University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 400

Replication: Do Coaches Stick With What Barely Worked? Evidence of Outcome Bias in Professional Sports

Pascal Flurin Meier, Raphael Flepp, and Egon Franck

February 2023

University of Zurich, Plattenstrasse 14, CH-8032 Zurich,
http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Pascal Flurin Meier

pascal-flurin.meier@business.uzh.ch

Raphael Flepp

raphael.flepp@business.uzh.ch

Egon Franck

egon.franck@business.uzh.ch

University of Zurich

Department of Business Administration

Plattenstrasse 14, CH-8032 Zurich, Switzerland

Replication: Do Coaches Stick With What Barely Worked? Evidence of Outcome Bias in Professional Sports

Pascal Flurin Meier, Raphael Flepp, and Egon Franck*

February 20, 2023

Abstract: Consistent with outcome bias, we replicate the finding of Lefgren et al. (2015) showing that professional basketball coaches in the NBA discontinuously change their starting lineup more often after narrow losses than after narrow wins, even though this outcome is conditionally uninformative. As our paper shows, this pattern is not restricted to the NBA; we find evidence of outcome bias in the top women's professional basketball league and college basketball. Finally, we show that outcome bias in coaching decisions generalizes to the National Football League (NFL). We conclude that outcome bias is credible and robust, although it has weakened over time.

Keywords: Outcome bias; Strategy revision; Regression discontinuity design; Replication

JEL Classification: D81; D83; D91; Z20

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest: None

Acknowledgments: The authors thank Michael Toren for excellent research assistance.

---

*Corresponding author: Pascal Flurin Meier, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland, pascal-flurin.meier@business.uzh.ch. Coauthors: Raphael Flepp, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland; and Egon Franck, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland.

1 INTRODUCTION

"Never change a winning game; always change a losing one."

Bill Tilden (n.d.), former American tennis player

Outcome bias describes the phenomenon in which decisions are evaluated more favorably in light of a positive outcome than a negative outcome even if the outcome is determined by factors outside of the agent's control (Baron & Hershey, 1988; Gino et al., 2008). The study by Lefgren et al. (2015, henceforth LPP) was the first to demonstrate outcome bias outside of laboratory experiments in a high-stakes field setting. In particular, LPP investigate the change in starting lineup decisions of National Basketball Association (NBA) coaches after narrow wins compared to narrow losses, which is an uninformative outcome regarding the team's performance. Using data from the 1991-2010 seasons, their main results show that coaches are between 5.2 and 6.1 percentage points less likely to change their starting lineup after narrow wins, which is consistent with coaches suffering from outcome bias.

In this paper, we reexamine the main analysis of LPP and extend it in several aspects. First, we expand the data horizon of professional NBA matches up to the 2019/20 season. Second, we investigate whether outcome bias in coaches' starting lineup decisions can also be observed in other basketball leagues, i.e., in the Women's National Basketball Association (WNBA) and in the first division of the National Collegiate Athletic Association (NCAA). Third, we test whether outcome bias in coaching decisions generalizes to the National Football League (NFL). Fourth, we test the persistence of outcome bias over time. Finally, we use a more sophisticated methodological approach following recent developments in regression discontinuity designs (Calonico et al., 2014a, 2014b; Calonico et al., 2017; Calonico et al., 2020). Thus, in the taxonomy of Mueller-Langer et al. (2019), we conduct a "scientific

replication" using new data and new methods to assess the robustness and credibility of outcome bias among sports coaches.

2 EMPIRICAL STRATEGY

We closely follow LPP in calculating the main variables and in setting up the regression discontinuity model. The regression discontinuity design (RDD) has emerged as one of the most credible nonexperimental strategies for the analysis of causal effects. It considers whether a score value of units lies on either side of a fixed threshold (Imbens & Lemieux, 2008), which assigns units to the treatment group if their score equals or exceeds the cutoff value and to the control group otherwise (Cattaneo et al., 2019). Relying on the assumption that all other determinants of the dependent variable vary smoothly through the threshold, the discontinuity in the dependent variable can be considered the local causal effect of the treatment (Cattaneo et al., 2019; Imbens & Lemieux, 2008).

Exploiting the quasi-experimental setting provided by narrow final match scores in high-scoring sports, we follow LPP and study whether coaches discontinuously change their strategy following a narrow loss compared to a narrow win. In high-scoring sports, it is quasi-random whether a team ultimately wins or loses by a narrow margin (e.g., by one point): Thus, narrow wins or losses are uninformative, and the outcome does not entail diagnostic information about the efficacy of a coach's strategy beyond the score itself. This suggests that a rational, non-outcome-biased coach should not be influenced by such uninformative performance signals.

To test outcome bias in coaching decisions, we estimate the following RD baseline regression model:

$$ChangingStarters_{i,g+1} = \alpha + \beta_1 Win_{i,g} + \beta_2 ScoreDiff_{i,g} + \beta_3 Win_{i,g} \times ScoreDiff_{i,g} + \varepsilon_{i,g} \qquad \textbf{(1)}$$

where i denotes the team and g denotes the match within a particular season. The dependent variable *ChangingStarters* takes the value of one if a coach changes the starting lineup of players in the subsequent match $g+1$ compared to the focal game g and zero if he or she fields the same players. Although coaches' strategic decisions are not limited to the starting lineup, the choice of the starting lineup is a key decision that is cleanly measurable LPP. *ScoreDiff* denotes the final point difference at the end of the match between the home and away teams. The score difference serves as the running variable and assigns teams to the control status if the point difference is negative (i.e., the focal team lost) and to the treatment status if the point difference is positive (i.e., the focal team won). The treatment status is denoted by *Win*, taking the value of one if a team won the game and zero if a team lost the game. To allow for different slopes on either side of the threshold, we interact the running variable *ScoreDiff* with the treatment indicator *Win*. The coefficient of interest is β_1, which is set to capture the (local) causal effect of the uninformative outcome of barely winning the game on the decision to change the starting lineup.

