Hierarchical Log Linear Models highlights associations of no practical significance between red cards and players' skin tone.

Michelangelo Vianello[1]*, Egidio Robusto[1]*, Pasquale Anselmi[1]*, Luca Stefanutti[1]*, Anna Dalla Rosa[1]*

[1]University of Padua.

*Correspondence to: Michelangelo.vianello@unipd.it.

Abstract

We found no clear evidence of an association between the tone of players' skin and the number of red cards received in their career. Comparing nested log-linear models, we found that the interaction between skin tone and red cards produces a very small increase in model fit ($\Delta G^2(4)=10.63$, $p=.031$, $OR_{max}=1.016$). Standardized parameters of the interaction are never significant ($Z_{max}=1.859$). When testing whether the mean implicit bias present in referees' countries of origin moderates the association effect, nested model comparisons highlighted a significant improvement in model fit, but the direction of the effect is not clearly interpretable in terms of racial bias. Indeed, lower levels of implicit bias are related to more red cards in very light-tone players and less red cards in light-tone players. When present, associations are so small in size that doubt is cast on their practical significance.

One Sentence Summary

Although some group of players with the same skin tone do show lower or higher than expected proportions of red cards, we found no clearly interpretable evidence of bias.

Results

Initial Approach

Given the current dataset and the three research questions, we firstly specified a negative binomial mixed model with players as a random factor and skin tone as fixed factor, but the estimation failed.

Hence, we decided to estimate models that are less computationally intensive. This led us to create a 3-way red cards frequency table with League (4 levels), Skintone (5 levels) and Matches (2 levels: with and without red cards) and to specify a first log-linear model of this form: $\ln(F_{ijk})=\lambda+ \lambda_L+ \lambda_S+ \lambda_M+ \lambda_{ML}+ \lambda_{SL,}$ which is nested in the second model: $\ln(F_{ijk})=\lambda+ \lambda_L+ \lambda_S+ \lambda_M+ \lambda_{ML}+ \lambda_{SL} + \lambda_{SM}$. A comparison between the models tests research question 1. A similar procedure has been followed for research questions 2a and 2b. For these questions, the models have been estimated on a frequency table in which the two moderators (dichotomized by means of a median split) have been added, thus generating a 5-way frequency table. As far as the skin tone ratings, we analyzed only players that were rated at the same level by the two raters. This was done because 1) raters rated a picture not a real player; 2) the picture could have been taken in different time periods across the year (e.g.) just before the league, after a holiday at the beach for only some players), resulting in insufficient or unclear information for making a judgment and because the percentage of perfect agreements was less than 80%. We thought it was important to analyze the data using a highly reliable measure of our dependent variable. We compared nested models using the difference in the likelihood ratio chi-square statistics of the models being compared. Results are provided in Table 1. Given the very high number of observations, it would be wise to adopt a conservative approach of alpha=.01. We found a significant interaction between Skintone, Matches, and IAT. Standardized parameter estimates and effect sizes of this interaction are provided in Table 2a and 2b.

Final Approach

We received one comment lamenting that dichotomization of the moderator variables should be avoided and that dropping ratings without agreement does not solve the problem of inaccurate ratings. We perfectly agree with these concerns, but we would like to note that 1) as we stated in the first part of this results section, log-linear models are our second choice because estimation procedures of a mixed model failed and 2) we believe that there is no risk of low power in these analyses, so using only perfect agreements increases the –otherwise low- reliability of our independent variable. So we made no changes to our initial approach.

Conclusion

These results highlight that there exist some association between players with similar skin tone, but 1) effect sizes are very small, casting doubt on the practical significance of these effects and 2) these association are not easily interpretable in terms of bias, which would reasonably hypothesize that dark and very dark skinned players get the highest proportion of red cards. We did not find evidence of this effect.

Tables

Table 1. Nested models comparisons for testing the three research questions.

