Expected Hypothetical Completion Probability

Sameer Deshpande* and Katherine Evans†

22 January 2019

1 Introduction

Consider two passing plays during the game between the Los Angeles Rams and visiting Indianapolis Colts in the first week of the 2017 season. The first passing play was a short pass in the first quarter from Colts quarterback Scott Tolzien intended for T.Y. Hilton which was intercepted by Trumaine Johnson and returned for a Rams touchdown. The second passing play was a long pass from Rams quarterback Jared Goff to Cooper Kupp, resulting in a Rams touchdown. In this work, we consider the question: which play had the better route(s)?

From one perspective, one could argue that Kupp's route was better than Hilton's; after all it resulted in the offense scoring while the first play resulted in a turnover and a defensive score. However, "resulting", or evaluating a decision based only on its outcome is not always appropriate or productive. Two recent examples come to mind: Pete Carroll's decision to pass the ball from the 1 yard line in Super Bowl XLIX and the "Philly Special" in Super Bowl LII. Had the results of these two plays been reversed, Pete Carroll might have been celebrated and Doug Pederson criticized.

If evaluating plays solely by their outcomes is inadequate, on what basis should we compare routes? Intuitive, we might tend to prefer routes which maximize the receiver's chance of catching the pass, or completion probability. If we let y be a binary indicator of whether a pass was caught and let $\mathbf{x}$ be a collection of covariates summarizing information about the

*CSAIL MIT, email: skdeshpande91@gmail.com
†Verily Life Sciences, email: causalkathy@gmail.com

pass, we can consider a logistic regression model of completion probability:

$$\log\left(\frac{\mathbb{P}(y=1|\mathbf{x})}{\mathbb{P}(y=0|\mathbf{x})}\right) = f(\mathbf{x}), \tag{1}$$

or equivalently $\mathbb{P}(y=1|\mathbf{x}) = \left[1+\mathrm{e}^{-f(\mathbf{x})}\right]^{-1}$, for some unknown function f.

If we know the function f, a first pass at assessing a route would be to plug in the relevant covariates $\mathbf{x}$ and see whether the forecasted completion probability exceeded some threshold, say 50%. If so, regardless of whether the receiver actually caught the actual pass, we could say that the route was run and ball was placed in such a way as to give the receiver a better chance than not of catching the pass. We could then directly compare the forecasted completion probabilities of the two plays mentioned above; if it turned out that the Tolzien interception had a higher completion probability than the Kupp touchdown, that play would not seem as bad, despite the much worse outcome. In a light of the fact there are often multiple eligible receivers running routes on a play, such a comparison is not completely satisfactory since it focuses only on a single player's chance of successfully catching a specific pass thrown to a single location along his route. Such a comparison does not, in particularly, answer the very natural follow-up question: was there another location along a possibly *different* receiver's route where the completion probability was higher? If so, one could argue that the quarterback ought to have thrown the ball to that spot.

Evaluating the completion probability at an arbitrary location along a different receiver's route is challenging for several fundamental reasons. First, even if we knew the true function f, we are essentially trying to deduce what might have happened in a *counterfactual* world where the quarterback had thrown the ball to a different player at a different time, with the defense reacting differently. On such a counterfactual pass, we do not observe many factors that are predictive of completion probability. Figure 1 illustrates this issues, showing schematics for an observed pass (left panel) and a hypothetical pass (right panel). In both passes, there are two receivers running routes; we have colored the route of the intended receiver on both passes blue and the route of the other receiver in gray.

Before proceeding, we pause for a moment to distinguish between our use of the term "counterfactual" and its use in causal inference. The general causal framework of counterfactuals supposes that we change some treatment or exposure variable and asks what happens to downstream outcomes. In contrast, in this work, we considering changing a midstream variable, the location of the intended receiver when the ball arrives, and then impute both upstream and downstream variables like the time of the pass and the receiver separation at the time the ball arrives. In this work, we use "counterfactual" interchangeably with "hypothetical" and hope our more liberal usage is not a source of further confusion below[1].

