Is missing a field goal always the fault of the placekicker?

"It's Good!" An Analysis of the Probability of Success for Placekicks

Christopher R. Bilder and Thomas M. Loughin

The outlook wasn't brilliant
for the KC eleven that day;
The score stood ten to seven
with but 4:12 to play.
And so when Gannon replaced Bono,
and Allen came up lame,
A sickly silence fell upon
the patrons of the game.

A struggling few got up to go
in deep despair. The rest...
clung to that hope which springs eternal
in the human breast.
They thought, if only Elliott
could get but one more aim.
We'd put up even money, now,
with Elliott in the game.

But Slaughter caught a pass,
to the wonderment of all,
And Gannon, the back-up,
ran for 14 yards with the ball;
And when the dust had risen and
the crowd could see for sure,
KC had the ball at the Indy 25
with but 56 seconds left to endure.

Then from eighty thousand throats
went up a lusty yell;
It rumbled in the Missouri valley,
it rattled through the dell;
It knocked upon the hilltops
and recoiled upon the parking lot,
For Elliott, placekicker Elliott, was
advancing to take his spot.

One hundred sixty thousand eyes were on him
as his hand marked the spot on the dirt;
One hundred sixty thousand hands applauded
as he wiped it on his shirt.
And when the offensive line
got into their stance,
Concentration gleamed from Elliott's eyes
as if he were in a trance.

Earlier in the day
Elliott wanted to hide,
Both from thirty-five and thirty-nine
were just wide.
Now there was an opportunity to erase
the memories of the past.
If Elliott makes the kick,
hero will be the memory that will last.

The center grips the ball
and sends it back;
The offensive line blocks
as the defense tries to sack.
And now, the holder places the ball down
and makes the ball ready to go,
And now the air is shattered by
the force of Elliott's blow.

Oh! Somewhere in the favored land
the sun is shinning bright;
Somewhere bands are playing,
and somewhere hearts are light;
Somewhere men are laughing,
and somewhere children play with deft;
But there is no joy in KC,
jobless Elliott misses wide left.

The preceding poem is an adaptation of Ernest L. Thayer's *Casey at the Bat*. The poem describes the January 7, 1996, playoff game between the Kansas City Chiefs and the Indianapolis Colts. Lin Elliott, Kansas City's placekicker, missed a field goal at the end of the game, and Kansas City's season ended with a three-point loss. Shortly after the game, in which Elliott missed two other field goals, the Chiefs released Elliott from the team.

In addition to the Kansas City-Indianapolis game, many other National Football League (NFL) games are often decided by the foot of a placekicker. With either a point after touchdown (PAT) or a field goal (FG), placekickers score points in nearly every game. Estimated probabilities associated with the success of a placekick are often thought of as proportions of the time a kicker has made a kick from a certain distance interval. The goal of this study is to examine the effects of many factors—for example, distance, game pressure, and weather—on the success probability of a placekick using logistic regression (see Sidebar 1 for more on logistic regression).

Background

The objective of a placekick is to kick the football over the crossbar and between the two uprights of the goalposts to score points. Two referees below the uprights determine if the placekick is a success or a failure.

PAT's and FG's are the two types of placekicks. PAT's are attempted after a touchdown has been scored. These placekicks are usually 20 yards in length and worth one point (if successful) to the kicking team. All other placekicks are field goals. These attempted field goals can vary greatly in length (18 to 66 yards in the 1995–96 NFL regular season) and are worth three points (if successful) to the kicking team.

Data

The study is based on data from the 1995–96 NFL regular season. The data for individual placekick results are from the NFL Web site (http://www.nfl.com), which provides play-by-play game summaries. Not all game summaries have complete records for all variables of interest, so some placekicks are excluded from the analysis. Despite this, over 1,700 placekick attempts are used for the analysis that follows. Data involving city altitude are gathered from the *Rand McNally Road Atlas* (1992) and the *World Almanac and Book of Facts* (1992).

