

Can Fresh Legs Kick?

Evaluating how last-minute substitutes perform in penalty shootouts

David Wolfe Bender*

Indiana University

August 31, 2023

Abstract

Penalty shootouts play a major role in some of the most notable association football tournaments in the world. The World Cup, the European Championships, and the UEFA Champions League are just a few of the competitions that use penalty shootouts to resolve a draw after 120 minutes of play. Many penalty shootouts occur in high-profile matches. For example, the 2022 World Cup and the 2020 European Championship both ended in a penalty shootout. Some team managers employ a strategy of bringing on a substitute — or “fresh legs” — to take a penalty kick; other managers name their starters — who have played 120 minutes by the end of extra time — to the penalty shootout lineup. What strategy is more effective? Are substitutes better or worse at taking penalties? **Or, in other words, can fresh legs kick?**

Introduction

The day is July 22, 2021. Two of the best international football teams in the world gather at Wembley Stadium in London. England, the fourth-best team in the world according to FIFA metrics, is up against Italy’s seventh-ranked squad. It’s the final match of the 2020 European Championships, a highly awaited tournament delayed a year due to the COVID-19 pandemic.

The paths to the final for each of the squads wasn’t easy. In the quarterfinals, Italy took down top-ranked Belgium in a demanding quarter final, 2-1. In the semifinal, Italy held a one-goal lead over the Spanish for 20 minutes in the second half before Alvaro Morata brilliantly placed the ball to the right of Gianluigi Donnarumma, Italy’s goalkeeper. In penalties, all but one of Italy’s five kickers scored. The Spanish didn’t have that luck; two of their players missed, Morata included.

The English forced their way out of a difficult group that included Croatia, a top 15 team named the runner up at the 2018 FIFA World Cup in Russia. Then, a date with

*I would like to offer my deepest appreciation to Professor Christopher DeSante of Indiana University for his guidance in my statistics career. He made this report possible.

Germany saw The Three Lions reign victorious by a two-goal margin. It took extra time for England to take down Denmark, but they nonetheless prevailed; star captain Harry Kane stepped to the line for a penalty kick in the 104th minute. The initial shot was stopped by the Danish goalkeeper, but the rebound came directly to Kane, bringing relief to English fans and sending his team through to the finals. Brits sang in the streets from London to Lancaster: “It’s coming home.”

In the final match, England got on the board in the 2nd minute. It took Italy until the 67th minute to score in front of the stadium’s 67,173 spectators. Neither team could find any success for the remainder of the second half, and neither team could find the back of the net in the 30-minute extra-time period.

The extra-time period wasn’t all boring, though. England Manager Gareth Southgate drew the attention of analysts and fans alike when he made a major move: bringing in two substitutes in the 120th minute, the last minute of the match.

Those two players were Marcus Rashford and Jadon Sancho, both young stars on the English team. When time expired and the team lined up for penalties, both of them were in the rotation, kicking third and fourth.

After two kicks for each squad, England is winning 2-1, with Jordan Pickford, England’s keeper, pulling an impressive stop on Italy’s Andrea Belotti. Italy ties it on their kick. Rashford steps back from the spot, takes two strides to his left, slows down as he approaches the ball, and sends it toward the left post. Rashford puts his hands to his head as it hits the left post and heads out of play.

Italy makes another. They now lead 3-2. Jadon Sancho’s slow approach to the ball also doesn’t work: Donnarumma, with thousands of English fans watching in the crowd, stops the kick. Italy’s lead is solidified.

A few kicks later, it’s over. Despite the songs convincing the world that “it’s coming home,” the trophy didn’t come home to England. Italy denied the Lions the chance to lift its first major trophy since 1966.

After Rashford and Sancho miss their kicks, English fans take to social media to express their frustrations. Many of the posts are blatantly racist (all three English kickers who missed were all Black: Rashford, Sancho, and Bukayo Saka). Those posts, of course, were deplorable. But some more reasonable individuals on social media raised an interesting quandary: are substitutes less likely to score on penalty kicks in penalty shootouts? Was Southgate’s strategy statistically ill-advised? Those are the question this study seeks to answer.

