Optimum Pass/Run Ratio in NFL

CMSAC Reproducible Research Competition

October 2020

Abstract

NFL passing plays have higher expected points added (EPA) than running plays in almost every game situation. A recent article by Lau Sze Yui[1] concludes, however, that passing on every play is not viable. Nevertheless, Yui's article suggests that the correct passing equilibrium is far from current run/pass ratios.

The relationship between EPA and Win Probability (WP) is the starting point for determining the optimal pass-run percentage. At low WPs, run plays have their highest EPAs and even have higher EPAs than pass plays. Presumably, this is mostly because run plays are unexpected in come-from-behind situations. Meanwhile, it is predictable for an offense to pass when its WP is low, making pass plays less effective.

The break-even WP—where pass plays start to have a higher EPA than run plays—was 13% in 2019. Theoretically, at this WP, pass plays became sufficiently unpredictable such that passing was more effective than running. Pass play percentage decreases as WP increases. Teams that are winning run more, and teams that are losing pass more. At the break-even WP of 13%, pass play percentage was 73%.

Maximizing passing is desirable because passing almost always provides more EPA than running. Teams, however, would not want to pass so often to make passing too predictable. Passing is too predictable at WPs less than 13% because running provides a higher EPA at these WPs. Teams' 2019 passing rates were more than 73% at these WPs. The optimum pass percentage in 2019, therefore, should have equaled, but not exceeded, 73%, to provide sufficient unpredictability.

In 2015–2018, optimal pass play percentages were 76%–83%. The actual pass play percentages, however, were 61%–62% in 2015–2019.

I. Data Set

NFL play-by-play data from `nflscrapR` was used for this research. Specifically, Ron Yurko provides play-by-play data for all regular seasons since 2009 in CSV files on GitHub.[2]

From this play-by-play data, the following attributes[3] were selected for inclusion in the data set:

> `posteam` (Team with possession)
> `down`
> `ydstogo`
> `play_type`

---

[1] Yui, L.S.(2020, March 23). *Finding the optimal pass/run ratio.* Football Outsiders. http://www.footballoutsiders.com/stat-analysis/2020/finding-optimal-passrun-ratio.

[2] Yurko, R. (2020, January 3). *NflscrapR regular season play-by-play data.* GitHub. http://github.com/ryurko/nflscrapR-data/tree/master/play__by__play__data/regular__season.

[3] Descriptions of all attributes in the nflscrapR play-by-play data set are available at: Horowitz, M. (n.d.). *Scrape JSON play-by-play.* RDocumentation. http://www.rdocumentation.org/packages/nflscrapR/versions/1.8.3/topics/scrape__json__play__by__play.

```
qb_kneel
qb_spike
qb_scramble
run_location (left, middle, or right)
run_gap (end, guard, or tackle)
epa (Expected Points Added for team with possession)
wp (Win Probability for team with possession at start of play)
third_down_converted
fourth_down_converted
sack
```

The `play_type` attribute classifies each play as a `pass` (including sacks) or a `run` (including scrambles). Other `play_type` indicators include: `punt`, `field_goal`, `kickoff`, `extra_point`, `qb_kneel`, `qb_spike`, and `no_play` (timeouts and penalties). The `play_type` is blank for rows indicating the end of play.

The data was filtered by `play_type` so that only `pass` plays and `run` plays were included. Also, 2-point conversion attempts (which have a `down`=NA), and any plays that have NAs for `epa` or `wp` were removed from the data set.

The `nflscrapR` data provides `run_location` and `run_gap` information that allows for an evaluation of running play success by run direction. Recent research has examined which run play locations have been most successful. In *The NFL's Run Gap Secret*,[4] Tej Seth shows that runs outside the tackles result in higher EPAs than inside runs. Similarly, in *Big, Useful Data Says Outside Rushes are the New Market Inefficiency*,[5] Dan Morse shows that rushing outside has been slightly more successful than rushing up the middle, except for short yardage situations (less than 5 yards to go).

To confirm and apply these previous findings, the data was split into three types of plays: outside runs, middle runs, and pass plays. "Middle runs" include runs having a `run_location` of `middle` or a `run_gap` of `guard`. Running plays toward the `run_gap` of `tackle` or `end` were classified as "outside runs."

Running plays having a `run_location`=NA are not included in the data set, as these plays are generally aborted plays such as fumbles by the quarterback at the snap. Also, because quarterback scrambles are often intended to be pass plays, scrambles were redefined as "pass plays."

