Causal Analysis of Tactics in Soccer: The Case of Throw-ins

Nirodha Epasinghege Dona and Tim B. Swartz *

Abstract

This paper investigates optimal target locations for throw-ins in soccer. The investigation is facilitated by the use of tracking data which provides the positioning of players measured at frequent intervals (i.e. 10 times per second). The methods for the investigation are necessarily causal since there are confounding variables that impact both the throw-in location and the result of the throw-in. A simple causal analysis indicates that on average, backwards throw-ins are beneficial and lead to an extra two shots per 100 throw-ins. We also observe that there is a benefit to long throw-ins where on average, they result in roughly four more shots per 100 throw-ins. These results are corroborated by a more complex causal analysis that relies on the spatial structure of throw-ins.

Keywords : big data, causal inference, player tracking data, soccer, spatio-temporal analyses.

*N. Epasinghege Dona (nirodha_epasinghege_dona@sfu.ca) is a PhD candidate and T. Swartz (tim@stat.sfu.ca) is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been partially supported by the Natural Sciences and Engineering Research Council of Canada. The work has been carried out with support from the CANSSI (Canadian Statistical Sciences Institute) Collaborative Research Team (CRT) in Sports Analytics. The authors thank Daniel Stenz, former Technical Director of Shandong Luneng Taishan FC who provided the data used in this paper. The data underlying this article will be shared on reasonable request to the corresponding author. We thank the Editor and two anonymous Reviewers whose comments have helped improve the manuscript.

1 INTRODUCTION

The investigation of cause and effect relationships is a fundamental research topic in both the sciences and the social sciences. Traditionally, cause and effect relationships are studied through experiments where randomization is the primary technical tool for investigation.

In sport, cause and effect relationships are also important. For example, teams and individuals want to know whether particular tactics are effective winning strategies. However, in sport, data do not typically arise through randomized experiments. Rather, data are usually collected from matches, in non-experimental settings.

Fortunately, the development of causal methods (Pearl 2009) has provided opportunities to investigate cause and effect relationships in non-experimental settings. Whereas the identification and measurement of relevant confounding variables is a necessary and challenging component of causal methods, the hurdle appears less imposing in sport. In sport, objectives are often clear (e.g. score more goals than the opponent), matches terminate in reasonable timeframes (e.g. often two to four hours), and rules are well defined. Most importantly, with the advent of detailed player tracking data (e.g. spatio-temporal data), our sporting intuition often permits the identification and measurement of relevant confounding variables.

Causal inference in sport assisted by player tracking data is a relatively new but potentially fruitful research area. Wu et al. (2021) provided a template for such analyses in soccer where the benefit of crossing the ball was investigated. In this investigation, the response variable Y (resultant shot) was binary, and the treatment X (crossing) was binary. Wu et al. (2021) generated conclusions that were contradictory to some of the existing literature, where they indicated that crossing is a valuable tactic. Epasinghege Dona and Swartz (2023) expanded on these ideas to carry out a causal analysis regarding pace of play in soccer. In this investigation, the response variable Y (excess shots) was discrete, and the treatment X (pace) was bivariate and continuous. Epasinghege Dona and Swartz (2023) established that playing with pace is a valuable strategy, a conclusion that had not been previously established in soccer.

This paper extends the causal investigations of Wu et al. (2021) and Epasinghege Dona and Swartz (2023). Our analysis investigates the optimal locations of throw-ins in soccer. In this investigation, the response variable Y (resultant shot) is binary, and the treatment

X (throw-in reception location) is spatial. Stone, Smith and Barry (2021) have previously studied the throw-in problem using data from the English Premier League where they obtained the surprising result that backward throw-ins are more successful in terms of shot creation. Notably, Stone, Smith and Barry (2021) did not have access to player tracking data. Without tracking data, it is not possible to assess the extent to which the recipient of a throw-in is open. Unlike our paper, Stone, Smith and Barry (2021) did not carry out a causal analysis.

As mentioned, the availability of player tracking data provides the opportunity for a deep-dive analysis of throw-ins in soccer. With player tracking data, the location coordinates for every player on the field are recorded frequently (e.g. 10 times per second in soccer). With such detailed data, the opportunity to explore novel questions in sport has never been greater. The massive datasets associated with player tracking also introduce data management issues and the need to develop modern data science methods beyond traditional statistical analyses. Gudmundsson and Horton (2017) provide a review of spatio-temporal analyses that have been used in invasion sports where player tracking data are available.

In Section 2, we describe the player tracking data and discuss related challenges. We then describe how we construct the throw-in datasets from the tracking data. The throw-in datasets are the source files which are used for the causal analysis. Some exploratory data analyses are also provided. In Section 3, we discuss the use of propensity scores in causal investigations. Propensity scores describe the probability of the treatment X (spatial location of throw-in) given underlying covariates W. In Section 4, we present a causal analysis concerning the optimal locations of throw-ins relative to the position on the pitch. This is done in three ways. The first two approaches are simple as they are based on defining a binary variable corresponding to backward/forward throw-ins and (2) defining a binary variable corresponding to short/long throw-ins. We then consider a more complex analysis based on the full spatial treatment X. The main result from the analyses is that both backward throw-ins and long throw-ins confer a competitive advantage. We provide some concluding remarks in Section 5.

Apart from tactics, there have been many recent investigations in the literature related to soccer. A sample of diverse topics include match fixing (Forrest and McHale 2019), the evaluation of passing (Håland et al. 2020), competitive balance (Manasis, Ntzoufras and

Reade 2022) and the forecasting of match results (Hubáček, Šourek and železný 2022).

2 DATA

For this investigation, we have a big data problem where both event data and player tracking data are available for 237 regular season matches (three matches missing) from the 2019 season of the Chinese Super League (CSL). The schedule is balanced where each of the 16 teams plays every opponent twice, once at home and once on the road.

Event data and tracking data were collected independently where event data consists of occurrences such as tackles and passes, and these are recorded along with auxiliary information whenever an "event" takes place. The events are manually recorded by technicians who view film. Both event data and tracking data have timestamps so that the two files can be compared for internal consistency. There are various ways in which tracking data are collected. One approach involves the use of Radio Frequency Identification (RFID) technology where each player and the ball have tags that allow for the accurate tracking of objects. In the CSL dataset, tracking data are obtained from video and the use of optical recognition software. The tracking data consists of roughly 1.3 million rows per match measured on 7 variables where the data are recorded every 1/10th of a second. Each row corresponds to a particular player at a given instant in time. Although the inferences gained via our analyses are specific to the CSL, we suggest that the methods are applicable to any soccer league which collects tracking data.