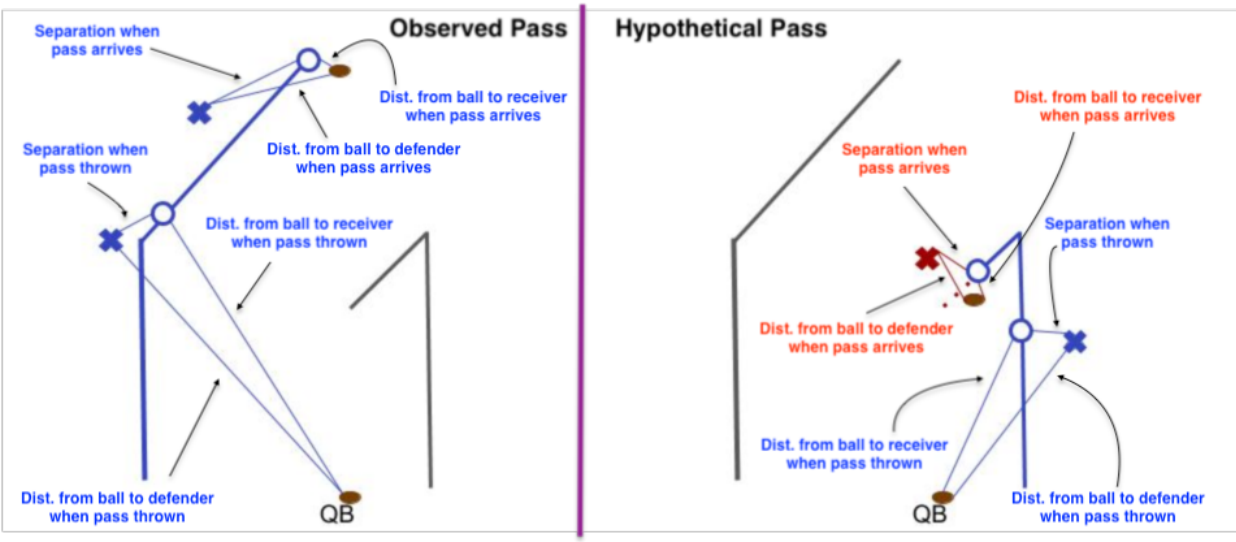

Figure 1: Schematic of what we directly observe on an actual pass (left panel) from our dataset and what we cannot observe for a hypothetical pass (right panel). In both passes, there are two receivers running routes.The targeted receiver is denoted with a circle and the defender closest to the receiver is denoted with an X. Unobservables are colored red while observables are colored blue.

For the observed pass, we directly observe all possible information about the pass including, for instance, the receiver's separation at the time the pass is made and at the time that the pass arrives. We similarly directly observe the receivers relative position to the ball when the pass arrives. In contrast, for the hypothetical play, we cannot directly observe the receiver's separation at the time the pass arrives nor can we observe his relative position to the ball at the same time.

Beyond the fact that we do not observe certain characteristics of the hypothetical pass, we typically do not know the true regression function f and must therefore estimate it using

[1]Author's note: We use the word "counterfactual" interchangeably with "hypothetical" because while an unobserved pass is hypothetical, the intended receiver of that pass is not.

the data. In doing so, estimation uncertainty about f propagates to the uncertainty about the hypothetical completion probabilities. We argue that an objective assessment of routes based on a completion probability must address the inherent uncertainty in the hypothetical inputs as well as uncertainty stemming from estimating the completion probability model.

In this work, we aim to overcome these challenges. Using tracking, play and game data from the first 6 weeks of the 2017 NFL season, we developed such an assessment, which we call Expected Hypothetical Completion Probability (EHCP). At a high-level, our framework consists of two steps. First, we estimate the log-odds of a catch as a function of several characteristics of each observed pass in our data. Then, we simulate the characteristics of the hypothetical pass that we do not directly observe and compute the average completion probability of the hypothetical pass. The rest of this paper is organized as follows. In Section 2, we describe our Bayesian procedure for fitting a catch probability model like in Equation (1) and outline the EHCP framework. We briefly discuss the results of our catch probability model and illustrate the EHCP framework on several routes in Section 3. We conclude with a discussion of potential methodological improvements and refinements and potential uses of our EHCP framework.

2 Methods

2.1 Estimating Completion Probability

The first step of the EHCP framework is to estimate completition probability. Table 1 lists several characteristics about the pass that we will use to predict completion probability.

