Evaluating Route Combinations for the 2020 NFL Season

By: Michael Carens & Michael Venit

Data Preparation

- Using the provided play by play data from SIS, we filtered for all pass plays.
- Next, we renamed SIS routes from the “SkillPositionPlayers” dataset to allow for more simplified and generalized route naming.
- Using the GameID and EventID, we then merged the play by play data with a transposed version of the SkillPositionPlayers route data and the new renamed routes. This allowed us to see all of the routes that took place for each observed play in the dataset.

SIS Route	Renamed Route
Beneath	Drag
Comeback	Comeback
Drag	Drag
Screen - Quick	Screen
Chip - Drag	Drag
Wheel	Wheel
Whip	Out
Screen - Tunnel	Screen
Chip - Curl	Curl
Over Ball	Snag
Pick	Other
Hitch & Go	Double Move
Quick	Flat
Angle	In
Post Corner	Double Move
Stick - Nod	Double Move
Chip - Seam	Vert
Out & Up	Double Move
Corner Post	Double Move
Chip	Other
Screen - Shovel	Screen
Jerk	In
Leak	Vert
Screen - Beneath	Screen
Screen - Drag	Screen

SIS Route	Renamed Route
Fade - Back Shoulder	Fade
Curl	Curl
Go/Fly	Vert
Out	Out
Dig	Dig
Check & Release	Other
Slant	Slant
Fade	Fade
Swing - Left	Swing
Jet Sweep Pass	Jet Sweep Pass
Blocking	Blocking
Swing - Right	Swing
Corner	Corner
Post	Post
Sluggo	Double Move
Chip - Flat	Flat
Flat - Left	Flat
Seam	Vert
Deep Cross	Cross
Flat - Right	Flat
Screen - RB	Screen
Screen - Bubble	Screen
Run Fake	Other
Screen - TE	Screen

Using these renamed routes, we can determine the most common route combinations in the NFL last season