2.1 The Throw-in Datasets

We pre-processed the CSL tracking and event data. Originally, the data were provided in xml files and we extracted content using the `read_xml` function from the `XML` package using R software. The resulting tracking and event data were written into csv file format.

Ultimately, we constructed throw-in dataframes for each match. These are comprehensive datasets that allows us to investigate various questions of interest related to throw-ins. A throw-in dataframe is a matrix where the rows correspond to throw-ins. Each throw-in has been translated and standardized such that throw-in angles and distances downfield are consistent according to direction that a team is attacking. The columns include the following basic variables: the identification of the throw-in team, the identification of the opponent team, the identification of the player who made the throw-in and the binary variable Y according to whether the end of possession from the throw-in resulted in a shot for the throw-in team. The variable Y serves as the response variable which indicates success related to a throw-in. An end of possession for the throw-in team occurs when the opponent gains possession, a whistle occurs, there is a stoppage in play or when the ball goes out of bounds. Although goals are a more direct measure of success, we note that goals are rare events in soccer with less than three goals per match on average in most professional leagues. We also record the relative spatial location $X = (r, \theta)$ of the received throw-in. By relative spatial location, we mean the length of the throw-in and its radius angle given the location on the field where the throw-in occurred to where the ball was received. The measurement is standardized with respect to the side of the field where the throw-in occurs. Using polar coordinates, the radius arm r is the length of the throw-in measured in metres and θ is the angle of the throw-in measured in degrees. For example, $X = (10, 90)$ describes a throw-in of length 10m that is thrown perpendicular to the touch line.

For the propensity scores described in Section 3, we wish to relate covariates W which have a potential impact on the relative spatial location X of the received throw-in. These variables are derived from our soccer intuition and are viewed as confounding variables in the causal analysis. In proposing covariates W, we take a broad perspective and introduce variables that may have even a hint of impacting the spatial locations of throw-ins. We introduce additional column variables $W = (t, d, f, o, b, r)$ to the throw-in dataframes where

$$
\begin{aligned}
t &\equiv \text{time of the throw-in in minutes, } t \in (0, 90) \\
d(t) &\equiv \text{score differential in favour of the throw-in team} \\
f(t) &\equiv \text{field location of the throw-in, } f(t) \in (0, 100) \\
o(t) &\equiv \text{openness of the receiver of the throw-in} \\
b &\equiv \text{pre-match betting odds corresponding to the throw-in team} \\
r &\equiv \text{red card variable corresponding to manpower advantage of throw-in team}
\end{aligned}
\tag{1}
$$

In (1), we define the time variable t such that throw-ins that occur during extra time in the first half are set to $t = 45$. For throw-ins that occur during extra time in the second half, we set $t = 90$. Therefore t is a mixed variable (both continuous and discrete).

The score differential $d(t)$ is a discrete variable and expresses the lead by the throw-in team. For example, $d(t) = -2$ indicates that the throw-in team is losing by two goals.

The field location variable $f(t)$ has been standardized to the interval $(0, 100)$ to account for fields of different length. For example, $f(t) = 50.0$ corresponds to a throw-in taken from midfield.

The openness variable $o(t)$ in (1) describes the degree to which the receiver of the throw-in is open. An open receiver is more likely to be targeted for a throw-in. To obtain $o(t)$, we first consider the shaded region in Figure 1 where only defenders in this region are assumed to pose a threat of intercepting the throw-in. The idea is that defenders who are behind the receiver can be "boxed out" by the receiver, and are not involved. Our experience is that a receiver with a defender on his back will move towards the ball, be able to keep the defender behind, and obtain possession. Thus, such a defender is not a threat to possession. We define a defender as "boxed out" if the defender is situated within 45 degrees from the perpindicular from the throw-in location to the receiver. Admittedly, the 45 degree angle is a bit arbitrary. The variable $o(t)$ is then calculated by taking the distance from the nearest defender in the shaded region to the receiver. In this way, longer distances convey greater openness of the receiver. If the throw-in is intercepted, $o(t) = 0.0$. Openness is related to the more complex notion of pitch control or field ownership. Pitch control was first introduced using Voronoi tesselations (Voronoi 1907, Kim 2004). More advanced metrics for field ownership are discussed in Wu and Swartz (2022).

For the fifth variable b in (1), we accessed pre-match betting odds available from the website https://www.oddsportal.com/soccer/china/super-league-2019/results/ . The betting odds (reported in decimal format and also known as European odds) provide us with the relative strength of the two teams. For simplicity, we temporarily ignore the vigorish imposed by the bookmaker. [1] In this case, the interpretation of betting odds b for a team is that the team has a pre-match probability $1/b$ of winning the match. Therefore, values of b slightly greater than 1.0 indicate a strong favourite whereas large values of b indicate an *underdog*. To better understand betting odds, consider fair odds b, a wager of

[1] For the actual analysis, consider bettings odds b_w, b_d and b_l corresponding to a team win, draw and loss respectively. For profitability, the bookmaker introduces a vigorish whereby $1/b_w + 1/b_d + 1/b_l > 1$. Therefore, the implied probability of a win is given by $p = (1/b_w)/(1/b_w + 1/b_d + 1/b_l)$. To measure the strength of a team, we instead base our analysis using betting odds defined as the reciprocal of p.

