Crowdstorming project on testing skin tone bias among referees: results without controlling for the club of the player

E. Bonnier*[1], K. Hederos Eriksson*[1,2], M. Johannesson[1], A. Sandberg*[3]

[1]Department of Economics, Stockholm School of Economics.

[2]Swedish Institute for Social Research (SOFI), Stockholm University.

[3]Institute for International Economic Studies (IIES), Stockholm University.

*Correspondence to: Evelina.Bonnier@hhs.se, Karin.HederosEriksson@hhs.se, and Anna.Sandberg@iies.su.se

Abstract

We examine whether a soccer player's likelihood of being awarded a red card is related to his skin color, and if the importance of skin color for this likelihood varies with the degree of racial discrimination in referees' home country. Our dependent variable is the number of red cards per game within each player/referee dyad, and our explanatory variable rates players' skin color on a 0 to 1 scale. We run weighted least squares regressions with referee fixed effects and clustered standard errors on player level, controlling for a player's height, weight, birth year, and position. In our original results we also controlled for the club of the player (measured once at the time of data generation); here we present results without controlling for the club of the player to test how that affects the results. Whether club is controlled for or not is important for the results. With a control for club the skin color variable was not significant, whereas without a control for club the skin color variable is significant in our "baseline model" (but the point estimate of the skin color variable on yellow cards still goes in the opposite direction). As in our original results the likelihood of receiving a red card is not significantly related to the degree of racial discrimination in the referee's home country.

One Sentence Summary

Whether the club of the player is controlled for is important for the results of the first research question; with a control for club the skin color variable is not significantly related to the likelihood of receiving a red card, whereas without a control for club the skin color variable is significant in our "baseline model".

Results

Estimation Strategy

Firstly, it is important to point out that the data set is insufficient to identify any causal effect of skin color on the probability of getting a red card. We can only estimate if the likelihood of getting a red card varies with skin color after controlling for as many variables as possible; but this variation may be correlated with unobserved factors affecting the likelihood of getting a red card. The most important unobserved variable here is "player style"; i.e. to what extent the player style is correlated with skin color (skin color could for instance be correlated with how physical a player plays, affecting the likelihood of getting yellow and red cards). The data set is also structured in a "non-optimal" way with the player/referee dyads. It would have been better if each game for each player was one observation.

Given the above limitations we have in the baseline used OLS regression (formally weighted least squares, see below) with the number of red cards per game as the dependent variable (i.e. the number of red cards for each player/referee dyad divided by the number of games in that dyad). This can be interpreted of as the likelihood of getting a red card in a game. We estimate the model separately for four categories of red/yellow cards: red card (a direct red card without getting a prior yellow card); yellow-red card (a red card after getting two yellow cards in the same game); direct red or yellow-red card (the sum of direct red cards and yellow-red cards); and yellow cards. We view direct red cards as the primary dependent variable; but the other variables serve as robustness tests (including also yellow cards is relevant here as if there is a "discrimination effect" it would be most plausible to observe this for both red cards and yellow cards).

To test if the likelihood of getting a red card varies with player skin color (research question 1), we include a skin color variable measured on a scale from 0 to 1 as an independent variable. The skin color variable indicates the average rating of a player's skin color given by rater 1 and rater 2, and it is included as a continuous variable. Player/referee dyads where skin color has not been rated are excluded from the analysis. We include referee fixed effects so that we estimate how the likelihood of getting a red card varies with skin color within referees. We also cluster the standard errors on player level to take into account that we have multiple observations for each player. To control for player characteristics we also control for the following variables in the regressions: height, weight, birth year, and position. As the number of games varies across player/referee dyads we weigh the observations by the number of games in each player/referee dyad (so that we formally use weighted least squares). In our original results we also controlled for the club of the player (measured once at the time of the data generation). Here we present results without controlling for club to test how this impacts the results. Club is a potentially important control variable as it controls for differences in the propensity of receiving a red card between leagues and between clubs within leagues (although it is measured imperfectly in the dataset as it is only measured at one point in time and does not capture switches between clubs and leagues).

As a robustness test we also run the models with two alternative, binary, specifications of the skin color variable. In the first of these specifications we code skin color as 1 if at least one of the two raters indicated the skin color to be larger than or equal to 0.75 (dark skin), and 0 otherwise. In the second specification we code skin color as 1 if at least one rater indicated the skin tone of the player to be equal to 1 (very dark skin), and 0 otherwise.

