European Football Leagues Show Little to No Skin-Tone Referee Bias

Bryson R. Pope[1], Nolan G. Pope[2]*

Abstract

Using data for over 300,000 player-referee-match football observations and standard logistical regression methods we look at if referees exhibit bias due to player's skin tone. We find that referees give more red cards to dark skin-toned players with an odds ratio of 1.179 (confidence interval $[0.952, 1.405]$). We find that referees give fewer yellow cards to dark skin-toned players with an odds ratio of 0.958 (confidence interval $[0.921, 0.994]$). These results should be taken cautiously due to many possible confounding variables. Using a good identification strategy that exploits the variation of different referees officiating a given player we look at if the implicit or explicit skin-tone prejudice of the referee's country affects the amount of skin-tone bias referees exhibit. We find that the implicit or explicit skin-tone prejudice of a referee's country does not affect referee's skin-tone bias.

One Sentence Summary

Small amounts of referee bias due to skin tone is found in red cards and no bias is found in yellow cards, however, these results have a poor identification strategy with no exogenous variation and therefore are likely confounded by unobserables such as playing style. With good identification we show that there is no relationship between referee country implicit or explicit skin-tone prejudice and red cards received by dark skin-toned players?

[1] Brigham Young University
[2] University of Chicago
* npope@uchicago.edu

Data

We changed the unit of observation in the data set from a player-referee dyad observation to a player-referee-game observation. Therefore, a player-referee dyad observation with 8 games was transformed into 8 different player-referee-game observations. This expanded the number of rows in the data set to equal the number of games played by players instead of the number of dyads that exist. In addition, certain control variables were transformed in minor ways. Namely we included squared values of age, height, and weight as addition controls and created an average rater value (a simple average of the two skin-tone ratings). We also create a standardized mean iat score and exp score with mean 0 and standard deviation 1 for each of the countries with iat and exp scores.

We did not exclude any observations. For 4 dyad observations, the dyad consisted of 1 game but 2 yellow cards were issued. Typically this should be coded as a yellowred or "soft" red card but shows up in the data as 2 yellow cards and no red cards. For these 4 observations we recoded the yellow card as 1 instead of 2 and left the redcards and yellowred variables as 0. Since this is such a small fraction of the cards award we see this as a trivial change.

Methodology

For question 1 our model is as follows:

$$Y_i = \alpha + \beta skintone_i + \delta X_i + \mu_i + \eta_i + \nu_i + \theta_i + \varepsilon_i \qquad (1)$$

where Y_i is the outcome of interest (e.g. number of yellow cards, hard red cards, or any red cards received) for a specific referee-match-player observation i. The variable $skintone_i$ is a categorical variable with values 1-5 ranging from "very light skin" to "very dark skin" with "neither dark nor light skin" as the center value. Additionally, X_i is a vector of player characteristics including height, height-squared, weight, weight-squared, age, and age-squared. The variables μ_i, η_i, ν_i, and θ_i are referee, league country, club, and player position fixed effects respectively.
Our logistic and nonlinear regressions include the same controls.
The controls are: height-squared, weight, weight-squared, age, age-squared, as well as referee, league country, club, and position fixed effects.

For question 2 our model is as follows:

$$Y_i = \alpha + \beta imp_i (or\ exp_i) + \rho_i + \varepsilon_i \qquad (2)$$

where imp_i and exp_i is the implicit and explicit normalized prejudice of the country from were the referee is from and ρ_i is a player fixed effect. Equations (2) is estimated separately for light skinned players (rating is 1 or 2) and dark skinned players (rating is 3, 4, or 5). The difference between these two estimations shows the relationship between referee country implicit or explicit skin-tone prejudice and red or yellow cards received by dark skin-toned players?

Results

The results can be shown in tables 1 through 12. All the results in the tables show the change in the probability of receiving the indicated card. While looking at the tables it should be

noted that red cards mean a straight red card. Any red cards include straight red cards and double yellow red cards.

Tables 1, 2 and 3 show that there is small positive effect of having dark skin on red cards, but these effects are coming from poorly identified variation and is likely excluding many confounding variables . Tables 4, 5, and 6 show similar results for any red card. Tables, 7, 8, and 9 shows a small negative effect for yellow cards. However, this effect is also poorly identified variation and is likely excluding many confounding variables. Tables 10, 11, and 12 show that there is no systematic difference in skin-tone bias exhibited by referees from countries from high and low implicitly or explicitly prejudice countries.