1 Methodology of Data Collection and Analysis

1.1 Penalty shootouts are surprisingly uncommon

Even the biggest football fans likely do not think about how rare penalty shootouts actually are in the sport. Most club leagues across the world do not use a playoff format; rather, teams play a set number of games and earn a ranking in a league table, with the top team

in that table being crowned a champion (the “point-ranked model”).

That structure is different from most American sports, where teams are traditionally ranked before playing in a series of games to “move on” in the tournament (the “rank and playoff model”). This is the structure found in the National Basketball Association, the National Hockey League, the National Football League, Major League Baseball, and Major League Soccer.

However, the rank and playoff model is rarely used in football internationally. The English Premier League — the most-watched sports league in the world — uses a point-ranked model, as do most professional football leagues.

Because of the lack of playoffs, there are remarkably few penalty shootouts each year. Many of the high-profile penalty shootouts come in international competitions, such as the FIFA World Cup or the UEFA European Championship. There are a few examples of tournaments between football clubs that feature penalty shootouts, but they are limited.

Therefore, we remain focused on the few club tournaments and many international tournaments or competitions that feature penalty shootouts.

1.2 Finding the right sample

The first step was finding a good group of tournaments/competitions to examine. High priority was placed on achieving a diversity in time period, type (club vs. international competition), confederation (which region of the world), and country.

We examined 11 different tournaments or competitions (for the purpose of this report, we’ll refer to them as “competitions”). Five of those international competitions were international, while the other six were club competitions.

1.3 Developing Hypotheses

We created our hypotheses before the data was collected. Additionally, we selected our sample *before* we completed any data analysis. We did this in the hopes of avoiding a situation where we develop a hypothesis based on analysis.

In a perfect world, we would have selected our sample before developing a hypothesis. This was not possible because the interest of this topic drew us to the data.

In total, those competitions occurred 159 times dating back to as early as 2000.

We selected these leagues and competitions based on an array of criteria. The most important criteria was that penalty kick and substitution data needed to be widely available through internet scraping or an Application Programming Interface (API).

For example, we emphasized diversity in location, confederation, and level of competition. UEFA — the famous European confederation — is the most frequently examined

¹International tournaments span countries and therefore are not applicable

²For the purposes of this study, We only reviewed penalty shootouts from the Third Round Proper and beyond. This decision was made due to the accessibility of data.

³Third-place matches in this competition were not reviewed in this study, as those matches were directly sent to a penalty shootout if teams were in a tie at the end of regulation.

Competition	Year Start	Type	Confederation	Country ¹	Occurrences
Champions League	2000	Club	UEFA	N/A	22
Europa League	2009	Club	UEFA	N/A	13
Community Shield	2000	Club	UEFA	United Kingdom	23
FA Cup ²	2009	Club	UEFA	United Kingdom	14
MLS Cup	2000	Club	CONCACAF	United States/Canada	23
World Cup	2002	International	FIFA	N/A	6
European Championships	2000	International	UEFA	N/A	6
UEFA Super Cup	2000	Club	UEFA	N/A	23
Gold Cup	2000	International	CONCACAF	N/A	12
African Cup of Nations ³	2000	International	CAF	N/A	12
AFC Asian Cup	2004	International	AFC	N/A	5

Table 1: List of Competitions Sampled

confederation, in line with the number of competitions under its brand. All but one of the confederations is examined (CONMEBOL).

1.4 Only men’s association football matches were reviewed

In the course of this study, we only reviewed data from men’s association football matches. This decision was for practical reasons.

A few of those practical reasons include but are not limited to:

- Too little data is available from women’s football matches in an accessible format. Our aspiration is that this will change in the future.
- The available data for women’s association football matches is predominately from North American and European football leagues and confederations, restricting the ability to value geographic diversity in our analysis.
- It is unclear whether or not having two groups separated on gender would compromise the data (for example, if the penalty kick success rate were higher among women than men, this could cause consistency issues for our sample).
- FIFA and other confederations have, at points in the past, used women’s football tournaments as opportunities to test other methods of determining the winner of a match (rather than using a penalty kick shootout). This could skew our data.⁴

That said, as the data for women’s football matches becomes more available across the world, it would be fascinating to see if the results of this study are replicated across the gender divide.