Table 1: Covariates used in our completion probability model. † = for these distance measures, we include the distance in the x and y direction in adding to the Euclidean distance. Δ represents the change in the variable while the pass is in the air. Rec. = receiver, Def. = nearest defender to receiver, Dist. = Distance

Observed at time of pass		Observable when pass arrives	
Time from snap to pass		Time from pass to play	Total time
Separation		Separation	Δ Separation
Rec. Speed		Rec. Speed	Δ Rec. Speed
Rec. Direction		Rec. Direction	Δ Rec. Direction
Ball – Rec. Dist.†	Ball – Def. Dist†	Ball – Rec. Dist†	Ball – Def. Dist.†
Cumulative Rec. Dist. Travelled		Cumulative Rec. Dist. Travelled	Δ Cumulative Rec. Dist. Travelled
		Cumulative Def. Dist. Travelled	

In addition to these covariates, we also included the number of seconds left in the half, down, yards to go, whether the offensive team is leading, and a categorical variable summarizing how many scores the offensive team is leading or trailing by (9+ points or more, 1 – 8 points, or 0 points).

Since we might reasonably expect that the log-odds of completing a pass is not linear in the covariates listed in Table 1, we use Bayesian Additive Regression Trees (BART) to estimate f. BART is a non-parametric regression technique that expresses f as the sum of several regression trees, each of which recursively partitions the covariate space and approximates the value of f over these smaller partitions. This is somewhat similar to random forests, which is also an ensemble-of-trees method. BART has shown great acuity in detecting non-linear and interaction effects, *without pre-specification of what those effects may be.*

BART begins by specifying a prior $\pi(f)$, meant to reflect all of our initial uncertainty about the unknown function f. We then update this prior distribution with the data to get the posterior distribution $\pi(f|\mathbf{y})$ using Bayes' theorem: $\pi(f|\mathbf{y}) \propto \pi(f)p(\mathbf{y}|f)$ where $p(\mathbf{y}|f)$ is the *likelihood* implied by the logistic model in Equation (1). Typically, this posterior distribution is not analytically tractable and we use a Markov Chain Monte Carlo (MCMC) simulation to generate draws $f^{(1)}, \ldots, f^{(N)}$ from it. For a review of Bayesian tree-based methods, please see Linero (2017) and for further details about the BART prior and MCMC procedure, please see Chipman et al. (2010). For a hypothetical pass with characteristics $\mathbf{x}^{\star}$, we can approximate the posterior predictive completion probability using these draws:

$$\mathbb{P}(y^{\star} = 1|\mathbf{x}^{\star}) = \frac{1}{N}\sum F^{(n)}(\mathbf{x}^{\star}) = \frac{1}{N}\sum_{n=1}^{N}\left[1 + e^{-f^{(n)}(\mathbf{x}^{\star})}\right]$$

where the $F^{(n)}(\mathbf{x}^{\star})$'s are draws from the posterior distribution of the forecasted completion probability.

2.2 Simulating Unobserved Covariates

As alluded to in Section 1 and Figure 1, when we consider hypothetical passes, we must account for the uncertainty in the covariates that summarize what happens after the pass was thrown. We do not know, for instance, how far a different receiver would have been from a defender had the ball been thrown to him. Similarly, we do not know far that receiver would be from the ball when he attempts to catch it. For each counterfactual pass, we first divide the covariates into two groups: those which we directly observe and those about which we are uncertain. Formally, let $\mathbf{x}^\star = (\mathbf{x}^\star_{\text{obs}}, \mathbf{x}^\star_{\text{miss}})$ be the partition of the counterfactual covariates into the observed and missing data. We propose to sample the values in $\mathbf{x}^\star_{\text{miss}}$ from the empirical distribution. For instance, since we cannot observe the vector from the receiver to the ball when the hypothetical pass arrives, we randomly sample this vector from the collection of all such vectors we actually observe in the dataset. So if we knew the true value of f, the log-odds of completion function, we could approximate