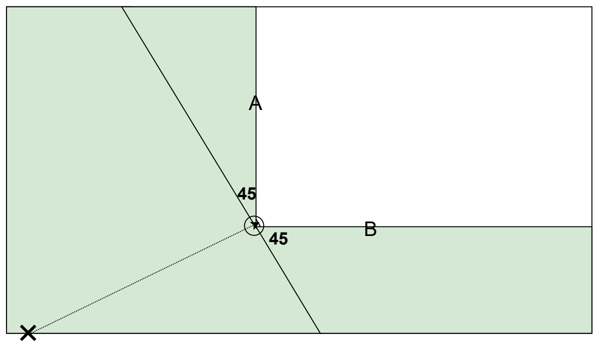

Figure 1: The figure depicts the line between thrower (X) and receiver (open dot), and two rays labelled A and B that are situated 45 degrees from the perpindicular to the line. The resultant shaded region corresponds to the area where defenders are assumed to pose a threat of intercepting the throw-in.

x dollars and probability p of winning the bet. The expected profit from such a wager is $-x*(1-p)+(xb-x)*p$ and setting this equal to zero yields $p = 1/b$. Our sporting intuition is that stronger teams may have different throw-in strategies than weaker teams. In general, stronger teams tend to play differently than weaker teams (see for instance, Silva and Swartz (2016)). Figure 2 depicts some of the variables described above.

The final variable r is binary and is set according to whether the throw-in team has a manpower advantage. One might expect the defensive team to behave differently (e.g. more players lined up behind the ball) in this scenario.

To create the throw-in dataframe, we looped frame by frame through the tracking data, where we matched events and time using the event data. The process required approximately 15 minutes of computation for all 237 matches.

To illustrate the propensity score variables W given by (1), consider a match where the score is 1-0. In the 70-th minute, teams are full-strength, a throw-in takes place at midfield for the leading team who are the favoured team with pre-match decimal betting odds $b = 1.5$ (with the vigorish removed). Further, suppose that the nearest defender to the receiver is standing along the sideline 8 metres away from the receiver. In this case,

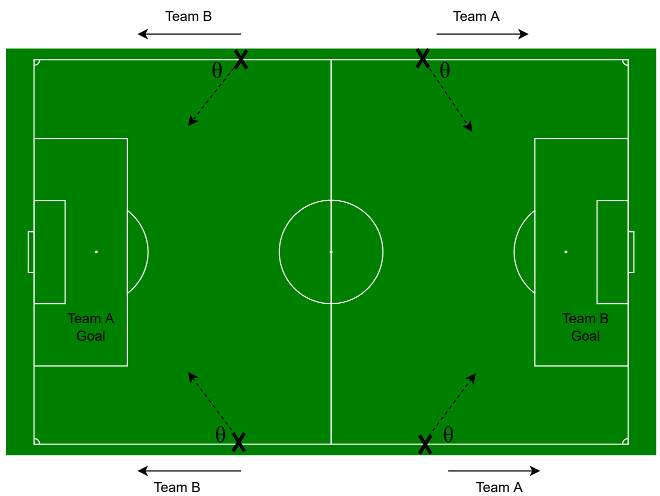

Figure 2: The plot illustrates some of the key variables corresponding to throw-ins. Four throw-ins are depicted where the angle θ is standardized accounting for the side of the field and the attacking direction. The four throw-ins each correspond to $r = 15$ metres and $f(t) = 60.0$.

$W = (t, d, f, o, b, r) = (70, 1, 50, 8.0, 1.5, 0)$.

The vector $W = (t, d, f, o, b, r)$ corresponds to our soccer intuition as a driver of the throw-in decision. We did consider other variables which did not have significant effects. For example, the speed of the target receiver was considered and this was obtained by taking the Euclidean distance of the player two frames before and two frames prior to the throw-in. We also experimented with the red card variable r in a categorical setting corresponding to manpower advantage, no advantage and manpower disadvantage. However, only manpower advantage proved significant, and hence it was reduced to a binary variable.

2.2 Data Management

With increasingly complex and large datasets in data science, the importance of data integrity cannot be overstated. Without accurate data, reliable inferences cannot be achieved. In this application, we constructed the throw-in datasets from the event and tracking data where the following issues presented challenges.

- Some throw-ins were identified as impossibly short (i.e. 0 metres in length). These were a consequence of foul throws (59 occurrences) and were removed from the dataset.
- Some throw-ins were identified as impossibly long (i.e. exceeding 40 metres in length). These were a consequence of an event happening during the throw-in which invalidated the throw-in (e.g. a foul), and from which the next event took place at a location other than the free-throw location. These events were rare (13 occurrences) and were removed from the dataset.
- In our event dataset, a throw-in is labelled before the throw-in occurred, and therefore, we should consider the next event to obtain the location where the throw-in was received. There are some events like player substitutions, red/yellow cards that occur before the throw-in. Therefore, we had to carefully eliminate these events before calculating the throw-in length r and angle θ.
- There were two throw-ins where the target receiver was not identified.

From the original 8467 throw-ins occurring in the matches, 8393 were useable for data analysis.

2.3 Exploratory Data Analysis

Exploratory data analyses often guide the development of formal statistical models. We present several plots related to our investigation.

In Figure 3, we provide a histogram of the direction variable θ associated with all throw-ins in the dataset. We observe that there are more forward throw-ins (i.e. $\theta < 90$) than backward throw-ins (i.e. $\theta > 90$). This corresponds to our intuition since the attacking team typically wishes to advance the ball downfield to a more threatening scoring position. The symmetric modes that we observe at approximately $\theta = 22.5$ and $\theta = 157.5$ are interesting. These throw-ins are close to the touch line.

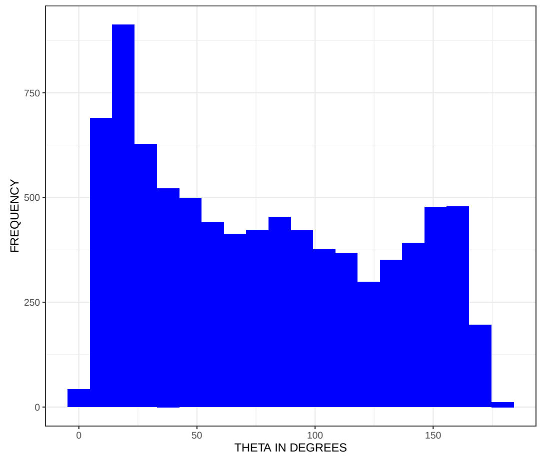

Figure 3: The histogram of the variable θ describes the throw-in angle relative to the sideline in the direction that the team is attacking.

In Figure 4, we provide a histogram of the radius arm r (or simply the length) associated with all throw-ins in the dataset. We observe a right-skewed histogram where the modal throw-in length is roughly 10 metres. There are several long throw-ins of approaching 40 metres in length. Whereas such throw-in distances may seem unlikely, these throw-ins may be the result of a ball that was not initially received and travelled for a period.

