.31	1.10	1.56
ImpBias *	107.90	1.16	10059.93
ExpBias	0.93	0.49	1.77
leagueFrance *	1.35	1.14	1.59
leagueGermany	1.11	0.95	1.29
leagueSpain *	1.55	1.33	1.80
SpecificPosCenter Back *	2.31	1.87	2.85
SpecificPosCenter Forward	1.13	0.90	1.43
SpecificPosCenter Midfielder *	1.61	1.19	2.18
SpecificPosDefensive Midfielder *	1.81	1.45	2.27
SpecificPosGoalkeeper	0.90	0.69	1.18
SpecificPosLeft Fullback *	1.60	1.24	2.06
SpecificPosLeft Midfielder *	1.39	1.05	1.83
SpecificPosLeft Winger	0.85	0.60	1.20
SpecificPosRight Fullback *	1.67	1.30	2.14
SpecificPosRight Midfielder	1.10	0.80	1.51
SpecificPosRight Winger	1.01	0.74	1.39
ContSkinTone:ImpBias	0.00	0.00	23.26
ContSkinTone:ExpBias	1.84	0.49	6.85

* 95% confidence intervals of estimated odds ratio do not overlap with 1

Table 3s. Supplemental analysis, omitting league. Odds ratios and 95% confidence intervals for all model variables (frequentist model using lme4)

Variable	(old estimate)	New Estimate	95% CI lower	95% CI upper
ContSkinTone *	1.31	1.31	1.10	1.56
ImpBias *	107.90	588.03	6.46	53488.49
ExpBias	0.93	0.98	0.52	1.83
leagueFrance *	1.35			
leagueGermany	1.11			
leagueSpain *	1.55			
SpecificPosCenter Back *	2.31	2.27	1.84	2.81
SpecificPosCenter Forward	1.13	1.11	0.87	1.40
SpecificPosCenter Midfielder *	1.61	1.52	1.13	2.06
SpecificPosDefensive Midfielder *	1.81	1.83	1.46	2.29
SpecificPosGoalkeeper	0.90	0.89	0.68	1.16
SpecificPosLeft Fullback *	1.60	1.59	1.24	2.05
SpecificPosLeft Midfielder *	1.39	1.41	1.07	1.86
SpecificPosLeft Winger	0.85	0.84	0.59	1.18
SpecificPosRight Fullback *	1.67	1.67	1.30	2.15
SpecificPosRight Midfielder	1.10	1.09	0.79	1.50
SpecificPosRight Winger	1.01	1.01	0.73	1.39
ContSkinTone:ImpBias	0.00	0.00	0.00	6.45
ContSkinTone:ExpBias	1.84	1.82	0.50	6.66

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.openciencefamework.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

McCullagh and Nelder (1989). Generalised Linear Models. Chapman and Hall

Hadfield JD (2010). MCMC methods for Multi-response Generalised Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software, 33(2), 1-22

Baayen, R. H. (2008). Analyzing linguistic data. A practical introduction to statistics using R.