THE QUARTERBACK PREDICTION PROBLEM: FORECASTING THE PERFORMANCE OF COLLEGE QUARTERBACKS SELECTED IN THE NFL DRAFT

By Julian Wolfson[*], Vittorio Addona[†] and Robert H. Schmicker[‡]

University of Minnesota, Macalester College, and University of Washington

National Football League (NFL) teams spend substantial time and money trying to predict which college quarterbacks eligible to be drafted into the NFL will have successful professional careers. But despite this investment of resources, it is common for quarterbacks to perform much better or worse than anticipated. Prior work on this "quarterback prediction problem" has concluded that NFL teams are poor at determining which quarterbacks are likely to be successful based on information available prior to the draft. However, these analyses have generally focused only on quarterbacks who played in the NFL, ignoring those who were drafted but did not appear in a professional game. Using data on all quarterbacks drafted since 1997, we considered the problem of predicting NFL success as defined by two metrics (games played and Net Points), based on when a quarterback was drafted and his performances in college and at the NFL Combine. Our analyses suggest that college and combine statistics have little value for predicting whether a quarterback will be successful in the NFL. Contrary to previous work, we conclude that NFL teams aggregate pre-draft information – including qualitative observations – quite effectively, and their inability to consistently identify college quarterbacks who will excel in the professional ranks is a consequence of random variability in future performance due to factors which are unlikely to be observable.

1. Introduction and Background. Quarterback is widely regarded as the most important position on a professional football team. Finding a good quarterback is difficult: In the National Football League (NFL), elite quarterbacks are rarely available via trade or free agency, and hence are

---

[*]Assistant Professor, Division of Biostatistics, University of Minnesota School of Public Health

[†]Assistant Professor, Department of Mathematics, Statistics and Computer Science, Macalester College

[‡]Research Scientist, Resuscitation Outcomes Consortium, University of Washington

1

most often acquired via the amateur draft. Briefly, the draft is a mechanism by which NFL teams select (in reverse order of their previous year's winning percentage) from a pool of eligible college players. Drafting a player gives a team exclusive rights to negotiate a contract with that player. Though players may be selected earlier or later in the draft for a variety of reasons, a player's draft position can generally be viewed as a team's assessment of his overall skill level.

Traditionally, quarterbacks command some of the largest contracts when entering the NFL via the draft. When drafting a quarterback, teams must therefore balance the substantial monetary investment required against the expected benefit derived from the quarterback's future performance. Given the high stakes involved, teams have a vital interest in predicting how successful an individual quarterback will be in the NFL. But in spite of the enormous volume of information available about draft-eligible quarterback prospects and the hundreds of person-hours spent assessing each player's abilities, it remains common for quarterbacks to perform dramatically better or worse than anticipated. Several current or recent NFL starting quarterbacks (e.g. Tom Brady, Matt Hasselbeck, Marc Bulger, Matt Cassel, Kyle Orton, and David Garrard) were drafted in the fourth round or later, meaning that at least 100 players, including a number of quarterbacks, were selected before them. Others (e.g. Kurt Warner and Tony Romo) went undrafted entirely. Moreover, several quarterbacks selected with one of the first five picks overall (e.g. JaMarcus Russell, Tim Couch, Akili Smith, Ryan Leaf, and Heath Shuler) have played very poorly in the NFL. The challenge of identifying college quarterback prospects who are most likely to succeed at the professional level is among the "Hilbert Problems" for football identified by Schatz (2005). In the remainder of this paper, we will refer to this challenge as the *quarterback prediction problem*.

The difficulty of predicting whether or not a college quarterback will be successful in the NFL was highlighted in a 2008 *New Yorker* article by Malcolm Gladwell (Gladwell, 2008). The article cited the work of Berri and Simmons (2009), who concluded that the draft position of a quarterback had a considerable impact on how much that quarterback played, but not on how well he performed in the NFL. Quinn et al. (2007) arrived at similar conclusions using different performance metrics. Lewin (2006) developed a projection system for future NFL quarterbacks, and concluded that games started and completion percentage in college were the only important predictors of later success, but did not provide the details of his methodology. Massey and Thaler (2010) considered whether the compensation of draft picks reflected their future performance, and concluded that teams were

overpaying early first-round draft picks.

If, as much of this work suggests, NFL teams are poor at identifying college prospects who are likely to succeed as NFL quarterbacks, two possible explanations are:

1. NFL teams may aggregate available information sub-optimally, emphasizing some attributes which do not correlate with NFL performance, and de-emphasizing other attributes which are more predictive of NFL success.
2. The variability in individual performance due to random, unmeasurable factors may make prediction inherently difficult, even if all available information were used optimally.

In this paper, we consider both of these explanations, and attempt to quantify how much each contributes to the quarterback prediction problem. Our work differs from previous research on this problem in two main ways. First, we base our analyses on all quarterbacks drafted into the NFL, not only on those who have played in at least one NFL game. Second, we explicitly estimate the predictive ability of our models to assess the inherent difficulty of the quarterback prediction problem.