For a variety of reasons, some explanatory variables are transformed prior to the analysis. For example, the wind speed inside dome stadiums is often not measured, although anyone who has been inside a dome stadium can attest that it is not 0. As a result, the wind variable is instead considered as a binary variable describing windy versus nonwindy conditions at game time. The 10th percentile of outdoor placekicks, 15 mph, is used as a cutoff point for windy versus nonwindy conditions. Because wind conditions inside dome stadiums are not likely to be too extreme, all placekicks in these stadiums are assumed to be attempted

Table 1—Explanatory Variables and Their Definitions

Variable	Definition	*p*-value*
Altitude	Continuous variable measuring the city elevation where the placekick took place	.36
Change	Binary variable for lead-change placekicks versus nonlead-change placekicks	<.01
Distance	Continuous variable for the distance in yards of the placekick	<.01
Dome	Binary variable for outside placekicks versus those inside a dome	.23
Home	Binary variable for whether the placekick is on the placekicker's home field or opponent's field	.49
PAT	Binary variable for whether the placekick is a PAT versus a FG	<.01
Precipitation	Binary variable for whether precipitation is falling (i.e., snow, rain, etc...) at game time versus no precipitation	.87
Surface	Binary variable for placekicks on grass versus artificial turf	.69
Temperature	Continuous variable for temperature at game time with dome placekicks assigned a value of 72 degrees	.77
Time	Continuous variable for the time remaining in the half with overtime placekicks assigned a value of 0	.01
Week	Assumed to be a continuous variable that ranges from the integer values of 1 to 17 denoting the week of the placekick attempt	.13
Wind	Binary variable for wind speed > 15 mph at game time versus ≤ 15 mph at game time or the placekick is attempted inside a dome	.10

* From the likelihood ratio test.

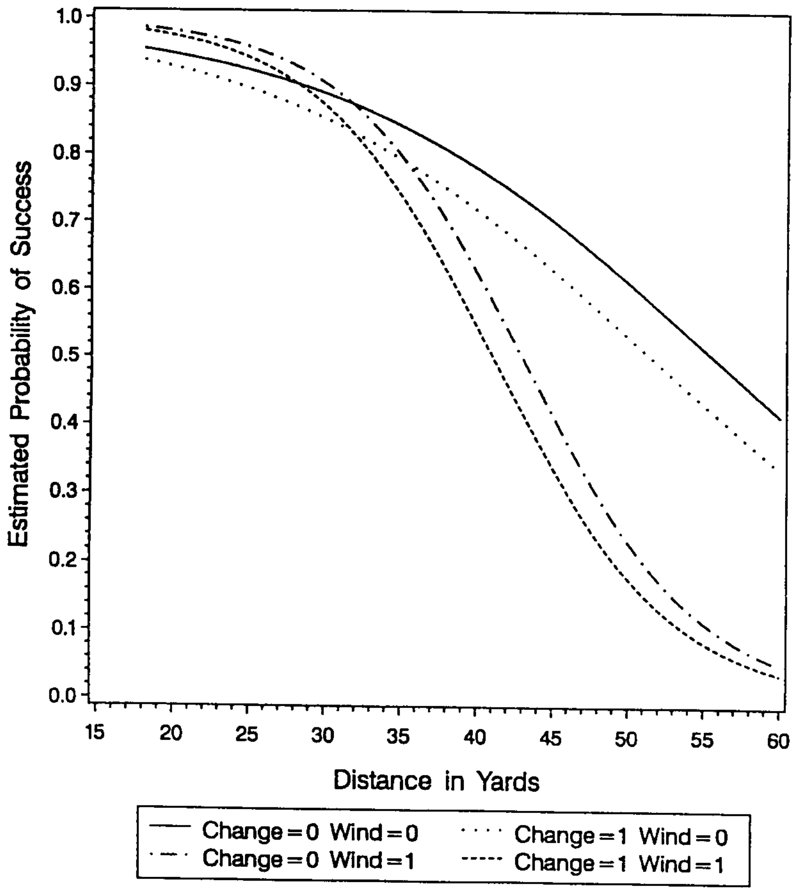

Figure 1. Estimated probability of success for different risk factors.

under nonwindy conditions. Wind direction would also seem to play an important role; however, this level of detail is not consistently available and thus not used in the analysis.