One problem with the above estimations is that there is a high fraction of zero observations. To test the robustness of our results we therefore also run a logit regression with the dependent variable coded as 0/1 for the above four dependent variables. More precisely, we code the four dependent variables as 1 if (i) at least one direct red card in that player/referee dyad has been awarded; (ii) at least one yellow-red card in that player/referee dyad has been awarded; (iii) at least one red or yellow-red card in that player/referee dyad has been awarded; and (iv) at least one yellow card in that player/referee dyad has been awarded. The logit model estimates the probability of a red card and makes sure that the predictions of the model are within 0-1 range. As control variables, we include player's height, weight, birth year, and position, as well as the number of games in the player/referee dyad. In our original logit results we also controlled for the club of the player, but here we present results without controlling for club to test how this impacts the results. A drawback of using the logit model in Stata (the software used for estimations) is that it is not straightforward to use referee fixed effect and cluster standard errors on player level. We therefore run the logit model without referee fixed effects but with clustered standard errors on player level.

As a final robustness check for research question 1, we test the importance of referee fixed effects by running an OLS model without referee fixed effects. In principle the number of red cards per player/referee dyad is a count variable, and therefore a count data model (e.g. the Poisson model) might have been appropriate. However, it is not straightforward to cluster standard errors on the player level in count data models in Stata and therefore we do not estimate this model.

To test the second research question (if the difference in the likelihood of getting a red card between players with different skin colors varies with the degree of discrimination in the home country of the referee) we interact the skin color variable with the variable for the discrimination in the home country of the referee. There are two different variables for this in the data set ('meanIAT' reports the mean implicit bias score using the race IAT, and 'meanExp' reports the mean explicit bias score using a racial thermometer task). We test each of these in separate regressions. Player/referee dyads where 'meanIAT' ('meanExp') is missing are excluded from the analyses addressing the effect of implicit (explicit) bias. Note that we do not run the logit model for this research question as interactions are not straightforward to interpret in the logit model.

Results

With a control for club we did not find a significant effect of the skin color variable in the OLS baseline models (for red cards the skin-color variable had a positive sign and a p-value of 1.72). However, as shown in Table 1, without controlling for club the skin color coefficient is statistically significant at the 5% level with a positive sign for direct red cards. The skin color coefficient is not significant for the yellow-red cards variable, but it is significant for the combined variable for total red cards (direct red cards plus yellow-red cards). For yellow cards the skin color variable is not significant and has a sign in the opposite direction, decreasing the plausibility that the estimated effect for red cards is related to referee discrimination.

We also calculate the odds ratio implied by the regression coefficient of the skin color variable, using direct red cards as the dependent variable. Our baseline regression estimates the number of red cards per game for a player, which can be interpreted as the probability of receiving a red card in a game. The predicted probability of receiving a red card in a game when skin color=0, at the mean of the other explanatory variables, is 0.00392. Adding the regression coefficient of 0.00124 gives us the predicted probability if skin color=1, which is 0.00516. These probabilities imply an odds ratio of 1.31687.[1] This can be compared to the non-significant odds ratio of 1.21 with a control for club in our original results. As shown in Table 1, The 95% confidence interval of the regression coefficient is [0.00027;0.00220]. This gives us an estimated probability when skin color=1 of 0.00421 (lower CI limit) and 0.00616 (upper CI limit). These estimates imply that the 95% confidence interval of the odds ratio is [1.06911;1.56513].[2]

When we categorize the skin color variable into two different categories (see Tables 2 and 3), the skin color variable is not significant for red cards anymore. So the results are sensitive to the categorization of the skin color variable. For total red cards (red or yellow-red cards) the skin color variable is still significant and for yellow cards there is still a non-significant point estimate in the other direction.

Dropping the referee fixed effects has little effect on the results as shown in Table 4. The logit model show similar results as the OLS (see Tables 5-7). Furthermore, as shown in Table 8, our OLS results are similar without weighing the observations according to the number of games within each player/referee dyad (although the point estimate of the skin color variable for red cards increases somewhat).

For the second research question we add an interaction term between the player's skin color and the degree of skin-color prejudice in the home country of the referee. The interaction is not significant in any of our estimations. When using the 'meanExp' variable to measure prejudice in the home country of the referee, the estimated coefficient of the interaction term is 0.00181 (with standard error 0.00232) for red cards. When instead using the 'meanIAT' variable

[1] The odds when skin color=1 is 0.00516/(1-0.00506)=0.00518. The odds when skin color=0 is 0.00392/(1-0.00392)=0.00394. Thus, the implied odds ratio is 0.00518/0.00394=1.31687.