Section 2 describes the data, while Section 3 introduces our outcome measures and predictors. Section 4 provides details of the methods we employ. In Section 5, we present the results of our analysis. We conclude with a brief discussion in Section 6.

2. Data. Draft position and most NFL statistics were obtained from Pro-Football-Reference.com for all quarterbacks drafted since 1997. Number of sacks and fumbles lost were obtained from NFL.com. College statistics back to the 2000-01 season were obtained from NCAA.com. Career college statistics for quarterbacks who played before 2000-01 were obtained from several other sources, including school websites. The NFL Scouting Combine is an annual week-long event held roughly two months prior to draft day, during which college football players undergo a variety of physical and mental evaluations at the request of NFL coaches, general managers, and scouts. Physical evaluations from the combine were obtained from nfl-draftscout.com.

In total, we obtained information on 160 quarterbacks. Brad Smith (who played quarterback for Missouri and was drafted in the fourth round of 2006 by the New York Jets) and Isaiah Stanback (who played quarterback for Washington and was drafted in the fourth round of 2007 by the Dallas Cowboys) were omitted from our analysis because they have played almost exclusively as wide receivers in the NFL.

The complete dataset is provided in the Supplementary Materials accompanying this paper.

3. Outcomes and predictors.

3.1. *Outcomes.* One fundamental challenge that arises in the quarterback prediction problem is how to quantify quarterback performance and thereby determine how "successful" a quarterback's professional career has been. Cumulative statistics (e.g. games played/started, pass attempts/yards, touchdowns etc.) are closely related to the number of opportunities given to a quarterback, opportunities which may be determined by factors other than on-field performance. For example, teams may be more reluctant to replace a player who is performing poorly if that player was drafted early (and hence highly paid); teams may be less tolerant of a poorly performing player if he was selected later in the draft. Figure 1 displays the number of games played by quarterbacks drafted since 1997, stratified by the round in which they were drafted.
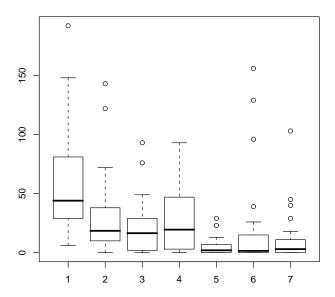

FIG 1. *Number of games played in the NFL by draft round*

In order to avoid this potential problem with cumulative statistics, Berri and Simmons (2009) quantified NFL performance by a variety of per-play

metrics, and concluded that a quarterback's NFL performance was not associated with when he was selected in the draft. However, in most of their analyses, quarterbacks with fewer than 100 plays of NFL experience were excluded. Many of the excluded players had never been involved in a single play in the NFL, and hence per-play metrics were undefined for these individuals.

Excluding quarterbacks with fewer than 100 plays from the analysis is problematic unless one assumes that these quarterbacks would have performed similarly, if given similar playing time, to those with more than 100 plays of experience. In other words, the results may be biased unless quarterbacks with fewer than 100 NFL plays are missing completely at random (Little and Rubin, 2002). But the missing completely at random (MCAR) assumption seems tenuous: Once a college quarterback has been drafted onto an NFL team, that team's coaches can observe his performance in training camp, team practices, and exhibition games before deciding whether or not to allow him to play in a regular season game. While one might assert that coaches and team personnel are beholden to draft status and other auxiliary factors when making these decisions, an alternative explanation for Berri and Simmons' surprising findings is that they reflect selection bias. That is, quarterback performance is unrelated to draft status *conditional* on an NFL coach deeming a quarterback sufficiently skilled to play professionally, but quarterbacks drafted in the earlier rounds are far more likely to possess this minimum skill level and reach the 100-play threshold.

We would argue that NFL teams, as well as casual fans, are generally interested in knowing whether one can predict the likelihood of NFL success for *all* drafted quarterbacks before they play an NFL game. Indeed, draft experts and fans often talk of a prospect's "bust potential," referring to the possibility that a highly-touted college quarterback will be drafted early, only to be judged incapable (presumably based on their performance in practice and exhibition games) of playing at the NFL level. It is clearly of interest to identify pre-draft information which might suggest that certain quarterbacks are more or less likely to be "busts".

For our analyses, we considered two cumulative statistics quantifying NFL performance:

1. **Games played.** Counts the total number of NFL games in which a quarterback has been involved in at least one play. In our analyses, we treated games played as an integer-valued random variable, and also considered three binary variants. Letting G be the number of games

played, we define

$$(1) \qquad G^{(K)} = \left\{ \begin{array}{lll} 1 & \text{if} & G \geq K \\ 0 & \text{if} & G < K \end{array} \right.$$

for $K = 1$, 16, and 48. These cutoffs correspond, informally, to a minimal, moderate, and substantial degree of NFL success. Quarterbacks with $G \geq 1$ (i.e. $G^{(1)} = 1$) can be thought of as having reached a minimum competence threshold: their team's coaching staff has judged them good enough to play in an NFL game. Similarly, quarterbacks with $G^{(48)} = 1$ are generally considered very good to excellent, as few poor quarterbacks are allowed to play in this many games (48 games corresponds to three complete seasons).