In all specifications, we cluster standard errors at the team-season level. Although covariates are conceptually not needed in an RDD, we follow LPP and additionally estimate our baseline model with covariates to improve precision (Calonico et al., 2019). We include an indicator variable, *Home*, which equals one if the team played at home and zero otherwise. Analogous to LPP, we control for the winning percentage in the last five games (*WinPerc5Games*) to account for team strength. Including this covariate slightly reduces the underlying sample since observations are missing for the first five games of each season. We also add team-season fixed effects to take into account remaining concerns of differences in team strength.

We mainly rely on two sets of results. We first investigate graphically whether there is discontinuity in changing the starting lineup between games that have been narrowly lost and

games that have been narrowly won. To this end, we plot the local sample means of *ChangingStarters* in nonoverlapping bins of the running variable, i.e., the point difference (*ScoreDiff*). Second, we quantitatively perform our RDD estimating Equation (1).

Due to its good balance between flexibility and simplicity, we rely on a local linear regression (i.e., a nonparametric approach), which has become the default approach in RDD analyses (Cattaneo et al., 2019). In particular, we use the nonparametric local polynomial estimation method with optimal bandwidth selection to obtain conventional RD estimates. This method estimates the bandwidth in a data-driven way that minimizes the mean squared error (MSE).[1] Additionally, we obtain bias-corrected RD estimates with robust confidence intervals as proposed by Calonico et al. (2014b) and implemented in Calonico et al. (2014a).

The RDD assumes that outcomes are quasi-randomly assigned around the cutoff. This assumption requires that teams on either side of the cutoff are comparable, i.e., there should be no difference with respect to team characteristics such as team strength (*WinPerc5Games*). Thus, teams that win narrowly should not discontinuously be more skilled around the cutoff. As suggested by Cattaneo et al. (2019), we analyze the covariate as if it were an actual outcome variable employing the same local linear method. Thus, in addition to the main results, we discuss the results from this balance check following the main results of each sport.[2]

In the following, we present the main results on a sport-by-sport basis. First, we extend the analysis by LPP and focus on matches played in the NBA, the premier basketball league in the world. Second, we remain within the sport of basketball but use different leagues. In particular, we repeat the analysis for the WNBA, the women's top basketball league, and the

[1] Following Cattaneo et al. (2019), we employ a triangular kernel that linearly downweighs observations further away from the cutoff value and yields to a point estimator with ideal properties when used in conjunction with a bandwidth that optimizes MSE.

[2] We also checked an alternative measure based on Klein Teeselink et al. (2022), who use the proportion of home games won by the home team and the proportion of away matches won by the away team as a proxy of skill difference. Our conclusions remain similar.

first division of the NCAA, the top American collegiate basketball league. Finally, we apply the same estimation approach for the NFL, the top American football league.

3 RESULTS

3.1 BASKETBALL

3.1.1 NBA

We first focus on the premier basketball competition in the world, the NBA, which is the underlying competition in LPP. We obtain data for the 1990/91 to 2019/20 seasons from https://www.basketball-reference.com/. In total, our sample consists of 37,409 games, including both regular season games and playoff games. Our final sample consists of 73,942 observations that include full information on our employed variables, i.e., the change in the lineup in the subsequent game and the point differential in the focal game.[3] Descriptive statistics of our employed variables are displayed in Table A1, Panel A in the appendix. On average, coaches in the NBA change their starting lineup in approximately 31.6% of subsequent matches, which is very similar to the average lineup change of 30.9% reported in LPP.

We start by illustrating the RDD graphically. We plot the mean likelihood of changing the starting lineup in the next game by the point difference in the last game. We fit a second-order polynomial using a window of -15 to +15 in the point difference. Figure 1 shows a visible discontinuity in the probability of changing the starting lineup, suggesting that coaches are more likely to change the strategy after a narrow defeat than after a narrow win.

----- *Insert Figure 1 about here* -----

[3] Because our analysis is within a season, we lack information on lineup decisions for certain games (e.g., the last game of the season). Thus, the number of observations in our analysis does not equal the number of games times two because the focal game may be the last game for team A but not team B. As a further consequence, the descriptive statistics for some variables may not be perfectly symmetrical.

Next, we proceed to estimate the magnitude of the effect using the nonparametric approach outlined in section 2. The baseline results are reported in Table 1. In addition to the conventional RD estimate with a conventional variance estimator and bias-corrected RD estimate with robust standard errors in parentheses (Calonico et al., 2014a, 2014b; Calonico et al., 2017), we report the number of observations to the left (#L) and right (#R) of the cutoff, the data-driven MSE-optimal bandwidth, and the number of observations underlying the whole sample.

