Replication: Do coaches stick with what barely worked? Evidence of outcome bias in sports

Meier, Pascal Flurin ; Flepp, Raphael ; Franck, Egon

Abstract: 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. This result is consistent with outcome bias because such narrow outcomes are conditionally uninformative. As our paper shows, this pattern is not restricted to the NBA; we also 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 in some instances.

DOI: https://doi.org/10.1016/j.joep.2023.102664

Journal of Economic Psychology 99 (2023) 102664

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Replication: Do coaches stick with what barely worked? Evidence of outcome bias in sports☆

Pascal Flurin Meier *, **Raphael Flepp**, **Egon Franck**

University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

A R T I C L E I N F O

Keywords:
Outcome bias
Strategy revision
Regression discontinuity design
Replication

A B S T R A C T

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. This result is consistent with outcome bias because such narrow outcomes are conditionally uninformative. As our paper shows, this pattern is not restricted to the NBA; we also 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 in some instances.

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 the agent's control (Baron & Hershey, 1988). The study by Lefgren et al. (2015, henceforth LPP) was the first to demonstrate outcome bias outside laboratory experiments in a high-stakes field setting. In particular, LPP investigates the decisions of National Basketball Association (NBA) coaches to change their starting lineup after narrow wins compared to narrow losses, which is an uninformative outcome regarding the team's performance.

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. Second, we investigate whether outcome bias in coaches' starting lineup decisions can also be observed in other basketball leagues. 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., 2020; Calonico et al., 2017; Calonico et al., 2014a, 2014b). 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.

☆ The data used in the study are available at https://data.mendeley.com/datasets/y7yc7dyvhn/4. The authors thank Michael Toren for excellent research assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
E-mail address: pascal-flurin.meier@business.uzh.ch (P.F. Meier).

https://doi.org/10.1016/j.joep.2023.102664
Received 27 February 2023; Received in revised form 17 July 2023; Accepted 1 August 2023
Available online 6 August 2023

2. Empirical strategy

We closely follow LPP in setting up the regression discontinuity model. We study whether coaches discontinuously change their strategy following a narrow loss compared to a narrow win. In high-scoring sports, it is quasirandom 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, nonoutcome-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:

$$ChangingStarters_{i,g+1} = \alpha + \beta_1 Win_{i,g} + \beta_2 ScoreDiff_{i,g} + \beta_3 Win_{i,g} \times ScoreDiff_{i,g} + \varepsilon_{i,g} \tag{1}$$

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. 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. 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 models 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. We also add team-season fixed effects to account for the remaining concerns of differences in team strength.

We mainly rely on two sets of results. We first investigate graphically whether there is a 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).

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 mean squared error (MSE) optimal bandwidth selection to obtain conventional RD estimates (Calonico et al., 2020; Calonico et al., 2017). 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).[1]

In the online appendix we conduct several examinations to assess the validity of our identification strategy for each sport.

3. Results

3.1. Basketball

We first focus on the premier basketball league in the world, the NBA, which is the underlying competition in LPP. LPP used the 1991–2010 seasons, obtaining 46,550 team game observations for regular season games. We obtain data for the 1990/91 to 2019/20 seasons from https://www.basketball-reference.com/. In total, our NBA sample consists of 37,409 games, including both regular season games and playoff games. Our final, cleaned sample consists of 73,942 observations.

In addition to LPP, we collect data on the top league in women's basketball, the WNBA, and Collegiate Athletic Association (NCAA) basketball from https://www.basketball-reference.com/. We obtain all games played for the period of 1997 to 2019 for the WNBA and games for the period of 2006 to 2019 for collegiate games.[2] In total, our sample consists of 5,243 WNBA and 80,814 collegiate games. Descriptive statistics showing the total number of observations after cleaning the data are displayed in Panels A, B & C of Table A1 in the online appendix.

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. Fig. 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. This also holds for the WNBA sample (Fig. 2) and the collegiate sample (Fig. 3). Consistent with outcome bias, this suggests that coaches are more likely to change their strategy after a narrow defeat than after a narrow win.

Next, we proceed to estimate the magnitude of the effect using the nonparametric approach outlined in section 2. The baseline

[1] We also performed an analysis based on the idea of local randomization, because the running variable ScoreDiff, is discrete. The estimated treatment effects across all leagues are very similar. As our results are not sensitive to this choice and the number of unique score values in our running variable is large enough (Cattaneo et al., 2023), we follow prior literature (Klein Teeselink et al., 2022; Meier et al., 2022) and mainly rely on the continuity-based RD framework.

[2] 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.

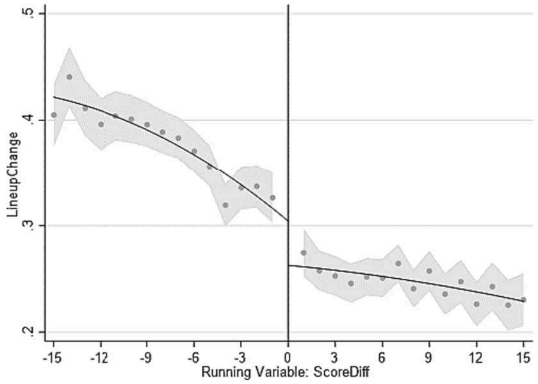

Fig. 1. RDD Plot – NBA.
Notes: The figure shows the regression discontinuity plot within the window of -15 to +15 of the running variable (ScoreDiff). Local sample means of our dependent variable (LineupChange) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals.

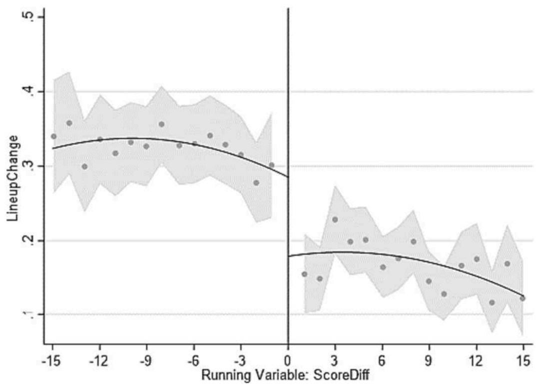

Fig. 2. RDD Plot – WNBA.
Notes: The figure shows the regression discontinuity plot within the window of -15 to +15 of the running variable (ScoreDiff). Local sample means of our dependent variable (LineupChange) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals.

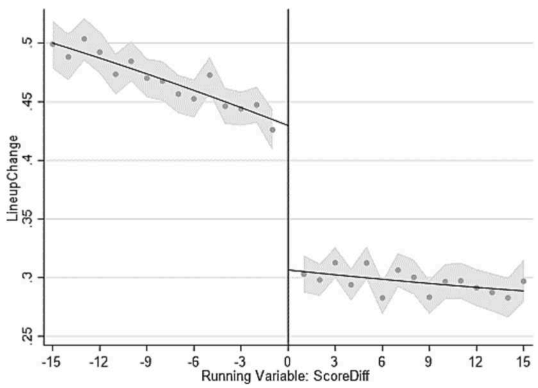

Fig. 3. RDD Plot – Collegiate Basketball.
Notes: The figure shows the regression discontinuity plot within the window of -15 to +15 of the running variable (ScoreDiff). Local sample means of our dependent variable (LineupChange) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals.

Table 1
Results - Basketball.

Dependent Variable: LineupChange					
Panel A: NBA Sample	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
Beta (Conventional)	−0.046***	−0.048***	−0.050***	−0.045**	−0.042*
	(0.012)	(0.013)	(0.011)	(0.015)	(0.019)
Beta (Robust)	−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
Panel B: WNBA Sample	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
Panel C: Collegiate Basketball Sample	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). The sample includes observations within the MSE-optimal bandwidth. #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.

results for all three leagues are reported in Table 1. We report the conventional RD estimate with a conventional variance estimator and the bias-corrected RD estimate with robust standard errors in parentheses (Calonico et al., 2017; Calonico et al., 2014a, 2014b).

In Columns I to III in Panels A, B, and C of Table 1, we estimate the RD model for all of the seasons in our data. 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. In Columns IV and V, we split the sample into two roughly equal subsamples, an early subsample and a late subsample, to examine the persistence of outcome bias over time.

For the NBA (Table 1, Panel A), 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 less likely to do so according to the robust RD estimator. This corresponds to a decrease of approximately 14.8% in the control mean of 31.1%.[3] 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

[3] Following 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.

in their baseline model.

For the WNBA sample (Table 1, Panel B), the effect size of 11.6 percentage points according to the conventional estimator is considerably larger than for the NBA sample. 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. For the collegiate basketball games (Table 1, Panel C), 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%.

Separating the samples into roughly two equally sized subsamples, we find that the effect size in the NBA 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. While 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 in the WNBA, the effect is persistent over time for collegiate matches, and the magnitude even increases slightly for the late period.

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 (NFL). Similar to basketball, American football is a relatively high-scoring sport and thus is suitable for our identification strategy.

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. We calculate whether a coach fields the same players in the subsequent match. Because teams have dedicated offensive and defensive lineups, we calculate our main variable LineupChange if both the offensive and defensive starting lineups exhibited a change from game g to game $g+1$ and 0 otherwise, constituting a major strategy change. 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.

Because the average change in the starting lineup in American football is considerably higher compared to basketball, we derive an additional measure of strategy adjustment equaling the number of changes in both the offensive and defensive lineups. We denote this variable LineupChangeNum. On average, coaches change 2.8 players per game in our sample. Descriptive statistics of our employed variables are displayed in Panel D of Table A1 in the online appendix.

First, we present graphical results of the RDD for the change in both lineups in Fig. 4. The figure portrays at a negative treatment effect. Our conclusions remain similar when examining the number of changes, which is displayed in Fig. 5, showing a clearly visible discontinuity around the threshold. More formal estimation results, as presented in Panel A of Table 2 for the binary decision to change both lineups, support this conjecture. The treatment coefficient is consistently negative. 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%). Again similar to the results from the NBA, the effect is smaller in the late subsample but remains statistically significant for the conventional RD estimator.

Examining the number of player changes in Panel B of Table 2, we again find consistent evidence of outcome bias across all specifications. On average, coaches who narrowly lost their game make approximately 0.361 to 0.515 more substitutions than coaches

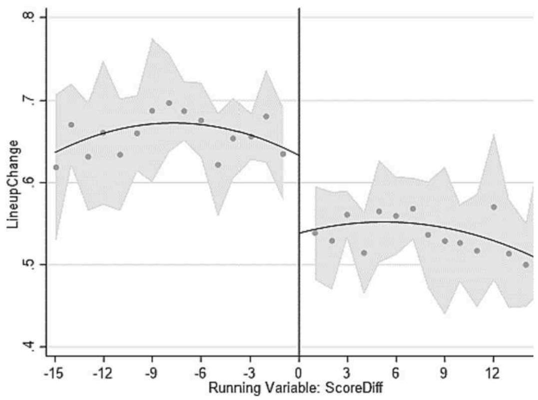

Fig. 4. RDD Plot – NFL (LineupChange).

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.

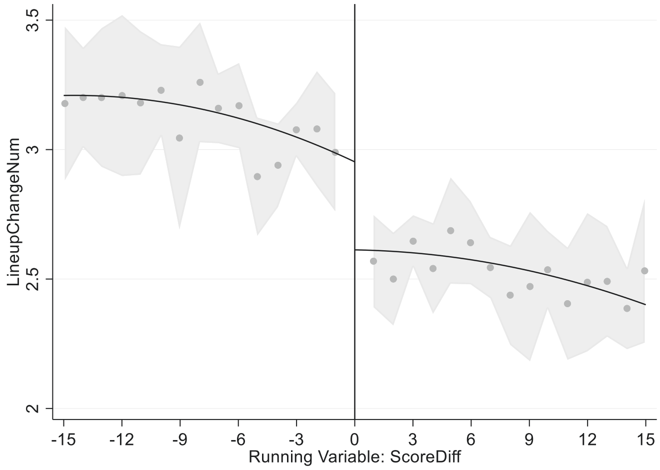

Fig. 5. RDD Plot – NFL (LineupChangeNum).

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 (LineupChangeNum) are plotted in the nonoverlapping bins of the running variable in steps of 1 goal difference. The shaded areas represent the 95% confidence intervals.

Table 2
Results – NFL.

	All 1990–2019	All 1990–2019	All 1990–2019	Early 1990–2004	Late 2005-2019
Panel A: LineupChange	I	II	III	IV	V
Beta (Conventional)	−0.099***	−0.113**	−0.104***	−0.111**	−0.088**
	(0.026)	(0.035)	(0.026)	(0.043)	(0.034)
Beta (Robust)	−0.102**	−0.120**	−0.099**	−0.124*	−0.080
	(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: LineupChangeNum	I	II	III	IV	V
Beta (Conventional)	−0.361***	−0.460***	−0.496***	−0.266*	−0.463***
	(0.092)	(0.125)	(0.086)	(0.122)	(0.134)
Beta (Robust)	−0.362**	−0.491**	−0.515***	−0.295*	−0.435**
	(0.115)	(0.156)	(0.105)	(0.149)	(0.166)
Bandwidth	11.69	10.32	10.05	11.02	11.59
#L	4,204	2,693	2,693	2,039	2,165
#R	4,416	2,897	2,897	2,144	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

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). The sample includes observations within the MSE-optimal bandwidth. #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.

who narrowly won the game.[4] As displayed in Panel B of Table 2, the effect in the late period (Column V) is larger than that in the early period (Column IV). This might be partly due to the increase in the average number of changes over the years from 2.41 before 2005 to 3.24 from 2005 onward. Indeed, in relation to the control mean, the decrease seems comparable (10.4 percentage points in the early

[4] We also examined the decision to change the offensive or defensive lineup separately as well as the number of changes for offensive and defensive lineup separately. We find evidence of outcome bias in both offensive and defensive lineup decisions although the effect size seems to be slightly stronger for the offensive lineup.

period vs. 13.6 percentage points in the late period).

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

4. Conclusion

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. Our findings complement existing studies that investigate outcome bias in both experiments (e.g., Baron & Hershey, 1988) and field settings (e.g., Gauriot & Page, 2019; Kausel et al., 2019; Meier et al., 2022).

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. Moreover, we respond to recent calls to use sports data to study economic decision-making (e.g., Balafoutas et al., 2019). 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, at least in the NBA and the WNBA, tends to be stronger in the earlier period that is 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. Second, the pattern could also be explained by coaches becoming increasingly aware of their biased decision-making.

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 in some instances. In light of the ongoing replication crisis in social sciences, it is reassuring to see that outcome bias replicates in the field.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joep.2023.102664.

References

Balafoutas, L., Chowdhury, S. M., & Plessner, H. (2019). Applications of sports data to study decision making. Journal of Economic Psychology, 75, Article 102153. https://doi.org/10.1016/j.joep.2019.02.009

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. (2019). & Titiunik, R [Rocío]. A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press.

Cattaneo, M. D., Keele, L., & Titiunik, R [Rocio]. (2023). A Guide to Regression Discontinuity Designs in Medical Applications. https://doi.org/10.48550/arXiv.2302.07413.

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

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, Article 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. (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.

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