1.5 We reviewed matches starting in the year 2000, but that changed slightly based on the competition to avoid oversampling bias

We wanted to review a large window of data. We identified the year 2000 as a good starting point for most competitions. That window was moved around slightly based on the availability of data.

For tournaments with a large number of matches, we started reviewing matches at a later year and later “stage”. For example, the Football Association Cup (often referred to as the FA Cup) occurs each year in the United Kingdom, and it features a large number of teams. Nearly 700 teams enter the event each year, resulting in a larger match pool.

To correct for this, we only reviewed matches from the Third Round Proper and beyond, avoiding oversampling of the FA Cup. Additionally, we start reviewing the FA Cup in 2009, not in 2000.

As a general rule, we wanted to avoid a situation in which more than 20% of our reviewed matches came from one competition. Table 1 displays the starting year for each competition.

1.6 Some matches are excluded in the name of fairness

We opted not to examine any tournament that uses a golden-goal method of determining a winner in extra time. This is because we assume managers might consider employing different strategies in that system than in the traditional system of determining a winner.

Additionally, we did not evaluate any match that went straight to penalty kicks after 90 minutes of regular time. This occurred, for example, in a few instances of third-place matches in the African Cup of Nations.

1.7 Only penalty shootout kicks were evaluated

Penalty kicks occur all of the time. Whenever a foul takes place inside of the 18-yard box, an opposing player takes a penalty kick toward the offending team’s goal.

But for the purposes of this study, we find little interest in any penalty kick that occurs during the match. Rather, we only evaluate penalty kicks that occur in the penalty shootouts.

All of that said, a future study may find it prudent to see if penalty kicks taken earlier in the game find more success than ones taken later, for example.

2 Process of Data Collection

2.1 We collected data from 162 matches. In total, there were 1,671 penalty kicks evaluated.

We developed a program that looked through every match from the 159 competition occurrences we wanted to evaluate. By the end of the evaluation, the program looked at 1,671

penalty kicks across 162 different penalty shootouts.

2.2 We collected six data points for each kick.

Our program scraped the data into a CSV spreadsheet that can be made available upon request. Each row contained six columns worth of information, which is displayed in Table 2.

Variable Name	Description
Player	The name of the player who took the penalty kick.
Date	The date of the penalty kick.
Team	The team for which the player shot the penalty kick.
Competition	The name of the competition in which the penalty kick was shot (according to the competition names outlined in Table 1).
Minutes Played	The number of minutes the shooting player played in the game, excluding stoppage time. The maximum value for this variable is 120 (90 minutes of regular time and 30 minutes of extra time). The minimum value is 1.
Make/Miss	A successful penalty kick is assigned a value of 1. An unsuccessful penalty kick is assigned a value of 0.

Table 2: Penalty kick dataset variables

3 Early Evaluations of Penalty Kick Success

3.1 Generally, kicking players perform well on penalty kicks

Of the 1,671 penalty kicks examined in this study, 1,209 of the kicks were converted. That’s a success rate of 72.35%. This study only tracked the binary state of goal: converted or unconverted. Our analysis does not track whether a goalkeeper stopped the ball or if a kick missed the goal entirely.

Along the same lines, we do not distinguish between a “good kick” and a “bad kick.” For the purposes of this study, a converted penalty is a good penalty kick, and an unconverted penalty is a bad penalty kick.

Of course, this is incorrect from a practical perspective. There are some players who take wonderful shots that are saved by a good goalkeeper. There are also instances in which players convert a bad shot.

⁴For example, Germany’s dominant women’s team won the 2006 FIFA Women’s World Cup 2-1 by winning in extra time. That year, knockout stage games ended in sudden death (also referred to as “golden goal”).

But it is too subjective to examine a penalty shot on any kind of scale. Therefore, in this study, we examine penalty kicks in a binary manner: a good kick is a made penalty, and a bad kick is a missed one.

In baseball, .300 is considered a “good” batting average. An average goalkeeper in this study saves less than 30% of penalty kicks, and that number includes shots that missed the goal entirely.

In other words, an above-average goalkeeper is preventing a penalty shot from being scored fewer than three times in every ten shots. Think about that the next time you say that hitting a major-league pitch is “the hardest thing to do in sports.”