$$\text{EHCP}(\mathbf{x}^\star_{\text{obs}}) = \mathbb{E}_{\mathbf{x}^\star_{\text{miss}}}[F(\mathbf{x}^\star_{\text{obs}}, \mathbf{x}^\star_{\text{miss}})] \approx \frac{1}{M}\sum_{m=1}^{M} F(\mathbf{x}^\star_{\text{obs}}, \mathbf{x}^{\star(m)}_{\text{miss}}), \tag{2}$$

where $\mathbf{x}^{\star(1)}_{\text{miss}}, \ldots, \mathbf{x}^{\star(M)}_{\text{miss}}$ are the draws of $\mathbf{x}_{\text{miss}}$ from the empirical distribution, $F(\cdot) = \left[1 + e^{-f(\cdot)}\right]^{-1}$ is the forecasted completion probability function, and the expectation is taken over the empirical distribution of $\mathbf{x}^\star_{\text{miss}}$. Rather than setting the value of $\mathbf{x}^\star_{\text{miss}}$ at some arbitrary fixed quantity, EHCP averages over the uncertainty in the unknown (and unobservable) values of $\mathbf{x}^\star_{\text{miss}}$. Importantly, since we are sampling the values of $\mathbf{x}^\star_{\text{miss}}$ from the set of values actually observed, EHCP is constructed using realistic values of the missing covariates.

Since we do not know f exactly but instead have only our MCMC samples, we can approximate EHCP for each posterior draw of f, thereby simulating draws from the posterior distribution of $\text{EHCP}(\mathbf{x}^\star_{\text{obs}})$. We can then report the posterior mean as a point estimate of the true EHCP on the hypothetical pass and also report the 95% interval, containing likely values of the EHCP. We can further consider all of the routes run on a given play and track these two quantities as the play develops to see which receiver-route combinations have the highest chance of pass completion.

3 Illustration

To illustrate our proposed framework, we return to the two plays from the introduction, the Kupp touchdown and the Tolzien interception.

3.1 Completion Probability Model

Figure 2 shows the histogram of the posterior draws of the forecasted completion probability F for the Kupp touchdown (blue) and the Tolzien interception (red). We see that there is substantial overlap in the bulk of these posterior distributions but the posterior for the Kupp touchdown is shifted slightly to the right of posterior for the Tolzien interception. Interestingly, on both of the these throws the receiver had less than 50% chance of catching the ball, with the posterior mean completion probability on the Kupp touchdown approximately 10 percentage points higher than the probability for the Tolzien interception (47% vs 37.1%).

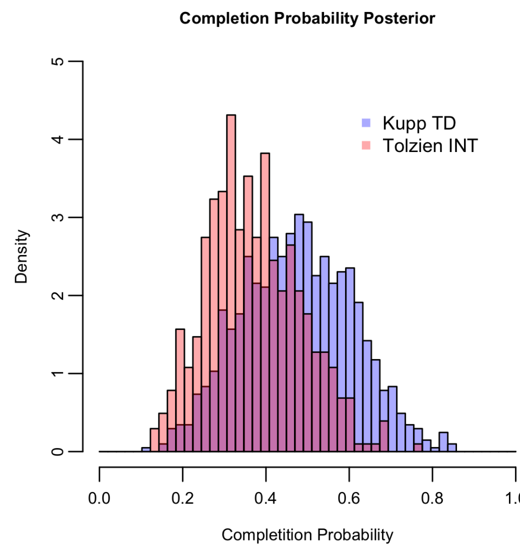

Figure 2: Histogram of posterior draws of completion probabilities for the Kupp touchdown (blue) and the Tolzien interception (red)

3.2 How EHCP Evolves Over A Route

Figure 3 shows the histogram of the posterior EHCP draws for Kupp and Hilton (the intended target on the Tolzien interception) at the times that the two passes actually arrived. As before, the posterior for the Kupp touchdown is shifted slightly to the right of the Tolzien interception. We find that the posterior mean EHCP for the Kupp touchdown is just around six percentage points higher than the posterior mean EHCP for the Tolzien interception (65.1% vs 59.0%).