2. **Net Points.** Berri and Simmons (2009) used a statistic, Net Points, which quantifies how many points a quarterback contributes to his team based on cumulative statistics. As per Berri (2008), Net Points is calculated as

Net Points = 0.08*Yards - 0.21*Plays - 2.7*Interceptions - 2.9*FumblesLost

where Yards = Passing Yards + Rushing Yards, and Plays = Pass Attempts + Rush Attempts + Sacks. Fractional Net Points are rounded to the nearest integer. Berri and Simmons computed Net Points only for quarterbacks who had accumulated statistics at the NFL level; for our analysis, we assigned zero Net Points to quarterbacks who have not played in the NFL, since they have not accumulated any of its component statistics. Thirty quarterbacks had small negative Net Points values (less than 10 in absolute value), which we set to zero. Figure 2 plots the distribution of Net Points.

Note that our outcome measures are defined for all drafted quarterbacks, and may be affected by the number of playing opportunities that a quarterback is afforded. The degree to which playing opportunities depend on factors other than on-field performance is unknown, but in Section 5, we present analyses contradicting the view that these factors play a major role in determining playing time for quarterbacks. We revisit this issue alongside our conclusions in Section 6.

3.2. *Predictors.* We considered the following predictors of NFL performance in our regression models: Draft position (Pick), year drafted (Year), passing statistics compiled during a quarterback's college career, and measurements from the NFL Scouting Combine (including Height and Weight).
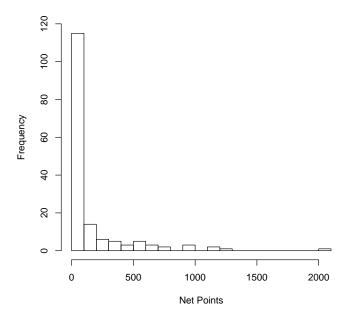
FIG 2. *Histogram of Net Points in the NFL*

In all our models, the Pick variable was log transformed. Table 1 presents summary statistics of the predictors in our analysis.

4. Methods.

4.1. *Regression models.* The binary variables $G^{(1)}$, $G^{(16)}$, and $G^{(48)}$ were modeled via logistic regression. Games played (G) was modeled via negative binomial (NB) regression (Agresti, 2002). Suppose that, given $\lambda > 0$, Y has a Poisson distribution with mean λ, and that $\lambda \sim Gamma(k, \mu)$. Then the marginal probability function of Y is negative binomial, taking the form

$$P(Y = y; k, \mu) = \frac{\Gamma(y + k)}{\Gamma(k)\Gamma(y + 1)} \left(\frac{k}{\mu + k}\right)^k \left(1 - \frac{k}{\mu + k}\right)^y$$

with $E(Y) = \mu$ and $var(Y) = \mu + \mu^2/k$. $\theta = 1/k$ reflects the degree of overdispersion of the counts; as $\theta \to 0$, the negative binomial distribution converges to the usual Poisson distribution. In our case, both games played and Net Points showed evidence of overdispersion: the negative binomial regression models we fit generally estimated $\theta \approx 2$, with standard errors less than 0.4.

Total $N = 160$

Predictor	Median [Min, Max]	# missing
ColGames	39 [12, 53]	19
Number of games played		
CompPerc	58.7 [40.9, 70.4]	13
Completion percentage = (# Completions) / (# pass attempts)		
YPA	7.7 [5.7, 10.1]	13
Yards per pass attempt = (Pass yards) / (# pass attempts)		
Int	28 [1, 64]	15
Number of interceptions		
TD	54 [0, 131]	11
Number of touchdowns		
Height	75 [70, 79]	51
Height (in inches)		
Weight	225 [192, 265]	51
Weight (in lbs.)		
40-yard dash	48.1 [43.3, 53.7]	51
Time to run 40 yards (in 0.1s of a second)		
Vertical jump	31.5 [21.5, 38.5]	79
Vertical leap height from a standing position (in inches)		
Cone drill	71.3 [67.2, 78.0]	82
Time to run a course marked by cones (in 0.1s of a second)		

TABLE 1

Summary statistics for predictors

As noted, quarterbacks could have zero Net Points either because they did not play in the NFL and were assigned zero points by definition, or because they did play and had their Net Points rounded to zero. Since zero values for this outcome can be viewed as having been generated by two separate processes, we modeled the Net Points outcome as a zero-inflated negative binomial (ZINB) random variable (Greene, 2008; Yau et al., 2003). The ZINB model extends the NB model by allowing extra probability mass to be placed on the value zero, with the probability that an observation is a structural or "excess" zero modeled by logistic regression. Although it is possible to use different predictors for the two components of a ZINB model, we used the same sets of predictors for both components in our analysis.

For each regression, we considered two primary models. The first model (Base) contained the college predictors (ColGames, CompPerc, YPA, Int,

and TD) listed in Table 1, along with Year; the second model contained all the Base predictors plus log(Pick), a term accounting for where a player was selected in the NFL draft. We also considered two secondary models with the same predictors as the primary models, but excluding quarterbacks selected in the first round. Due to the financial investment required to sign first-round draft selections, one could reasonably argue that the playing opportunities for these quarterbacks are most heavily influenced by external factors unrelated to their on-field performance. An analysis which excludes these players may indicate whether the predictors of success differ for more "disposable" quarterbacks who were selected later in the draft and did not command a large contract. Finally, we refit these four models using the combine measurements (Height, Weight, 40-yard dash, Vertical jump, and Cone drill) from Table 1 in place of college statistics.