No attempts are made to distinguish between any individual placekicker's effect on the probability of success. One reason there is no differentiation between placekickers is that NFL-caliber placekickers are often thought of as "interchangeable parts" by teams. NFL teams regularly allow their free-agent placekickers, who are demanding more money, to leave for other teams because other placekickers are available. In the past few years this has happened to Chris Jacke, Chris Boniol, and Nick Lowery, for example.

Another reason there is no differentiation between placekickers is that, as Morrison and Kalwani (1993) conclude, there does not seem to be significant skill differences among placekickers. They do believe that some placekickers are better than others, "but mostly because of how far they kick compared to the typical NFL kicker." NFL-caliber placekickers are given the same opportunities to attempt placekicks from distances of 55 yards or less. Not all NFL-caliber placekickers, however, are given the same opportunities to attempt placekicks greater than 55 yards. In fact, only eight placekickers were given the opportunity to attempt placekicks greater than 55 yards. With so little data available, any model is unreliable for attempts at these extremely long distances.

Modeling

Following Hosmer and Lemeshow (1989), separate logistic regression models are fit for each variable as an initial screening process to reduce the variable pool. Table 1 contains a description of variables considered along with the results.

Among the variables that appear to be important are distance and PAT. This is, of course, not very surprising. It is well known that the probability of success decreases as placekick lengths increase, and most PAT's are attempted from the relatively short distance of 20 yards.

Other variables that appear to be important are time and change. Both of these variables may be thought of as a measurement of the "pressure" of the placekick attempt. End-of-the-game or lead-changing (meaning that the game

Sidebar 1

Let $X_i = [1, X_{i1}, X_{i2}, \ldots, X_{ik}]$, for $i = 1, \ldots, N$, be a row vector of explanatory variables. Let Y_i be a binary response coded as a 1 to denote a success and a 0 to denote a failure. The Y_i are independent Bernoulli random variables with success probabilities $P(Y_i = 1 \mid X_i) = p_i$ for $0 \leq p_i \leq 1$. The logistic regression response function is

$$p_i = E(Y_i \mid X_i) = \frac{\exp[\beta' X_i]}{1 + \exp[\beta' X_i]} \qquad (1)$$

where β is a vector of unknown parameters. The logit transformation is

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta' X_i \qquad (2)$$

where Equation (1) is solved for $\beta' X_i$. All modeling in this study is performed using PROC LOGISTIC in SAS (SAS Institute 1989).

becomes tied or the kicking team takes the lead if successful) placekicks are typically thought of by most football fans as being more difficult. Buffalo Bills fans surely remember placekicker Scott Norwood's famous "wide right" field goal at the end of Super Bowl XXV (Bills lost 20–19).

The wind and week variables appear to be potentially important with p-values under .15 in their respective models. Under windy conditions, placekickers need to compensate for the wind speed and direction. Stadiums often cause swirling winds that could further affect success probabilities. The potential importance of week may suggest a fatigue factor or pressure factor as the season progresses.

Further interpretation of "how much" these variables affect the probability of success is deferred until a final model, including all important variables, is developed.

Some of the most interesting findings here are the variables that do not have a significant effect. For example, surface and dome are two of the variables that most football fans often consider important. Placekickers typically prefer to kick off of artificial turf and inside a dome stadium. The screening analysis suggests, however, that these variables do not significantly affect the probability of success.

A few of the variables that the analysis did not find important may be because of lack of information or data. The Denver Broncos' Mile High Stadium is most likely the only place where altitude may help success probabilities of long placekicks. The altitude variable may be nonsignificant because not many long placekicks are attempted there in comparison to all the other stadiums combined. The precipitation variable measures the playing conditions of the game, but the variable does not make a distinction between a blizzard or a light drizzle. In both of those cases, the precipitation variable is equal to 1. Moreover, the precipitation variable only measures conditions at kickoff. Because football games last about three hours, these conditions may change.