	Model	G^2	*df*	*p*	*AIC*	*Change*	E_L	*OR=MAX(exp(λ))*
1	RedCards*League (YL) + Skintone*League (SL)	343039.38	136	0.0000	344344.6		.075977	
2	RedCards*League (YL) + Skintone*League (SL) + **Skintone*Redcards** (SY)	343028.75	132	0.0000	344342.0	$\Delta G^2(4)=10.63$, p=.031;	.075974	**OR**$_{SY}$=1.016; 95% CI [.999, 1.033]
3	RedCards*League (YL) + RedCards*IAT (YI) + Skintone*League (SL) + SKintone*IAT (SI)	321747.31	130	0.0000	323064.5		0.069679	
4	RedCards*League (YL) + RedCards*IAT (YI) + Skintone*League (SL) + SKintone*IAT (SI)+ **Skintone*Redcards** (SY)	321722.25	126	0.0000	323047.4	$\Delta G^2(4)=25.03$, p<.001	0.070195	**OR**$_{SYI}$=.972; 95% CI [.961, .982] *
	IAT moderates the association between Skintone and redcards (at α<.01) such that referees coming from nations with a high implicit preference for light-skinned people tend to give less redcards to skintone=1 (Very light) and more to skintone=2 (light). Skintone categories=3,4,5 show n.s. association.							
5	RedCards*League (YL) + RedCards*EXP (YE) + Skintone*League (SL) + SKintone* EXP (SE)	328613.03	130	0.0000	318809.9		0.068864	
6	RedCards*League (YL) + RedCards* EXP (YE) + Skintone*League (SL) + SKintone* EXP (SE)+ **Skintone*Redcards** (SY)	328603.00	126	0.0000	318807.3	$\Delta G^2(4)=10.03$, p=.04	0.068665	**OR**$_{SYE}$=.96; 95% CI [.943, .976] **

Notes: E_L is the effect size of the likelihood ratio chi square statistic (G^2) as defined by Johnston, Berry & Mielke (2006). Odds Ratios provided in the last column have been computed by means of the exponentiation rule of the lambda parameters. The value provided is the strongest among 5 (see Table 2b).
* LOWER implicit bias=LOWER association between red cards and skin tone = 2 (Light skin tone)
** LOWER explicit bias=HIGHER association between red cards and Skin tone categories 4 and 5 (dark and very dark tone)

Table 2a Standardized lambda parameters measuring the association between Skin tone categories, IAT and Matches on the DV Redcards

			Skintone category				
			Very light	Light	Medium	Dark	Very dark
IAT		***Matches***					
Low		*With redcards*	5.259	-5.708	-1.476	2.596	-1.184
		Without redcards	-5.259	5.708	1.476	-2.596	1.184
High		*With redcards*	-5.259	5.708	1.476	-2.596	1.184
		Without redcards	5.259	-5.708	-1.476	2.596	-1.184

Table 2b Odds ratios (exp(lambda)) measuring the association between Skin tone categories, IAT and Matches on the DV Red cards

			Skintone category				
			Very light	Light	Medium	Dark	Very dark
IAT		***Matches***					
Low		*With redcards*	1.028	0.972	0.988	1.023	0.99
		Without redcards	0.972	1.029	1.012	0.977	1.01
High		*With redcards*	0.972	1.029	1.012	0.977	1.01
		Without redcards	1.028	0.972	0.988	1.023	0.99

Data and Output

Input and output BMDP files are publicly available at https://osf.io/haefk/

References and Notes

Johnston, J. E., Berry, K. J., & Mielke Jr, P. W. (2006). Measures Of Effect Size For Chi-Squared And Likelihood-Ratio Goodness-Of-Fit Tests. *Perceptual and motor skills, 103*(2), 412-414.

