Evaluating the practice of “icing” kickers in the NFL using play-by-play data from 2001 to 2022

Datasci 203: Lab 2

Alberto Lopez Rueda, Arjun Kandaswamy, Jared Feldman

Contents

1 Introduction

Icing the kicker is a football strategy whereby the defensive team calls a timeout prior to a field goal kick with the aim of forcing a mistake. Icing the kicker is an extended practice and is also a highly-debated topic among coaches, fans, and commentators. The strategy is often deployed in high-pressure situations at the end of games, where a field goal can lead to victory or defeat.

Given the increasingly popular usage of analytics in sports, the goal of the study is to use a data-driven approach to understand if icing the kicker has an impact on the make rate of field goals. To accomplish this, we leverage NFL data and apply a set of regression models to estimate the potential impact when controlling for confounders. Hence, this analysis aims to address the following research question:

Does icing affect the probability of a kicker making or missing a field goal?

2 Data and Operationalization

The data for this study is sourced from the nflreadr R package, which contains NFL play-by-play data from 1999 until 2022. However, we evaluate field goals from 2001 onwards due to some data quality issues in 1999 and 2000 (e.g., consecutive rows with same or similar data). The full dataset contains 1.1m rows, with each row representing one play, and 372 features. We perform an initial filter to select all field goals, yielding a total of 22,415 plays.

Classifying which field goals are iced is a subjective matter. We operationalize iced kicks as those field goals where a timeout was intentionally called in the previous play by the defensive team. This means we include iced kicks in non-clutch situations and exclude timeouts related to injuries, as well as timeouts called by the offensive team. This methodology returns a total of 1,358 instances (6% of total field goals) with the number of iced kicks, as well as field goals, evenly spread across years (~60 and ~1,000 events per year, respectively). Initially, we considered adding more logic to our iced kick operationalization, such as time remaining or score differential. However, this would have subjectively filtered out ~300-400 field goals that could have been perfectly impacted by icing.

To operationalize field goal rate, we leverage the variable named `field_goal_result` to create a new binary feature `fg_outcome` which takes the value of 1 if the field goal was made and 0 otherwise. Overall, it is important to highlight that all the six columns needed to operationalize both field goal outcome and iced kick were completely populated with reasonable values.

In our research we use ten additional dependent variables, mainly grouped into three broad categories as described in the table below. Half of these variables were taken directly from the dataset (e.g., kick distance, wind, and temperature) while the rest were feature-engineered.

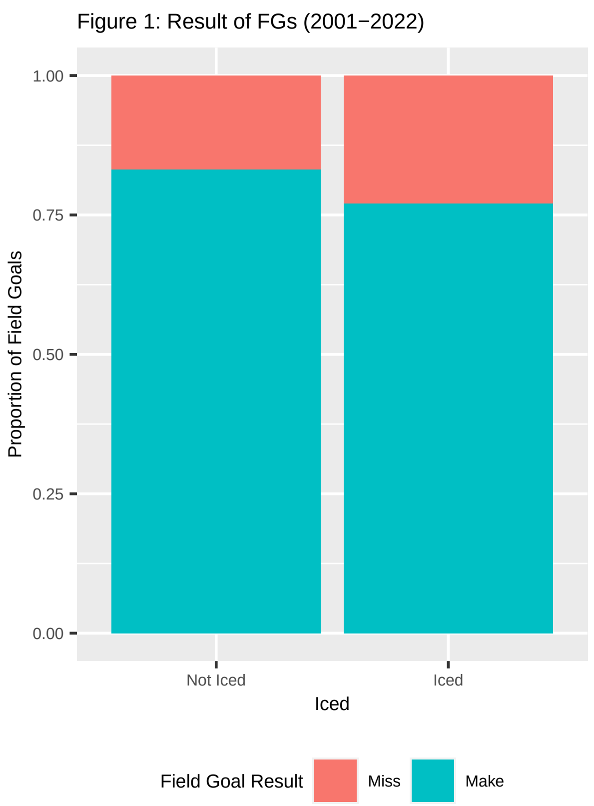
Variable group	Description of variables
Situational	`kick_distance`, `high_pressure` (if the field goal happens in the last two minutes of each half or in OT), `score_differential` and `defteam_timeouts_remaining` (number of timeouts available for defensive team)
Environmental	`wind`, `temperature`, `is_covered` for type of stadium (i.e., indoor vs. outdoor) and `surface_2` (i.e., synthetic vs. natural grass)
Performance	`accuracy_5` (% kicks made from the last five field goals) and `pro_kicker` if the kicker was ever a Pro Bowler or All Pro prior to the kick