4.2. *Assessing predictive accuracy.* Predictions for each quarterback in the dataset were generated based on the fitted models:

- For the logistic regressions, predictions $\hat{G}_i^{(K)}$ were obtained as

$$\hat{G}_i^{(K)} = \mathbf{1}[\hat{\pi}_i^{(K)} \geq 0.5]$$

  with $\hat{\pi}_i^{(K)}$ representing the estimated probability that $G_i^{(K)} = 1$. For the "Intercept only" model where $\hat{\pi}^{(K)}$ is the same for all individuals, predictions were derived via a biased coin-toss method, so that $\hat{G}_i^{(K)}$ was generated as a Bernoulli random variable with success probability equal to $\hat{\pi}^{(K)}$.

For the integer-valued outcomes, we label our predictions as $\hat{Y}_i$, referring either to predicted games played (NB models) or Net Points (ZINB models).

- In the NB regressions, predictions $\hat{Y}_i$ were obtained from the fitted values for each individual i.
- In the ZINB regressions, predictions were obtained for each individual i via

$$\hat{Y}_i = \left\{ \begin{array}{ll} 0 & \text{if } \hat{\phi}_i < 0.5 \\ \hat{\theta}_i & \text{if } \hat{\phi}_i \geq 0.5 \end{array} \right.$$

  where $\hat{\phi}_i$ is the estimated probability that individual i represents a structural zero, and $\hat{\theta}_i$ is the estimated mean for individual i given that he/she is not a structural zero.

Predictive accuracy for binary outcomes was quantified by the misclassification rate

$$MR = \frac{1}{n} \sum_i \mathbf{1}[|\hat{G}_i^{(K)} - G_i^{(K)}| > 0.5] \ ,$$

and predictive accuracy for integer-valued outcomes was quantified via the absolute prediction error

$$APE = \frac{1}{n} \sum_i |\hat{Y}_i - Y_i| \ ,$$

where Y_i refers to either games played or Net Points. Both the misclassification rate and absolute prediction error were estimated via 5-fold cross-validation using the original data (Efron and Gong, 1983).

5. Results.

5.1. *Games played.* Table 2 reports the results of the eight regression models, associated with the integer-valued games played variable G described in Section 4.1. The values in Table 2 represent the percent increase in the mean of G (and corresponding 95% confidence intervals) associated with one-unit increases in each predictor. Tables 3, 4, and 5 give the percent increases in the odds of $P(G^{(K)} = 1)$ (and corresponding 95% confidence intervals) for $K = 1$, 16, and 48, respectively. Confidence intervals which exclude zero are highlighted in bold.

	All quarterbacks		Rounds 2-7 only	
Variable	Base	Base+Pick	Base	Base+Pick
ColGames	1 (-3,5)	1 (-3,5)	2 (-4,8)	2 (-3,7)
CompPerc	**8 (0,15)**	5 (-2,11)	4 (-4,13)	2 (-6,11)
YPA	-6 (-34,35)	-23 (-45,8)	-28 (-55,18)	-17 (-47,30)
Int	1 (-2,5)	0 (-3,3)	0 (-4,4)	0 (-2,2)
TD	0 (-2,2)	0 (-2,1)	0 (-2,3)	0 (-3,4)
Year	**-19 (-26,-11)**	**-18 (-25,-11)**	**-22 (-32,-11)**	**-21 (-30,-10)**
log(Pick)	—	**-33 (-43,-22)**	—	**-47 (-71,-2)**
Height	0 (-22, 28)	1 (-20, 28)	-12 (-35, 19)	-4 (-30, 31)
Weight	1 (-3, 5)	0 (-3, 4)	0 (-4, 4)	0 (-4,4)
40-yard dash	5 (-21, 45)	-1 (-25, 33)	12 (-28, 70)	3 (-33,56)
Vertical jump	4 (-10, 22)	-1 (-13, 14)	6 (-13, 28)	1 (-17, 22)
Cone drill	0 (-15, 19)	11 (-6, 32)	14 (-7, 40)	16 (-5, 42)
Year	**-20 (-31,-9)**	**-22 (-32,-11)**	**-21 (-34, -5)**	**-23 (-36,-8)**
log(Pick)	—	**-37 (-54,-16)**	—	-47 (-73,11)

TABLE 2

Percent change in number of NFL games played (with 95% confidence intervals) associated with one-unit differences in college and combine statistics, year drafted, and draft position.