[2] For the lower [upper] CI limit, the odds when skin color=1 is 0.00419/(1-0.00419)=0.00421 [0.00612/(1-0.00612)=0.00616]. As shown in footnote 1, the odds when skin color=0 is 0.00394. Thus, the implied odds ratio is 0.00421/0.00394=1.06911 [0.00616/0.00394=1.56513].

to measure prejudice, the estimated interaction coefficient is 0.00459 (standard error 0.01416). These results are reported in Tables 9 and 10. Moreover, in Tables 11-14, we show that the results are robust to excluding player/referee dyads where ‘meanExp’ and ‘meanIAT’ are based on small samples (n<100 and n<1000). Thus, for the second research question, we conclude that we cannot reject the null hypothesis of no effect.

Differences between initial and final approach

In our initial analytical approach we used red or yellow-red cards per game, instead of red cards per game, as the main dependent variable. We also used a different scale of the skin color rating (1-5 instead of 0-1). As mentioned above we also controlled for club in our preferred original specification.

Conclusion

To test if the likelihood of being awarded a red card varies with soccer players' skin color, we estimate a weighted least squares model using the number of red cards per game within each player/referee dyad as the main dependent variable. We control for players' height, weight, birth year, and position, including referee fixed effects, and clustering standard errors on player level. In our original results we also controlled for the club of the player, which is our preferred specification. Here we present results without controlling for the club of the player to test how that affects the results. Whether club is controlled for or not is important for the results. With a control for club the skin color variable was not significant, whereas without a control for club the skin color variable is significant in our baseline analysis for red cards. But this result is sensitive to the categorization of the skin color variable (categorizing the skin color variable into two categories rather than using a continuous measure lead to non-significant results for direct red cards). If yellow cards are used as the dependent variable the point estimate of the skin color variable also goes in the opposite direction. Overall we find no robust evidence of soccer players' skin color being related to the likelihood of being awarded a red card.

To test if the difference in the likelihood of getting a red card between players with different skin colors varies with the degree of discrimination in the home country of the referee, we interact the skin color variable with two different measures of racial discrimination in the referees' home country. This interaction is not significant in any of our estimations, and we cannot reject the null hypothesis of no effect.

Tables

Table 1. OLS regressions: Effect of skin color.
Skin color is included as a continuous variable, taking on the average value of rater 1 and rater 2 on a 0-1 scale.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00124*	0.00197**	0.00074	-0.00374
	(0.00049)	(0.00076)	(0.00052)	(0.00539)
	[0.00027 - 0.00220]	[0.00049 - 0.00346]	[-0.00027 - 0.00175]	[-0.01431 - 0.00683]
Constant	0.09592	-0.00140	-0.09732	0.02224
	(0.07076)	(0.11308)	(0.07066)	(0.85026)
	[-0.04288 - 0.23472]	[-0.22320 - 0.22041]	[-0.23591 - 0.04128]	[-1.64554 - 1.69002]
R^2	0.03	0.03	0.03	0.08
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parentheses clustered on player. Confidence interval in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 2. OLS regressions: Effect of skin color.
Skin color is included as a binary variable ("Dark skin"), taking on the value 1 if at least one of the raters indicated the skin tone to be equal to 0.75 or more, and 0 if both raters indicated the skin tone to be below 0.75.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Dark skin	0.00068	0.00117*	0.00049	-0.00558
	(0.00038)	(0.00056)	(0.00038)	(0.00395)
	[-0.00007 - 0.00143]	[0.00007 - 0.00228]	[-0.00026 - 0.00124]	[-0.01332 - 0.00216]
Constant	0.09805	0.00349	-0.09456	-0.04571
	(0.07069)	(0.11288)	(0.07072)	(0.84975)
	[-0.04061 - 0.23671]	[-0.21793 - 0.22490]	[-0.23328 - 0.04415]	[-1.71248 - 1.62107]
R^2	0.03	0.03	0.03	0.08
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence interval in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 3. OLS regressions: Effect of skin color
Skin color is included as a binary variable (“Very dark skin”), taking on the value 1 if at least one of the raters indicated the skin tone to be equal to 1, and 0 if both raters indicated the skin tone to be below 1.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Very dark skin	0.00090	0.00171*	0.00081	-0.00377
	(0.00052)	(0.00076)	(0.00054)	(0.00532)
	[-0.00011 - 0.00191]	[0.00021 - 0.00320]	[-0.00024 - 0.00186]	[-0.01421 - 0.00668]
Constant	0.09075	-0.00828	-0.09902	0.03260
	(0.07109)	(0.11328)	(0.07045)	(0.84991)
	[-0.04870 - 0.23019]	[-0.23048 - 0.21392]	[-0.23721 - 0.03916]	[-1.63449 - 1.69969]
R^2	0.03	0.03	0.03	0.08
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence interval in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 4. OLS regressions: Effect of skin color (continuous: 0-1).
No referee fixed effects.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00129**	0.00158*	0.00029	-0.01070
	(0.00048)	(0.00076)	(0.00051)	(0.00556)
	[0.00036 - 0.00222]	[0.00010 - 0.00306]	[-0.00071 - 0.00130]	[-0.02161 - 0.00021]
Constant	0.03275	-0.06528	-0.09803	-0.29442
	(0.06857)	(0.10880)	(0.06178)	(0.83354)
	[-0.10175 - 0.16725]	[-0.27870 - 0.14813]	[-0.21922 - 0.02316]	[-1.92939 - 1.34055]
R^2	0.00	0.00	0.00	0.04
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Standard errors in parenthesis clustered on player. Confidence Intervals in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 5. Logit regression (binary dependent variables): Effect of skin color.
Skin color is included as a continuous variable, taking on the average value of rater 1 and rater 2 on a 0-1 scale.