After the initial variable screening process, the most parsimonious model that best estimates the probability of success is developed using modeling techniques described by Hosmer and Lemeshow. Table 2 shows this final model. Note that during the model-building process, all nontwenty-yard PAT's are removed from the dataset due to a couple of large residuals for these types of placekicks. Because very few PAT's are not 20 yards, there are not enough data to determine whether the results of these observations are real trends or improbable events. Thus, the removal of nontwenty yard PAT's slightly reduces the population of inference; a more detailed study would be needed to describe these anomalies.

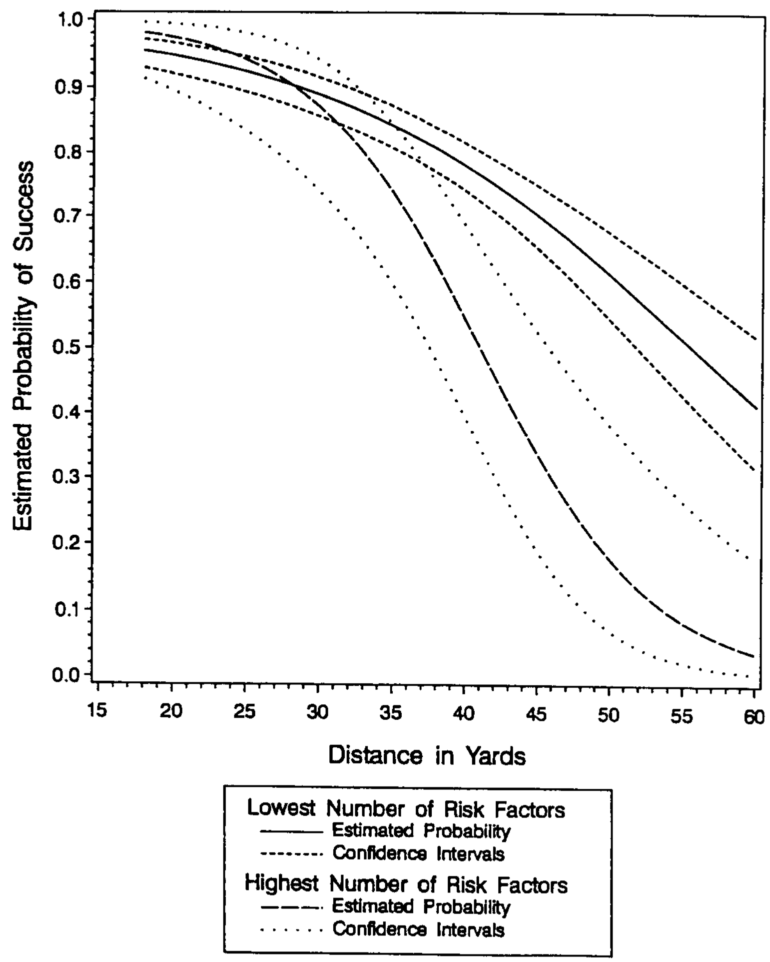

Figure 2. Estimated probability of success with 90% confidence intervals for lowest and highest number of risk factors.

Table 2—Final Model

$$\text{logit}(\hat{p}_i) = 4.4984 - .0807\text{distance} + 1.2592\text{PAT} - .3306\text{change} + 2.8778\text{wind} - .0907\text{distance}*\text{wind}$$

where	Distance	= the distance in yards
	PAT	= 1 if the placekick is a PAT, 0 if the placekick is an FG
	Change	= 1 if a successful placekick will cause a lead change, 0 otherwise
	Wind	= 1 if the wind speed is > 15 mph, 0 if the wind speed is ≤ 15 mph or the placekick is attempted inside a dome

Table 3—Estimated Odds Ratios

Variable interval	Odds ratio	90% Wald confidence interval
Change	.72	(.52, .99)
PAT	3.52	(1.86, 6.66)
10-yard decrease in distance with wind = 1	5.55	(2.65, 11.61)
10-yard decrease in distance with wind = 0	2.24	(1.86, 2.70)
wind with distance = 20	2.90	(.65, 12.96)
wind with distance = 30	1.17	(.49, 2.77)
wind with distance = 40	.47	(.26, .87)
wind with distance = 50	.19	(.07, .55)
wind with distance = 60	.08	(.01, .44)

Final Model Interpretation

The odds ratios in Table 3 measure the effect a variable has on the probability of success while holding the other variables constant. [See Sidebar 2 for further details on the construction of the odds ratios.]