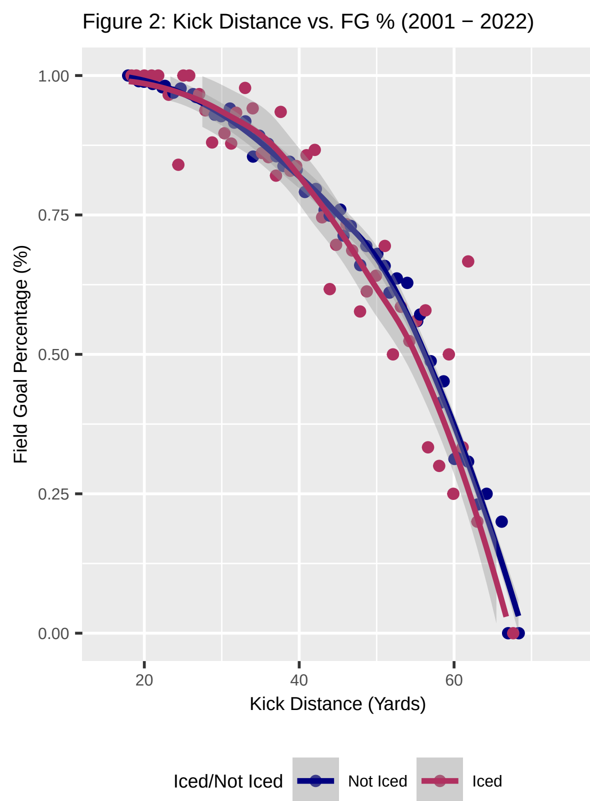
One of the new variables is `high_pressure`, denoting kicks happening after the two-minute warning in each half (i.e., end of Q2 and Q4) or at any time during Overtime. During these periods, we found that games tend to witness more iced kicks. Additionally, field goal accuracy is roughly 6 percentage points lower on average compared to other periods (e.g., Q1/Q3 and early Q2/Q4). We felt this impact was not properly captured in the readily-available time-related variables, such as time remaining per quarter or quarter. Likewise, we believe that kickers' skill can be a confounder in our model, so we created two new performance variables (`accuracy_5` and `pro_kicker`) as defined in the table above.

3 Data Wrangling and Visualization

In Figure 1 we show a naive analysis of iced kick success, plotting the proportion of kicks made that were iced or not iced in our dataset. This shows that iced kicks have lower accuracy, although this approach does not account for confounders such as kick distance, for example.

According to existing research, kick distance is the variable that most impacts field goal accuracy [1]. In Figure 2 we can observe that there is a negative relationship between distance and field goal accuracy that seems to be polynomial in nature. As a result, we decide to include a kick-distance-squared variable in our regression models. We can also observe that the percentage of field goals made seems to be lower for iced-kicks, although this could potentially be the result of confounders.