	red card (binary)	red or yellow-red card (binary)	yellow-red card (binary)	yellow card (binary)
Skin color (0-1)	0.28767**	0.19590*	0.07430	-0.05089
	(0.10515)	(0.08768)	(0.12140)	(0.05489)
	[0.00289]	[0.00372]	[0.00061]	[-0.00959]
	<0.00105>	<0.00166>	<0.00100>	<0.01034>
	{0.00082 - 0.00496}	{0.00047 - 0.00697}	{-0.00134 - 0.00256}	{-0.02986 - 0.01068}
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Standard errors of coefficients in parentheses. Marginal effects (holding all other variables at their mean value) in square brackets. Standard errors of marginal effects in angle brackets. Confidence interval of marginal effects in curly brackets. Binary dependent variable: =1 if >0 cards of type 'X' within the player/referee dyad, =0 otherwise. Controls for: player birth year ('birth_year'), number of games the referee and player encountered each other ('games'), position FE ('position'), height and weight. Standard errors clustered on player.

Table 6. Logit regression (binary dependent variables): Effect of skin color.
Skin color is included as a binary variable ("Dark skin"), taking on the value 1 if at least one of the raters indicated the skin tone to be equal to 0.75 or more, and 0 if both raters indicated the skin tone to be below 0.75.

	red card (binary)	red or yellow-red card (binary)	yellow-red card (binary)	yellow card (binary)
Dark skin	0.10179	0.04508	-0.02555	-0.11035**
	(0.08532)	(0.06800)	(0.09331)	(0.04126)
	[0.00103]	[0.00086]	[-0.00021]	[-0.02079]
	<0.00086>	<0.00129>	<0.00076>	<0.00778>
	{-0.00066 - 0.00271}	{-0.00167 - 0.00339}	{-0.00171 - 0.00129}	{-0.03604 - -0.00553}
N	123,868	123,868	123,868	123,868

* *p*<0.05; ** *p*<0.01

Standard errors of coefficients in parentheses. Marginal effects (holding all other variables at their mean value) in square brackets. Standard errors of marginal effects in angle brackets. Confidence interval of marginal effects in curly brackets. Binary dependent variable: =1 if >0 cards of type 'X' within the player/referee dyad, =0 otherwise. Controls for: player birth year ('birth_year'), number of games the referee and player encountered each other ('games'), position FE ('position'), height and weight. Standard errors clustered on player.

Table 7. Logit regression (binary dependent variables): Effect of skin color.
Skin color is included as a binary variable ("Very dark skin"), taking on the value 1 if at least one of the raters indicated the skin tone to be equal to 1, and 0 if both raters indicated the skin tone to be below 1.

	red card (binary)	red or yellow-red card (binary)	yellow-red card (binary)	yellow card (binary)
Very dark skin	0.10699	0.08004	0.04086	-0.09149
	(0.10853)	(0.09079)	(0.12830)	(0.05976)
	[0.00108]	[0.00152]	[0.00033]	[-0.01724]
	<0.00109>	<0.00172>	<0.00105>	<0.01126>
	{-0.00106 - 0.00322}	{-0.00186 - 0.00490}	{-0.00173 - 0.00240}	{-0.03930 - 0.00482}
N	123,868	123,868	123,868	123,868

* *p*<0.05; ** *p*<0.01

Standard errors of coefficients in parentheses. Marginal effects (holding all other variables at their mean value) in square brackets. Standard errors of marginal effects in angle brackets. Confidence interval of marginal effects in curly brackets. Binary dependent variable: =1 if >0 cards of type 'X' within the player/referee dyad, =0 otherwise. Controls for: player birth year ('birth_year'), number of games the referee and player encountered each other ('games'), position FE ('position'), height and weight. Standard errors clustered on player.