In Figure 5, we provide a histogram of the openness variable o which gives the distance from the receiver to the receiver's nearest opponent who is positioned in the shaded region of Figure 1. We observe that the median distance is roughly 10 metres and that the histogram is right-skewed. From a practical point of view, a player is comfortably open whether $o = 10$ metres or $o = 30$ metres, for example. Note that there were 126 cases in the throw-in dataset where there were no opponents in the shaded area (see Figure 1). These observations are not reflected in Figure 5, although they were retained for the causal analysis.

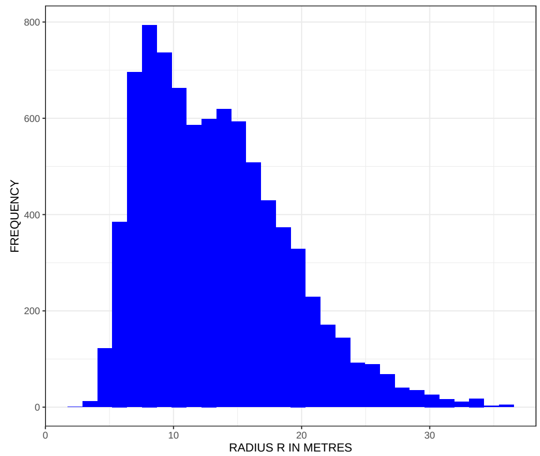

Figure 4: The histogram of the variable r describes the throw-in length.

3 PROPENSITY SCORES

Imagine temporarily that the throw-in problem was designed as a randomized experiment. For each throw-in, we would randomize the location $X = (r, \theta)$ of the received throw-in, and we would then relate Y (whether the completion of the possession resulted in a shot) to X. This would allow us to determine an optimal X. With randomization, the idea is that the underlying conditions leading to X would be nearly uniform across different realizations of X.

Of course, with match data, X is not randomized. And it is quite possible that not only does X depends on the covariate W in (1) but also Y depends on W. In other words, W is a confounding variable when investigating the relationship between Y and X.

We therefore wish to obtain propensity scores $\mathrm{P}(X \mid W)$ that describe how the probability of the treatment X (i.e. relative spatial location of the throw-in) is related to the confounding variable W. If we are able to do this, then through matching, we can compare Y_1 under a treatment X_1 relative to Y_2 under a treatment X_2 if X_1 and X_2 have similar propensity scores. This is the essential logic of the causal approach where propensity scores are used as a substitute for randomization.

Whereas we utilize a propensity score matching (PSM) approach, it is also possible to carry out analyses based on weighted propensity scores (PSW). The latter has the advantage

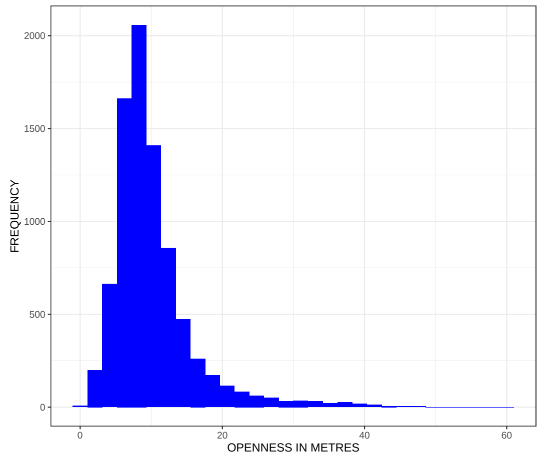

Figure 5: The histogram of the openness variable o describes the degree to which the receiver is open (see formal definition in Section 2.1).

that all observations can be used in the analysis. Narita, Tena and Detotto (2023) provide an insightful tutorial on the use of propensity score analyses with particular attention to PSW analyses.

We consider three causal analyses in Section 4: (1) backward throw-ins versus forward throw-ins, (2) long throw-ins versus short throw-ins and (3) a composite analysis based on the full spatial variable $X = (r, \theta)$. The propensity score models that we use in these three analyses are logistic, logistic and random forests, respectively.

4 CAUSAL ANALYSES

We return to the primary question concerning the optimal locations involving throw-ins. In Section 3, we have developed a propensity score model which yields scores $\mathrm{P}(X \mid W)$. Our objective now is to investigate the causal relationship between the binary response variable Y (whether the throw-in possession results in a shot) and the spatial treatment variable $X = (r, \theta)$.

We present three analyses where the first two analyses address the following simple questions: (1) Is a forward throw-in preferable to a backward throw-in? (2) Is a long throw-in preferable to a short throw-in? For the third analysis, we use more sophisticated methods to investigate the impact of the full spatial variable $X = (r, \theta)$ on Y.

For the purposes of the simple causal analyses in Sections 4.1 and 4.2, we make some adjustments to the confounding variable $W = (t, d, f, o, b, r)$ presented in equation (1). First, we discretize the time variable t according to the two categories $t < 45$ minutes and $t > 45$ minutes. This has been done since we are doubtful that the corresponding response variables are linear with respect to t, and we believe that the two halves of a match reflect different playing styles. We also categorize the score differential to $d(t) = -2, -1, 0, 1, 2$ where $d(t) = -2$ indicates that the throw-in team is losing by a large margin (two or more goals) and $d(t) = 2$ indicates that the throw-in team is winning by a large margin (two or more goals). With respect to goal differential, we attempted expanding the categories to $d(t) = -3, -2, -1, 0, 1, 2, 3$. However, we observed several insignificant effects and we believe this was due to few observations corresponding to the cells $d = -3$ and $d = 3$. We discretize the field location variable f according to $f < 67$ and $f \geq 67$ since there are tactical differences in the final third of the pitch. Third, we truncate the openness variable o such that values of $o > 10$ metres are set according to $o = 10$. This is done because we believe there is a meaningful difference in openness between $o = 1$ metres and $o = 2$ metres, for example. However, for $o > 10$, all throw-in receivers are effectively open. We did experiment with different thresholds of openness (e.g. $o > 12$ metres) but found that this made little difference in the causal analyses.