Existing research also suggests that wind and temperature can meaningfully impact field goal accuracy [2]. Contrary to most other variables in our dataset, roughly 30% of `wind` and `temp` values were not populated. To mitigate this issue, we fill empty rows by extracting data from a separate column named `weather` that contains a long description (e.g., "Fair Temp: 44° F, Humidity: 40%, Wind: NW 7 mph"). In addition, we input wind and temperature data for indoor stadiums using the average for each type of stadium (e.g., closed and domes). Lastly, we clean the data by handling wind outliers above 35 mph and fixing some data-quality issues (e.g., zero recorded wind speed when the stadium is outdoors and the weather column is empty). After these steps, the percentage of rows with missing `wind` or `temp` values was below 4%.

We condense the four existing categories of stadiums contained in the `roof` variable into a new variable named `is_covered` which takes the "Yes" value if the stadium is a dome or closed (e.g., indoor). Likewise, we map all the existing types of grass into two categories ("natural" and "synthetic") in our `surface_2` variable.

4 Modeling and Results

We divide our field goal dataset into an exploration set by randomly selecting 30% of all iced kicks and 30% of non-iced kicks, where we perform model building. The remaining rows are assigned to a confirmation set to generate the statistics for the report.

Our first regression is a baseline model that takes into account the three most important variables according to existing literature: kick distance and its square, wind, and temperature. Since weather conditions in indoor vs. outdoor stadiums are very different, and we input a lot of missing data for wind and temperature for indoor stadiums, we add the variable `is_covered` as well as two interaction terms to the regression. In our stargazer table, β_1 represents the change in `fg_outcome` if a kick is iced in outdoor stadiums, leaving all other variables constant.

In our second model, we incorporate covariates in order to reduce omitted variable bias, namely: `high_pressure`, `score_differential`, `surface_2`, `accuracy_5` and `pro_kicker`. Based on our causal graph, we believe all these variables can be potential confounders since they can affect both field goal outcome and the decision to ice.

Lastly, given our dataset is imbalanced, we use propensity score matching in our third model using all the variables described earlier plus the number of timeouts remaining available by the defensive team. Propensity score matching effectively subsets our dataset to include non-iced kicks with comparable propensity scores to iced kicks, allowing us to better isolate the potential causal effect of icing the kicker. We set hyperparameters to "nearest neighbor" with a 2 to 1 proportion and no replacement, since these settings minimize standardized mean differences (i.e., most balanced dataset). After conducting this process, our matched confirmatory dataset contains 2,628 field goals, which we use to run the same regression as in our second model.

Table 2: Iced Field Goals - Regression Results

	Dependent variable:		
	Field Goal Outcome		
	Model 1 (Baseline)	Model 2 (Additional Variables)	Model 2 with PSM
	(1)	(2)	(3)
Iced	−0.029**	−0.013	−0.012
	(0.013)	(0.014)	(0.016)
Kick Distance	0.011***	0.012***	0.019***
	(0.002)	(0.002)	(0.004)
Kick Distance Squared	−0.0003***	−0.0003***	−0.0004***
	(0.00003)	(0.00003)	(0.0001)
Wind	−0.003***	−0.002***	−0.001
	(0.001)	(0.001)	(0.002)
Temperature	0.001***	0.001***	0.001***
	(0.0002)	(0.0002)	(0.001)
Covered Stadium	0.054	0.052	−0.073
	(0.045)	(0.045)	(0.121)
Score Differential		−0.0003	0.001
		(0.0003)	(0.001)
High Pressure Situation		−0.020***	−0.010
		(0.008)	(0.017)
Synthetic Surface		0.017**	0.017
		(0.007)	(0.017)
Accuracy of Last 5 FGs		0.045***	0.031
		(0.017)	(0.043)
Pro Bowl/All Pro		0.015**	0.022
		(0.006)	(0.016)
Wind * Covered Stadium	0.005**	0.005**	0.016**
	(0.002)	(0.002)	(0.006)
Temperature * Covered Stadium	−0.001*	−0.001*	−0.001
	(0.001)	(0.001)	(0.002)
Constant	0.868***	0.803***	0.656***
	(0.034)	(0.037)	(0.089)
Observations	15,070	14,568	2,614
R^2	0.130	0.133	0.179
Adjusted R^2	0.130	0.132	0.174
Residual Std. Error	0.354 (df = 15061)	0.352 (df = 14554)	0.381 (df = 2600)
F Statistic	281.985*** (df = 8; 15061)	171.911*** (df = 13; 14554)	43.466*** (df = 13; 2600)
Note:			*p<0.1; **p<0.05; ***p<0.01