That the EHCP and forecasted completion probabilities are somewhat different is not surprising, as they measuring two different quantities: the forecasted completion probability model uses the exact information about what actually happened after the ball was thrown while EHCP averages over the uncertainty in what might have happened after the ball was thrown. We also note that often EHCP posteriors seem to have less variance than the posterior completion probability. This is also not surprising; EHCP represents an *average* probability over several possible realizations of the pass while the forecasted completion probability considers only a single pass. In a certain sense, because EHCP averages over many passes, it somewhat mitigates uncertainty introduced in our estimation of f.

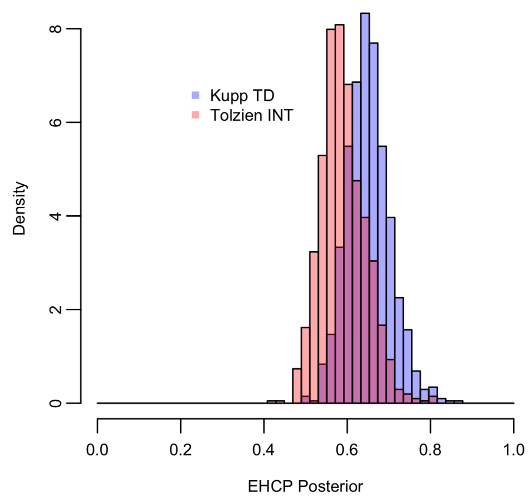

Figure 3: Histogram of posterior draws of EHCP for Kupp touchdown (blue) and the Tolzien interception (red)

While comparing the EHCP for the two receivers actually targeted in the two plays at the times that the actual passes arrived is interesting, the real power of EHCP lines in projecting what might have happened had the ball been delivered to other receivers earlier in the play. Figures 4 and 5 show the posterior mean of the EHCP for each receiver at various points in his route for the Kupp touchdown and Tolzien interception.

We see that Kupp's posterior mean EHCP at the time the actual pass arrived (location A in the figure) was 65.1%. Almost two seconds earlier, however, his posterior mean EHCP was 85.1% (location B in the figure). Looking at the full posterior distributions of the EHCP at these two locations, we find that the 95% intervals are nearly disjoint. So we may conclude with reasonable certainty that Kupp's EHCP would have been higher had the pass been delivered earlier along his route.

Even more interesting, we find that of all of the receiver during this play, Sammy Watkins actually had the highest posterior mean EHCP 1.5 seconds after the snap (92.2% at location C). At that time, Kupp's posterior mean EHCP was 91.9% and his 95% interval was (85.5%, 96.8%), virtually identical to Watkins'. Our analysis suggests that while the actual play resulted in a touchdown, there were times earlier in the play where the receivers would have had substantially larger expected completion probabilities. That being said, there are many reasons that the pass was not actually thrown to Watkins at location C. We will return to this point in Section 4.