4.1 Causal Analysis based on Throw-in Direction

We simplify the problem involving the spatial causation variable $X = (r, \theta)$ to a binary context such that the control $0 < \theta < 90$ corresponds to a forward throw-in and the treatment $90 < \theta < 180$ corresponds to a backward throw-in. Therefore, the corresponding propensity score becomes $P(90 < \theta < 180 \mid W)$ which we fit using logistic regression. In this framework, there are $n_0 = 5023$ control observations and $n_1 = 3370$ treatment observations. Of course, there may be other classes of interest with respect to throw-in direction (e.g. sideways throw-ins); the full spectrum of throw-in directions are analyzed in Section 4.3.

In Table 1, we provide the results of logistic regression based on the variable $W = (t, d, f, o, b, r)$ described in Section 2.1. We observe that the time t of the match is significant where more throw-ins go backward as the time progresses. This may be a function of teams tiring and less willing to move forward up the pitch. The goal differential d is highly significant where we observe that greater leads are associated with forward throw-ins. This may be a consequence of the leading team playing better, having more confidence and energy, and consequently moving downfield more frequently. The throw-in position variable f is highly significant. As the throw-in location moves into the attacking third, there are more backwards throw-ins due to the constraints of the endlines. The openness variable o is significant and this aligns with our intuition. With teams defending their own goal, we expect more openness with backwards throw-ins. The betting odds variable b is highly significant and indicates that weaker teams tend to have more forward throw-ins. This is an interesting result and may be explained that these teams have less confidence and perhaps feel that the only way for them to succeed is to move forward. Conversely, stronger teams are typically more comfortable on the ball, willing to build up play through structured passing and possession, and thus, more likely to throw backwards. With respect to the red card variable, we observe that the throw-in team (with a manpower advantage) tends to have more backward throw-ins. This may be explained by the defensive team playing a more cautious style with more defenders behind the ball.

Variable	Estimate	Std Error	p-value
intercept	-0.184	0.134	0.171
time $t(45, 90)$	0.108	0.050	0.030 *
goal differential $d(-1)$	-0.208	0.091	0.021 *
goal differential $d(0)$	-0.575	0.088	5e-11 ***
goal differential $d(1)$	-1.185	0.101	2e-16 ***
goal differential $d(2)$	-1.250	0.122	2e-16 ***
field location $f(\geq 67)$	0.706	0.047	2e-16 ***
openness o	0.026	0.011	0.018 *
betting odds b	-0.050	0.008	5e-10 ***
red card r	0.641	0.176	3e-04 ***

Table 1: Results from logistic regression in Section 4.1 which determines the propensity scores $P(90 < \theta < 180 \mid W)$.

Since $n_1 < n_0$, the matching concept (Austin 2011, Imbens 2004) is that we attempt to match each of the n_1 treatment cases with a corresponding control case so that each pair has a similar estimated propensity score based on the underlying match circumstances W. Then the intention is that the resulting two groups (controls and treatments) are similar in the match characteristics, and the differences between the two groups can be attributed to the treatment (i.e. backward throw-in). There are many ways that the matching of propensity scores can be carried out (Stuart 2010). For example, matching may be carried out either with or without replacement. Matching may also be greedy (where each treatment case is matched with the closest eligible control case) or performed to optimize some global criterion. Further, randomization can be introduced in the matching procedure so that sensitivity due to the matching can be assessed. There are some downsides of propensity score matching (Guo, Fraser and Chen 2020). For example, with unequal treatment and control groups, matching results in a loss of data. Also, propensity score matching reduces the dimensionality of the covariate vector to a single dimension.

In our application, we begin with the n_1 cases where the throw-ins are backward, and we use a nearest neighbor method for selecting the matched cases where the throw-ins are forward. Specifically, we use the *Matching* package (Sekhon 2011) in the statistical programming language R to randomly select (with replacement) control cases that fall within a specified tolerance of the propensity scores for the treatment cases. Sampling with replacement tends to increase the quality of matching when compared to sampling without replacement. Unlike deterministic matching procedures, the random aspect of the nearest neighbor procedure allows us to repeat analyses to check the sensitivity of the inferences. We repeated the analyses 1000 times. To get a sense of the matching, for each of the 1000 analyses, we recorded the maximum absolute difference in propensity scores. This quantity was then averaged over the 1000 analyses and yielded the difference 0.0048.

We tested for balance in the covariate distributions $W = (t, d, f, o, b, r)$ of the matched treatment and control groups using the two-sample t-test (Rosenbaum and Rubin 1985). For a particular matching (selected randomly), the p-values corresponding to t, d, f, o, b, r were 0.222, 0.396, 0.116, 0.962, 0.190 and 0.640, respectively. The lack of significance suggests that there is balance in the matching across the confounding variables.

Following the implementation of the matching procedure, we calculate the average treatment effect ATE $= \bar{Y}(1) - \bar{Y}(0)$ where $\bar{Y}(1)$ is the average number of resultant shots from a backward throw-in and $\bar{Y}(0)$ is the average number of resultant shots from a forward throw-in. We obtained ATE $= 0.018$ with standard error 0.006 leading to p-value 0.001.

This was based on the 1000 iterations of the matching procedure using $n_1 = 3370$ matched pairs. The result is significant and suggests that backward throw-ins are beneficial. This corroborates the findings of Stone, Smith and Barry (2021). Specifically, our causal analysis indicates that from 100 backward throw-ins, roughly two more will result in a shot than if the throw-ins had been forward. This is a meaningful result in terms of gaining a competitive advantage.

In Figure 6, we present a more nuanced view of the situation for a randomly selected case involving matching. For each group (treatment and control), we smooth the variable Y with respect to the propensity score. On average, under our model's specifications, we observe that there is no advantage to executing a forward throw-in. As the propensity scores increase (i.e. conditions more favorable to making a backward throw-in), the benefit of the backward throw-in (in terms of shots) increases compared to making a forward throw-in. This implies that players tend to make the correct decisions with respect to the direction of throw-ins. As backward throw-ins become more probable, backward throw-ins have higher probabilities of successful outcomes.