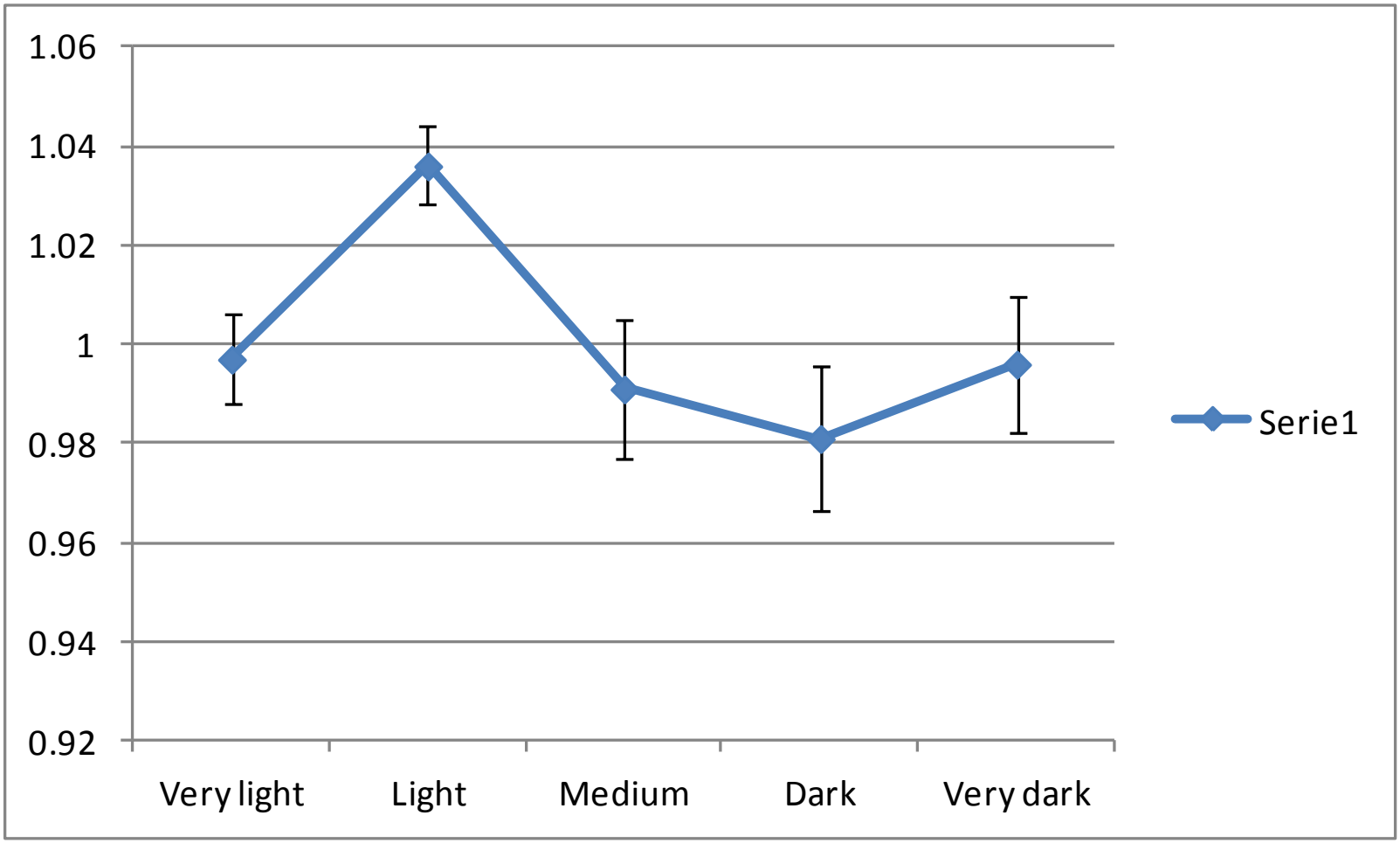
Supplement

Replication of the analysis without covariates (parameter estimates and confidence intervals)

	Model	G^2	*df*	*p*	*OR=MAX(exp(λ))*	*Association Involved in OR*
1	Skintone*RedCards (SY)	42172.3	30	0.00	**OR**$_{SY}$=.965; 95% CI [.958, .973]	**Skintone cat=2 (light)*RedCards=No**
2	Skintone*RedCards (SY) + Skintone*IAT (SI) + **Skintone*Redcards*IAT** (SYI)	20785.29	20	0.00	**OR**$_{SYI}$=.972; 95% CI [.961, .982]	**Skintone cat=2(light)*Redcards=NO*IAT=High**
3	RedCards*League (YL) + RedCards*IAT (YI) + Skintone*League (SL) + Skintone*IAT (SI)	27689.00	20	0.00	**OR**$_{SYE}$=.969; 95% CI [.962, .977]	**Skintone cat=2(light)*Redcards=NO*EXP=High**

Notes: the direction of the effects provides support to the hypothesis that Skintone=Light receives MORE red cards than expected.

Figure S1. Odds ratio by Skin Tone category. Higher values represent higher frequency of red cards than expected under the assumption of independence.



Error bars represent 95% symmetric confidence intervals