From Table 3, the estimated odds of a successful placekick are between 1.01 and 1.92 (1/.52 and 1/.99) times higher for nonlead-change attempts than lead-change attempts. When a successful placekick would put the kicking team ahead or tied, more pressure may be placed on the placekicker to make the placekick. Thus, the probability of success decreases for these lead-change attempts. This result may indicate a decrease in performance for placekickers in "clutch" situations.

An extra first down (10 yards) improves the chances of a successful placekick. When placekicks are attempted under windy conditions (wind = 1), the estimated odds of a successful placekick are between 2.65 and 11.61 times greater for each 10-yard decrease in distance. On the other hand, when placekicks are attempted under nonwindy conditions (wind = 0), the estimated odds of a successful placekick are between 1.86 and 2.70 times greater for each 10-yard decrease in distance. Windy placekick attempts are helped more by the reduction in distance than nonwindy placekick attempts.

The last five odds ratios in Table 3 further illustrate the effect that wind and distance have on the probability of success for specific distances. For shorter placekicks, the wind does not seem to play an important role in the success or failure of the placekick; however, for longer placekicks, the wind does play a role. For instance, the estimated odds of success for a 50-yard placekick attempt under nonwindy conditions are between 1.82 and 14.29 times (1/0.55 and 1/0.07) higher than under windy conditions. This illustrates that wind plays an important factor for longer placekicks.

Figure 1 shows a plot of the estimated FG probability of success versus distance for different combinations of the wind and change variables. The lines illustrate how the estimated probability of success decreases as the distance of the placekick increases. The plot further shows how the probability of success decreases as more "risk factors" are present for the placekick attempt. The "risk factors" associated with a placekick describe the addition of windy and/or lead-change conditions to the attempt.

Figure 2 illustrates the difference in estimated success probabilities

continued on page 30

Sidebar 2

Logistic regression odds ratios measure how much greater the odds of success are for one level of an explanatory variable when compared to another level. For example, suppose a binary variable, X, is coded by 1s and 0s. The table below contains a contingency table that summarizes the corresponding simple logistic regression model probabilities. The table shows that $p(1)$ represents the probability $Y = 1$ if $X = 1$, and $1 - p(1)$ represents the probability that $Y = 0$ if $X = 1$. Similarly, the table shows $p(0)$ and $1 - p(0)$.

	$Y = 1$	$Y = 0$
$X = 1$	$p(1) = \dfrac{\exp[\beta_o + \beta_1 X]}{1 + \exp[\beta_o + \beta_1 X]}$	$1 - p(1) = \dfrac{1}{1 + \exp[\beta_o + \beta_1 X]}$
$X = 0$	$p(0) = \dfrac{\exp[\beta_o]}{1 + \exp[\beta_o]}$	$1 - p(0) = \dfrac{1}{1 + \exp[\beta_o]}$

Using the table, odds are constructed as $O(0) = p(0)/(1 - p(0)) = \exp[\beta_o]$ and $O(1) = p(1)/(1 - p(1)) = \exp[\beta_o + \beta_1]$. Thus, the odds ratio is $O(1)/O(0) = \exp[\beta_1]$. Estimated odds ratios are found by replacing the parameters with their estimates. Similar odds ratios can be constructed when there is more than one variable in the model or when the variable is continuous.

words." Sure smokers die in greater numbers than nonsmokers, but by how many? What does this mean in terms of length of life? What we need are a few indications of this effect at ages at which mortality begins to really matter, say age 60. Figure 1 uses the maxim, "Keep the axes away from anything that counts."

We can improve things by dumping these principles of bad graphics. Return to mortality, defined as a percent of the total sample at each age, plot using logarithmic spacing on the left where we expect to look for effect sizes, add the corresponding log values on the right for more detailed inspection, use base 10 for logs, try evocative symbols, and put in a few light lines to highlight the effect at age 60 in terms of both mortality and age. The result is Fig. 2.