For question 1: A one point increase skin-tone rating is associated with a 0.000203 (0.000111) increase in the probability of getting a red card on a base of 0.00430. In other words a one point increase in skin-tone rating is associated with a 4.72% increase in the likelihood of getting a hard red card. Similarly for any red card, a one point increase skin-tone rating is associated with a 0.000340 (0.000152) increase in the probability of getting a red card on a base of 0.00430. In other words a one point increase in skin-tone rating is associated with a 4.20% increase in the likelihood of getting any red card. For a yellow card, a one point increase skin-tone rating is associated with a 0.000827 (0.000551) decrease in the probability of getting a red card on a base of 0.00430. In other words a one point increase in skin-tone rating is associated with a 0.63% decrease in the likelihood of getting a yellow card. The 95 percent confidence interval for a hard red card is [-0.000019, 0.000425]. The 95 percent confidence interval for any red card is [0.000036, 0.000644]. The 95 percent confidence interval for a yellow card is [-0.000275, 0.001929].

For question 2a: The size of the effect for hard red cards is 0.000554 (p-value = 0.313). The size of the effect for any red card is 0.000252 (p-value = 0.436). The size of the effect for yellow cards is -0.003798 (p-value = 0.25). The 95 percent confidence interval for the between referee country implicit skin-tone prejudice and a hard red cards received by dark skin-tones players is [-0.001716, 0.002814]. The 95 percent confidence interval for the between referee country implicit skin-tone prejudice and any red cards received by dark skin-tones players is [-0.002854, 0.003358]. The 95 percent confidence interval for the between referee country implicit skin-tone prejudice and yellow cards received by dark skin-tones players is [-0.015062, 0.007469].

For question 2b: The size of the effect for hard red cards is 0.001249 (p-value = 0.115). The size of the effect for any red card is 0.001444 (p-value = 0.155). The size of the effect for yellow cards is -0.005669 (p-value = 0.135). The 95 percent confidence interval for the between referee country explicit skin-tone prejudice and a hard red cards received by dark skin-tones players is [-0.000835, 0.003333]. The 95 percent confidence interval for the between referee country explicit skin-tone prejudice and any red cards received by dark skin-tones players is [-0.001400, 0.004288]. The 95 percent confidence interval for the between referee country explicit skin-tone prejudice and yellow cards received by dark skin-tones players is [-0.004619, 0.015957].

Question 1: Are soccer referees more likely to give red cards to dark skin toned players than light skin toned players?

Table 1: Linear Analysis of Skin Tone and Red Cards

Variable	(1)	(2)	(3)	(4)	(5)
Skin Tone Rating	0.000323***	0.000328***	0.000270***	0.000187*	0.000203*
	[0.000097]	[0.000098]	[0.000105]	[0.000109]	[0.000111]
Height		0.000691	0.000490	0.000712	0.001254
		[0.000775]	[0.000793]	[0.000830]	[0.000835]
Height Squared		-0.000002	-0.000001	-0.000002	-0.000003
		[0.000002]	[0.000002]	[0.000002]	[0.000002]
Weight		0.000264	0.000213	0.000208	0.000070
		[0.000283]	[0.000292]	[0.000312]	[0.000317]
Weight Squared		-0.000002	-0.000001	-0.000001	0.000000
		[0.000002]	[0.000002]	[0.000002]	[0.000002]
Age		-0.000009	-0.000134	-0.000159	-0.000162
		[0.000276]	[0.000286]	[0.000292]	[0.000311]
Age Squared		0.000001	0.000002	0.000003	0.000004
		[0.000005]	[0.000005]	[0.000005]	[0.000005]
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.00	0.00	0.01

Note: Results use the linear probability model version of equation (1).

Table 2: Logistical Regression Model Analysis of Skin Tone and Red Cards

Variable	(1)	(2)	(3)	(4)
Skin Tone Rating	0.000307***	0.000309***	0.000242***	0.000159*
	[0.000089]	[0.000089]	[0.000090]	[0.000093]
Height		X	X	X
Weight		X	X	X
Age		X	X	X
League FE			X	X
Position FE			X	X
Club FE				X
Referee FE				
Observations	373,067	371,813	371,813	369,789
R-squared	0.00	0.00	0.01	0.02

Note: Results use the logistical regression version of equation (1).

Table 3: Non-Linear Analysis of Skin Tone and Red Cards

Variable	(1)	(2)	(3)	(4)	(5)
Skin Rating 2	0.000610**	0.000639***	0.000306	0.000178	0.00011
	[0.000241]	[0.000246]	[0.000248]	[0.000270]	[0.000273]
Skin Rating 3	0.000730*	0.000812**	0.000597	0.000368	0.000175
	[0.000378]	[0.000382]	[0.000390]	[0.000407]	[0.000408]
Skin Rating 4	0.000859*	0.000926**	0.000687	0.00048	0.000454
	[0.000439]	[0.000443]	[0.000468]	[0.000486]	[0.000491]
Skin Rating 5	0.001379***	0.001349***	0.001220**	0.000973*	0.001216**
	[0.000495]	[0.000500]	[0.000517]	[0.000537]	[0.000547]
Height		X	X	X	X
Weight		X	X	X	X
Age		X	X	X	X
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.00	0.00	0.01

Note: Results use the linear probability model version of equation (1).

