

PREDICTING WIN PERCENTAGE FROM QB PLAY

By Cole Yant & Beau McAndrew

INTRODUCTION

Often regarded as the most important position in football, the quarterback is responsible for coordinating the offense and distributing the ball to his teammates. They touch the ball on nearly every play and need to be careful with it at all times. The best of the best typically convert attempts into yardage, turn the ball over minimally, and throw tons of touchdowns, among many more aspects of the game. Our goal with this study is to see how these statistics really translate to wins among quarterbacks in the NFL.

Motivation for this investigation stems from our interest in the sport. As longtime fans of the game, we have been following the performances of these teams over the years and as both statisticians and avid fantasy football players, we have been keeping up with the statistics of the quarterbacks over their careers. We know that if a quarterback puts up good stats, the team is likely in a better place to win. However, we feel that there has not been enough discussion publicly as to which statistics are most important when judging QB play. We hoped that this report would lead us to an interesting conclusion as football fans, but also figure that the findings could be, although unlikely, applicable for an NFL team to use when making a decision about which QB to add to their team to give them the best chance of winning. By finding which quarterbacking metrics are influential to team success, NFL front-offices could put extra weight on these specific stats when making decisions about which QB to add to a roster or when just evaluating a quarterback, a la Billy Beane and the A's with on-base percentage in Moneyball. Stripped down to the bare bones, our report will try to answer one question: Are interception rate, touchdown rate, and yards per attempt predictive of wins?

METHODS

The data used can be found at www.advancedsportsanalytics.com/nfl-raw-data. We played no part in the collection of this information. Contained in the data were game logs, game info, and player info for every NFL player who recorded passing, rushing, or receiving statistics. We chose to only include games where the individual played at least 90% of the offensive snaps in order to

remove games that may provide misleading results. Additionally, we limited our dataset to only include QBs with at least 30 games played to ensure an adequate sample size, since the plan was to group by player. The most important variables to our question were the passing stats and game outcome. Pass attempts, passing yards, passing touchdowns, interceptions, player, team, home score, and visitor score were the variables we used to arrive at the final dataset.

Modifications made to the data were extensive. Our idea for game analysis was to convert completions, yards, touchdowns, and interceptions to rates based on the number of attempts. This idea came about because we were worried that game script or pacing of the game would cause raw numbers to vary too much game-to-game. This resulted in yards per attempt, touchdowns per attempt, and interceptions per attempt as our level 1 variables. We also created a binary win/loss variable using the player's team in conjunction with the home/away team information and scores. This was done by matching the team to either home or away and comparing that to the difference in team score. Finally, all game stats were then standardized due to the nature of the numbers. Interception, completion, and touchdown rates are on a (0, 1) scale while yards per attempt are on a (-100, 100) scale, although most games fall in the (6, 9) range. Due to the large disparity in possible outcomes, we felt it was necessary to standardize all variables. Finally, we aggregated the level 1 variables across all data to come up with averages for each player as level 2 variables.

The modeling process included using a two-level logarithmic mixed model using games as level 1 units and players as level 2 units. The response variable for this model was game result, coded as (0=loss, 1=win), and model output was a predicted win percentage.

RESULTS

```
Random effects:
  Groups Name      Variance Std.Dev. Corr
  player (Intercept) 0.08085  0.2843
           int_att   0.13613  0.3690   0.27
Number of obs: 955, groups:  player, 22

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.35368    0.09799   3.609 0.000307 ***
yds_att          0.63139    0.09710   6.502 7.91e-11 ***
int_att_avg      0.40090    0.64306   0.623 0.533010
yds_att_avg     -0.90701    0.42689  -2.125 0.033613 *
tds_att_avg      1.42071    0.62511   2.273 0.023042 *
tds_att          0.41692    0.09427   4.423 9.75e-06 ***
int_att         -0.48611    0.11731  -4.144 3.41e-05 ***
yds_att:int_att_avg 1.26086    0.45150   2.793 0.005228 **
---
```

Figure 1. Final model

Level 1 units: Games

Variables: yds/att (yards per attempt), td/att (touchdowns per attempt), int/att (interceptions per attempt)

Level 2 units: Players

Variables: avg yds/att, avg td/att, avg int/att across all data

Interpretations:

Our adjusted ICC value was .062. Only 6.2% of our variance could be explained by player to player variation.