Table 8. OLS regressions: Effect of skin color (continuous: 0-1).
No weights.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00156**	0.00250**	0.00093	-0.00129
	(0.00059)	(0.00089)	(0.00063)	(0.00542)
	[0.00040 - 0.00273]	[0.00075 - 0.00424]	[-0.00030 - 0.00216]	[-0.01192 - 0.00933]
Constant	0.11785	0.01416	-0.10369	-0.28634
	(0.08698)	(0.12256)	(0.07895)	(0.81475)
	[-0.05276 - 0.28847]	[-0.22625 - 0.25456]	[-0.25855 - 0.05116]	[-1.88445 - 1.31178]
R^2	0.04	0.04	0.04	0.06
N	123,868	123,868	123,868	123,868

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects.
Standard errors in parenthesis clustered on player. Confidence Intervals in square brackets.

Table 9. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean explicit bias in the referee's home country.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00047	-0.00015	-0.00063	-0.00925
	(0.00109)	(0.00157)	(0.00100)	(0.00898)
	[-0.00166 - 0.00260]	[-0.00324 - 0.00293]	[-0.00259 - 0.00134]	[-0.02687 - 0.00838]
Skin col X Mean Exp	0.00181	0.00505	0.00324	0.01324
	(0.00232)	(0.00357)	(0.00255)	(0.01778)
	[-0.00274 - 0.00636]	[-0.00195 - 0.01204]	[-0.00176 - 0.00823]	[-0.02164 - 0.04812]
Constant	0.09527	-0.00339	-0.09866	0.02569
	(0.07069)	(0.11289)	(0.07061)	(0.85026)
	[-0.04339 - 0.23393]	[-0.22481 - 0.21804]	[-0.23716 - 0.03984]	[-1.64208 - 1.69345]
R^2	0.03	0.03	0.03	0.08
N	123,715	123,715	123,715	123,715

* p<0.05; ** p<0.01

Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervlls in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 10. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean implicit bias in the referee's home country.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	-0.00033	-0.00860	-0.00826	-0.02345
	(0.00482)	(0.00733)	(0.00494)	(0.04737)
	[-0.00979 - 0.00912]	[-0.02298 - 0.00578]	[-0.01795 - 0.00142]	[-0.11637 - 0.06947]
Skin col X Mean IAT	0.00459	0.03089	0.02630	0.05777
	(0.01416)	(0.02172)	(0.01490)	(0.13866)
	[-0.02318 - 0.03236]	[-0.01171 - 0.07349]	[-0.00293 - 0.05553]	[-0.21421 - 0.32975]
Constant	0.09567	-0.00332	-0.09899	0.02721
	(0.07075)	(0.11287)	(0.07053)	(0.85047)
	[-0.04310 - 0.23444]	[-0.22471 - 0.21807]	[-0.23734 - 0.03935]	[-1.64097 - 1.69540]
R^2	0.03	0.03	0.03	0.08
N	123,715	123,715	123,715	123,715

* p<0.05; ** p<0.01
Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position (FE for 'position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervals in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games).

Table 11. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean explicit bias in the referee's home country. Excluding observations with nexp<100.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00034	-0.00015	-0.00049	-0.01183
	(0.00120)	(0.00173)	(0.00109)	(0.00973)
	[-0.00202 - 0.00270]	[-0.00354 - 0.00325]	[-0.00261 - 0.00164]	[-0.03091 - 0.00725]
Skin col X Mean Exp	0.00224	0.00509	0.00285	0.01946
	(0.00261)	(0.00391)	(0.00271)	(0.02016)
	[-0.00288 - 0.00737]	[-0.00258 - 0.01276]	[-0.00246 - 0.00815]	[-0.02009 - 0.05900]
Constant	0.09869	0.00080	-0.09789	0.04637
	(0.07080)	(0.11321)	(0.07070)	(0.85107)
	[-0.04018 - 0.23756]	[-0.22126 - 0.22286]	[-0.23656 - 0.04078]	[-1.62300 - 1.71574]
R^2	0.03	0.03	0.03	0.08
N	121,943	121,943	121,943	121,943

* p<0.05; ** p<0.01

Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervals in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games). Only using observations for which mean_exp was determined from a sample with nexp >=100.