Table 4: Linear Analysis of Skin Tone and All Red Card Types

Variable	(1)	(2)	(3)	(4)	(5)
Skin Tone Rating	0.000470***	0.000430***	0.000412***	0.000275*	0.000340**
	[0.000133]	[0.000134]	[0.000143]	[0.000149]	[0.000152]
Height		0.001946*	0.000331	0.000201	0.000701
		[0.001028]	[0.001053]	[0.001115]	[0.001124]
Height Squared		-0.000005*	-0.000001	-0.000001	-0.000002
		[0.000003]	[0.000003]	[0.000003]	[0.000003]
Weight		0.001264***	0.000726*	0.000787*	0.000612
		[0.000379]	[0.000391]	[0.000417]	[0.000421]
Weight Squared		-0.000007***	-0.000004	-0.000004	-0.000003
		[0.000002]	[0.000003]	[0.000003]	[0.000003]
Age		0.000706*	0.000057	0.000081	0.000203
		[0.000362]	[0.000376]	[0.000386]	[0.000413]
Age Squared		-0.000013**	-0.000002	-0.000003	-0.000004
		[0.000006]	[0.000006]	[0.000007]	[0.000007]
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.00	0.00	0.01

Note: Results use the linear probability model version of equation (1).

Table 5: Logistical Regression Model Analysis of Skin Tone and All Red Card Types

Variable	(1)	(2)	(3)	(4)
Skin Tone Rating	0.000452***	0.000408***	0.000371***	0.00026**
	[0.000124]	[0.000125]	[0.000124]	[0.000129]
Height		X	X	X
Weight		X	X	X
Age		X	X	X
League FE			X	X
Position FE			X	X
Club FE				X
Referee FE				
Observations	373,067	371,813	371,813	370,152
R-squared	0.00	0.00	0.02	0.02

Note: Results use the logistical regression version of equation (1).

Table 6: Non-Linear Analysis of Skin Tone and All Red Card Types

Variable	(1)	(2)	(3)	(4)	(5)
Skin Rating 2	0.001191***	0.001272***	0.000600*	0.00044	0.000371
	[0.000336]	[0.000342]	[0.000344]	[0.000374]	[0.000378]
Skin Rating 3	0.000962*	0.001018*	0.000588	0.000288	0.000169
	[0.000515]	[0.000521]	[0.000531]	[0.000560]	[0.000564]
Skin Rating 4	0.001057*	0.000997*	0.000928	0.000542	0.000667
	[0.000593]	[0.000599]	[0.000625]	[0.000647]	[0.000654]
Skin Rating 5	0.002278***	0.002067***	0.002085***	0.001710**	0.002153***
	[0.000677]	[0.000683]	[0.000706]	[0.000737]	[0.000750]
Height		X	X	X	X
Weight		X	X	X	X
Age		X	X	X	X
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.00	0.00	0.01

Note: Results use the linear probability model version of equation (1).

Table 7: Linear Analysis of Skin Tone and Yellow Cards

Variable	(1)	(2)	(3)	(4)	(5)
Skin Tone Rating	-0.001004**	-0.002251***	-0.001243**	-0.001373**	-0.000827
	[0.000480]	[0.000485]	[0.000513]	[0.000540]	[0.000551]
Height		0.043702***	0.016172***	0.014497***	0.016387***
		[0.003797]	[0.003857]	[0.004063]	[0.004086]
Height Squared		-0.000126***	-0.000048***	-0.000043***	-0.000048***
		[0.000010]	[0.000011]	[0.000011]	[0.000011]
Weight		0.020456***	0.009204***	0.008889***	0.008289***
		[0.001374]	[0.001392]	[0.001481]	[0.001495]
Weight Squared		-0.000127***	-0.000050***	-0.000049***	-0.000045***
		[0.000009]	[0.000009]	[0.000009]	[0.000010]
Age		0.019040***	0.010029***	0.008936***	0.008794***
		[0.001322]	[0.001359]	[0.001400]	[0.001482]
Age Squared		-0.000339***	-0.000178***	-0.000159***	-0.000155***
		[0.000023]	[0.000023]	[0.000024]	[0.000025]
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.02	0.02	0.03

Note: Results use the linear probability model version of equation (1).