----- *Insert Table 1 about here* -----

In Columns I to III of Table 1, we estimate the RDD for all of the seasons in our data, i.e., from 1990/91 to 2019/20. Starting in Column I with no covariates and fixed effects, we add covariates in Column II and estimate the full specification including team-season fixed effects in Column III. The estimates show that coaches who barely won their previous game are 4.6 percentage points less likely to adjust the starting lineup in the next game according to the conventional RD estimator and 4.3 percentage points according to the robust RD estimator. This corresponds to a decrease of approximately 14.8% in the control mean of 31.1%.[4] The coefficients remain robust and similar in terms of magnitude and statistical significance when including covariates and fixed effects. These results align well with LPP, although our point estimates are moderately smaller than the effect of 5.3 percentage points reported in their baseline model.

In Columns IV and V, we split the sample into two roughly equal subsamples to examine the persistence of outcome bias over time. In Column IV, we examine the 1990/91 to 2004/05 seasons, which we denote as the "early" subsample. With beta coefficients of -0.045 and -

[4] Following Flepp (2021) and Ludwig and Miller (2007), we define the control mean as the conventional local estimate of the likelihood of changing the lineup just below the threshold of zero, representing the nontreatment counterfactual.

0.041, the results are statistically significant under both the conventional and robust RD estimators. In Column V, we turn to the 2005/06 to 2019/20 seasons, denoted as the "late" subsample. In this subsample, the effect size decreases to approximately 4.2 (3.5) percentage points using the conventional (robust) RD estimator, and the effect is only significant for the conventional point estimator.

Finally, we discuss the validity of the RDD by examining the identifying assumption. The results presented in Column I of Table A2 in the appendix show no evidence of a discontinuity in the winning percentages in the last five games around the cutoff. Thus, teams that narrowly won do not have a discontinuously higher winning percentage than teams that narrowly lost, suggesting that the identification strategy is valid.

Overall, we replicate the finding that outcome bias is present in the starting lineup decisions of NBA coaches, although outcome bias decreases from the earlier time period to the later period.

3.1.2 WNBA

Second, we focus on the top league in women's basketball, the WNBA. We obtain all games played, i.e., both regular season and play-off games, for the period of 1997 to 2019 from https://www.basketball-reference.com/.[5] In total, our sample consists of 5,243 games; it is therefore considerably smaller than the NBA sample due to fewer teams and fewer games per team. Our final sample consists of 10,194 observations for which we have full information on our employed variables, i.e., the change in the lineup and the point differential in the last game. Descriptive statistics are displayed in Table A1 Panel B. With an average of 25.0% in *LineupChange*, coaches in the WNBA change their starting lineup less frequently compared to the NBA (31.6%).

[5] Data on the WNBA are only available from 1997.

Again, we first show a graphical illustration followed by the formal estimation of our RDD. The graphical illustration in Figure 2 hints at a treatment effect of winning the game: coaches who won their previous game seem to change the starting lineup in the following game less often than coaches who lost their previous game.

----- *Insert Figure 2 about here* -----

This suggestive evidence of outcome bias is supported by the more formal estimation results presented in Table 2 for the whole period from 1997 to 2019. With an effect size of 11.6 percentage points according to the conventional estimator (11.9 percentage points according to the robust estimator), the effect is considerably larger than in the NBA sample. The effect sizes remain robust under the different specifications (Columns I to III), although they slightly decrease in size when estimating the full model in Column III. The baseline effect corresponds to a decrease of approximately 41.6% in the control mean of 27.9%., i.e., the effect is also larger in relative terms compared to the NBA.

Again, the results are robust for the early subsample, covering the 1997 to 2009 seasons (Column IV). However, the effect is considerably smaller for the late subsample, and we fail to reject the null hypothesis of no effect for the late subsample covering the 2010 to 2019 seasons (Column V).

----- *Insert Table 2 about here* -----

Finally, we examine the identifying assumption, i.e., whether the winning percentage is continuous at the cutoff. The results presented in Column II of Table A2 in the appendix show no evidence of a discontinuity in the winning percentages in the last five games around the cutoff.

Taken together, the results for the WNBA are consistent with the findings for the NBA, suggesting that outcome bias is present, at least in the earlier time period.

3.1.3 COLLEGIATE BASKETBALL

Third, we collect data on all collegiate basketball games from https://www.basketball-reference.com/ from 2006 to 2019.[6] Our sample consists of 80,814 games, and the final sample (i.e., for which we have full information on our employed variables, the change in lineup in and the point differential in the last game) consists of 149,867 observations. Descriptive statistics are displayed in Table A1 Panel C. With an average of 38.3%, the lineup is changed more frequently in collegiate basketball than in both the NBA and the WNBA.

----- *Insert Figure 3 about here* -----

Again, we start with a graphical illustration in Figure 3, which also hints at the outcome-biased behavior of coaches in collegiate basketball matches. More formal estimations support this conjecture, as reported in Table 3. The results suggest that coaches who narrowly won the game have a 12.5 percentage point lower probability of changing the starting lineup than coaches who narrowly lost. This corresponds to a decrease of approximately 29.0% in the control mean of 43.1%. The results remain similar when using the robust RD estimator or when including covariates and team-season fixed effects. Moreover, comparing the estimates from the early period from 2006/07 to 2012/13 (Column IV of Table 3) to the late period from 2006/07 to 2012/13 (Column IV) reveals that the effect is persistent over time and that the magnitude even increases slightly for the late period.