Year was negatively associated with games played in nearly all models; predictably, more years in the league generally leads to more games played. The only other predictor which was consistently associated with games

Variable	All quarterbacks		Rounds 2-7 only	
	Base	Base+Pick	Base	Base+Pick
ColGames	0 (-6, 6)	0 (-6, 7)	2 (-5, 9)	1 (-6, 8)
CompPerc	3 (-8, 17)	-1 (-13, 13)	1 (-11, 14)	-1 (-13, 13)
YPA	36 (-29, 172)	33 (-33, 171)	27 (-36, 157)	32 (-33, 169)
Int	2 (-3, 8)	1 (-4, 7)	1 (-4, 7)	1 (-4, 7)
TD	0 (-3, 3)	0 (-3, 3)	0 (-3, 3)	0 (-3, 3)
Year	**-19 (-32, -4)**	**-21 (-36, -4)**	**-21 (-36, -5)**	**-21 (-36, -4)**
log(Pick)	—	**-75 (-91, -47)**	—	**-68 (-90, -13)**
Height	-2 (-38, 54)	4 (-36, 71)	-2 (-40, 58)	3 (-37, 69)
Weight	3 (-4, 11)	2 (-6, 10)	2 (-5, 10)	2 (-6, 10)
40-yard dash	3 (-35, 72)	-2 (-42, 70)	-6 (-45, 65)	-5 (-45, 68)
Vertical jump	5 (-17, 32)	-1 (-22, 26)	0 (-23, 27)	-2 (-23, 26)
Cone drill	1 (-23, 32)	9 (-17, 45)	5 (-19, 39)	9 (-18, 45)
Year	-10 (-29, 13)	-15 (-35, 10)	-13 (-34, 11)	-15 (-35, 10)
log(Pick)	—	**-65 (-90, -19)**	—	-56 (-88, 39)

TABLE 3

Percent change in odds of playing ≥ 1 NFL game (with 95% confidence intervals) associated with one-unit differences in college and combine statistics, year drafted, and draft position.

Variable	All quarterbacks		Rounds 2-7 only	
	Base	Base+Pick	Base	Base+Pick
ColGames	1 (-4, 7)	1 (-5, 8)	2 (-5, 9)	0 (-7, 8)
CompPerc	7 (-4, 19)	4 (-7, 17)	3 (-9, 16)	1 (-11, 15)
YPA	-5 (-45, 64)	-27 (-62, 39)	-25 (-62, 43)	-22 (-61, 54)
Int	0 (-4, 5)	-1 (-6, 5)	0 (-5, 6)	0 (-5, 6)
TD	1 (-1, 4)	1 (-2, 4)	1 (-2, 4)	1 (-2, 4)
Year	**-21 (-33, -10)**	**-26 (-39, -12)**	**-21 (-35, -5)**	**-19 (-34,-3)**
log(Pick)	—	**-62 (-76, -45)**	—	**-65 (-86,-15)**
Height	22 (-20, 90)	37 (-14, 128)	4 (-35, 69)	20 (-29, 109)
Weight	0 (-7, 6)	-3 (-9, 4)	-1 (-7, 6)	-3 (-10, 5)
40-yard dash	-9 (-42, 40)	-19 (-50, 32)	-10 (-49, 56)	-14 (-54, 58)
Vertical jump	2 (-17, 27)	-3 (-23, 21)	0 (-22, 28)	-4 (-26, 24)
Cone drill	14 (-11, 48)	31 (-1, 77)	26 (-4, 71)	**36 (1, 89)**
Year	**-27 (-43, -10)**	**-34 (-51, -16)**	**-25 (-44, -4)**	**-29 (-49, -7)**
log(Pick)	—	**-60 (-82, -28)**	—	**-75 (-95, -8)**

TABLE 4

Percent change in odds of playing ≥ 16 NFL games (with 95% confidence intervals) associated with one-unit differences in college and combine statistics, year drafted, and draft position.

played was draft position (with quarterbacks drafted in the later rounds playing fewer games). The influence of draft status was relatively consistent across models: one log differences in draft pick number (eg. the difference between the first overall selection and the third, or the tenth overall selection and the twenty-seventh) were associated with 30-60% fewer games

	All quarterbacks		Rounds 2-7 only	
Variable	Base	Base+Pick	Base	Base+Pick
ColGames	**10 (1,21)**	**14 (2, 29)**	**18 (0,49)**	16 (-3,46)
CompPerc	**16 (0,37)**	**25 (4, 56)**	**34 (3,86)**	**34 (3,88)**
YPA	-33 (-71,47)	**-72 (-91,-23)**	-73 (-95,2)	-72 (-94,14)
Int	-1 (-8,6)	-2 (-11,8)	0 (-15,16)	-1 (-17,16)
TD	0 (-3,4)	-2 (-7,3)	-4 (-13,3)	-4 (-13,4)
Year	**-38 (-52,-23)**	**-45 (-61,-28)**	**-56 (-79,-29)**	**-53 (-78,-25)**
log(Pick)	—	**-65 (-80,-45)**	—	-59 (-94,140)
Height	38 (-28, 181)	28 (-37, 173)	5 (-52, 133)	40 (-42, 284)
Weight	6 (-3, 18)	8 (-4, 25)	6 (-5, 21)	6 (-7, 23)
40-yard dash	-3 (-53, 77)	-4 (-56, 84)	43 (-40, 258)	39 (-54, 362)
Vertical jump	2 (-28, 43)	-7 (-39, 38)	7 (-30, 67)	-3 (-49, 69)
Cone drill	-8 (-40, 38)	-1 (-40, 62)	-13 (-47, 41)	-14 (-52, 46)
Year	**-45 (-67, -19)**	**-51 (-74, -24)**	-32 (-63, 9)	-37 (-69, 5)
log(Pick)	—	**-58 (-82, -21)**	—	-87 (-99, 11)