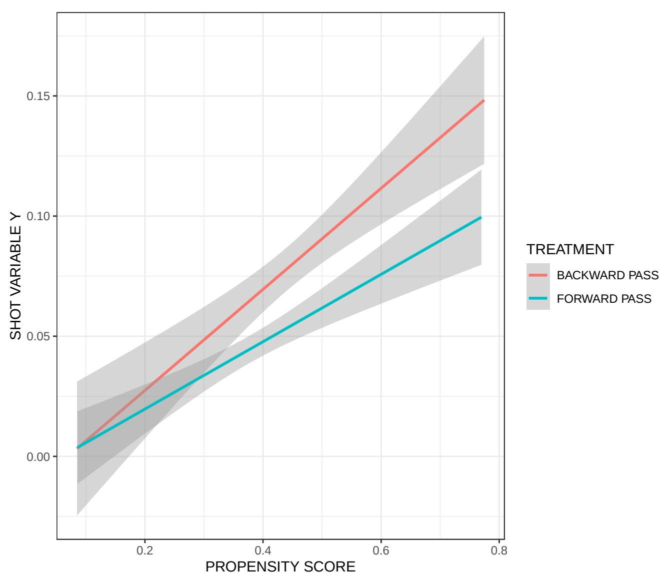

Figure 6: Smoothed plots of the shot variable Y with respect to the propensity score for backward throw-ins (treatment red) and for forward throw-ins (control blue).

4.2 Causal Analysis based on Throw-in Length

In this section, we consider a second inferential question involving the length of throw-ins. In recent years, the "long" throw-in has gained popularity in professional soccer, and we wish to investigate whether the long throw-in confers an advantage. We again simplify the problem involving the causation variable $X = (r, \theta)$ to a binary context. In this case, we define the treatment of a long throw-in as $r > 15$ metres and $60 < \theta < 120$. The dual condition is imposed so that long throw-ins do not correspond to throw-ins along the touchline (i.e. θ near 0 or 180). Rather our interest is focused on long throw-ins that are directed towards the middle of the pitch. Such long throw-ins appear to be an increasingly common tactic. Therefore, the corresponding propensity score becomes $P((r > 15) \cap (60 < \theta < 120) \mid W)$ which we fit using logistic regression. In this framework, there were $n_0 = 7831$ control observations and $n_1 = 562$ treatment observations.

In Table 2, we provide the results of logistic regression based on the variable $W = (t, d, f, o, b)$. Note that the red card variable r was not included in the analysis of Table 2 since it was not statistically significant. In this analysis, we do not have as many statistically significant terms as in Table 1. However, we do note that the coefficient estimates generally correspond to our soccer intuition. For example, we expect more longer throw-ins in the second half where one of the teams may be desperate and in need of a goal. This same pattern appears with respect to the goal differential where a team trailing by one goal (desperate) is more likely to make a long throw-in. When a team is tied or leading (i.e. $d = 0, 1, 2$), they are less likely to make a long throw-in. The positive coefficient for f is also sensible as long throws are more common in the attacking third. The openness variable o is highly significant; in order to retain possession on the throw-in, it is reasonable that long throw-ins require receivers to be more open than with short throw-ins.

We carry out the matching procedure as described in Section 4.1. To check the sensitivity of the inferences, we repeated the analyses 1000 times. For each of the 1000 analyses, we recorded the maximum absolute difference in propensity scores. This quantity was then averaged over the 1000 analyses and yielded the difference 0.0002. We then obtained the average treatment effect ATE $= \bar{Y}(1) - \bar{Y}(0) = 0.042$ with standard error 0.016 and corresponding p-value 0.004. This was based on the 1000 iterations of the matching procedure using $n_1 = 562$ matched pairs. Here, $\bar{Y}(1)$ is the average number of resultant shots from

Variable	Estimate	Std Error	p-value
intercept	-3.613	0.274	2e-16 ***
time $t(45, 90)$	0.355	0.095	2e-04 ***
goal differential $d(-1)$	0.116	0.161	0.472
goal differential $d(0)$	-0.113	0.160	0.477
goal differential $d(1)$	-0.602	0.194	0.002 **
goal differential $d(2)$	-0.444	0.228	0.052
field location $f(\geq 67)$	0.212	0.089	0.017 *
openness o	0.111	0.023	2e-06 ***
betting odds b	-0.024	0.015	0.135

Table 2: Results from logistic regression in Section 4.2 which determines the propensity scores $P((r > 15) \cap (60 < \theta < 120))$.

a long throw-in and $\bar{Y}(0)$ is the average number of resultant shots otherwise. The result indicates that long throw-ins are beneficial as they lead to approximately four more shots per 100 throw-ins. In this analysis, the p-value is larger than in Section 4.1 but is still significant. We note that the distance analysis presented here involves fewer observations than the directional analysis of Section 4.1. The result indicates that the recent trend involving more long throw-ins is a sound tactic.

In Figure 7, we consider a randomly selected case involving matching. We smooth the variable Y with respect to the propensity score for each group (treatment and control). We again see that professional soccer players are making the correct decisions. As it becomes more likely for executing a longer throw-in, the benefits of doing so increase. Inspecting the propensity scores, we observe that long throw-ins are relatively rare. This suggests that teams may consider increasing the frequency of long throw-ins.

4.3 Causal Analysis based on Full Spatial Location of Throw-in

In this section, we utilize the full spatial variable $X = (r, \theta)$ to gain insight on the causal relationship between X and the shot variable Y. Gelman and Meng (2004) consider structures beyond the simple binary X as analyzed in Sections 4.1 and 4.2. Here, we use machine learning methods to obtain the propensity scores.

A rationale for machine learning methods in prediction is that complex phenomena

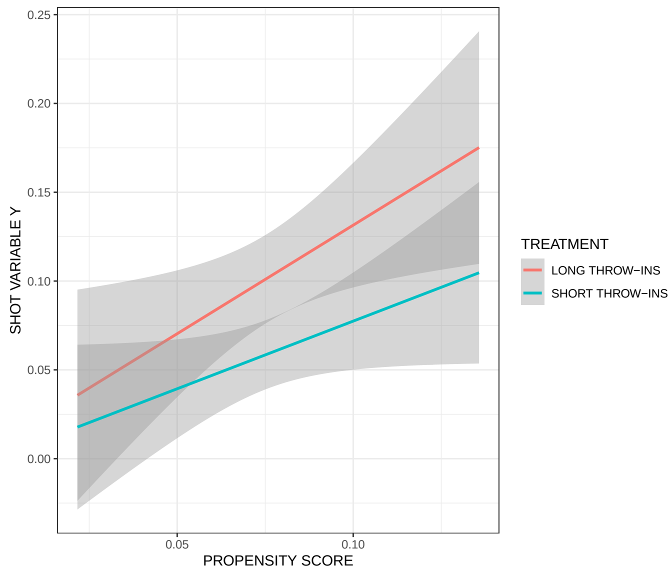

Figure 7: Smoothed plots of the shot variable Y with respect to the propensity score for the treatment group involving long throw-ins (red) and for the control group (blue).

are often difficult to model explicitly. Here, we have a two-dimensional spatial response variable $X = (r, \theta)$, and an explanatory vector W described by (1). We have little apriori knowledge about the relationship between X and W. For example, the relationship may only involve a subset of the variables W, the components of W may be correlated, and most importantly, the relationship $X \approx g(W)$ involves an unknown and possibly complex function g. In addition, the stochastic aspect of the relationship is typically unknown.