Our baseline regression shows a significant impact of icing on field goal outcome (0.017 p-value). However, our second and third models, which reduce the potential effect of confounders compared to our baseline model, show no relationship with p-values above 0.30. It is important to highlight that the coefficients of `iced` are meaningfully lower in our second and third models, meaning our first model had higher omitted variable bias. This suggests to us that icing the kicker does not cause a difference in field goal outcome.

In addition, kick distance and temperature are also highly significant, with fairly consistent coefficients across all models. There are a number of variables that are significant in model two but not in propensity score matching, perhaps due to discarding good matches that are "unneeded" although we believe this requires further investigation.

5 Limitations

Consistent regression estimates require an assumption of independent and identically distributed (IID) data. Because field goals take place over time, there is a possibility of autocorrelation due to improvements in kicking strategies and/or better equipment. While indeed the yearly field goal accuracy has been trending up [3], the interaction term `iced*season` is statistically insignificant across all models (p-values > 0.60). In addition, there could be autocorrelation due to players' streaks, although we try to somewhat mitigate this by adding the `accuracy_5` momentum indicator. Lastly, there could be some geographic clustering when it comes to weather conditions. Although all these could be important issues, overall we believe the assumption of IID data is sufficiently met.

Consistent regression estimates also require the existence of a unique best linear predictor. Based on our EDA, the vast majority of covariates do not have a heavy tailed distribution, barring perhaps wind speed to a relatively low degree. Overall, we consider the distribution of data allows us to assume the existence of a best linear predictor.

Regarding structural limitations, several omitted variables may bias our results. Within the group of environmental variables, our models do not account for precipitation/snow and altitude. This group of variables may be positively correlated with being iced and negatively correlated with field goal outcome, resulting in a negative omitted variable bias. The main effect is therefore being driven away from zero, making the tests overconfident. Likewise, our model does not account for wind direction, which can be as important as wind speed. For example, the presence of crosswinds could increase the likelihood of being iced and decrease field goal accuracy as well, resulting in negative bias.

It is also possible to speak about potential pseudo reverse causality. This is because the outcome of prior field goals can have an impact on the likelihood of being iced. This potential relationship is depicted by our variable `accuracy_5` (i.e., % kicks made from the last five field goals). We add the term "pseudo" because although the coefficients associated with icing are negative across all models, the p-values are not statistically significant.

We do not think `iced` is an outcome variable located in the RHS because it does not cause a change in the rest of the variables. However, this could be an issue interpreting the coefficients of `wind` and `temp` given the temporal nature of the data. In other words, both of these variables affect `fg_outcome` and `accuracy_5` (particularly if field goals happen in the same game) and this last variable can in turn also affect `fg_outcome` on the play through player's psychology (e.g., hot or cold streak).

6 Conclusion

This research aims to analyze the efficacy of the strategy of icing a kicker prior to a field goal in the NFL. Our analysis shows that when accounting for potential confounders, there is no causal link between icing and field goal outcome. This conclusion is consistent with existing literature, although the team believes NFL coaches will continue trying to ice the kicker "just in case". Future research on the topic should be focused on inputting wind and temperature data for indoor stadiums based on the stadium and date of game, as well as adding more potential confounders, particularly direction of wind and precipitation.

7 References

- [1] https://www.kaggle.com/code/pvabish/kick-accuracy-beyond-field-goal-percentage/
- [2] http://www.advancedfootballanalytics.com/2012/01/temperature-and-field-goals.html
- [3] https://www.pro-football-reference.com/years/NFL/kicking.htm