Does race color our evaluation of others? We provide new evidence on racial biases in evaluation, by examining how the number of fouls awarded against black and white NBA players varies with the racial composition of the refereeing crew. Our setting provides intriguing insights into own-race bias: relative to social, judicial, or labor market settings, the evaluators in our sample (NBA referees) are a particularly expert group, with substantial experience, continual feedback, and they face robust career incentives to be accurate. Indeed NBA Commissioner Stern has claimed that these referees are “the most ranked, rated, reviewed and statistically analyzed and mentored group of employees of any company in any place of the world.”

NBA referees are effectively randomly assigned to each game. Moreover the number of games played is large so we can assess both a very clear baseline rate at which individual players commit fouls and also a clear baseline for the number of fouls called by different referees. Against this baseline we fine systematic evidence of own-race bias. Players earn up to 4 percent fewer fouls or score up to 2% more points when they are recipients of a positive own-race bias rather than a negative opposite race effect.

In this regression, the dependent variable is fouls per 48 minutes played. The independent variable is percentage of white referees in the game (there are three referees). The data included all NBA games played from 1991-2004 yielding over a quarter of a million player-game combinations. For black players the coefficient of the regression of fouls per 48 minutes on % of white referees (0 . .33 .67, 1) is 0.182 (0.066).

(From Price and Wolfers, Racial Discrimination among NBA Referees, preliminary version, May, 2007.)

1. What are the individuals and variables in this study? For categorical variables, name the levels. For quantitative variables, name the units.

2. Do the data result from a random sample from some population and/or a randomized comparative experiment? Or are they from an observational study? If a random sample, is it a simple random sample or is it some more complicated design such as a stratified random sample? If a randomized comparative experiment, describe the design. If from an observational study, is the analysis treating the data as if it came from a random sample or from a comparative experiment?

3. What is the parameter about which we are making inferences? (Which Greek letters are customarily used for these parameters and what do they correspond to in terms of the population?)

4. Are the inferences confidence intervals, hypothesis tests, or some other kind of inference?

5. What is the standard statistical inference technique for an inference in this situation? Is there any evidence as to which technique was used in this paper? What statistics are reported and computed in the report (and what are the customary symbolic names for these statistics)?

6. The statistical technique used makes assumptions about the population distribution. What are these assumptions? Are we justified in using this technique in this case?

7. Explain clearly the meaning of the $p$-value or confidence interval in the context of the particular research question.