3.2 Penalty kick conversion rates didn’t vary too much from competition to competition.

Players performed the best in penalty kicks in the Europa League, with a success rate of 80.17%, while they performed the worst in the AFC Asian Cup at 66.99%. The range was 13.18.

There were no outliers in the dataset in terms of success rate.

Table 3 shows this data. Perhaps unsurprisingly, the conversion rates between international matches and inter-club matches is not too different. In international matches, players in our sample converted kicks 71.2% of the time; in club matches, that number was only slightly higher at 73.4%.

Competition	Occurrences ⁵	Penalty Kicks	Success Rate
Champions League	22	149	71.14%
Europa League	13	121	80.17%
Community Shield	23	57	73.68%
FA Cup	14	222	75.23%
MLS Cup	23	266	69.17%
World Cup	6	186	68.28%
European Championships	6	138	69.57%
UEFA Super Cup	23	39	79.49%
Gold Cup	12	121	70.25%
African Cup of Nations	12	269	76.21%
AFC Asian Cup	5	103	66.99%

Table 3: Penalty Kick Conversion Rates by Competition

⁵This number represents how many times that competition occurred in the timeframe we examined. For example, there were six FIFA World Cups that we examined, and 186 penalty kicks were taken in shootouts across those six occurrences.

3.3 UEFA is the crown jewel of association football, but they aren't special on penalty kicks

UEFA is Europe's confederation. It is the most powerful confederation within FIFA. All but two FIFA presidents have been European.⁶ There is a reason why: UEFA is generally seen as the stepping stone to FIFA power.

Gianni Infantino — the Swiss-born FIFA president — rose to power after Sepp Blatter resigned in disgrace. What was his previous post? He stood as UEFA's Secretary General, a high-ranking position within the powerful European confederation.

Stanley Rous served as FIFA president for more than a decade. Born in East Suffolk, he literally wrote the book on the rules of association football.

UEFA also boasts the best leagues: winning England's Premier League, Spain's La Liga, or France's Serie A is seen as the crown jewel in club competition.⁷ Few other countries boast such strong leagues.

It doesn't stay within a country's borders, however. UEFA hosts the Champions League and the Europa League — club competitions that span all across Europe. Teams qualify from their domestic leagues and compete in a tournament held each year (we review penalty kicks from both of those tournaments).

Additionally, UEFA hosts the second-biggest international tournament: the European Championships. It's held every four years, halfway between the quadrennial World Cup. Aside from it being a major honor in the sport, the Euros are also a momentum builder for the World Cup.

All of that is to say that the best players in the world play in Europe. Only two confederations have ever won World Cups: UEFA and CONMEBOL. Teams from UEFA have won 12 out of the 22 finals. UEFA teams have appeared more times than any other confederation: 29 times.

Since the World Cup first appeared in 1930, only two finals two non-European teams matching up against one another. Conversely, nine World Cup finals featured two European teams fighting for football's highest honor.

Why is this important? One would think that better players would shoot better. It turns out that they don't shoot all that better. In the sample of our UEFA-organized club competitions and UEFA-organized international competitions, players shot with a 74.24% conversion rate, not all that much better than the 71.54% seen at non-UEFA-organized events.⁸

4 Hypotheses

Based on anecdotal and observed evidence, our hypothesis is that players substituted into the match for the sole purpose of taking a penalty kick will perform worse than players who play for more of the match.

$$H_0 : \hat{p}(\textit{LastMinuteSubstitutes}) = \hat{p}(\textit{OtherSubstitutes} + \textit{FullTimePlayers})$$

$$H_1 : \hat{p}(\text{LastMinuteSubstitutes}) < \hat{p}(\text{OtherSubstitutes} + \text{FullTimePlayers})$$

5 Last-Minute Substitutions

5.1 In this study, any substitution made in the last five minutes of a game is considered to be a “last-minute substitution.”

We can’t live in the minds of the managers who make substitutions in a game. Some managers might sub a player in because they think they’ll put a shot on goal in the last five minutes, for example; others might be making the substitution specifically for the purpose of taking a penalty shot in the shootout.

Therefore, we are forced to assign an arbitrary number to what should be considered a “last-minute substitution.” The extra-time period is 30 minutes long. Substitutions can only be made when the ball leaves play, so we give the managers five minutes to make these decisions.