----- *Insert Table 3 about here* -----

Finally, we check the validity of the RDD by examining the identifying assumption of continuity at the cutoff in Column III of Table A2. Again, we fail to reject the null hypothesis of no treatment in the winning percentage of the last five games, suggesting that the RDD is valid.

[6] Due to computational requirements when including team-season fixed effects, we limit our analysis to the 2006 to 2019 seasons although data are also available for the 2004 and 2005 seasons.

In summary, these results provide compelling evidence that outcome bias in coaching decisions also includes collegiate basketball.

3.2 AMERICAN FOOTBALL

We have presented evidence of outcome bias in different leagues in basketball. To test whether outcome bias is basketball-specific or is a more general phenomenon of sports coaches, we also consider American football and analyze matches from the National Football League (NFL). Similar to basketball, American football is a relatively high-scoring sport and thus is suitable for our identification strategy.

Football players are typically dedicated either to the offense or to the defense. After a change in possession (e.g., touchdown, field goal, interception, failed fourth down), offensive and defensive players switch. Thus, compared to basketball, we are able to observe a change in both starting lineup decisions.

We obtain data for the period from 1990 to 2019 from https://www.pro-football-reference.com/. In total, our sample consists of 7,762 games, including both regular season and play-off games. Similar to basketball, we calculate whether a coach fields the same players in the subsequent match. Because teams have dedicated offensive and defensive lineups, we calculate two variables, i.e., *LineupChangeOff* and *LineupChangeDef*. The former (latter) equals 1 if the offensive (defensive) starting lineup is changed from game g to game $g+1$ and 0 otherwise. We drop games with a point differential of zero, i.e., those that ended in a tie. Our final sample consists of 14,468 observations for which we have full information on our employed variables, i.e., the lineups in the next game and the point differential in the last game.

Given its high intensity and the higher number of players in the starting lineup compared to basketball, the average change in the starting lineup from one game is also considerably higher: 82.8% for the offensive lineup and 71.1% for the defensive lineup.

This complicates the identification of outcome bias since in 8 (7) out of 10 games, on average, the offensive (defensive) lineup is changed. Moreover, we argue that a complete lineup strategy change by the coach occurs only if adjustments are made in both the offensive and the defensive lineup. Thus, our main measure of strategy adjustment is a dummy variable equaling one when a coach changes both the offensive and defensive lineups and zero otherwise. We denote this variable *LineupChangeBoth*. On average, coaches change both lineups in 60.4% of games in our sample. Descriptive statistics of our employed variables are displayed in Table A1 Panel D.

First, we present graphical results of the RDD for the change in both lineups in Figure 4. We also separately show the change in the offensive and defensive lineups in Figure 5 and Figure 6, respectively. The figures hint at a negative treatment effect, i.e., that coaches change the starting lineup less often after a narrow win compared to a narrow defeat.

----- *Insert Figure 4,* Figure 5, & Figure 6 *about here* -----

More formal estimation results, as presented in Table 4 Panel A for the decision to change both lineups, support this conjecture. The treatment coefficient is consistently negative across Columns I to III, which again is consistent with outcome bias. The effect size ranges from -0.120 to -0.099 depending on the model choice. The baseline effect corresponds to a decrease of approximately 15.4% in the control mean of 64.3% and thus closely aligns with the effect from the NBA (14.8%). Similar to the NBA and the WNBA, the effect is larger in the early subsample (1990 to 2004 seasons) than in the late subsample (2005 to 2019 seasons).

Examining the lineup decisions separately, we find that the results are mainly driven by a change in the offensive lineup, although there is also weak statistical evidence with regard to a change in the defensive lineup (see Table 4 Panel B & C).

----- *Insert Table 4 about here* -----

Finally, we examine continuity in the winning percentage in the last five games at the cutoff to strengthen the validity of our design. Column IV of Table A2 suggests that there is no significant evidence of a discontinuous jump in the winning percentage.

Overall, these findings are consistent with the previous evidence from basketball and suggest that outcome bias generalizes to American football.

4 CONCLUSION

We use quasi-random wins and defeats, which are uninformative about a coach's strategy, to study outcome bias in the field. We replicate LPP for the NBA and extend the analysis in several key aspects. In particular, we use more recent data, alternative leagues and sports, and more advanced RDD techniques. Our findings across these multiple leagues and sports confirm that coaches fall prey to outcome bias and adjust the starting lineup after a narrow defeat more often than after a narrow win. As a result, coaches may misinterpret favorable (unfavorable) outcomes as evidence for maintaining (changing) the chosen strategy even when more subtle evidence suggests otherwise.

Our findings complement existing studies that investigate outcome bias in both laboratory (e.g., Baron & Hershey, 1988; Gino et al., 2008; Marshall & Mowen, 1993) and field settings (e.g., Gauriot & Page, 2019; Kausel et al., 2019; Meier et al., 2022). Most importantly, Gauriot and Page (2019) show that outcome bias is also present in the context of European soccer, where soccer players' subsequent playing times and ratings from journalists are influenced by lucky goal scoring successes. Finding robust evidence across different leagues and sports supports the explanation that outcome bias is a general tendency of decision-makers, at least in the context of sports, and is neither specific to the NBA nor a statistical artifact. Given that professional sports coaches, who are highly incentivized and experienced,

exhibit the tendency to overvalue uninformative outcome information in their decision-making, the effect might also be widespread in other economic decision-making contexts.