II. Expected Points Added (Pass Plays vs. Running Plays)

As shown in the following table, on average in 2019, passing plays gained 0.07 EPA/play, while running plays to the outside lost 0.01 EPA/play, and running plays up the middle lost 0.06 EPA/play.
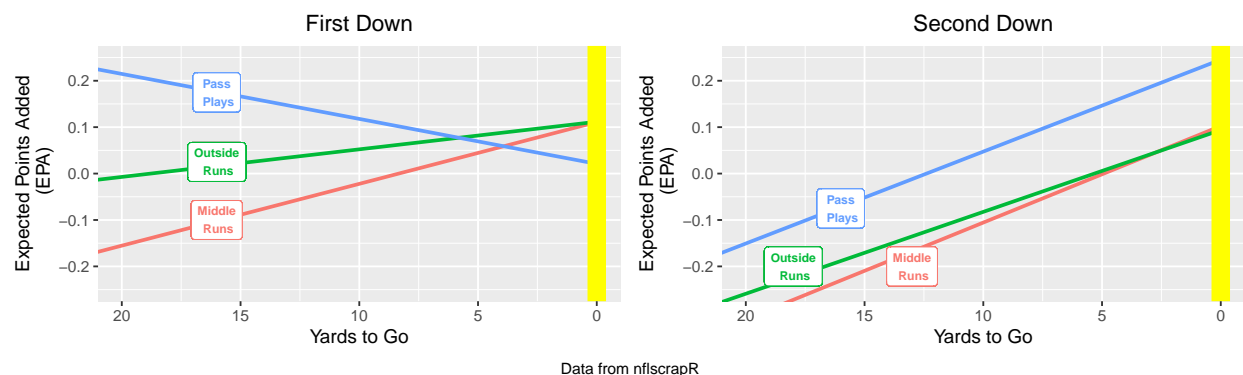
Table 1: EPA/play in 2019.

	EPA/play (2019)
Pass Plays	0.0686
Outside Runs	-0.0122
Middle Runs	-0.0578

Summarizing the data at each down and distance scenario gives further insight into the comparative values of running and passing. On first and second downs, all plays (whether or not a first down was converted) are relevant. The following graphs are linear regression models that show how average EPA varied based on yards-to-go on first and second down for running and passing plays in 2019.

[4]Seth, Tej. (2020, July 15). *The NFL's run gap secret.* M-Fans. The Michigan Football Analytics Society. http://mfootballanalytics.com/2020/07/15/the-nfls-run-gap-secret.

[5]Morse, D. (2019, October 22). *Big, useful data says outside rushes are the new market inefficiency.* Sports Illustrated. http://www.si.com/nfl/washingtonfootball/gm-report/epa-week-8-nfl-power-rankings-where-to-run.

Figure 1: EPA by Yards-to-Go & Play Type on 1st & 2nd Downs (2019).


In these graphs, the yellow line represents the first down marker. In 2019, passing plays almost always had higher EPAs than running plays on first and second downs. For the specific data point of first down and 10 yards to go (not shown), on average, passing plays obtained an EPA of 0.109, running outside obtained an EPA of -0.021, and running up the middle obtained an EPA of -0.095.

One exception where running plays obtain higher EPAs than passing plays is at first down and less than 6 yards to go (for example, 1st and 5, after a penalty). At this point, running became more effective than passing on first down.

Generally, outside runs provided higher EPAs than running up the middle. Running up the middle, however, becomes as effective as running outside as the offense gets closer to a first down. On second down, running up the middle and running outside became equally effective at 3 yards to go and less.

Next, in comparing third and fourth downs, it is best to determine the effectiveness of passing plays and running plays only in cases where a first down conversion was successful. Again, the following linear regression models show how average EPA varied based on yards-to-go on third and fourth down conversions for running and passing plays in 2019.