The predicted win rate is $e^{.35368}$ or 58.8% for an average quarterback with average interception rate, yards per attempt, and touchdown rate, as well as average aggregated interception rate, aggregated yards per attempt, and aggregated touchdown rate.

Each 1 standard deviation increase in yards per attempt is associated with an $e^{.63139}$ or 88% increase in median win rate for an average quarterback with average interception rate and touchdown rate, as well as average aggregated interception rate, aggregated yards per attempt, and aggregated touchdown rate.

Each 1 standard deviation increase in a player's aggregated interception rate is associated with an $e^{.4009}$ or 49.3% increase in median win rate for an average quarterback with average interception rate, yards per attempt, and touchdown rate, as well as average aggregated yards per attempt and aggregated touchdown rate.

Each 1 standard deviation increase in a player's aggregated yards per attempt is associated with an $e^{-.90701}$ or 59.6% decrease in median win rate for an average quarterback with average interception rate, yards per attempt, and touchdown rate, as well as average aggregated interception rate and aggregated touchdown rate.

Each 1 standard deviation increase in a player's aggregated touchdown rate is associated with an $e^{1.42071}$ or 414% increase in median win rate for an average quarterback with average interception rate, yards per attempt, and touchdown rate, as well as average aggregated interception rate and aggregated yards per attempt. (Note: No QBs in the dataset were at least 1 SD apart in aggregated TD rate, hence the seemingly large number.)

Each 1 standard deviation increase in touchdown rate is associated with an $e^{.41692}$ or 51.7% increase in median win rate for an average quarterback with average interception rate and yards per attempt, as well as average aggregated interception rate, aggregated yards per attempt, and aggregated touchdown rate.

Each 1 standard deviation increase in interception rate is associated with an $e^{-.48611}$ or 39.5% decrease in median win rate for an average quarterback with average touchdown rate and yards per attempt, as well as average aggregated interception rate, aggregated yards per attempt, and aggregated touchdown rate.

Recording an above average yards per attempt while having an above average aggregated interception rate results in a larger positive effect of the two variables on the predicted win rate. Recording below average in both variables results in a smaller negative effect on the predicted win rate. Recording below average in one variable and above average in another results in a larger negative effect on the predicted win rate.



Figure 2. Examining win percentages among all QBs in dataset

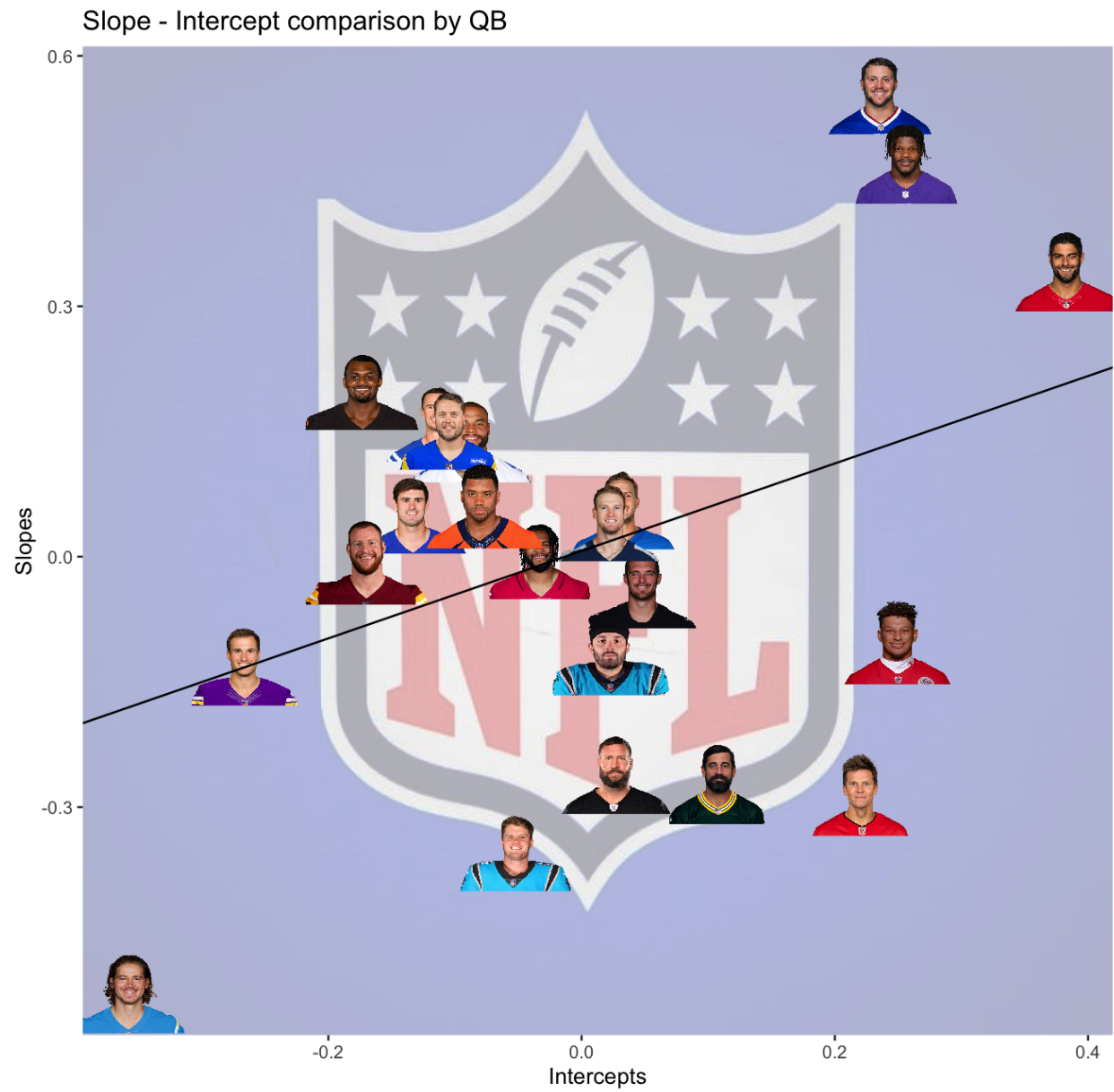


Figure 3. Comparing slopes and intercepts among all QBs

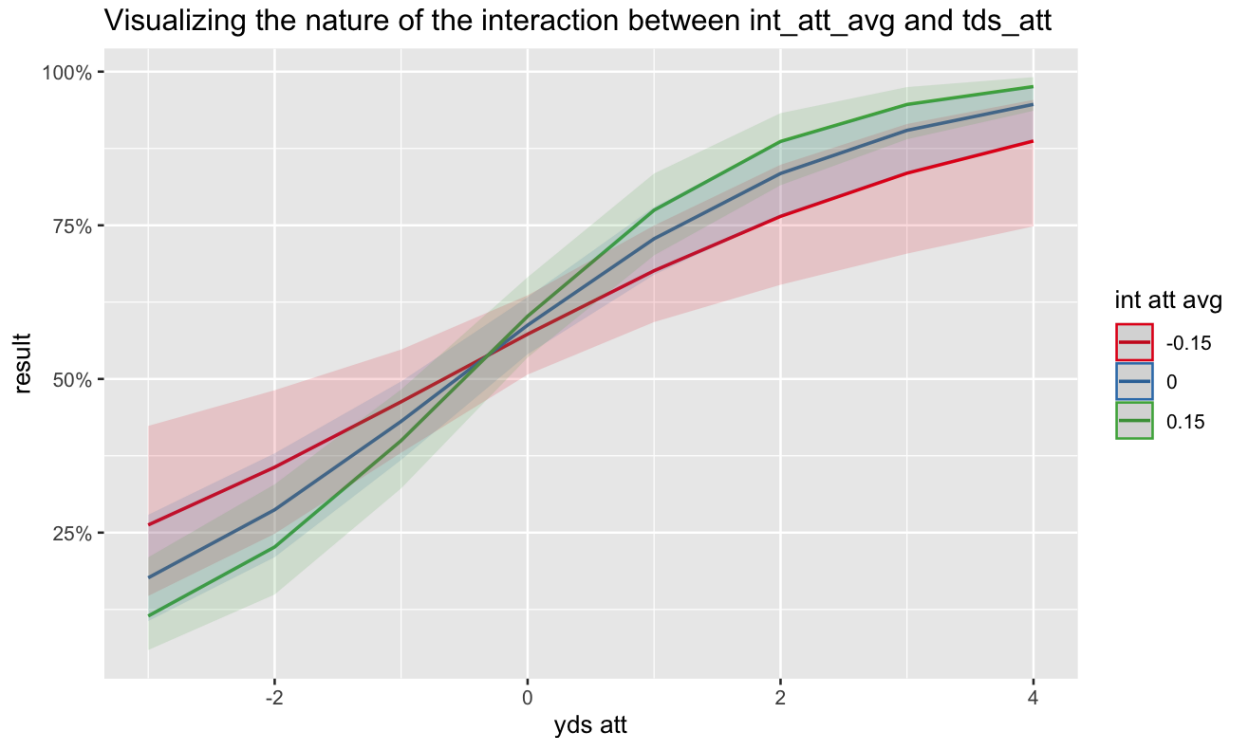


Figure 4. Interaction effect plot between aggregated INT rate (level 2) and TD rate (level 1)

Residual plots and general discussion about this process for a logarithmic mixed model can be found below in the appendix.

DISCUSSION

The results of this model indicate that the variables we chose to investigate are quite predictive of team wins. Our process of model building allowed us to compare more complex models with previous models to identify whether the newer model was going in the right direction, as discussed more in depth within the appendix. This process revealed each successive layer of complexity accounted for a significant amount of more variance. With that being said, not all variables proved to be beneficial in predicting wins. In fact, we tested every combination of cross-level interaction and random slopes and came to the conclusion that the only ones that really benefited our model are the ones we included in this report. Interestingly, interception rate was the only variable that accounted for much variance when added as a random slope. Comparing these interception slopes between the different quarterbacks was interesting to us as football fans because we could come up with theories as to why certain QBs have high slopes

and others unexpectedly lower. For instance, we reasoned that Josh Allen, Lamar Jackson, and Jimmy Garoppolo all have high slopes because they have historically played on teams with tremendous defenses, boast interception rates that are higher than usual, and tend to be fairly average in passing stats despite winning a lot of games. On the other hand, players like Patrick Mahomes and Aaron Rodgers have very small interception rates, and tend to lose when they commit those turnovers. This can be seen in Figure 3, as they are very low in slopes despite being among the highest in win rate as seen in Figure 2. One last thing we would like to touch on is the positive interaction seen between aggregated interception rate and yards per attempt. This relationship makes sense when digging deeper into theory about the game. For example, QBs who don't take risks or lack the ability to throw the ball downfield will carry a lower yards per attempt. Conversely, players who take risks and throw the ball downfield will have higher yards per attempt. These profiles tend to be predictive of wins in different ways. This can be observed in Figure 4, where the red line represents predicted win percentage for players -0.15 sds below average in aggregated interception rate, blue line exactly average, and green line $.15$ sds above average in aggregated interception rate. Essentially, when risk-taking QBs aren't getting yards, their team tends to lose more than safer QBs, while safer QBs don't win at as high of a rate as risk-takers when they produce efficient yardage games.

Our findings here are surface-level, but could be useful in breaking down the game in a basic way. We have shown that quarterbacks performing well in our statistics tend to correspond with their team winning football games. This information may be used by coaches to decide which player will perform best, although it is unlikely that our findings are anything groundbreaking to NFL coaches. Additionally, this report could be cited as a source for teaching people about the game and overall importance of the quarterback position.

While we did find our model to be predictive of wins, we are aware that there may be a few limitations to our approach. First and foremost, our sample of only 22 quarterbacks is not very large. A few of these players may be in unique situations that are not at all predictive of how QB play in the future will lead to wins. The game is also changing every year, with new data and strategies being employed to get the most out of players. Data from 2019 may be considered too old to use in just a few years if there is a change in offensive approach by the league.

Extrapolating these limited past results to future ones may not be particularly useful. Finally, the fact that these QBs face off against each other surely has some bearing on results, especially in the statistics. When one QB plays poorly, it means the opposing quarterback doesn't need to play well to win the game. Oftentimes teams will turn to a conservative gameplan to retain their lead instead of trying to throw the ball and risk interceptions.

If somebody would like to take our research further there is definitely room for improvement. The most obvious option would be to address the simplicity of the model; there are so many different ingredients that go into winning a football game that extend beyond QB play that went unchecked. If someone were to dig further into modeling win probabilities, we would suggest looking to add more context to the solution. This may include more levels, such as breaking up the games by season, or observing the team as a level 2 unit rather than player. Another option could be incorporating more data about the team matchup into the mix, like offensive or defensive strengths of both teams. There are countless more aspects of the game to incorporate past the QB situation if somebody wants to get close to accurately predicting win probabilities.

APPENDIX

Data source: <https://www.advancedsportsanalytics.com/nfl-raw-data>

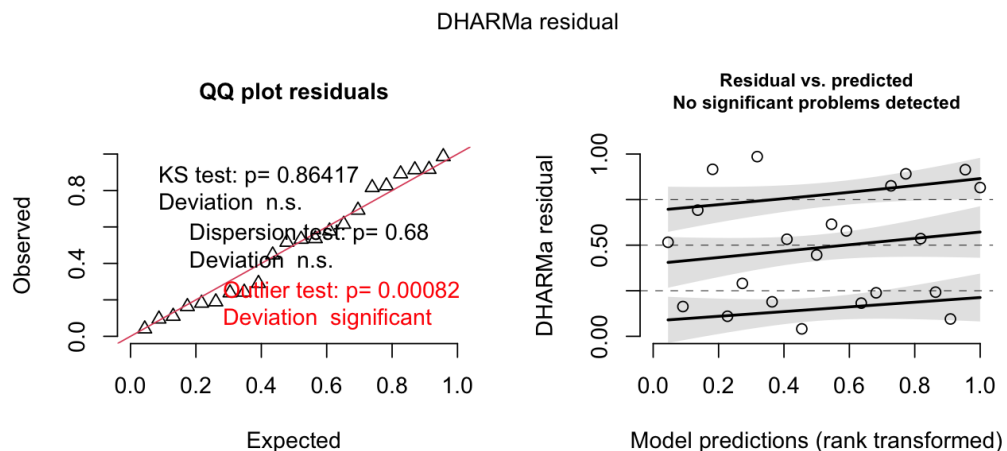
Actual data used for model building is attached.

Model building:

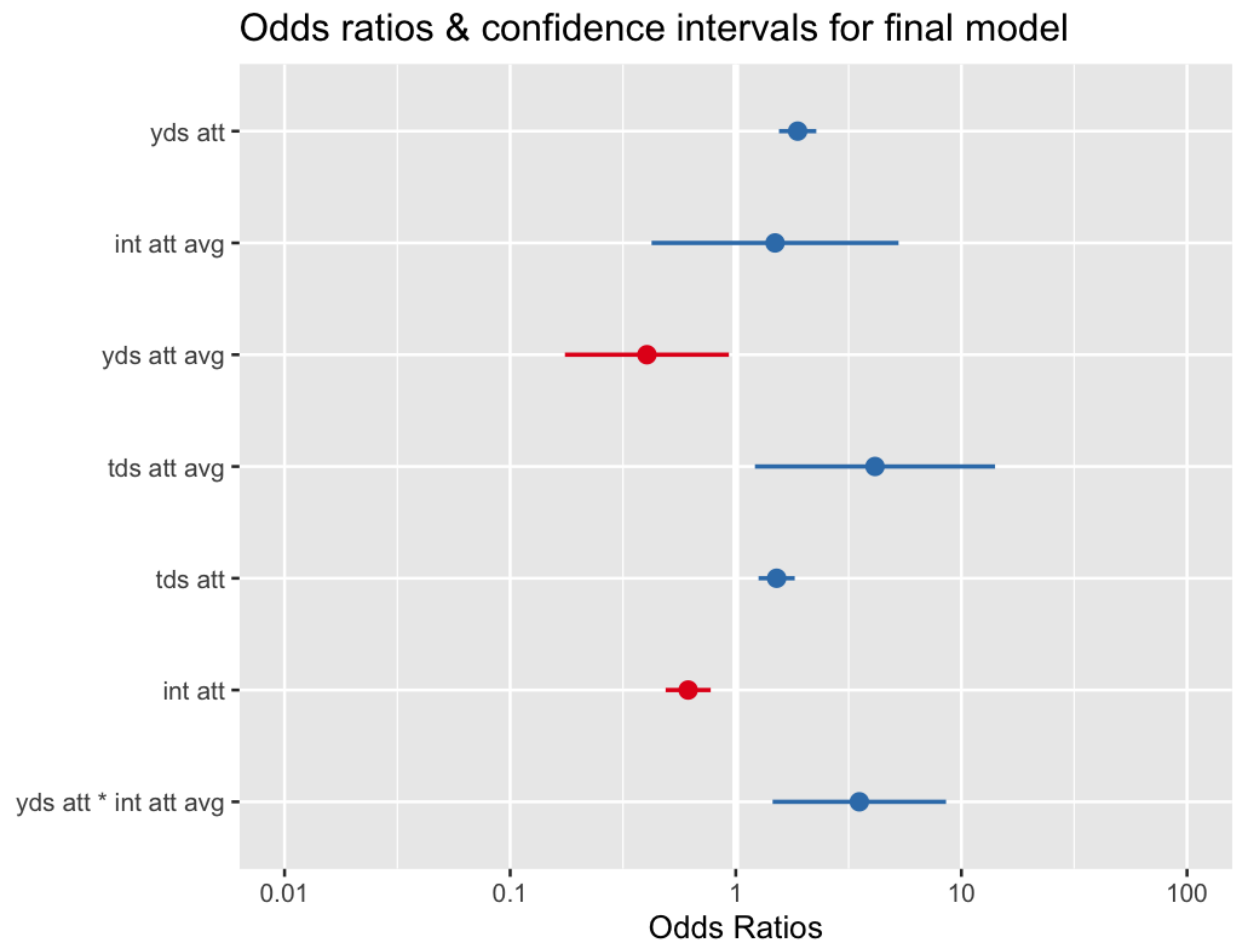
1. `model1 <- glm(result~player, data=qb_final_s)`
2. `model2 <- glmer(result~1+(1|player), data = qb_final_s, family = 'binomial') #done`
3. `model3 <- glmer(result~(1|player) + yds_att + tds_att + int_att, data = qb_final_s, family = 'binomial')`
4. `model4 <- glmer(result~ (1|player) + yds_att_avg + tds_att_avg + yds_att + tds_att + int_att, data = qb_final_s, family = 'binomial')`
5. `model5 <- glmer(result~ (1+int_att|player) +yds_att_avg + tds_att_avg + yds_att + tds_att + int_att, data = qb_final_s, family = 'binomial')`
6. `model6 <- glmer(result~ yds_att*int_att_avg+(1+int_att|player) +yds_att_avg + tds_att_avg + yds_att + tds_att + int_att, data = qb_final_s, family = 'binomial')`

The method we used to hypothesis test our models involved using the `anova()` command to compare new models to the previous model. Additionally, terms with insignificant p-values were dropped from the model, except for the `int_att_avg` variable which was used in a significant interaction despite being insignificant on its own.

One notable exclusion from our report is any discussion of checking residuals. This is because there aren't exactly residuals to visualize with a binary response variable while the model predicts numbers in between 0 and 1. However, we discovered an R package that uses simulation to generate residuals for multi-level models that are not linear. More information can be found here: [DHARMa Documentation](#).



This package generated these plots. We have no problems with linearity, equal variance, or normality in our model. However, there apparently is an outlier among the players. We chose to move forward including the outlier because we don't believe it is truly an outlier for our population of QBs, and our model should include all possible QBs since it is a relatively small, finite population.



Unused figure 1. Odds ratio confidence interval for each term included in final model. These help visualize the effects and confidence of each variable.