Our findings also suggest that outcome bias is stronger in the earlier time period studied. We envisage two complementary explanations that could lead to this pattern. First, data analytics increasingly gained a foothold in professional sports in the 2010s. Currently, a dedicated data analytics team is the norm rather than the exception in professional sports, and advanced analytics is employed to analyze the strategies for team composition and team improvement (Sarlis & Tjortjis, 2020). Second, the pattern could also be explained by coaches becoming increasingly aware of their biased decision-making. For instance, Pope et al. (2018) showed that awareness of the racial bias of referees may subsequently reduce that bias, exploiting the widespread media attention that occurred in May 2007 following the release of an academic study (Price & Wolfers, 2010). Certainly, outcome bias is much less of a (societal) challenge than racial bias, suggesting that (media) attention is likely to have been lower; nevertheless, the awareness of coaches might have impacted their decision-making through the dissemination of academic knowledge.

In conclusion, we successfully replicate the main finding of LPP and consider the documented effect of outcome bias in sporting coaches' lineup decisions to be robust and credible, although with a tendency to weaken over time.

REFERENCES

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569. https://doi.org/10.1037/0022-3514.54.4.569

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192–210. https://doi.org/10.1093/ectj/utz022

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R [Rocio] (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404. https://doi.org/10.1177/1536867X1701700208

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R [Rocio] (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, *101*(3), 442–451. https://doi.org/10.1162/rest_a_00760

Calonico, S., Cattaneo, M. D., & Titiunik, R [Rocio] (2014a). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, *14*(4), 909–946. https://doi.org/10.1177/1536867X1401400413

Calonico, S., Cattaneo, M. D., & Titiunik, R [Rocio] (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Cattaneo, M. D., Idrobo, N., & Titiunik, R [Rocío]. (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Flepp, R. (2021). Uninformative performance signals and forced CEO turnover. *Available at SSRN 3904056*.

Gauriot, R., & Page, L. (2019). Fooled by performance randomness: Overrewarding luck. *Review of Economics and Statistics*, *101*(4), 658–666. https://doi.org/10.1162/rest_a_00783

Gino, F., Moore, D. A., & Bazerman, M. H. (2008). *No harm, no foul: The outcome bias in ethical judgments*. Citeseer.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Kausel, E. E., Ventura, S., & Rodríguez, A. (2019). Outcome bias in subjective ratings of performance: Evidence from the (football) field. *Journal of Economic Psychology*, *75*, 102132. https://doi.org/10.1016/j.joep.2018.12.006

Klein Teeselink, B., van den Assem, M. J., & van Dolder, D. (2022). Does losing Lead to winning? An empirical analysis for four sports. *Management Science.* Advance online publication. https://doi.org/10.1287/mnsc.2022.4372

Lefgren, L., Platt, B., & Price, J [Joseph] (2015). Sticking with what (barely) worked: A test of outcome bias. *Management Science*, *61*(5), 1121–1136. https://doi.org/10.1287/mnsc.2014.1966

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*(1), 159–208.

Marshall, G. W., & Mowen, J. C. (1993). An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management*, *13*(3), 31–47. https://doi.org/10.1080/08853134.1993.10753956

Meier, P. F., Flepp, R., Meier, P., & Franck, E. P. (2022). Outcome bias in self-evaluations: Quasi-experimental field evidence of Swiss driving license exams. *Journal of Economic Behavior & Organization*(201), 292–309. https://doi.org/10.1016/j.jebo.2022.07.013

Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics—How many and which papers are chosen for replication, and why? *Research Policy*, *48*(1), 62–83. https://doi.org/10.1016/j.respol.2018.07.019

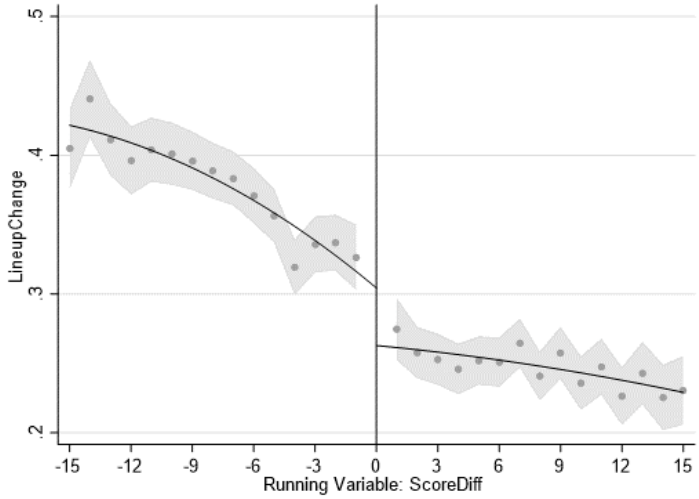
Pope, D. G., Price, J [Joseph], & Wolfers, J [Justin] (2018). Awareness reduces racial bias. *Management Science*, *64*(11), 4988–4995. https://doi.org/10.1287/mnsc.2017.2901

Price, J [J.], & Wolfers, J [J.] (2010). Racial Discrimination Among NBA Referees. *The Quarterly Journal of Economics*, *125*(4), 1859–1887. https://doi.org/10.1162/qjec.2010.125.4.1859

Sarlis, V., & Tjortjis, C. (2020). Sports analytics — Evaluation of basketball players and team performance. *Information Systems*, *93*, 101562. https://doi.org/10.1016/j.is.2020.101562