Table 8: Logistical Regression Model Analysis of Skin Tone and Yellow Cards

Variable	(1)	(2)	(3)	(4)
Skin Tone Rating	-0.001008**	-0.002263***	-0.001216**	-0.001182**
	[0.000491]	[0.000492]	[0.000501]	[0.000525]
Height	X	X	X	X
Weight		X	X	X
Age		X	X	X
League FE			X	X
Position FE			X	X
Club FE				X
Referee FE				
Observations	373,067	371,813	371,813	371,813
R-squared	0.00	0.00	0.02	0.03

Note: Results use the logistical regression version of equation (1).

Table 9: Non-Linear Analysis of Skin Tone and Yellow Cards

Variable	(1)	(2)	(3)	(4)	(5)
Skin Rating 2	0.010764***	0.010692***	0.006504***	0.006273***	0.005547***
	[0.001287]	[0.001302]	[0.001301]	[0.001403]	[0.001415]
Skin Rating 3	0.001275	-0.00154	-0.004509**	-0.004361**	-0.004545**
	[0.001928]	[0.001943]	[0.001979]	[0.002091]	[0.002115]
Skin Rating 4	-0.004806**	-0.008376***	-0.00282	-0.002482	-0.001395
	[0.002165]	[0.002185]	[0.002281]	[0.002363]	[0.002392]
Skin Rating 5	-0.001367	-0.006257***	-0.004223*	-0.004640*	-0.001684
	[0.002350]	[0.002369]	[0.002427]	[0.002545]	[0.002586]
Height		X	X	X	X
Weight		X	X	X	X
Age		X	X	X	X
League FE			X	X	X
Position FE			X	X	X
Club FE				X	X
Referee FE					X
Observations	373,067	371,813	371,813	371,813	371,813
R-squared	0.00	0.00	0.02	0.02	0.03

Note: Results use the linear probability model version of equation (1).

Question 2: Are soccer referees from countries high in skin-tone prejudice more likely to award red cards to dark skin toned players?

Table 10: Red Cards and the Skin-Tone Prejudice of a Referee's Country

Variable	IAT	IAT	EXP	EXP
Panel A. Light Skin-Tone Players				
Bias Rating	0.002271***	0.000459	0.001572***	-0.00001
	[0.000545]	[0.000678]	[0.000472]	[0.000577]
Player FE		X		X
Observations	283,132	283,132	283,132	283,132
R-Squared	0.00	0.01	0.00	0.01
Panel B. Dark Skin-Tone Players				
Bias Rating	0.00136*	0.001013	0.001346*	0.001239
	[0.000811]	[0.000911]	[0.000763]	[0.000868]
Player FE		X		X
Observations	89,748	89,748	89,748	89,748
R-Squared	0.00	0.01	0.00	0.01
Panel C. Difference: Dark Minus Light				
Bias Rating	-0.000911	0.000554	-0.000227	0.001249
P-Value	0.176	0.313	0.400	0.115

Note: Results use the linear probability model version of equation (2).

Table 11: Yellow Cards and the Skin-Tone Prejudice of a Referee's Country

Variable	IAT	IAT	EXP	EXP
Panel A. Light Skin-Tone Players				
Bias Rating	0.074663***	0.012939***	0.054699***	0.005865*
	[0.002908]	[0.003551]	[0.002522]	[0.003020]
Player FE		X		X
Observations	283,132	283,132	283,132	283,132
R-Squared	0.00	0.05	0.00	0.05
Panel B. Dark Skin-Tone Players				
Bias Rating	0.033167***	0.009141**	0.031566***	0.011535***
	[0.003942]	[0.004372]	[0.003708]	[0.004164]
Player FE		X		X
Observations	89,748	89,748	89,748	89,748
R-Squared	0.00	0.03	0.00	0.03
Panel C. Difference: Dark Minus Light				
Bias Rating	-0.041496***	-0.003798	-0.023133***	0.005669
P-Value	0.000	0.250	0.000	0.135

Note: Results use the linear probability model version of equation (2).

Table 12: Any Type of Red Cards and the Skin-Tone Prejudice of a Referee's Country

Variable	IAT	IAT	EXP	EXP
Panel A. Light Skin-Tone Players				
Bias Rating	0.006921***	0.002292**	0.005098***	0.001208
	[0.000759]	[0.000943]	[0.000658]	[0.000802]
Player FE		X		X
Observations	283,132	283,132	283,132***	283,132
R-Squared	0.00	0.01	0.00	0.01
Panel B. Dark Skin-Tone Players				
Bias Rating	0.004194***	0.002544**	0.003699***	0.002653**
	[0.001099]	[0.001234]	[0.001033]	[0.001175]
Player FE		X		X
Observations	89,748	89,748	89,748	89,748
R-Squared	0.00	0.01	0.00	0.01
Panel C. Difference: Dark Minus Light				
Bias Rating	-0.002727**	0.000252	-0.001399	0.001444
P-Value	0.021	0.436	0.127	0.155

Note: Results use the linear probability model version of equation (2).