Visual Revelations is about communication in the broad sense. Graphs play a key role, but words are not neglected. If you intend to buy one book to help you get across the results of your statistical investigations to your readers, this should probably be it.

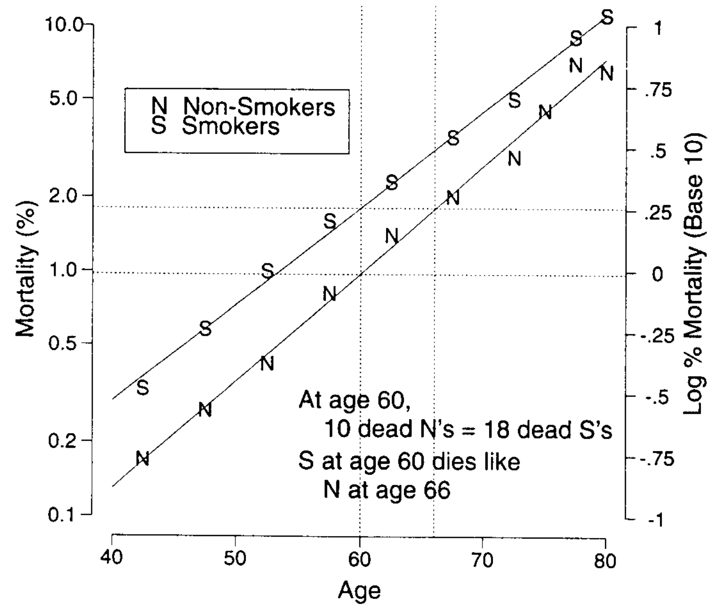

Figure 2. The mortality rates for smokers and nonsmokers at various ages. At around age 60 smoking is equivalent to 6 years of aging, and smokers die nearly twice as often.

References and Further Reading

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, New York: Wiley.

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

——— (1990), *Envisioning Information*, Cheshire, CT: Graphics Press.

——— (1997), *Visual Explanations*, Cheshire, CT: Graphics Press.

"It's Good!"

continued from page 24

between the case with the lowest number of risk factors (change = 0 and wind = 0) and the case with the highest number of risk factors (change = 1 and wind = 1). The 90% Wald confidence intervals for the probability of success are included on the plot. For placekicks greater than 38 yards, the confidence intervals do not overlap.

Conclusion

What is the estimated probability of success for Elliott's end-of-the-game FG discussed in the introduction? Assuming that play off and regular-season placekicks do not have different success probabilities, the model produces what is shown in Table 4. The estimated probability shows that about two out of every three FG's on average are made in this situation.

Table 4—Elliott's End-of-the-Game FG

Distance	PAT	Change	Wind	$\hat{p}$	90% Wald confidence interval
42	0	1	0	.69	(.63, .74)

Should Kansas City forgive Elliott for the failure? After all, one out of three FG's on average are missed in this situation. Maybe Elliott's FG is that one out of three, maybe not. We cannot answer that question here; however, Kansas City needs to remember that, for binary outcomes, "You're going to win some and you're going to lose some."

References and Further Reading

Collett, D. (1991), *Modelling Binary Data*, London: Chapman & Hall.

Hoffman, M. S. (Ed.) (1991), *The World Almanac and Book of Facts 1992*, New York: Pharos Books.

Hosmer, D. W., and Lemeshow, S. (1989), *Applied Logistic Regression*, New York: Wiley.

Morrison, D. G., and Kalwani, M. U. (1993), "The Best NFL Field Goal Kickers: Are They Lucky or Good?" *Chance*, 6(3), 30–37.

Rand McNally Road Atlas (1992), Chicago: Rand McNally & Company.

SAS Institute, Inc. (1989), *SAS/STAT User Guide*, Version 6 (4th ed., Vol. 2), Cary, NC: SAS Institute Inc., 846.

Thayer, E. L. (with illustrations by Bill Jackson) (1984), *Casey at the Bat*, Wichita, KS: The Oak Park Press.