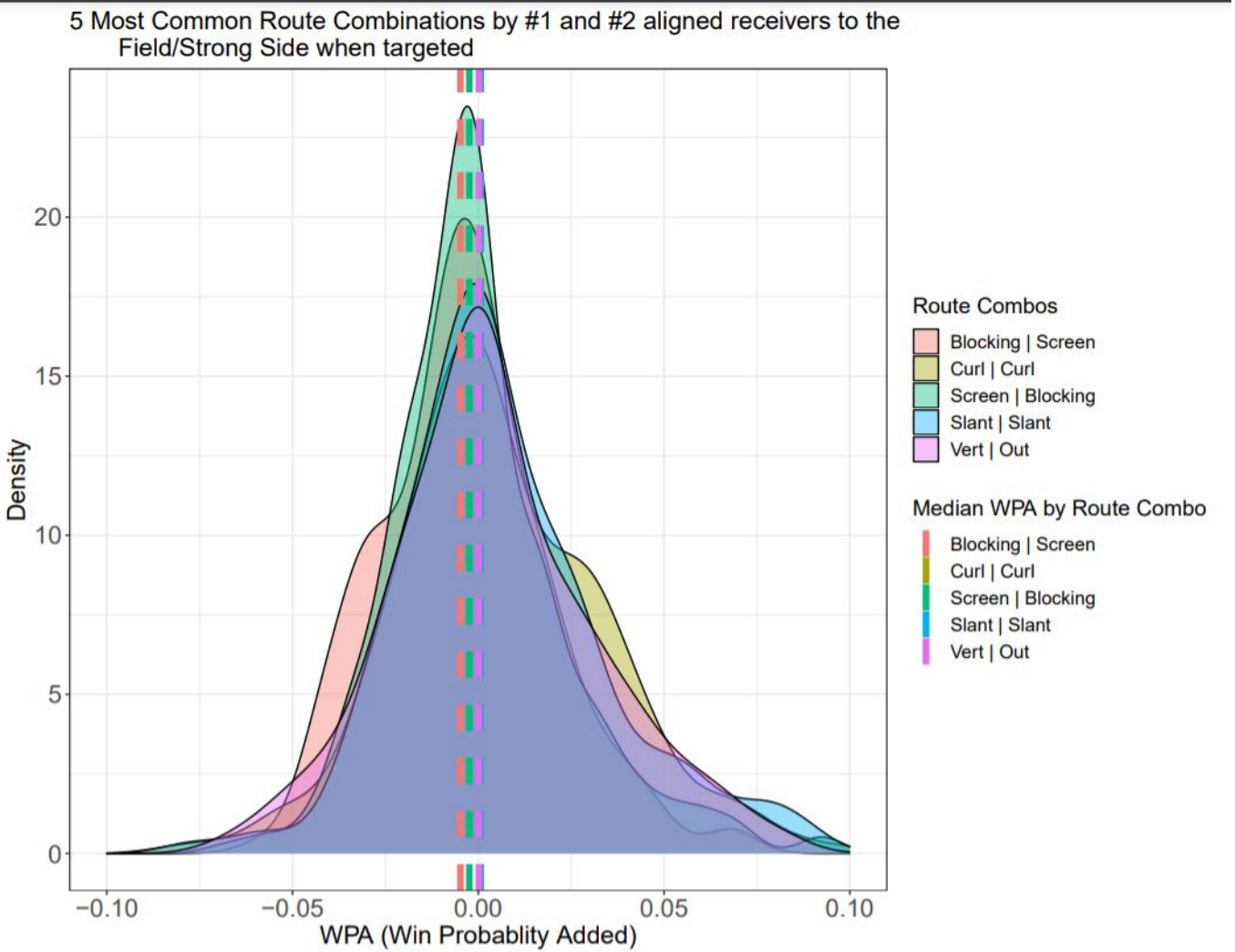
Most Common Route Combinations min. 250 plays ("denotes #1 aligned WR route" | "denotes #2 aligned WR route")	
Route Combo	**Number of Plays**
Vert | Out	586
Curl | Flat	563
Curl | Curl	534
Blocking | Blocking	400
Slant | Slant	388
Fade | Out	379
Out | Out	375
Curl | Vert	364
Out | Curl	310
Slant | Flat	309
Screen | Blocking	290
Vert | Curl	275
In | In	273

Finding the most effective route combinations

- We took a two phase approach to finding the most effective route combination.
- The first phase was a descriptive approach where we examined the top 5 most common route combinations by particular alignments and coverages and using nflfastR's win probability added to find the most effective route combinations.
- The second phase was a more predictive approach where we used an OLS regression to predict Expected Points Added per Play using certain route combinations and coverages in different areas of the field.