Figure 2: EPA by Yards-to-Go & Play Type for 3rd & 4th Down Conversions (2019).
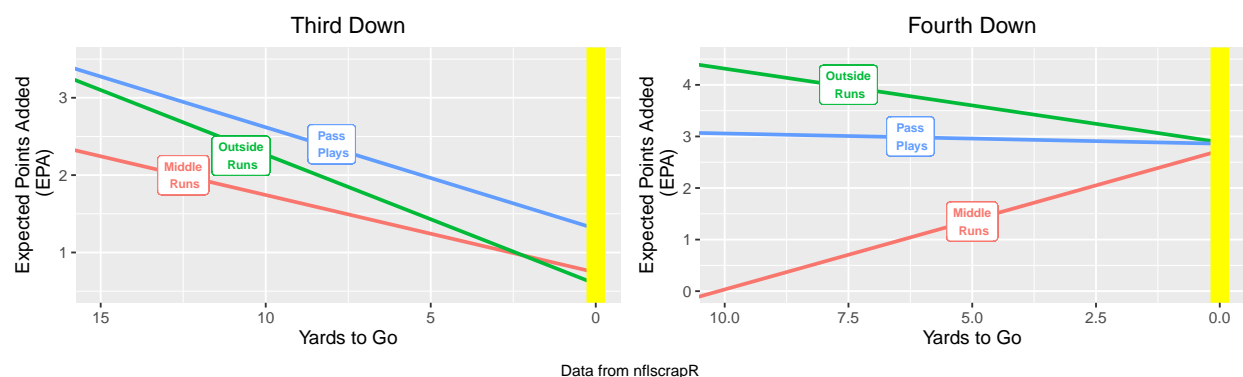
On third down, passing plays continued to get higher EPAs than running plays. Outside runs were more effective than middle runs until about 2½ yards to go.

On fourth down, it appears at first that outside running was more effective than passing. These results, however, may driven by a small sample size because there were only 34 outside runs that resulted in fourth down conversions in 2019. Furthermore, the overall average of pass plays' EPA for fourth down conversions was 3.01, which is higher than the overall average of outside running plays' EPA of 2.95.

The R^2 values for each of the linear regression models are as follows:

Table 2: R^2 Values.

Play Type	1st Down	2nd Down	3rd Down Conversions	4th Down Conversions
Pass Plays	0.0541	0.3326	0.8089	0.0254
Outside Runs	0.0115	0.3303	0.4327	0.4161
Middle Runs	0.1335	0.3617	0.3865	1.0

These R^2 values suggest that overall, the models best describe second and third down situations. As discussed above, there are insufficient sample sizes at fourth down in a single season to produce an accurate fit. For example, the R^2 value of 1 for fourth down conversions on runs up the middle is a result of only having two data points for successful fourth down conversions resulting from runs up the middle (one at 4th and 1 and one at 4th and 2).

Similarly, it is uncommon to have first down and something other than 10 yards to go, so the first down model likewise suffers from a lack of sufficient data for situations other than 1st and 10.

Key Point 1

- *In 2019, passing plays had higher EPAs than running plays in almost every game situation.*
  - *The only exception was on first down at less than 6 yards to go, where running plays had higher EPAs than passing plays.*

Key Point 2

- *When running the ball in 2019, running to the outside resulted in higher EPAs than running up the middle, except in short yardage situations on second and third downs.*
  - *On second down, running up the middle became as or more effective than running outside at 3 yards to go or less.*
  - *On third down conversions, running up the middle became as or more effective than running outside at 2½ yards to go or less.*

Key Point 3

- *There is insufficient data to fully support the overall models for first and fourth downs. However, the 2019 data does show that:*
  - *On first down and 10 yards to go, passing plays obtained 0.13 more EPA than running outside, and 0.20 more EPA than running up the middle.*
  - *On fourth down, pass plays obtained a higher EPA on average than running plays.*

III. Pass/Run Ratios in 2019

For the 2019 data set, there were 19,717 passing plays and 12,087 running plays. This corresponds to a passing play percentage of 62% and a pass/run ratio of 1.63.

This is a higher pass/run ratio than reported on *Pro-Football-Reference.com* and *NFL Ops Stats Central*. *Pro-Football-Reference.com* indicates that there were 17,853 league-wide passing attempts,[6] and 13,387 rushing attempts.[7] Likewise, *NFL Ops Stats Central* indicates that there were 54.9% passing plays and 41.2% rushing

---

[6]NFL season by season passing (2020). Pro Football Reference. http://www.pro-football-reference.com/years/NFL/passing.htm.

[7]NFL season by season rushing & receiving (2020). Pro Football Reference. http://www.pro-football-reference.com/years/NFL/rushing.htm.

plays in 2019.[8] Each of these statistics provides a pass/run ratio of 1.33 (17,853/13,387 = 1.33 and 54.9/41.2 = 1.33).