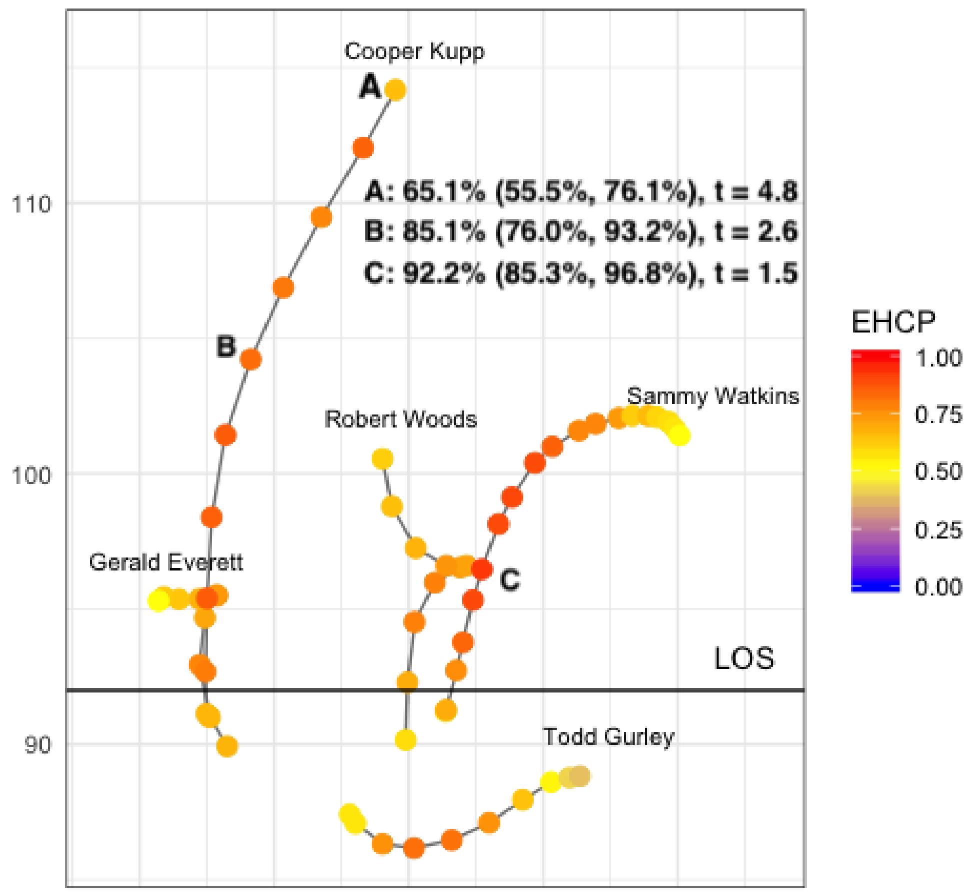

Figure 4: Posterior mean EHCP for each receiver on the Kupp touchdown. 95% posterior intervals are shown in parentheses. t lists the time in seconds after the snap

Turning our attention to the the Tolzien interception, we find that T.Y. Hilton, the targeted receiver, had an EHCP of 59.0% at the time the actual pass arrived (location A in the figure).

Similar to the Kupp touchdown, almost two seconds earlier, his EHCP was substantially higher (89% at location B). Further, Donte Moncrief had the highest EHCP of all receivers at location C, 2.4 seconds after the snap. The substantial overlap in the 95% intervals for Hilton and Moncrief at this time means that we cannot tell with much certainty which of the two receivers had the higher EHCP.

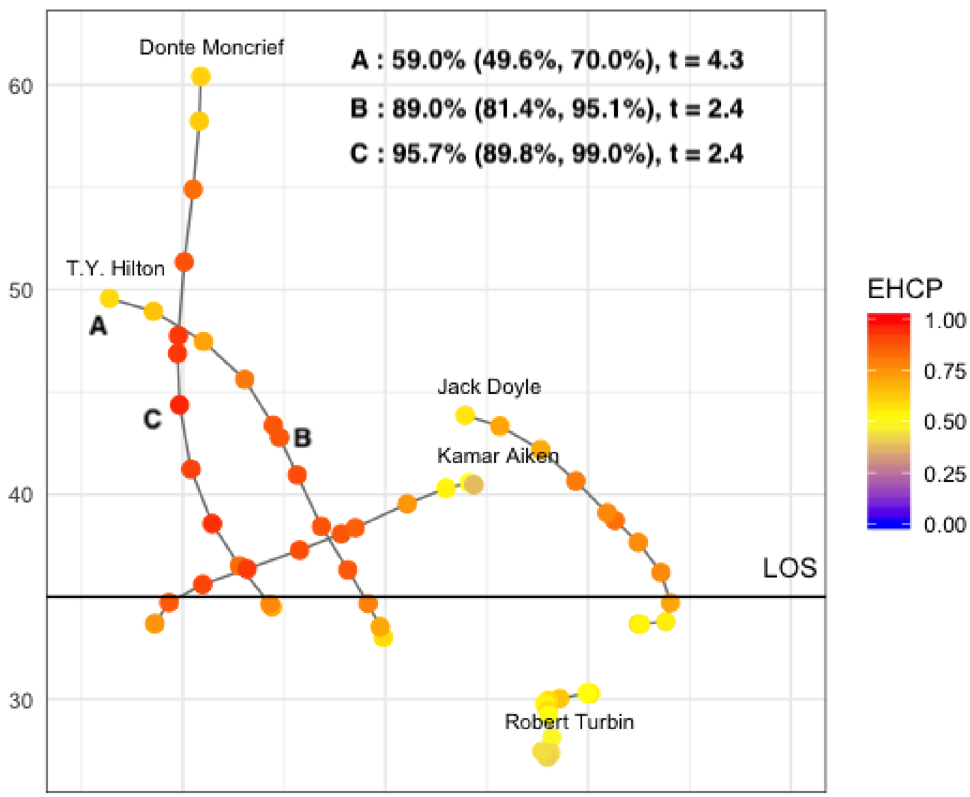

Figure 5: Posterior mean EHCP for each receiver on the Tolzien touchdown. 95% posterior intervals are shown in parentheses. t lists the time in seconds after the snap