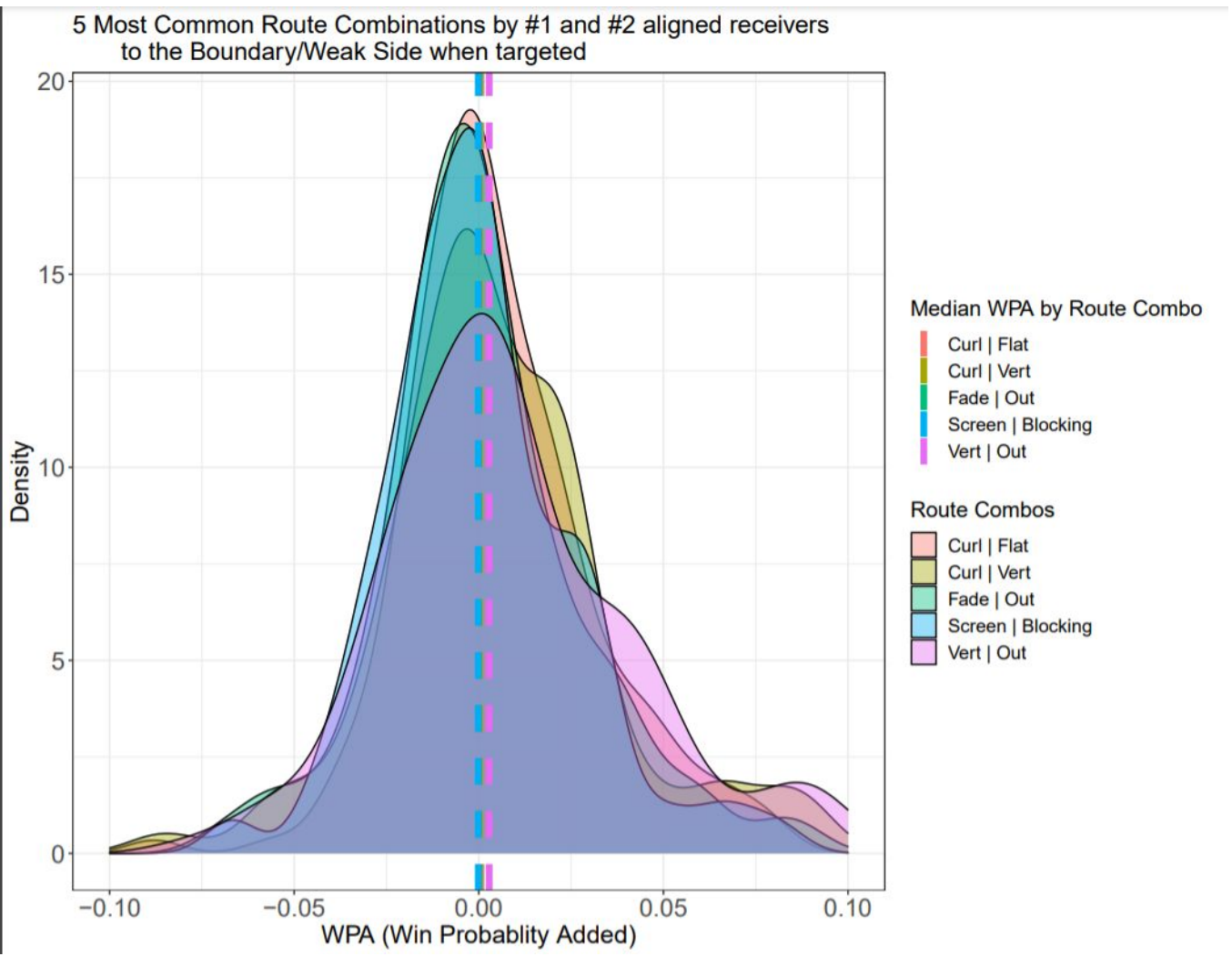
Assumptions and Outlining the descriptive approach

- While reusing the dataset that we previously prepared, we spliced the data for targets to routes to the strong and weak sides of the field.
- The strong side is defined as the “field”, which is the side of the field opposite of where the ball is snapped from if it is on a hash. (e.g. ball on the left hash then the field is to the right). If the ball is not on a hash, the strong side is defined as whichever side has more eligible receivers. If there are an equal number of receivers, the right side is the default strong side since offenses have a tendency to target the right side more than the left.
- The weak side is defined as the “boundary”, which is the same side of the field where the ball is snapped from (e.g. ball on the left hash then the boundary is to the left). If the ball is not on a hash, the weak side is defined as whichever side has fewer eligible receivers.
- For each segment of targets to the field and boundary, we then looked at the most common coverages and most common route combinations that are targeted against these coverages.