The pass/run ratio of 1.63, however, classifies each `qb_scramble` and `sack` as "pass plays." There were 781 scrambles and 1,268 sacks included in the passing play total. Also, the data subset excludes each `qb_kneel` and `qb_spike` because `nflscrapR` does not classify them as either a `pass` or a `run`. There are 399 kneel-downs and 75 spikes excluded from the data.

Making these adjustments would result in a pass/run ratio of 1.3. Specifically, the number of pass plays would be:

$$19,717 - 781 \; scrambles - 1,268 \; sacks + 75 \; spikes = 17,743.$$

The number of run plays would be:

$$12,087 + 781 \; scrambles + 399 \; knees = 13,267.$$

Then, the pass/run ratio would be:

$$\frac{17,743 \; passes}{13,267 \; runs} = 1.34.$$

Note that the total numbers of pass plays and run plays do not include penalties or aborted plays, and are less than the totals reported on *Pro-Football-Reference.com.* Nevertheless, the pass/run ratio of 1.3 is equal to that reported in *Pro-Football-Reference.com* and *NFL Ops Stats Central.*

The pass/run ratio of 1.63, however, is a better measure, because it is designed to represent the play call, not the play result. That is, sacks and quarterback scrambles are intended to be passing plays even if they are not recorded as pass attempts. Also, penalties are excluded so that the yardage and EPA values on each play are not skewed by penalty yardage.

Key Point 4

- *In 2019, teams called 62% passing plays.*
  - *This percentage classifies sacks and scrambles as passing plays, and does not include spikes, kneel-downs, penalties, or aborted plays.*

IV. Relationship Between WP & EPA

Ben Baldwin recently compared the relationship of EPA/play to WP for passing and running plays.[9]

(Cont'd on Next Page)

---

[8]Chart the data (2020). NFL Football Operations. http://operations.nfl.com/stats-central/chart-the-data. Use the interactive dropdown menus to compare % Rushing Plays to % Passing Plays.

[9]Baldwin, B. [@benbbaldwin]. (2019, December 16). *Expected Points Added (EPA) per play vs Win Probability* [Tweet]. Twitter. http://twitter.com/benbbaldwin/status/1206640759899918336.
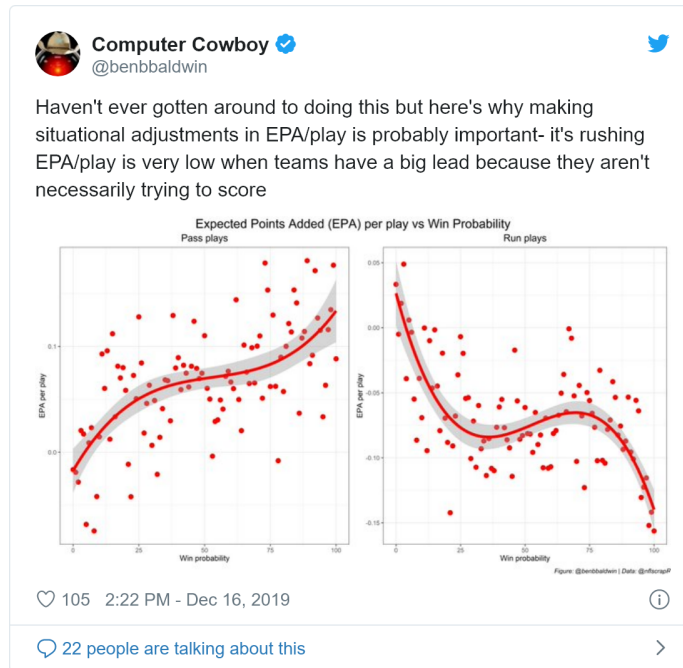
It is also helpful to see the running and passing curves on the same plot as follows.