TABLES AND FIGURES

Figure 1: RDD Plot - NBA



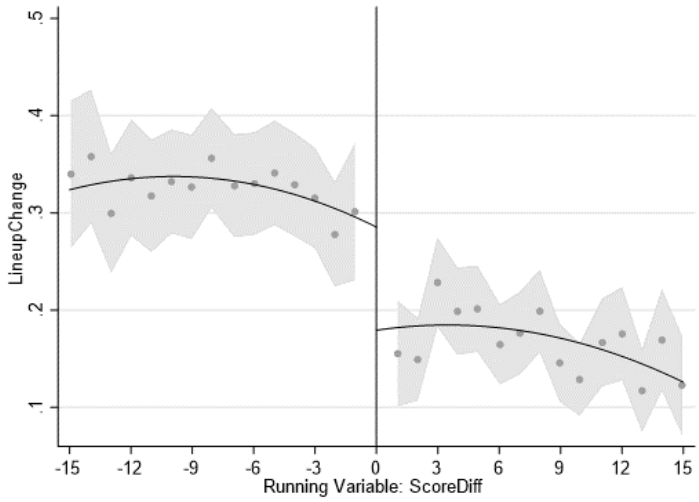
Notes: The figure shows the regression discontinuity plot within the window of -15 to +15 of the running variable (*ScoreDiff*). Local sample means of our dependent variable (*LineupChange*) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals. The model includes a second-order polynomial, which is allowed to differ on either side of the cutoff.

Table 1: Results - NBA

		Dependent Variable: *LineupChange*				
		I All (1990/91-2019/20)	II All (1990/91-2019/20)	III All (1990/91-2019/20)	IV Early (1990/91-2004/05)	V Late (2005/06-2019/20)
Conventional	Beta	-0.046***	-0.048***	-0.050***	-0.045**	-0.042*
		(0.012)	(0.013)	(0.011)	(0.015)	(0.019)
Robust	Beta	-0.043**	-0.045**	-0.047***	-0.041*	-0.035
		(0.014)	(0.016)	(0.014)	(0.017)	(0.023)
	Bandwidth	10.66	9.864	8.195	12.47	9.293
	#L	21,061	18,074	16,104	11,852	9,895
	#R	21,335	18,326	16,331	12,016	10,013
	Total Observations	73,942	69,562	69,562	35,808	38,134
	Team-Season FE	No	No	Yes	No	No
	Covariates	No	Yes	Yes	No	No

Notes: *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are reported in parentheses and clustered at the team-season level. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). #L (#R) denotes the number of observations used to the left (right) side of the cutoff. Total observations denotes the original number of observations in the underlying sample. The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff. The following covariates are included if indicated: *Home* & *WinPerc5Games*.

Figure 2: RDD Plot - WNBA



Notes: The figure show the regression discontinuity plot within the window of -15 to +15 of the running variable (*ScoreDiff*). Local sample means of our dependent variable (*LineupChange*) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals. The model includes a second-order polynomial, which is allowed to differ on either side of the cutoff.

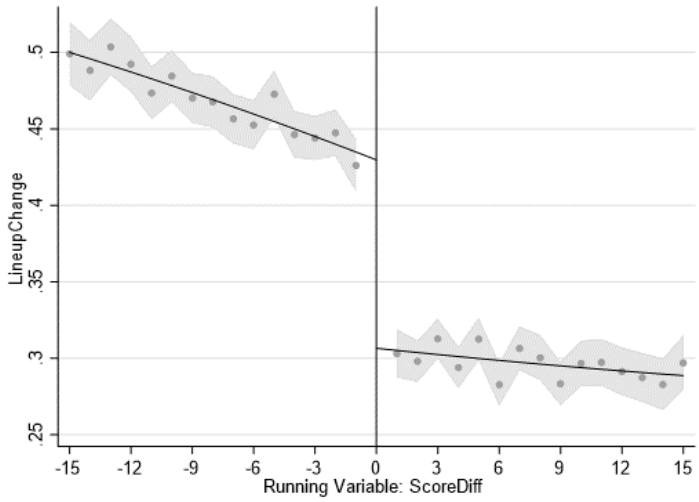
Table 2: Results - WNBA

Dependent Variable: *LineupChange*

	I All 1997-2019	II All 1997-2019	III All 1997-2019	IV Early 1997-2009	V Late 2010-2019
Beta Conventional	-0.116**	-0.109**	-0.096**	-0.144**	-0.084
	(0.036)	(0.037)	(0.033)	(0.048)	(0.048)
Beta Robust	-0.119**	-0.109*	-0.088*	-0.156**	-0.074
	(0.045)	(0.046)	(0.040)	(0.060)	(0.060)
Bandwidth	7.892	8.334	6.780	8.597	8.127
#L	1,970	1,977	1,427	1,106	1,198
#R	2,026	2,045	1,477	1,137	1,231
Total Observations	10,194	8,734	8,734	4,913	5,281
Team-Season FE	No	No	Yes	No	No
Covariates	No	Yes	Yes	No	No

Notes: *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are reported in parentheses and clustered at the team-season level. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). #L (#R) denotes the number of observations used to the left (right) side of the cutoff. Total observations denotes the original number of observations in the underlying sample. The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff. The following covariates are included if indicated: *Home* & *WinPerc5Games*.