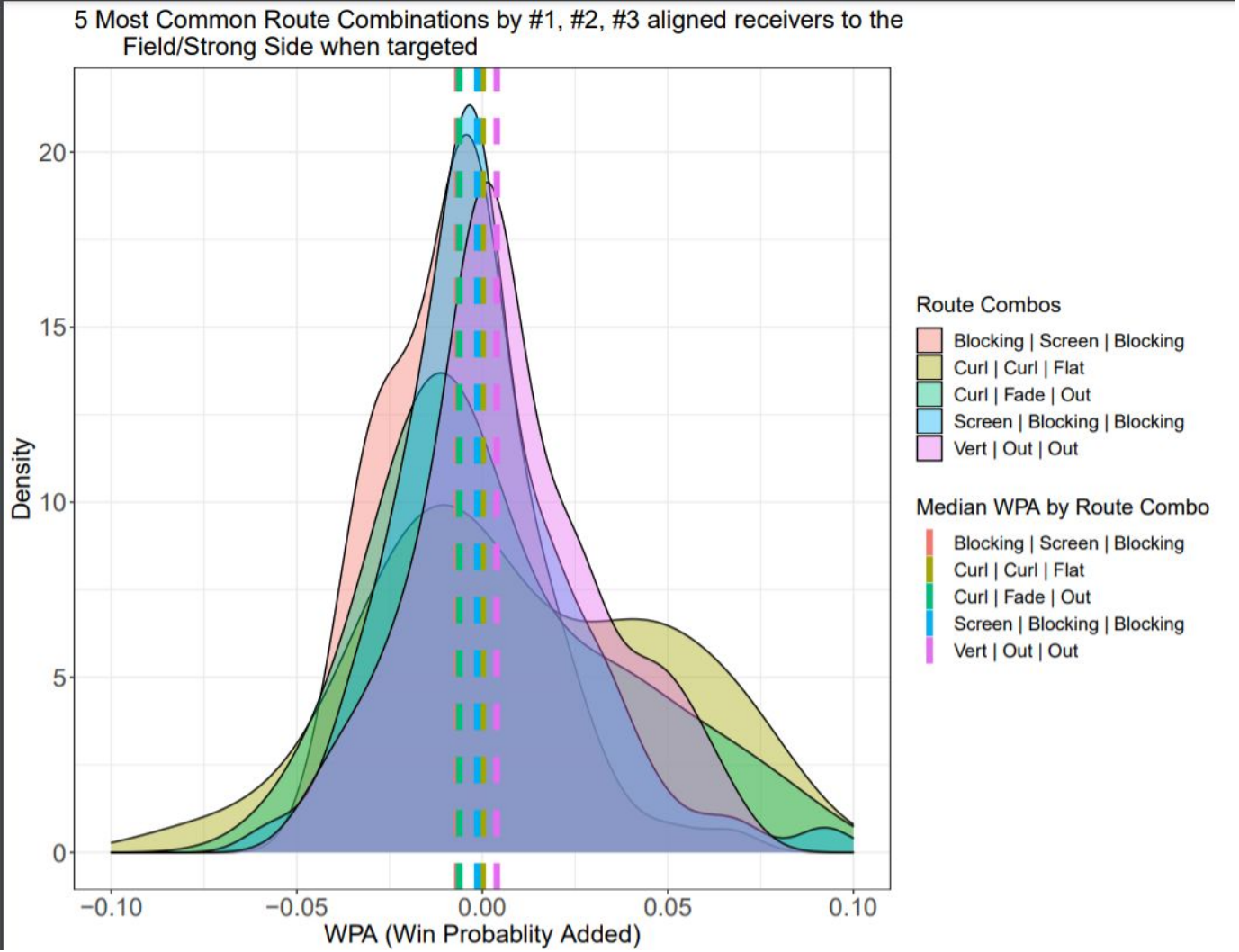
Analysis:

Without filtering for specific coverages, it does not appear one particular route combination has an advantage over another when targeting the strong side of the field.