*Figure 3: WP vs. EPA for Passing & Running Plays (2019).*
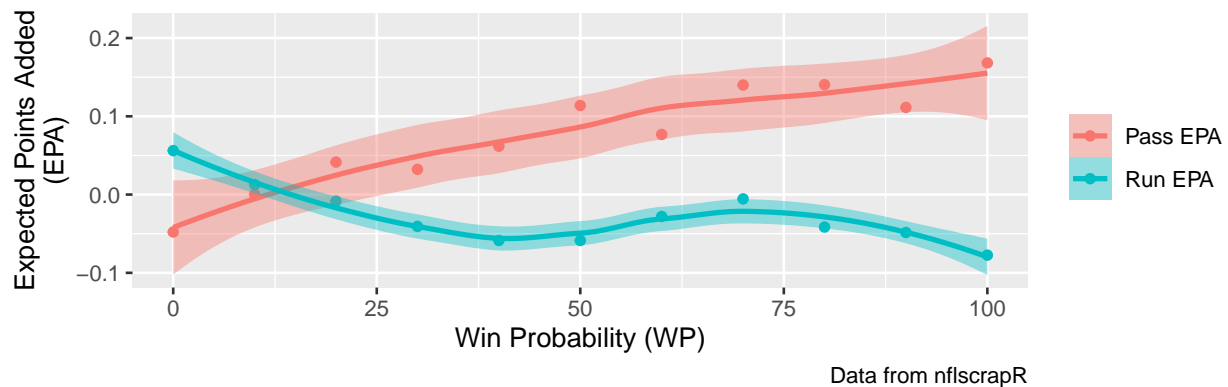
In this plot, each WP is rounded to the nearest 10. This allows for discrete WP data points at which the EPAs are averaged. Using multiples of 10 also provides larger sample sizes at each WP data point, and reduces the noise of the plot by only including 11 data points.

The resulting plot shows that at low WPs—when passes are expected—running plays have higher EPAs than passing plays. The two lines intersect at the "break-even point"—the WP where passing gets higher EPAs than running.

For 2019, the intersection point was at a WP of 13%. In other words, passes are predictable at a win probability less than 13% and consequently have lower EPAs than run plays.

*Figure 4: WP vs. EPA Break-Even Point (2019).*

Data from nflscrapR
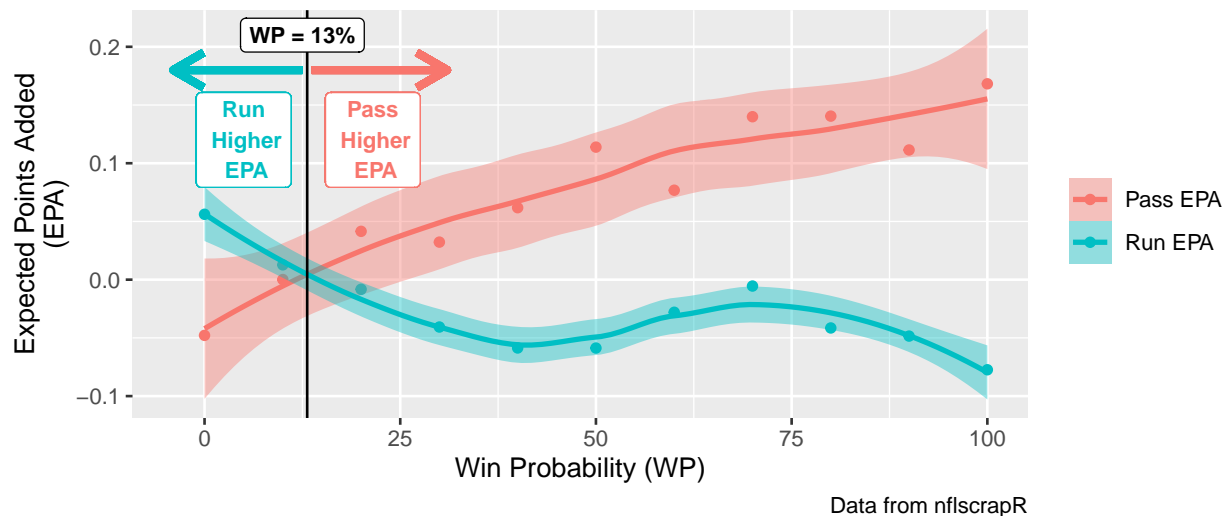
Key Point 5

- *In 2019, running plays had higher EPAs than passing plays at win probabilities of 13% or less.*
  - *Running plays have higher EPAs than passing plays when passing is predictable.*

V. Relationship Between WP & Passing Plays

There is also a relationship between WP and the percentage of passing plays that an offense calls. That is, when the offense has a low probability of winning, the offense calls more passing plays. When the offense has a high probability of winning, the offense runs the ball more. The relationship between the percentage of passing plays that an offense calls to the offense's WP is shown as follows.