Figure 3: RDD Plot – Collegiate Basketball



Notes: The figure show the regression discontinuity plot within the window of -15 to +15 of the running variable (*ScoreDiff*). Local sample means of our dependent variable (*LineupChange*) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals. The model includes a second-order polynomial, which is allowed to differ on either side of the cutoff.

Table 3: Results - Collegiate Basketball

Dependent Variable: *LineupChange*

	I All 2006/07- 2019/20	II All 2006/07- 2019/20	III All 2006/07- 2019/20	IV Early 2006/07- 2012/13	V Late 2013/14- 2019/20
Beta Conventional	-0.125***	-0.122***	-0.126***	-0.105***	-0.143***
	(0.008)	(0.009)	(0.006)	(0.012)	(0.012)
Beta Robust	-0.124***	-0.121***	-0.124***	-0.104***	-0.145***
	(0.010)	(0.010)	(0.006)	(0.015)	(0.014)
Bandwidth	12.96	12.89	12.99	13.14	10.73
#L	43,261	37,655	37,655	22,154	19,292
#R	46,733	40,755	40,755	24,165	20,643
Total Observations	149,867	125,826	125,826	72,204	77,663
Team-Season FE	No	No	Yes	No	No
Covariates	No	Yes	Yes	No	No

Notes: *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are reported in parentheses and clustered at the team-season level. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). #L (#R) denotes the number of observations used to the left (right) side of the cutoff. Total observations denotes the original number of observations in the underlying sample. The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff. The following covariates are included if indicated: *Home* & *WinPerc5Games*.

Figure 4: RDD Plot – NFL (*LineupChangeBoth*)

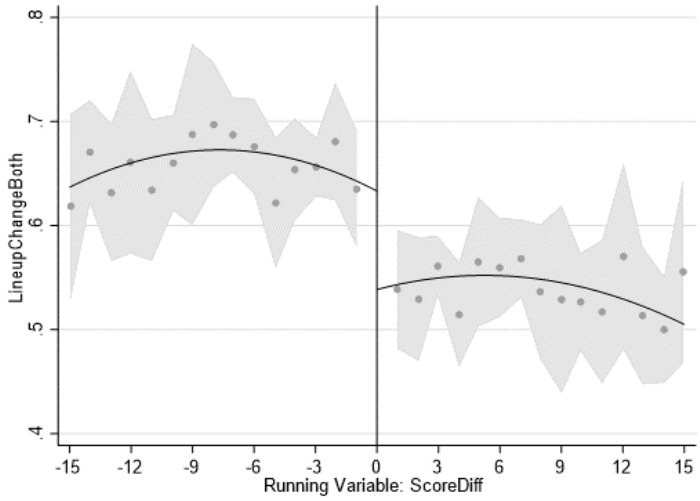


Figure 5: RDD Plot – NFL (*LineupChangeOff*)

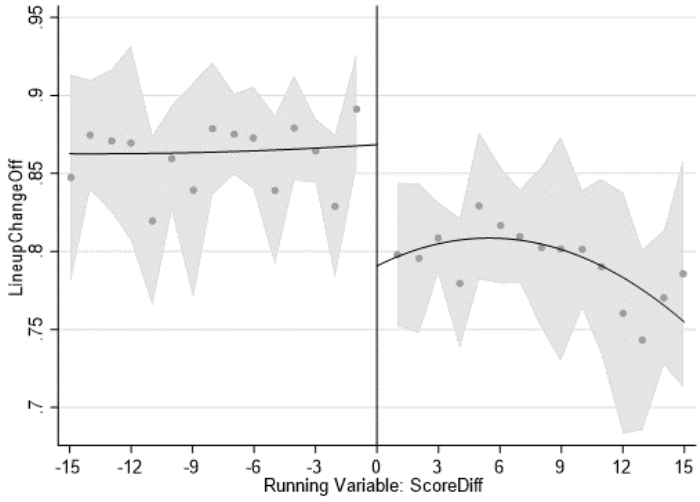
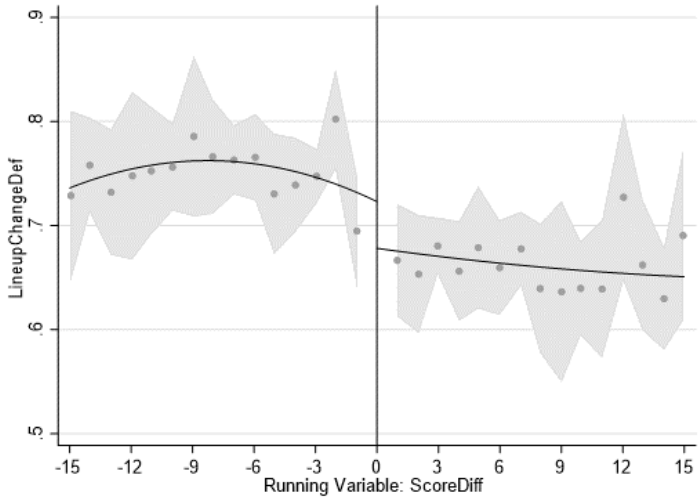


Figure 6: RDD Plot – NFL
(*LineupChangeDef*)



Notes: The figure show the regression discontinuity plot within the window of -15 to +15 of the running variable (*ScoreDiff*). Local sample means of our dependent variable (*LineupChangeX*) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals. The model includes a second-order polynomial, which is allowed to differ on either side of the cutoff.