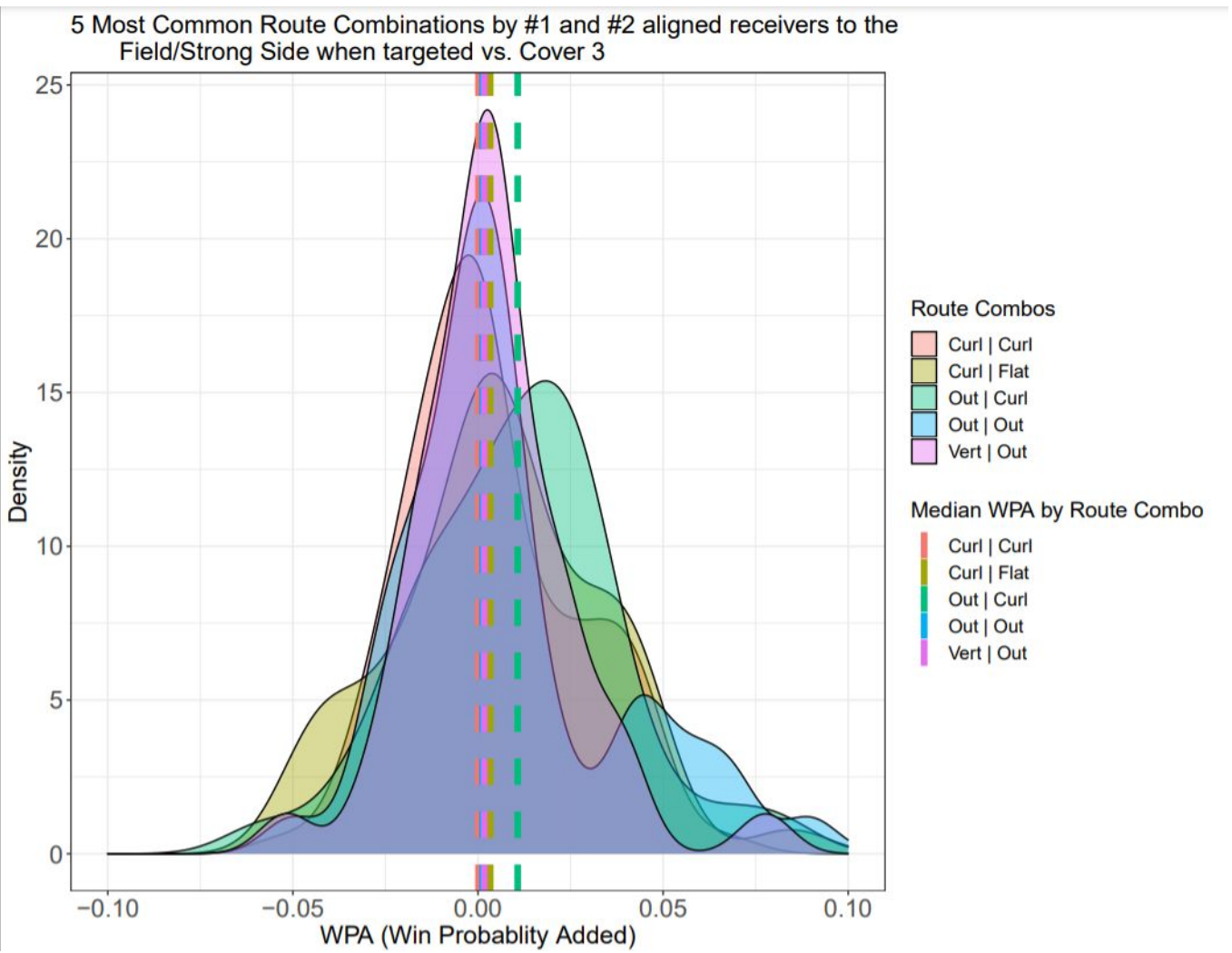
Analysis:

Similar to the previous section, there does not appear to be a distinct advantage in any particular route combination when targeting the boundary when not considering specific coverages.