TABLE 5

Percent change in odds of playing $\geq$ 48 NFL games (with 95% confidence intervals) associated with one-unit differences in college and combine statistics, year drafted, and draft position.

played and similar decreases in the odds of achieving the previously defined games played thresholds. Neither college nor combine statistics were associated with number of games played, playing in $\geq$ 1, or playing in $\geq$ 16 NFL games. However, completion percentage and number of games played in college were positively associated with playing in $\geq$ 48 games in the NFL, even after adjusting for draft status.

Models fitted to quarterbacks drafted after the first round yielded generally similar results to models fitted to all drafted quarterbacks. The only notable differences between models including and excluding first-round quarterbacks were for $G^{(48)}$, the indicator of playing at least 48 NFL games. For $G^{(48)}$, confidence intervals for College Games, YPA, Year and log(Pick) excluded zero in models using all quarterbacks but included zero when first-round picks were omitted. However, the point estimates for these covariates did not change substantially.

5.2. *Net Points.* Table 6 summarizes the results of the NB count portions of the ZINB models for Net Points, as before, reporting percent increases in the mean for a one-unit increase in each predictor, along with 95% confidence intervals. For the sake of brevity, we do not report coefficient estimates from the "excess zeros" portions of the ZINB models. Briefly, log(Pick) attained significance in all of these models, with later selections having a higher chance of being a zero. Faster cone drill times were associ-

ated with a higher probability of being zero in two of the four models. No other combine measure or college statistic, nor year drafted, was associated with the probability of being an excess zero.

Variable	All quarterbacks		Rounds 2-7 only	
	Base	Base+Pick	Base	Base+Pick
ColGames	3 (-2,8)	2 (-2,7)	2 (-8,14)	-1 (-11,10)
CompPerc	**11 (1,21)**	**9 (1,17)**	1 (-11,15)	-1 (-13,13)
YPA	-14 (-44,32)	**-36 (-59,-2)**	**-61 (-80,-23)**	**-72 (-87,-38)**
Int	1 (-4,6)	1 (-3,5)	-1 (-7,6)	-1 (-8,7)
TD	-1 (-3,2)	-1 (-3,1)	1 (-3,4)	2 (-2,7)
Year	**-20 (-28,-11)**	**-20 (-28,-12)**	**-28 (-41,-11)**	**-28 (-42,-11)**
log(Pick)	—	**-27 (-39,-14)**	—	53 (-42,309)
Height	-12 (-32, 15)	-9 (-29, 18)	-26 (-49, 7)	-21 (-48, 20)
Weight	**5 (1, 10)**	**4 (0, 8)**	**5 (0, 10)**	4 (-1, 9)
40-yard dash	6 (-28, 56)	6 (-26, 53)	61 (-20, 225)	53 (-22, 199)
Vertical jump	-8 (-24, 12)	-9 (-23, 9)	8 (-22, 49)	6 (-22, 43)
Cone drill	-10 (-24, 8)	-2 (-18, 17)	-3 (-25, 24)	-2 (-24, 26)
Year	**-27 (-38, -14)**	**-27 (-38, -15)**	**-23 (-41, 0)**	**-23 (-41, -1)**
log(Pick)	—	-24 (-43, 1)	—	-25 (-71, 93)

TABLE 6

Percent change in NFL Net Points (with 95% confidence intervals) associated with one-unit differences in college and combine statistics, year drafted, and draft position.

Generally, the conclusions for Net Points are very similar to those for NFL games played: Year and draft position were negatively associated with Net Points (i.e. quarterbacks drafted more recently and later in the draft produced fewer Net Points), and college/combine statistics were generally not associated with this outcome. The one exception to this rule was YPA, which was negatively associated with Net Points in three of the four models in which it was incorporated, including both models adjusting for draft position. We note, however, that the direction of this relationship is contrary to conventional wisdom (which would dictate that quarterbacks with higher college YPA will tend to have more success in the NFL). We discuss the interpretation of this counter-intuitive result in Section 6.

As with games played, models for Net Points fitted to all quarterbacks did not differ greatly from models fitted to quarterbacks drafted in rounds 2-7. Point estimates for the (negative) effect of YPA on Net Points were larger in magnitude for models excluding first-round quarterbacks, as were estimates of the (positive) effect of 40-yard dash time, although the very wide confidence intervals for the latter should be noted.