Because we track how many minutes a player plays in a game in our analysis, those with 5 or less minutes played will be counted as a “last-minute substitution.”

5.2 Of the 1,671 penalty kicks examined, only 38 were taken by last-minute substitutions. The conversion rate on those 38 kicks was 57.89%.

In our study, we refer to this group as the “substitution group.” Those who played at least six minutes of the game were in a group we named the “standard group.” For further clarity, this includes any player who played the entire game, players who were sent into the game after 90 minutes, and even players who were subbed in after the first period of extra time.

This data is displayed in Table 4.

	Standard Group (StdG)	Substitution Group (SubG)
Success Rate	72.69%	57.89%
Kicks Taken	1633	38

Table 4: Success Rate by Group

The Standard Group’s success rate was 14.80% higher than the Substitution Group.

⁶One of the two non-European FIFA presidents was Issa Hayatou, a Cameroonian who served on an acting basis for 141 days.

⁷Only a few players have been crowned champions in all three of those leagues. The most notable example is Portugal’s Cristiano Ronaldo. He did it with Manchester United three times (2007, 2008, 2009), Real Madrid twice (2012, 2017), and Juventus twice (2019, 2020).

⁸For the purposes of this part of the analysis, we excluded World Cup matches.

5.3 The question we must answer: what are the chances that happened randomly? Is this statistically significant? Because we only selected a sample, we have to create a hypothesis test to determine the answer.

In our next step, we will evaluate our hypotheses by using a two-group, one-tail proportion test (also known as a Z-Test).

The formulas for that calculation are as follows:

$$Z = \frac{\hat{p}_{SubG} - \hat{p}_{StdG}}{StdError}$$

$$StdError = \sqrt{\frac{\hat{p}_{SubG}(1 - \hat{p}_{SubG})}{n_{SubG}} + \frac{\hat{p}_{StdG}(1 - \hat{p}_{StdG})}{n_{StdG}}}$$

Table 5 shows the value for each of these variables.

	Control Group (StdG)	Treatment Group (SubG)
\hat{p}	0.7268830374	0.5789473684
n	1633	38

Table 5: Proportions and Number of Each Group

After we substitute the values into our Standard Error equation, it looks like this:

$$StdError = \sqrt{\frac{.72688(1 - .72688)}{1633} + \frac{.57895(1 - .57895)}{38}}$$

$$StdError = .08084$$

After calculating our Standard Error, we can find our two-group Z Score using the aforementioned formula. That looks like this:

$$Z = \frac{.57895 - .72688}{.08084}$$

$$Z = -1.82991$$

6 Conclusions

In order to test if our data is significant, we test at the $p = .05$ significance level. This is sometimes referred to as the “95% Confidence Interval.”

We use a one-tailed Z test because we are testing if the mean of the Treatment Group (SubG) is significantly smaller than the Control Group (StdG).

Using the one-tail Z test, we first find a critical value. If our Z score is smaller than our critical value (represented by Z^*), then we reject the null hypothesis.

At the $p = .05$ significance level, this is our critical value:

$$Z^* = -1.6449$$

On a graph, the Z Value is represented by the dotted black line. The red area is the area in which the null hypothesis would be rejected (values less than the critical Z value). This graph is shown in Figure 1.