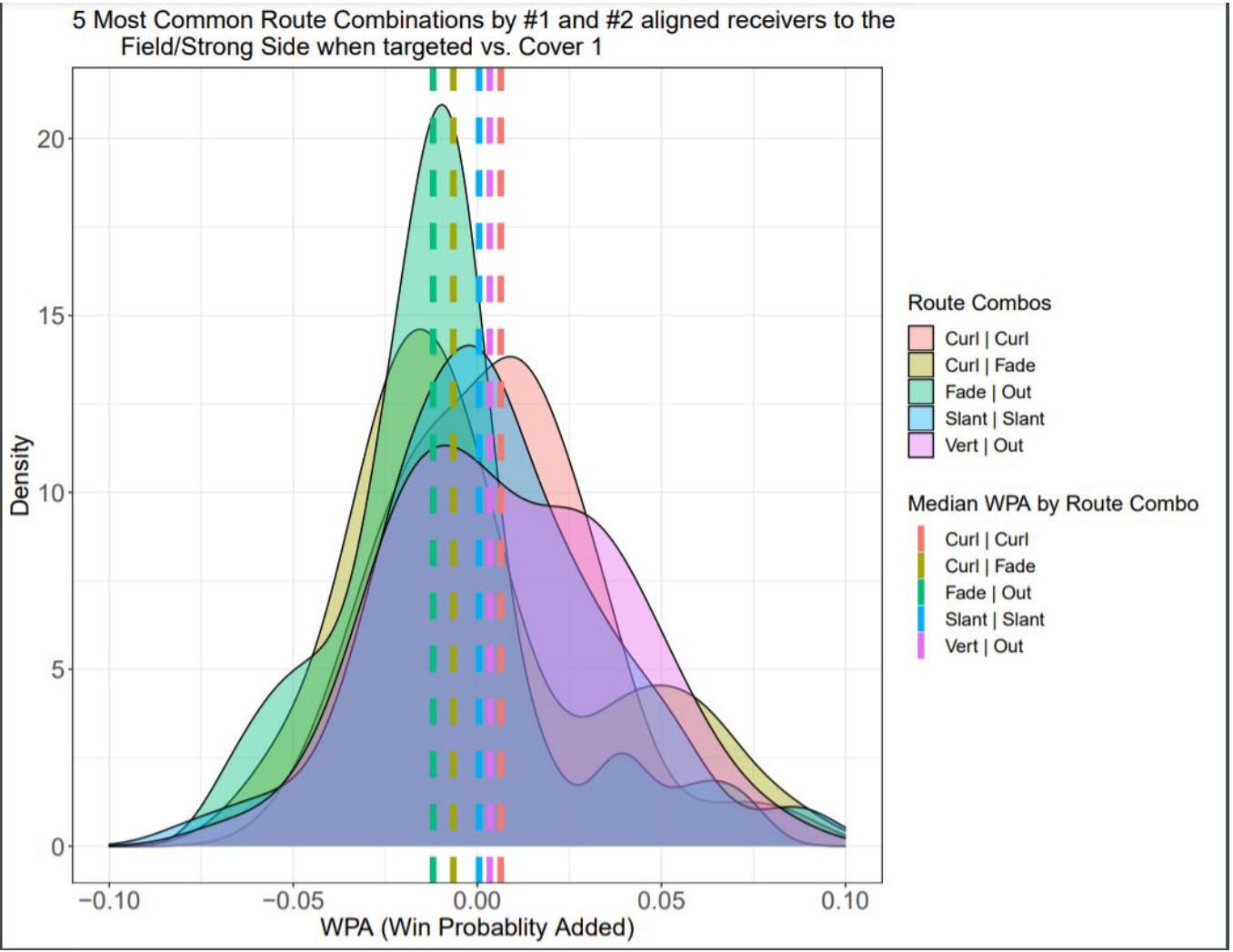
Analysis:

For three receiver combinations, it appears a “Vert | Out | Out” combination is the only route combination whose median win probability added is distinctly greater than 0%. Interestingly, both screen combinations possess median win probabilities added below zero. Unfortunately, the sample sizes for unique three receiver combinations were too low to filter for specific coverages and draw any further conclusions.