5.3. *Predictive accuracy.* We compared the predictive performance of nine models:

1. **Intercept only:** A naive model which uses no predictor information, estimating a common intercept term for the entire population.
2. **Year:** A model including draft year as the sole predictor.
3. **+ log(Pick):** A model including Year and log(Pick).
4. **+ College Stats:** A model including Year and the college statistics listed in Table 1.
5. **+ Combine Stats:** A model including Year and the combine statistics listed in Table 1.
6. **+ College + Combine:** A model including Year, and college and combine statistics.
7. **+ log(Pick) + College:** A model including Year, log(Pick), and college statistics.
8. **+ log(Pick) + Combine:** A model including Year, log(Pick), and combine statistics.
9. **+ log(Pick) + College + Combine:** A model including all of the available predictors.

Figure 3 summarizes the misclassification rate estimates for $G^{(1)}, G^{(16)}$, and $G^{(48)}$ from 100 runs of 5-fold cross-validation. Results (not shown) were similar when first-round picks were excluded from the analysis.

From Figure 3, we observe that, for $G^{(1)}$, the model containing only information on what year a quarterback was drafted had the smallest misclassification rate. For $G^{(16)}$ and $G^{(48)}$, the model which additionally incorporated information on a quarterback's draft position performed best. College and combine measurements provided no additional predictive value beyond Year and Pick; all the models including college and combine statistics misclassified quarterbacks at a higher rate than the simpler models. Indeed, these models generally offered no improvement in misclassification rate over models with Year as the sole predictor. The model including college and combine statistics but not log(Pick) (sixth row, for each of $G^{(1)}$, $G^{(16)}$, and $G^{(48)}$, in Figure 3) had worse classification performance than all but the naive Intercept Only model.

Figures 4 and 5 summarize the cross-validation estimates (based on 100 runs of 5-fold cross-validation) of the absolute prediction error for games played and Net Points, respectively. As with the binary outcomes, results were similar when quarterbacks drafted in the first round were excluded from the analysis.

For the integer-valued games played outcome, models with Year and Year + log(Pick) appeared to predict slightly better than the Intercept Only model, and models incorporating college and combine statistics performed substantially worse. The decrease in prediction error due to including Year

and log(Pick) was greater for the Net Points outcome, while the models using college and combine statistics did not seem to yield better prediction of Net Points than the Intercept Only model.

6. Discussion. Based on the preceding analyses, we draw the following conclusions:

NFL teams appear to use pre-draft information intelligently. Year drafted and draft position were by far the most important predictors of future NFL success. We found some evidence that quarterbacks with higher college YPA are likely to produce fewer Net Points in the NFL, indicating that NFL teams may be drafting college quarterbacks with high YPA earlier than their talent level would dictate. YPA may be inflated for quarterbacks who play at large colleges with elite surrounding talent or in systems designed to emphasize the passing game. But, overall, it does not appear that NFL teams are systematically under- or over-emphasizing particular quantitative measures.

Our results also suggest that draft position provides information not contained in college and combine statistics. This is not surprising, since NFL teams possess a plethora of qualitative information on quarterback prospects not related to in-game performance. For example, reports on player attributes compiled by professional scouts, observations obtained at "Pro Days" organized by individual colleges and universities, knowledge of how strength of college opponents/teammates (or the "system" in which the quarterback played) may have affected traditional statistics, injury status, and personal interactions may all provide crucial knowledge to an NFL team.

A competing interpretation of our results is that NFL teams are using pre-draft information sub-optimally and reinforcing these decisions by systematically denying or awarding playing time to quarterbacks based on their draft position without regard to on-field performance. Previous work has focused on this possibility, but the resulting approach (considering per-play data, and thereby excluding quarterbacks who have not played in the NFL) is vulnerable to selection bias, which may be severe in this case. In our analyses, we chose to consider outcomes which are dependent on the amount of playing time a quarterback is given. We investigated the plausibility of the hypothesis that playing time is awarded to highly-drafted quarterbacks without regard to performance by fitting models which excluded quarterbacks drafted in the first round, precisely those one would expect to benefit most from a policy of awarding opportunities based on status rather than

merit. Neither the effect of draft position nor any of the other predictors we considered was appreciably different in the analyses of this subset of quarterbacks. Though this finding does not rule out the possibility that external factors influence playing time decisions, it suggests that the role of such factors may be exaggerated.

College and combine statistics for drafted quarterbacks are not reliably associated with, or predictive of, success in the NFL.
In sports statistics circles, much has been made about a projection system (Lewin, 2006) for quarterbacks which uses the number of games started in college and college completion percentage to predict future NFL success. In our analyses, these variables were only associated with an indicator of playing at least 48 NFL games, but they were not related to any of our other outcome measures. Generally, college and combine performance statistics provided no additional predictive ability beyond year drafted and draft position. Indeed, in most cases, including college/combine measurements degraded predictive performance, suggesting that the amount of statistical noise in these predictors overwhelms any predictive value they might have.

The quarterback prediction problem is inherently difficult. Though it appears that NFL teams do have some ability to discriminate between quarterbacks who are likely to be successful in the NFL and those who are not, there remains substantial uncertainty in predicting the future performance of college quarterback prospects. Even the best-performing predictive model for the indicator of playing at least 16 NFL games had a misclassification rate over 30%. Similarly, the smallest estimated prediction error for the integer-valued games played outcome was nearly 20 games, over one seasons' worth. The smallest estimated prediction error for Net Points was greater than 125 points, a threshold achieved by fewer than 30% of the quarterbacks in our dataset.