Table 4: Results - NFL

	All 1990-2019	All 1990-2019	All 1990-2019	Early 1990-2004	Late 2004-2019
Panel A:					
LineupChangeBoth	I	II	III	IV	V
Beta	-0.099***	-0.113**	-0.104***	-0.111**	-0.088**
Conventional	(0.026)	(0.035)	(0.026)	(0.043)	(0.034)
Beta	-0.102**	-0.120**	-0.099**	-0.124*	-0.080
Robust	(0.033)	(0.044)	(0.033)	(0.052)	(0.041)
Bandwidth	11.12	9.907	8.199	10.12	11.02
#L	4,204	2,402	2,332	1,933	2,165
#R	4,416	2,577	2,498	2,034	2,272
Total Observations	14,468	9,870	9,870	6,994	7,474
Team-Season FE	No	No	Yes	No	No
Covariates	No	Yes	Yes	No	No
Panel B:					
LineupChangeOff	I	II	III	IV	V
Beta	-0.068**	-0.113***	-0.099***	-0.096**	-0.038
Conventional	(0.021)	(0.029)	(0.018)	(0.034)	(0.024)
Beta	-0.077**	-0.128***	-0.106***	-0.108**	-0.042
Robust	(0.026)	(0.035)	(0.022)	(0.040)	(0.030)
Bandwidth	9.509	7.820	8.725	9.936	10.58
#L	3,604	2,173	2,332	1,733	2,077
#R	3,778	2,341	2,498	1,813	2,177
Total Observations	14,468	9,870	9,870	6,994	7,474
Team-Season FE	No	No	Yes	No	No
Covariates	No	Yes	Yes	No	No
Panel C:					
LineupChangeDef	I	II	III	IV	V
Beta	-0.057*	-0.042	-0.011	-0.047	-0.067*
Conventional	(0.023)	(0.030)	(0.028)	(0.032)	(0.032)
Beta	-0.055	-0.039	0.003	-0.047	-0.056
Robust	(0.029)	(0.037)	(0.034)	(0.040)	(0.039)
Bandwidth	11.90	11.04	7.184	13.59	10.62
#L	4,204	2,818	2,173	2,206	2,077
#R	4,416	3,037	2,341	2,320	2,177
Total Observations	14,468	9,870	9,870	6,994	7,474
Team-Season FE	No	No	Yes	No	No
Covariates	No	Yes	Yes	No	No

Notes: *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are reported in parentheses and clustered at the team-season level. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). #L (#R) denotes the number of observations used to the left (right) side of the cutoff. Total observations denotes the original number of observations in the underlying sample. The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff. The following covariates are included if indicated: *Home* & *WinPerc5Games*.

Appendix

Table A1: Descriptive Statistics

Variable	N	Mean	SD	P25	Median	P75
Panel A: NBA Sample						
LineupChange	73,942	0.316	0.465	0.000	0.000	1.000
ScoreDiff	73,942	0.078	13.575	-9.000	1.000	9.000
Home	73,942	0.500	0.500	0.000	1.000	1.000
WinPerc5Games	69,562	0.508	0.259	0.400	0.600	0.600
Panel B: WNBA Sample						
LineupChange	10,194	0.250	0.433	0.000	0.000	1.000
ScoreDiff	10,194	0.165	13.001	-9.000	1.000	9.000
Home	10,194	0.500	0.500	0.000	1.000	1.000
WinPerc5Games	8,734	0.514	0.256	0.400	0.600	0.800
Panel C: College Basketball Sample						
LineupChange	149,867	0.383	0.486	0.000	0.000	1.000
ScoreDiff	149,867	1.457	16.105	-9.000	2.000	11.000
Home	149,867	0.520	0.500	0.000	1.000	1.000
WinPerc5Games	125,826	0.527	0.270	0.400	0.600	0.800
Panel D: NFL Sample						
LineupChangeBoth	14,468	0.604	0.489	0.000	1.000	1.000
LineupChangeOff	14,468	0.828	0.378	1.000	1.000	1.000
LineupChangeDef	14,468	0.711	0.453	0.000	1.000	1.000
ScoreDiff	14,468	0.416	14.841	-8.000	1.000	10.000
Home	14,468	0.504	0.500	0.000	1.000	1.000
WinPerc5Games	9,870	0.513	0.262	0.400	0.600	0.800

This table presents the descriptive statistics for the regression variables of interest.

Table A2: Winning Percentage in Prior Five Games

Dependent Variable: *WinPerc5Games*

		I NBA	II WNBA	III Collegiate	IV NFL
Conventional	Beta	0.010	0.035	0.004	0.038
		(0.009)	(0.026)	(0.005)	(0.021)
Robust	Beta	0.009	0.041	0.002	0.035
		(0.011)	(0.032)	(0.006)	(0.026)
	Bandwidth	10.90	7.860	11.64	9.943
	#L	19,788	1,686	35,134	2,402
	#R	20,061	1,743	38,007	2,577
	Total Observations	69,562	8,734	125,826	9,870
	Team-Season FE	No	No	No	No
	Covariates	No	No	No	No

Notes: *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are reported in parentheses and clustered at the team-season level. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). #L (#R) denotes the number of observations used to the left (right) side of the cutoff. Total observations denotes the original number of observations in the underlying sample. The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff.