A Database of NBA Refereeing

Fernando De Stefanis, Amanda Li, Connor Turkatte, Steven Wei[1]

Contents

Introduction

Referees are supposed to be impartial, bringing sound judgment amid the chaotic environment of a sporting event. However, a substantial body of literature[2][3][4] shows that the impartiality of officials is often questionable. In an effort to increase transparency on these issues, the National Basketball Association, in March 2015, started releasing its Last Two Minute reports.[5] These documents include an assessment of all executed calls and notable non-calls for the last two minutes of any close game[6] during the regular season and the playoffs.

While the NBA's effort to increase transparency is commendable, seeing as it is one of very few sports league willing to publicly acknowledge officiating mistakes, the format in which the reports are published does not lend itself to statistical analysis. The reports are published as PDF files on a game-by-game basis, making their systematic use for statistical research a daunting task. It is not surprising to see, therefore, that as of today only one

---

[1] All authors are Undergraduates at the University of Chicago. For suggestions or to obtain a copy of the dataset, please contact destefanis@uchicago.edu

[2] Boyko, Ryan H., Adam R. Boyko, and Mark G. Boyko. "Referee bias contributes to home advantage in English Premiership football." Journal of sports sciences 25, no. 11 (2007): 1185-1194.

[3] Johnston, Ron. "On referee bias, crowd size, and home advantage in the English soccer Premiership." Journal of Sports Sciences 26, no. 6 (2008): 563-568.

[4] Rodenberg, Ryan M., and Choong Hoon Lim. "Payback calls: A starting point for measuring basketball referee bias and impact on team performance." European Sport Management Quarterly 9, no. 4 (2009): 375-387.

[5] The reports are available on the Official Website of the NBA

[6] Close games are defined by the NBA as games "which were within five points at the two-minute mark (and during overtime, where applicable)" where the two-minute mark is the 4th quarter's.

paper[7] has used them as a data source.[8] Furthermore, another issue with the data is that the Association does not release the name of the officials in charge of each individual game, making it even more burdensome to connect each call to the trio of referees that executed it, or to evaluate the performance of individual officials over multiple games.

Each Game Report includes an analysis of all the calls in the last two minutes, with the names of the player who committed the foul and the one who was disadvantaged by the foul, and an assessment of the correctness of the call. Each call can be one of four categories: CC, Correct Call, IC, Incorrect Call, CNC, Correct Non Call, and INC, Incorrect Non Call. Additionally, each call is commented with a description of the action and the reason for the assessment. Calls that could only be assessed by using technology (slow motion, different camera angles, etc.) are not given a comment, but without the assessment. This explains the relatively large number of calls in our dataset that do not have a corresponding comment.

In this paper, we analyze the Last Two Minute reports published between March 1, 2015, and January 12, 2017. In the first section, we discuss the structure of our database and the process we followed in its creation. In the second, we provide statistical analysis on the data we obtained. In the third, we conclude.

1   The Database

In order to create the database, our first objective was to turn the PDF files from the last two minute reports into a usable format. We used ABBYY PDF converter[9] to obtain a text file out of each game report. We then aggregated the reports into a single text file and parsed through it in order to retain the information we needed. The next step was to match each game in our dataset to the three referees in charge of officiating. In order to do so, we matched the games to the corresponding box scores from Basketball Reference, one of the foremost providers of NBA data. We retrieved the names of the officials from the box score and linked them to the games.

The resulting database contains 733 games, 72 referees and 11301 calls, involving 473 distinct players committing a foul and 442 players receiving a foul. The difference in the number of players committing *versus* receiving a foul is justified by the fact that in 'crunch time' teams tend to play more around their top scorers, and the NBA rules are engineered to disincentivize fouling on players that are not in possession on the ball during the last two

---

[7]Deutscher, Christian. "No referee bias in the NBA: New evidence with leagues' assessment data." Journal of Sports Analytics 1, no. 2 (2015): 91-96.

[8]In addition, this paper only worked on a subsection of the data that is currently available, since it was published only one year after the beginning of the program in 2015.

[9]The software can be downloaded here.

minutes.[10] In terms of calls, the database includes 3921 Correct Calls, 141 Incorrect Calls, 5700 Correct Non-Calls, 1035 Incorrect Non-Calls and 471 Unassessed Calls.

In the next page, we provided the list of the Top 10 committing players and disadvantaged players in terms of total number of fouls. These lists justify the theory that we provided above to explain the discrepancy in the total number of players on the two sides. As we can see, the list of top committing players includes nine of the most aggressive rim protectors in the league, who are expected to play defense aggressively in the key moments of the game. The Top 30 list for the same category only includes 5 guards. The only outlier in this list is James Harden, a point guard infamous for his weak defensive play. Our hypothesis is that due to his relatively low number of fouls (in the 2015-16 season he ranked 331st in the league in fouls per 36 minutes[11]) he is used by his team to provide intentional fouls in the key moments of close games. On the other hand, the Top 10 disadvantaged list is very straightforward: these are the top players in their respective teams, and are expected to be running plays in the last two minutes of a game. This list is also largely constituted by guards and small forwards, the primary ball handlers for any team, with only four 'big men' in the top 30.

Top Committing Players

Player	Number of Fouls
Al Horford	111
Brook Lopez	110
Draymond Green	110
James Harden	110
Marcin Gortat	110
Karl-Anthony Towns	107
Cody Zeller	104
Steven Adams	102
Marc Gasol	101
Serge Ibaka	99

---

[10]A foul on a player who is not holding the ball during the last two minutes of a game is punished with two free throws *and* a possession, while a foul on the ball handler is punished as a common foul, i.e. with the possession *or* the free throws, but not both.

[11]Source: Basketball Reference

Top Disadvantaged Players:

Player	Number of Fouls
James Harden	177
Russell Westbrook	148
LeBron James	130
DeMar DeRozan	129
Kemba Walker	120
Isaiah Thomas	119
Andrew Wiggins	118
DeMarcus Cousins	117
John Wall	114
Reggie Jackson	113

Top Referees by Number of Calls:

| Referee	Number of Calls
Jason Phillips	913
Monty McCutchen	824
Scott Foster	792
Josh Tiven	790
Kane Fitzgerald	781
Eric Lewis	779
Marc Davis	765
Ken Mauer	729
Tom Washington	702
Tony Brothers	686

2   Statistical Analysis

We will try to answer the following questions:

1. How often are referees wrong?

2. Do referees err on the side of caution? i.e. are there more incorrect non-calls versus incorrect calls?

3. Who are the best refs in the NBA?

4. Does the NBA give its best refs to the best games?

5. Do star players get better refereeing?

2.1   The Best Referees

Ranking the best referees in the league poses a methodological issue. A simple ranking in terms of error rate is methodologically incorrect due to the great variance in the total number of calls, which implies that the referees with fewer calls will be the ones with the most extreme results. The issue is visible in the following graph:
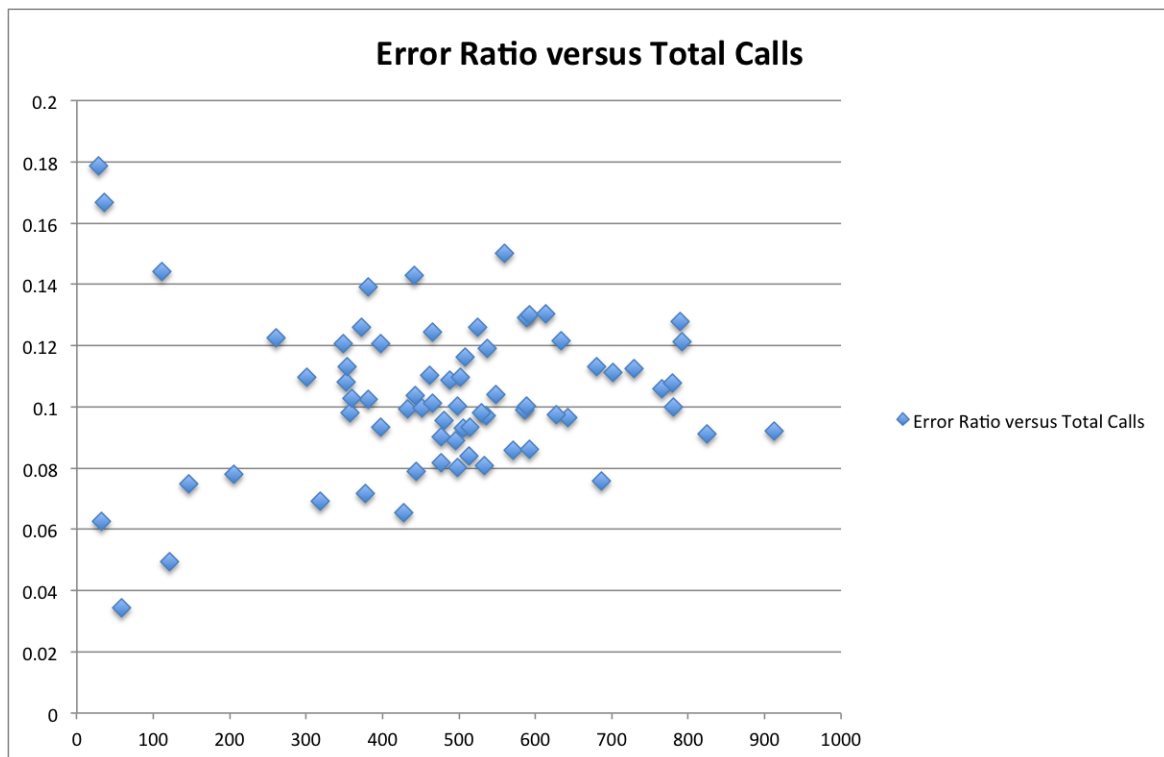
In order to obtain a meaningful ranking, therefore, we need to establish a cutoff. The error rate in the original dataset has the following statistics: $\sigma_0 = 0.02467404290499483$, $\mu_0 = 0.103690161047$, $c_v = 0.237959345957$, where c_v is the coefficient of variation. This already answers the first of the six questions that we posed: referees are wrong about 10% of the time, an error rate that is pretty substantial. However, what is more interesting is the split of this error rate: only $\frac{141}{141 + 1035} * 100 = 11.98\%$ of officiating mistakes are incorrect calls (IC). This validates the popular wisdom that referees err on the side of caution, i.e. they let the teams play the possession without whistling in the event of minor transgressions.[12] Just 1% of calls in the NBA are type 1 errors, i.e. instances of the referee blowing the whistle when they shouldn't have. Methodologically, to create our ranking we need to establish a cutoff that maintains a large portion of our data and minimizes σ while leaving μ unchanged (i.e. cuts off outliers equally on both extremes). Visually, from the previous graph, the two obvious cutoffs seem to be 300 and 400. We obtain the following results:

Statistic	Full Dataset	300 Cutoff	400 Cutoff
σ	0.02467404290499483	0.01816929587329971	0.017861323524206565
μ	0.103690161047	0.104041380372	0.10359995662
c_v	0.237959345957	0.174635282696	0.172406669915
Dataset Size	72	62	49

As we can see, there is a compelling methodological reason to prefer the 300 cutoff, since both the 300 and 400 cutoff result in a nearly identical change in standard deviation and variation rate, while leaving the mean untouched. Thus, since the 300 cutoff preserves a substantially larger amount of data, it is preferable. Here is the Top 10:[13][14]

Name	IC	INC	Total Incorrect Calls	Total Calls	Error Ratio
Mark Lindsay	1	27	28	428	0.0654205607476635
Tony Brothers	6	46	52	686	0.075801749271137
Lauren Holtkamp	5	30	35	444	0.0788288288288288
Steve Anderson	9	31	40	498	0.0803212851405622
Kevin Scott	6	37	43	533	0.0806754221388368
Bill Kennedy	7	32	39	477	0.0817610062893082
J.T. Orr	5	38	43	513	0.0838206627680312
Nick Buchert	4	45	49	571	0.0858143607705779
James Capers	7	44	51	592	0.0861486486486486
Derek Richardson	2	42	44	495	0.0888888888888889

[12]While this is a commonly acknowledged fact of basketball refereeing, it is still interesting to see it present itself in such a staggering fashion once it is backed by data

[13]The full ranking for the 300 and 400 cutoff can be accessed here and here.

[14]Lauren Holtkamp is the only woman in our dataset. Her incredible performance relative to her peers more than justifies the NBA's commitment to increase female participation at all levels of the league, including refereeing.

The striking feature of this list is that only of the Top 10 referees by number of calls, Tony Brothers, is included in it. Therefore, while these referees are the most accurate by our dataset, they are not the ones who are in charge of most games. There are multiple justifications that we can provide for this. First and perhaps foremost, the NBA does not evaluate the performance of its referees based solely on their accuracy. Other factors may include the ability to manage the game throughout its entirety, to control player behavior, and in general to enforce discipline on the court. Second, the less experienced referees may be in charge of relatively easier-to-officiate games, where players may not be trying to aggressively draw fouls.[15] However, if this were the case, we would expect to see a different proportion of shooting fouls over total fouls for the less experienced referees. We tried to verify if this is the case, by comparing the mean and standard deviation of the aforementioned ratio for the two Top 10 groups, but we obtained the following results (the acc subscript denotes the top 10 by accuracy, while the top subscript denotes the top 10 by number of calls):

Statistic	Value
μ_{10}	0.26609368803995859
σ_{10}	0.01704123262411838
μ_{acc}	0.25995995334064753
σ_{acc}	0.030250139208445232

As far as our test is concerned, therefore, we cannot identify whether top referees have 'harder' games to officiate; however, due to the lack of sophistication of our test, we still believe that this is a likely explanation.

2.2 Better Referees to Better Games?

The first issue with assessing whether better games get better referees is to define what constitutes a better game. The top 8 teams in the standings for the 2015-16 season were the following: Golden Sate, San Antonio, Cleveland, Toronto, Oklahoma City, Los Angeles Clipper, Atlanta and Boston. We take games involving any of these teams as a proxy for 'better' games. Our dataset includes 394 games played by any of these teams.

In order to investigate whether the top performing referees get to officiate a larger proportion of the games involving these teams, we compute the ratio of the games involving said teams by the top and bottom 10 performers in our ranking with cutoff at 300. We would expect the referees in the top 10 to officiate a somewhat larger proportion of these games due to their superior performance. However, the results suggest the opposite:

---

[15]Players may take advantage of an overly aggressive defense by seeking contact during a shot in order to gain free throws. A more detailed explanation of the practice is provided in this ESPN article.

Statistic	Value
μ_{top}	0.54796954314720814
σ_{top}	0.005793872645774596
μ_{bottom}	0.55456852791878175
σ_{bottom}	0.008136795288355024

As we can see, there is no difference whatsoever in the proportion of games officiated by top and bottom performers. Therefore, we conclude that the NBA, in its officiating decisions, opts to maintain a consistent quality of refereeing across the board, rather than providing top-quality officials to its marquee matchups.

2.3 Star Player Advantage

Since our dataset spans multiple years, we consider as Star Players those who started in at least one All-Star Game over the three seasons we are analyzing.[16] We compute the mean and standard deviation for the ratio of incorrect calls and non calls to total calls for star players, and obtain the following results:

Statistic	Value
μ	0.10166362916798044
σ	0.027509540916309735

Therefore, the performance of referees on these star players is remarkably consistent with their performance on the entire dataset, which has $\mu = 0.1037$. We can conclude that NBA referees are remarkably capable of adjusting for star player bias.

3 Conclusion

The key takeaways from our research are the following:

1. NBA referees are remarkably good. Considering the fast-paced nature of the game, it is incredible that the type-1 error rate is just 1%. As a term of comparison, the rate of false positives for ELIZA tests for Hepatitis C is about three times as high.[17] In fairness, however, the astoundingly low type-1 error rate can be partly attributed to the much higher (almost 10%) type-2 error rate. NBA referees seem to follow the old adage: 'Let the players decide the game'.

---

[16]We exclude Kobe Bryant from the dataset since his participation in the 2015-16 All-Star game was largely motivated by the fact that he had announced his retirement for the end of the season. Since the starters for the All-Star Game are decided by popular vote, Bryant's candidacy was widely supported by longtime fans; however, his status as an All-Star was not backed by the quality of his play. He had his second-worst season in terms of Value Over Replacement Player, posting a negative number in this metric, which indicates that his performance was inferior to the league's average player's.

[17]Chou R, Clark E, Helfand M. Screening for Hepatitis C Virus Infection [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2004 Mar. (Systematic Evidence Reviews, No. 24.) 3, Results. Available here.

2. We were not able to detect any relevant form of bias. Furthermore, the league does a good job in terms of ensuring consistent officiating quality across different games.

3. There was no correlation whatsoever between the performance of officials in our dataset and the relevance of the games in which they were involved. Excluding Bill Kennedy, James Capers, and Tony Brothers, the remaining officials in our Top 10 have only officiated 2 playoff games total.

This last point is perhaps the most disappointing takeaway of our analysis. The NBA has promoted its Last Two Minutes reports as a way to increase transparency. However, there seems to be no correlation between the performance according to these reports and the 'importance' (for lack of a better term) of a referee within the league. We believe, therefore, that the word transparency is misused in this instance. The reports do not allow us to catch a glimpse of how NBA officials are evaluated and held accountable for their mistakes. We are not able to discern from this dataset any relevant reason why the top referees in the NBA are who they are.

In conclusion, we endorse the position of many basketball greats in criticizing these reports.[18] As referees petitioned to eliminate the reports entirely,[19] NBA Commissioner Adam Silver stated "We do (Last Two-Minute Reports) to be as transparent as possible with the public in terms of how we think critical moments of the game are being officiated." In addition to the importance of transparency, Silver also stressed the importance of consistency. "In terms of building confidence in the public, they want to see consistency," Silver said. "So they want to understand if we call something a foul, why we called it a foul, and we often give explanations for why we believe something was a foul, whether it was correctly called or incorrectly called". While we fully believe that the reports achieve their second stated goal, consistency, in a remarkable fashion, we also wish to emphasize that they mostly fail in terms of transparency.

Therefore, we wish to see data that could actually give us an opportunity to understand what criteria determine the relative standing of referees and the assignments of the more prestigious games, especially in the playoffs. The solution is more data, not less, but insofar as the reports maintain their current format, it seems that their only contribution is to increase the social media buzz after games.[20]

---

[18]Gregg Popovich, LeBron James, Dwyane Wade, and Kevin Durant have been harsh critics of the reports in the past.

[19]Source: USA Today

[20]On March 9th 2017, a Google News research for the query 'last two minute reports NBA officials' returns 7760 results, over 1000 per day