Given the poor predictive performance of models incorporating a variety of quantitative measures, it seems unlikely that collecting more statistics on the performance of college quarterbacks will yield a clearer picture about their likelihood of success in the NFL. Indeed, one might reasonably argue that there are few observable factors, either quantitative or qualitative, which are not already being used in a near-optimal way to predict quarterback performance. Though NFL draft "experts" at the major sports networks may object, it appears that factors which are inherently unmeasurable and/or random play a major role in determining whether a quarterback will

succeed at the professional level.

REFERENCES

AGRESTI, A. (2002). *Categorical Data Analysis (Wiley Series in Probability and Statistics)*. 2nd ed. Wiley-Interscience. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471360937.

BERRI, D. and SIMMONS, R. (2009). Catching a draft: on the process of selecting quarterbacks in the national football league amateur draft. *Journal of Productivity Analysis*. URL http://dx.doi.org/10.1007/s11123-009-0154-6.

BERRI, D. J. (2008). *Back to back evaluations on the gridiron*, chap. 14. 1st ed. Chapman & Hall/CRC, Boca Raton, FL, 241–261.

EFRON, B. and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, **37** 36–48. URL http://dx.doi.org/10.2307/2685844.

GLADWELL, M. (2008). Most likely to succeed. *The New Yorker*. URL http://www.newyorker.com/reporting/2008/12/15/081215fa_fact%_gladwell.

GREENE, W. H. (2008). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. *Social Science Research Network Working Paper Series*. URL http://ssrn.com/abstract=1293115.

LEWIN, D. (2006). College quarterbacks through the prism of statistics. URL http://www.footballoutsiders.com/stat-analysis/2006/college-quarterbacks-through-prism-statistics.

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. 2nd ed. Wiley-Interscience. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471183865.

MASSEY, C. and THALER, R. H. (2010). The loser's curse: Overconfidence vs. market efficiency in the national football league draft. Tech. rep., Social Science Research Network.

QUINN, K. G., GEIER, M. and BERKOVITZ, A. (2007). Passing on success? productivity outcomes for quarterbacks chosen in the 1999-2004 national football league player entry drafts. Tech. Rep. 0711, International Association of Sports Economists. URL http://ideas.repec.org/p/spe/wpaper/0711.html.

SCHATZ, A. (2005). Football's hilbert problems. *Journal of Quantitative Analysis in Sports*, **1**. URL http://ideas.repec.org/a/bpj/jqsprt/v1y2005i1n2.html.

YAU, K. K. W., WANG, K. and LEE, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, **45** 437–452. URL http://dx.doi.org/10.1002/bimj.200390024.

CORRESPONDING AUTHOR:
JULIAN WOLFSON
DIVISION OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA
A460 MAYO BUILDING,
MMC 303
420 DELAWARE ST. S.E.
MINNEAPOLIS, MN 55455
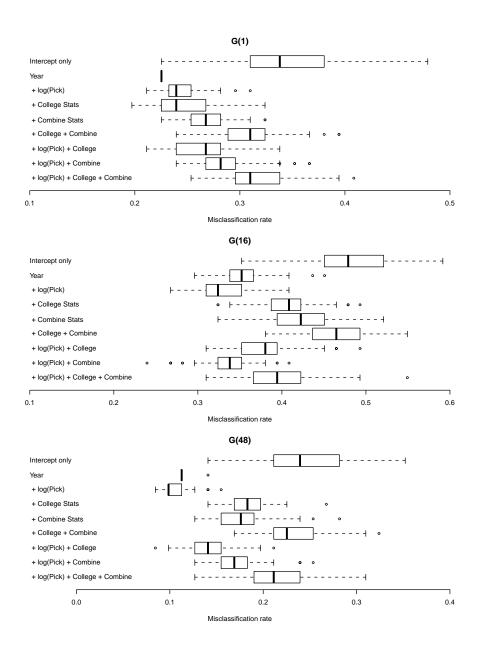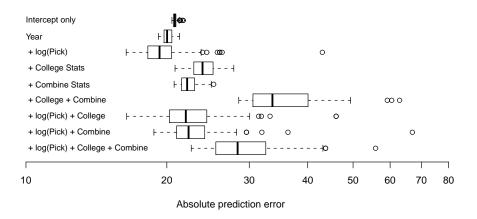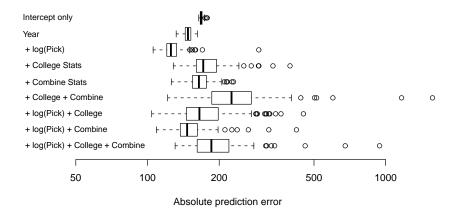E-MAIL: julianw@umn.edu

FIG 3. *Misclassification rates from 100 runs of 5-fold cross-validation*

FIG 4. *Absolute prediction error estimates for NFL games played from 100 runs of 5-fold cross-validation*

FIG 5. *Absolute prediction error estimates for Net Points from 100 runs of 5-fold cross-validation*