*Figure 5: WP vs. Pass Play Percentage (2019).*

Data from nflscrapR
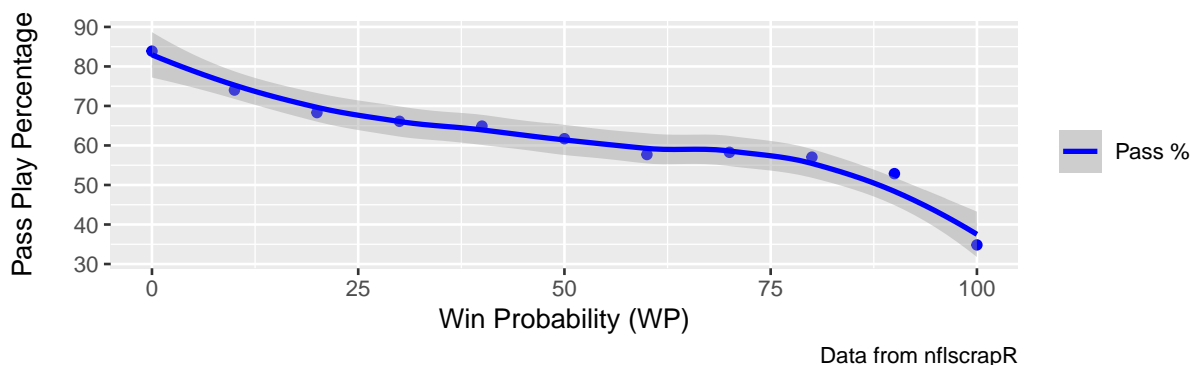
This plot shows that pass play percentage steadily decreases as winning percentage increases. There is a steep decrease in the pass play percentage after the win probability reaches 80%.

Meanwhile, the optimal pass play percentage will be the pass play percentage at a win probability of 13%. At this point, the number of running plays are large enough such that passing is unpredictable, and the passing plays' EPAs are not lower than running plays' EPAs. The typical passing play percentage at a 13% WP is 73%.

Accordingly, because it is best to maximize passing in almost all game situations, the optimal pass percentage in 2019 was 73%. This was 11% higher than teams' actual pass play percentage of 62% in 2019.

The following plot illustrates that 13% was the break-even WP at which passing becomes more effective than running, and that the average passing percentage at the break-even win probability is 73%.

Figure 6: Optimal Passing Percentage (2019).