We do note, however, that they are very close to one another on the field, which could partially explain the similarity in EHCP at that point in time. It is interesting to note that the posterior mean EHCPs at the time the pass actually arrived to Hilton (4.3 seconds after the snap) hovered between 40% and 60% for all receivers on the field. This would seem to suggest that the entire play design itself may not have been optimal.

4 Discussion

As presented here, EHCP provides an objective way to evaluate offensive plays retrospectively. Specifically, we can track how the completion probability evolves for each receiver over the course of a play in a way that accounts for the uncertainty about missing covariates. The EHCP framework can also be used prospectively. A defensive coordinator might, for instance, ask how best to cover a particular set of routes being run. She may fix some of the unobserved covariates like the defender's position relative to targeted receiver and then average over the uncertainty in the remaining covariates to derive the EHCP for that particular combination of receiver-defender positioning. Repeating this for various defender locations would enable her to construct optimal defender trajectories that minimize the intended receiver's EHCP.

Our completion probability model and the EHCP framework can also be used to provide more nuanced broadcast commentary. In particular, if there was a play where the forecasted completion probability and EHCP were high and the receiver failed to the catch the ball, one may reasonably assign some amount of blame to the receiver for not catching the ball; after all, the route was run and the ball was delivered to give him a high probability of catching it. On the other hand, if the receiver catches a ball with very low forecasted completion probability and EHCP, it would be worthwhile to point out that receiver is succeeding despite the route design and pass delivery. Finally, one could aggregate the discrepancy between outcome and EHCP over all of a receiver's targeted routes to measure how the receiver is executing his assigned routes.

We note that the NFL's Next Gen Stats include a Completion Probability metric that is similar to our forecasted completion probability but uses different input variables than us. Notably, Completion Probability includes a number of quarterback-centric features such as speed of and distance to the nearest pass rusher at the time of the throw. As a result, Next Gen Stats frames Completion Probability primarily as a quarterback statistic whereas EHCP is focused more on the receiver and route. Since quarterback pressure affects where the pass ends up (e.g. if it is over- or under-thrown), EHCP accounts for it rather indirectly in averaging over the uncertainty in the ball's position relative to the receiver. That said, incorporating variables about the delivery of observed passes directly into the completion probability model is straightforward as is simulating the unobserved values of these variables for counterfactual passes in the EHCP calculation. Doing so would result in an EHCP that better accounts for why balls were thrown when they were and would enable more nuanced assessment of the hypothetical passes. We hope that our method, and our transparency about how we developed it, will facilitate further iterations that combine information about the quarterback and all receivers.

There are several potential areas of methodological and modeling improvement. It is quite straightforward to include more covariates about the individual players involved in the pass completion model. Though we have not done so in this paper, we could use a variant of BART (Linero, 2018) that performs variable selection to identify the main drivers of successfully completing passes. Similarly, while we have focused on completion probability, we can construct analogous measures by considering different outcomes. For instance, we can track whether the play resulted in a first down and derive the expected hypothetical first down probability.

More substantively, we could develop a more sophisticated imputation model of $\mathbf{x}^{\star}_{\text{miss}}$. In the present paper, we have taken by far the simplest approach and sampled $\mathbf{x}^{\star}_{\text{miss}}$ from the observed distribution from *all passes* in our dataset. It would be interesting to construct predictive models of $\mathbf{x}^{\star}_{\text{miss}}$ using the observed covariates $\mathbf{x}^{\star}_{\text{obs}}$ and to feed forecasts from these models into the EHCP calculation in Equation (2).

References

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266 – 298.

Linero, A. R. (2017). A review of tree-based Bayesian methods. Technical report, Florida State University.

Linero, A. R. (2018). Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, 113(522):626–636.