For this application, we use random forests as the chosen machine learning algorithm. Random forests (Genuer and Poggi 2020) are particularly easy to implement using the *randomForest* package (Liaw and Wiener 2002) in the R programming language. The basic idea is that a random forest is a collection of many decision trees where prediction results are aggregrated over trees. The use of multiple trees improves prediction and makes inference less reliant on a single tree. The splits in the trees accommodate non-linear relationships and terminal nodes provide the estimated probabilities of discretized values of X.

A feature of random forest procedures is that the cutpoints for component variables in W (which determine nodes in trees) are obtained optimally by the algorithm. The random forests procedure provides us with propensity scores $P(X \mid W)$ for data (X, W) which is a necessary ingredient of the spatial causal analysis. In this analysis, $P(X \mid W)$ is the probability that the ball is received at spatial location $X = (r, \theta)$ given the match situation W.

In choosing the tuning parameters of the random forest procedure, we aimed for predictive accuracy. We used the grid search method to obtain the optimized hyperparameters of the random forest model. From that, we obtained ntrees = 300, mtry = 1 and nodesize = 2. All the other hyperparameters were set to their default values. For the evaluation of model performance, we used 10-fold cross validation.

In Figure 8, we present the feature importance plot of the variables $W = (t, d, f, o, b, r)$ used in the random forest procedure. The plot is provided as part of the *randomForest* package. As in Section 4.1 and Section 4.2, we observe that the variables t, d, f, o, b are important. In particular, t, f, o, b are roughly of the same importance with goal differential d slightly less important. The red card variable r does not appear important in the full spatial analysis.

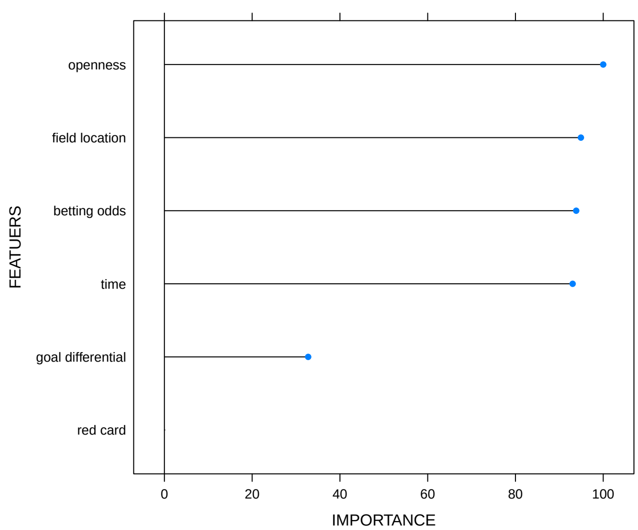

Figure 8: A plot of feature importance for the random forest procedure of Section 4.3.

Our causal investigation begins by discretizing the two-dimensional space X where throw-ins may be received. The region is truncated such that we only include observations extending 18 metres vertically from the throw-in location and 20 metres from the throw-in location horizontally (both left and right). Therefore, the area of the region is 720 squared metres (i.e. 18 metres by 40 metres). The region is then divided into rectangles of dimension 4 metres (horizontally) by 2 metres (vertically). This leads to 720/(2*4) = 90 rectangles. Based on $n = 7704$ throw-ins in the truncated region, we would expect $7704/90 \approx 85$ throw-ins on average, per rectangle. We amalgamate neighbouring rectangles when the number of observations in a rectangle is less than 30. There were 19 rectangles with fewer than 30 observations. After joining rectangles with insufficient observations, we were left with 71 rectangles. For every throw-in received, there is a propensity score $P(X \mid W)$ obtained using the machine learning methods based on random forests.

The matching idea used previously in the binary analyses of Sections 4.1 and 4.2 is now extended for the grid structure. We begin by randomly selecting a throw-in and noting its propensity score $p = P(X \mid W)$. Within each of the remaining 70 rectangles, we then select the throw-in whose propensity score is closest to p; these are the matching observations.

The process in the preceding paragraph is repeated $M = 30$ times. This means that there are 30 observations within a given rectangle, and each of these observations is matched to an observation in each of the remaining 70 rectangles. For each rectangle, we calculate the average number of shots $\bar{Y}$ generated by the throw-ins within the rectangle.

There is variability in the procedure due to the initial $M = 30$ observations that were randomly selected. Therefore, the entire procedure is repeated 100 times with $\bar{Y}$ for each rectangle averaged over the 100 iterations. We investigated the matching of propensity scores by calculating the maximum absolute difference of propensity scores across the $\binom{71}{2}$ pairs. This was then averaged over the $M = 30$ observations and the 100 iterations giving a value of 0.003. Therefore, the small difference suggests that the matching was successful.

To investigate the causal effect of X on Y, we produce a smoothed heat map of the average treatment effect $\bar{Y}$. In Figure 9, the heatmap is smoothed using the function *interp.loess* in the R package *tgp*. We observe darker regions (i.e. larger $\bar{Y}$) to the left (backward throw-ins) that are not too long (i.e. less than 5 metres outward). We also observe darker regions near the top (longish throw-ins from roughly 10 metres to 17 metres). This corroborates our findings from Section 4.1 and Section 4.2.