Data from nflscrapR
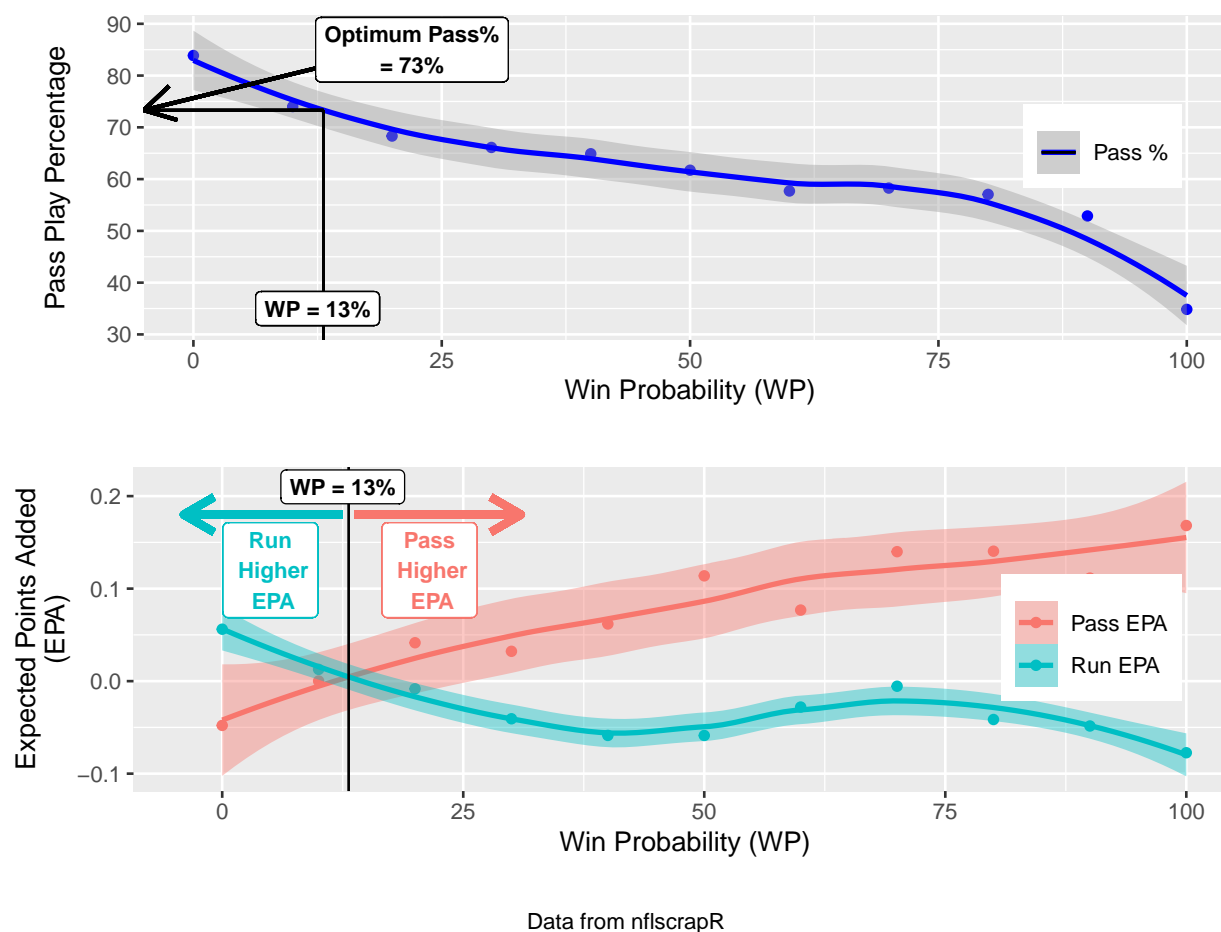
Key Point 6

- *The percentage of passing plays decreases as WP increases.*
  - *Pass play percentage decreases fastest at about an 80% WP and above.*

Key Point 7

- *The passing play percentage corresponding to a 13% WP is 73%.*
- *Accordingly, the optimum pass play percentage in 2019 was 73%*
  - *The optimal pass play percentage balances maximizing pass plays while keeping pass plays sufficiently unpredictable such that passing plays would not have lower EPAs than running plays.*

VI. Years 2015–2018

Play-by-play data from `nflscrapR` for each of the 2015–2018 seasons provide similar results to the 2019 season.

First, the percentage of pass plays called in 2015–2018 were nearly identical to 2019. Specifically, the pass play percentages were 62% in each of 2015, 2016, 2018, and 2019. The pass play percentage was 60% in 2017.

The relationship between pass play percentage and WP has also remained steady over the last five years as shown below:

Figure 7: WP vs. Pass Play Percentage in 2015–2019.
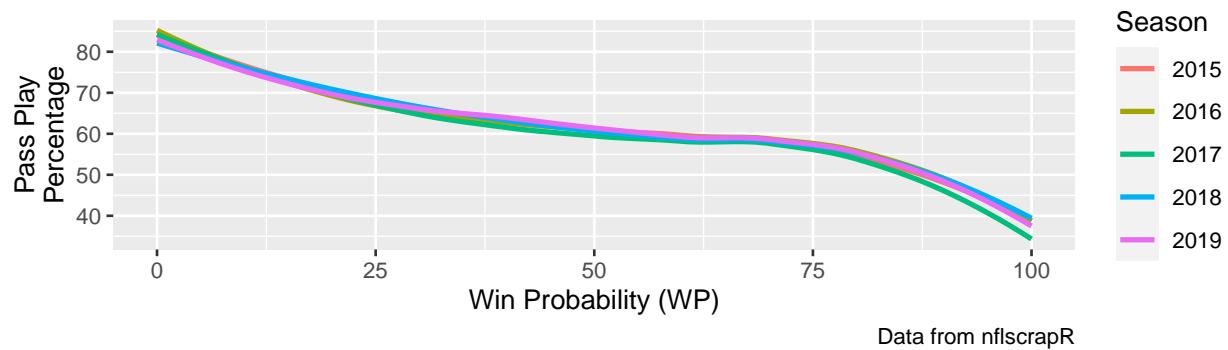
These plots of pass play percentage at each WP are very similar from year to year. The percentage of pass plays called at each WP changed little over the past 5 years.