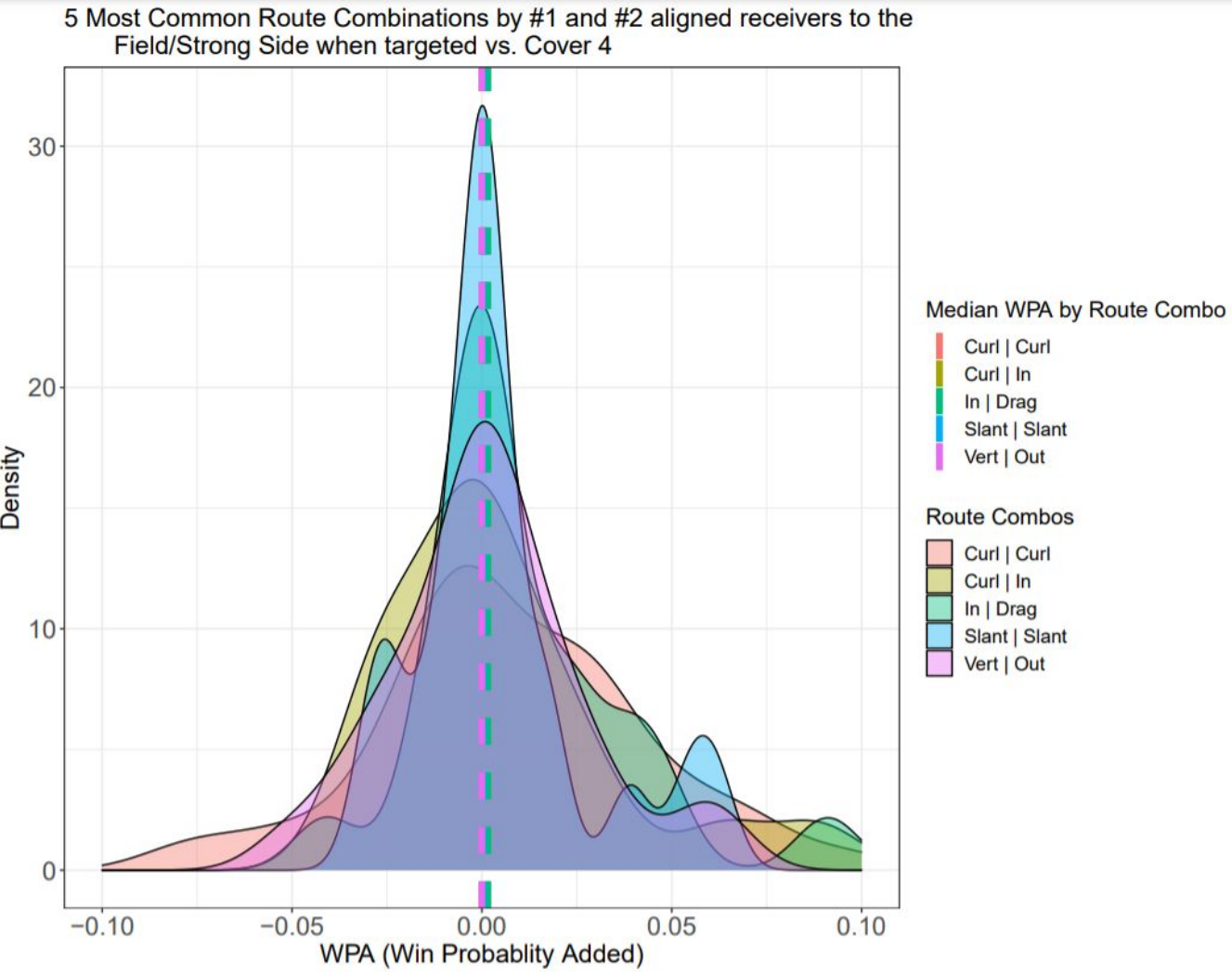
Analysis:

The “Out | Curl” route combination clearly stands out from all other common route combinations to the strong side of the field against Cover 3.