Our investigation of variability associated with the matching procedure involves calculating the standard deviation $s(\bar{Y})$ for each rectangle. For a particular rectangle, we obtain $\bar{Y}_i$ for the i-th iteration, $i = 1, \ldots, 100$. The quantity $s(\bar{Y})$ is the resultant sample standard deviation corresponding to the $\bar{Y}_i$ values. Then $s(\bar{Y})$ is averaged over the 71 rectangles where we obtain $\bar{s}(\bar{Y}) = 0.064$. From the color coding legend in Figure 9, it is apparent that the variability due to matching does not lead to maps with meaningful colour differences.

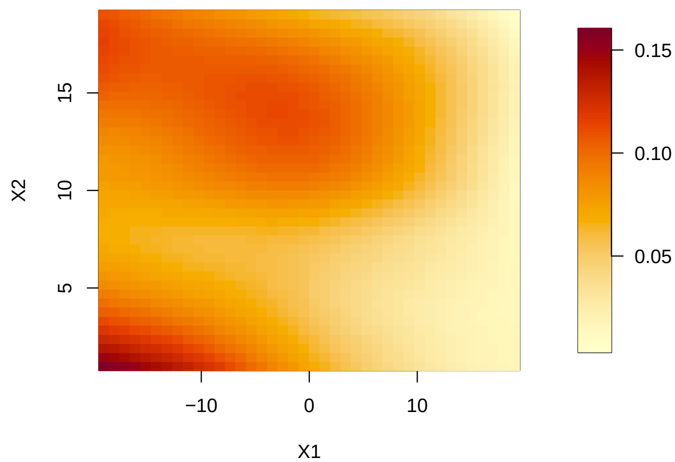

Figure 9: A smoothed heat map of the average treatment effect $\bar{Y}$ over the grid of the reception locations of throw-ins according to the algorithm of Section 4.3. The point $(X1, X2) = (0, 0)$ refers to the throw-in location.

5 CONCLUSIONS

The evaluation of tactics is a difficult and important problem for teams seeking to gain a competitive edge. This paper uses causal methods facilitated by tracking data to investigate throw-ins in soccer. Our results suggest the surprising result that backward throw-ins are more effective than forward throw-ins. It is surprising since the receptor of a backward throw-in is in a less threatening offensive position on the field. We also demonstrate the benefit of the long throw-in, a tactic that appears to be increasing in usage but whose benefits have not been previously quantified.

We see this work as a template for the use of causal methods in sport to assess tactics. Of course, a limitation of causal methods is the latency of confounding variables W that effect the tactic X and the response Y. An underlying premise of our work is that sport specific knowledge and tracking data permit the identification of the important confounding variables.

There is an important and practical question related to our work. Given a particular game situation $W^* = (t, d, f, o, b, r)$, how should the throw-in be executed X to optimize Y? Let's assume that all confounding variables have been identified. Then we wish to compare a throw-in tactic X_0 against a throw-in tactic X_1 both occurring under W^*. There would be n_0 observations under X_0 and n_1 observations under X_1. Unfortunately, n_0 and n_1 would be small (likely 0 or 1), and therefore a meaningful comparison could not be carried out. Perhaps there is some way around this, maybe by categorizing W^*, X_0 and X_1 into larger classes of interest. With the provision of more data, this may be a future research direction.

Whereas tactics related to set plays are perhaps the easiest and first investigations that come to mind, we also wish to continue the analyses of tactics to more complex scenarios and across various sports.

6 REFERENCES

Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46, 399-424.

Epasinghege Dona, N. and Swartz, T.B. (2023). A causal investigation of pace of play in soccer. *Statistica Applicata - Italian Journal of Applied Statistics*, To appear.

Forrest, D. and McHale, I.G. (2019). Using statistics to detect match fixing in sport. *IMA Journal of Management Mathematics*, 30(4), 431-449.

Gelman, A. and Meng, X-L., Editors (2004). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley: New York.

Genuer, R. and Poggi, J-M. (2020). *Random Forests with R*, Springer: New York.

Gudmundsson, J. and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), Article 22.

Guo, S., Fraser, M. and Chen, Q. (2020). Propensity score analysis: Recent debate and discussion. *Journal of the Society for Social Work and Research*, 11(3), 463-482.

Håland, E.M., Salte Wiig, A., Stålhane, M. and Hvattum, L.M. (2020). Evaluating passing ability in association football. *IMA Journal of Management Mathematics*, 31(1), 91-116.

Hubáček, O., Šourek, G. and železný, L.M. (2022). Forty years of score-based soccer match outcome prediction: An experimental review. *IMA Journal of Management Mathematics*, 33(1), 1-18.

Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86, 4-29.

Kim, S. (2004). Voronoi analysis of a soccer game. *Nonlinear Analysis: Modelling and Control*, 9(3), 233-240.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomForest. *R News, The Newsletter of the R Project*, Vol 2/3, 18-22.

Manasis, V., Ntzoufras, I. and Reade, J.J. (2022). Competitive balance measures and the uncertainty of outcome hypothesis in European football. *IMA Journal of Management Mathematics*, 33(1), 19-52.

Narita, K., Tena, J.D. and Detotto, C. (2023). Causal inference with observational data: A tutorial on propensity score analysis. *The Leadership Quarterly*, Accessed May 16, 2023 at https://doi.org/10.1016/j.leaqua.2023.101678

Pearl, J. (2009). *Causality: Models, Reasoning and Inference, Second Edition*, Cambridge University Press: Cambridge.

Rosenbaum, P.R. and Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Sekhon, J.S. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software*, 42, 1-52.

Silva, R. and Swartz, T.B. (2016). Analysis of substitution times in soccer. *Journal of Quantitative Analysis in Sports*, 12(3), 113-122.

Stone, J.A., Smith, A. and Barry, A. (2021). The undervalued set piece: Analysis of soccer throw-ins during the English Premier League 2018-2019 season. *International Journal of Sports Science and Coaching*, 16(3), 830-839.

Stuart, E.A. (2010). Matching methods for causal inference. A review and a look forward. *Statistical Science*, 25(1), 1-21.

Wu, Y., Danielson, A., Hu, J. and Swartz, T.B. (2021). A contextual analysis of crossing the ball in soccer. *Journal of Quantitative Analysis in Sports*, 17(1), 57-66.

Wu, Y. and Swartz, T.B. (2022). A new metric for pitch control based on an intuitive motion model. *Manuscript under review.*