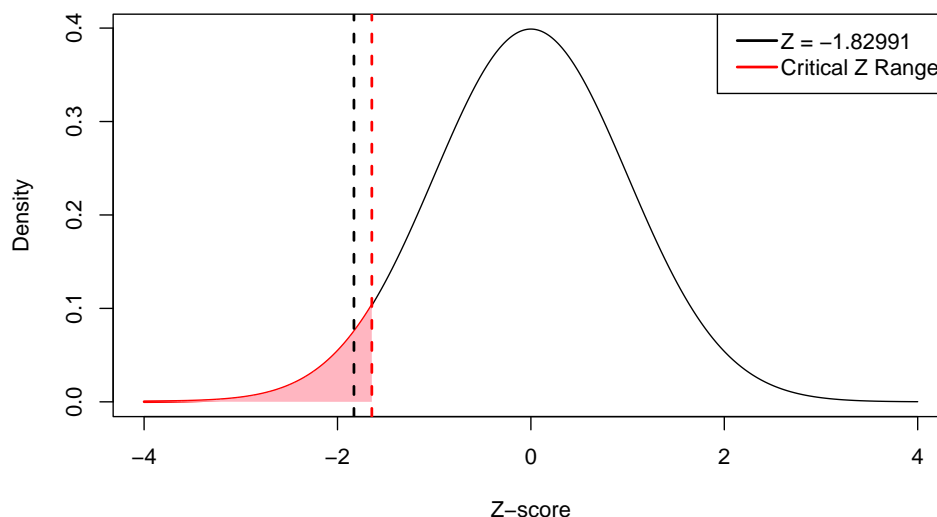


Figure 1: Z Value of Data and Critical Z Value of the Dataset

6.1 Last-minute substitutes perform worse in penalty kicks to a statistically significant degree

He's probably used to this by now, but it's bad news for England Manager Gareth Southgate. Our Z Score of -1.82991 is lower than our critical Z value of -1.6449. This means that **we reject the null hypothesis**. Our data is significant.

Southgate is not the only manager to fall victim to this, and the 2020 European Championships weren't the first time it's affected England's national team.

In the 2006 World Cup, Portugal knocked England out in the quarterfinals. By the end of extra time, neither team scored a goal. Portugal kicked first, scoring on England goalkeeper Paul Robinson. Almost immediately, England couldn't grasp onto good news. Frank Lampard — the longtime English midfielder — couldn't convert the shot.

After three shots for each team, they were tied 1-1 on penalties. Portugal's Helder Postiga stepped up and placed the ball into the back of net. England's next kicker was Jamie Carragher, a well-capped player. He had played more than two dozen or so games for the Lions by the time he stepped up as the team's fourth kicker.

Carragher was a last-minute substitute in the match. He only played one minute of the game.

He lines up and places a brilliant shot toward the net. Right after it hits the net, the whistle of referee Horacio Elizondo rang loudly. Because Elizondo didn't blow his whistle before the kick, it needed to be retaken. On the second shot, Carragher missed.

Portugal's shot to win the game came at the foot of an upstart Portuguese player. The player with the number 17 on the back of his shirt loudly struck the ball to the right of Robinson, securing a win. That player would later go on to drop the first digit of his kit number and become the best goal scorer in Portugal's storied football history: Cristiano Ronaldo.

Carragher later retired from football before coming out of retirement to play for England's 2010 World Cup squad. That team would lose in the Round of 16 to Germany. It was a year of losing for Carragher; even Andy Burnham, Carragher's preferred pick for Labor Party leader, would lose in the party's 2010 leadership election.

6.2 Should this change a manager's strategy?

It's unclear exactly what causes this phenomenon. Is it nerves of having just stepped onto the pitch? Do they perform worse because their legs really are so fresh to the field? We don't really know. We also don't explicitly know that this is the cause, though the data does not lie.

Or, is there another explanation? Maybe, for example, your starting lineup is always your best lineup. So players being subbed into the match are less strong than your Starting 11. Though, that suggestion is best eliminated by the statistics for a player who played anything less than 120 minutes. The conversion rates of substitutes and starters are nearly identical. The difference is less than one percent, and it favors the substitutes, for what it's worth.

As FIFA moves the rules to allow five substitutions instead of three, managers in knockout round games may find it prudent to remember this analysis. Is it worth it to sub in a few players at the last minute? The statistics say no. Why? **Because Fresh legs clearly can't kick.**

6.3 Future studies

There are a myriad of fascinating questions that appeared during this analysis. For example, it would be interesting to look into whether or not penalty kicks are more likely to be converted earlier in the match than later.

Additional questions form around how players perform in penalty kicks in different rounds of a tournament. For example, are conversion rates lower in final matches than they are in

quarterfinals? One would think the nerves of a final match would be higher than that of a quarterfinal.

Finally, our data also tracked the date of various penalty kicks. As more and more teams use data to help their goalkeepers save penalty kicks, are conversion rates decreasing over time? There are a number of examples of coaches handing "cheat sheets" to goalkeepers ahead of a penalty shootout to help provide context on potential shooters ahead of a shot. It's perfectly legal, though it's controversial among some fans. Regardless of your opinions, it's an interesting question.