Table 12. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean explicit bias in the referee's home country. Excluding observations with nexp<1000.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00105	0.00002	-0.00103	-0.01670
	(0.00166)	(0.00247)	(0.00164)	(0.01427)
	[-0.00220 - 0.00430]	[-0.00482 - 0.00486]	[-0.00425 - 0.00220]	[-0.04469 - 0.01129]
Skin col x Mean Exp	0.00056	0.00453	0.00397	0.03032
	(0.00375)	(0.00598)	(0.00432)	(0.03394)
	[-0.00679 - 0.00791]	[-0.00720 - 0.01626]	[-0.00450 - 0.01244]	[-0.03626 - 0.09690]
Constant	0.09442	-0.01430	-0.10871	0.01806
	(0.07141)	(0.11548)	(0.07320)	(0.87118)
	[-0.04565 - 0.23449]	[-0.24081 - 0.21222]	[-0.25230 - 0.03487]	[-1.69075 - 1.72687]
R^2	0.03	0.03	0.02	0.08
N	104,880	104,880	104,880	104,880

* $p<0.05$; ** $p<0.01$

Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position FE ('position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervalls in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games). Only using observations for which mean_exp was determined from a sample with nexp>=1000.

Table 13. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean implicit bias in the referee's home country. Excluding observations with niat<100.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	-0.00054	-0.00975	-0.00921	-0.04802
	(0.00601)	(0.00902)	(0.00601)	(0.05833)
	[-0.01232 - 0.01124]	[-0.02746 - 0.00795]	[-0.02101 - 0.00258]	[-0.16243 - 0.06640]
Skin col X Mean IAT	0.00533	0.03432	0.02899	0.12968
	(0.01762)	(0.02664)	(0.01801)	(0.17095)
	[-0.02923 - 0.03988]	[-0.01794 - 0.08658]	[-0.00633 - 0.06432]	[-0.20563 - 0.46499]
Constant	0.09922	0.00066	-0.09856	0.04448
	(0.07087)	(0.11317)	(0.07059)	(0.85105)
	[-0.03979 - 0.23823]	[-0.22132 - 0.22264]	[-0.23702 - 0.03990]	[-1.62484 - 1.71379]
R^2	0.03	0.03	0.03	0.08
N	121,911	121,911	121,911	121,911

* p<0.05; ** p<0.01

Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position (FE for 'position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervals in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games). Only using observations for which mean_iat was determined from a sample with niat >=100.

Table 14. OLS Regressions: Effect of skin color (continuous: 0-1) interacted with mean implicit bias in the referee's home country. Excluding observations with niat<1000.

	red cards per game	red or yellow-red cards per game	yellow-red cards per game	yellow cards per game
Skin color (0-1)	0.00207	-0.00646	-0.00852	-0.08806
	(0.00852)	(0.01327)	(0.00901)	(0.08607)
	[-0.01464 - 0.01878]	[-0.03249 - 0.01957]	[-0.02620 - 0.00915]	[-0.25688 - 0.08075]
Skin col x Mean IAT	-0.00231	0.02441	0.02672	0.24534
	(0.02484)	(0.03910)	(0.02692)	(0.25374)
	[-0.05103 - 0.04642]	[-0.05229 - 0.10111]	[-0.02608 - 0.07952]	[-0.25237 - 0.74305]
Constant	0.09478	-0.01401	-0.10879	0.01495
	(0.07152)	(0.11547)	(0.07308)	(0.87066)
	[-0.04550 - 0.23506]	[-0.24049 - 0.21248]	[-0.25214 - 0.03457]	[-1.69284 - 1.72274]
R^2	0.03	0.03	0.02	0.08
N	104,880	104,880	104,880	104,880

* p<0.05; ** p<0.01

Dependent variable: cards per game ('X Cards'/'games'). Controls for: player birth year ('birth_year'), position (FE for 'position'), height and weight. Referee fixed effects. Standard errors in parenthesis clustered on player. Confidence Intervall in square brackets. Weight: number of games for the given player/referee dyad (STATA command: aweight=games). Only using observations for which mean_iat was determined from a sample with niat >=1000.