Next, comparisons of EPAs at the different WPs for running and passing plays in 2015–2018 show the following:

Figure 8: WP vs. EPA for Running & Passing Plays in 2015–2018.
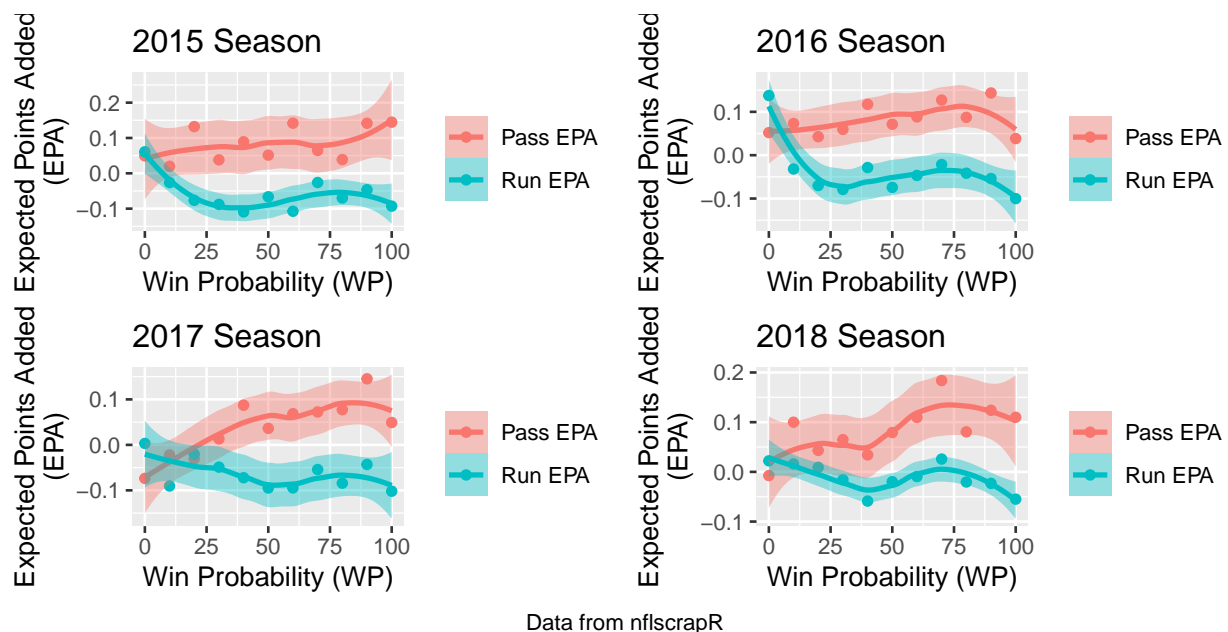
There is variability among the WP vs. EPA curves for running and passing plays from year to year. That is, the break-even intersection points change each year. Because the percentages of pass plays have not changed much at each WP, there is presumably more than unpredictable play calling alone that causes the success of run plays at low WPs.

The break-even WPs, and corresponding optimum pass play percentages, in each of the past five years have been:

Table 3: Break-Even WPs & Optimum Pass Play Percentages in 2015–2019.

Year	Break-Even WP	Optimum Pass Play %
2015	1.43%	83.0%
2016	4.93%	80.4%
2017	10.6%	75.7%
2018	1.79%	81.0%
2019	13.1%	73.3%

Over the past five years, the optimal pass play percentages have ranged from 73%–83%. In every year, the optimum pass play percentage has been higher than the actual pass play percentage of 61%–62%.

Key Point 8

- *The percentage of passing plays has not changed over the past 5 years.*
  - *The percentage of passing plays called at each win probability has also stayed about the same over the past 5 years.*

Key Point 9

- *The EPAs of run and plass plays at each win probability does change from year to year.*
  - *As a result, the optimal pass percentage changes from year to year.*

Key Point 10

- *The optimal pass play percentage has ranged from 73%–83% over the past 5 years.*
  - *The actual pass play percentage of 61%–62% has always been lower than the optimal pass play percentage for each of the past 5 years.*

Conclusion & Limitations

In conclusion, while it is known that passing provides higher EPAs than running, and it is generally assumed that teams should pass the ball more, this research quantifies how much more teams should pass to achieve optimal play calling for maximizing EPA. Specifically, actual pass play percentages have been 61%–62% over the past five years, but optimal pass play percentages would have been 73%–83%.

This research identifies the optimum pass play percentage on a league-wide basis, and should vary from year to year as defenses adjust their expectations based on the number of passing plays being called. Furthermore, the league-wide optimum pass play percentage would correspond to a team having an average passing offense and average rushing offense. Each individual team will have a different optimum pass play percentage based on the relative strengths and weaknesses of their passing and running games.