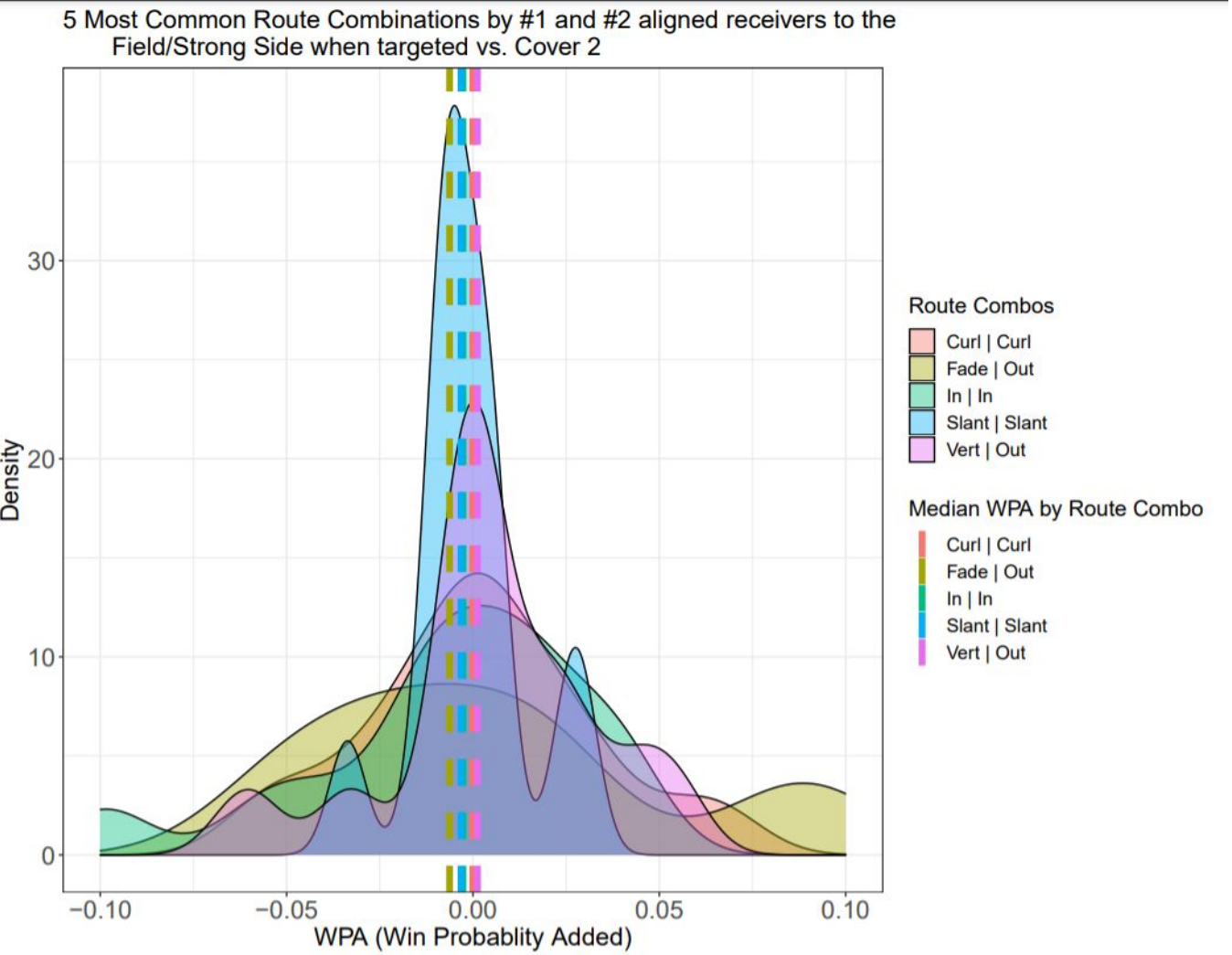
Analysis:

In this chart, we find that “Curl | Curl” and “Vert | Out” are the most effective routes against Cover 1 to the strong side of the field.



Analysis:

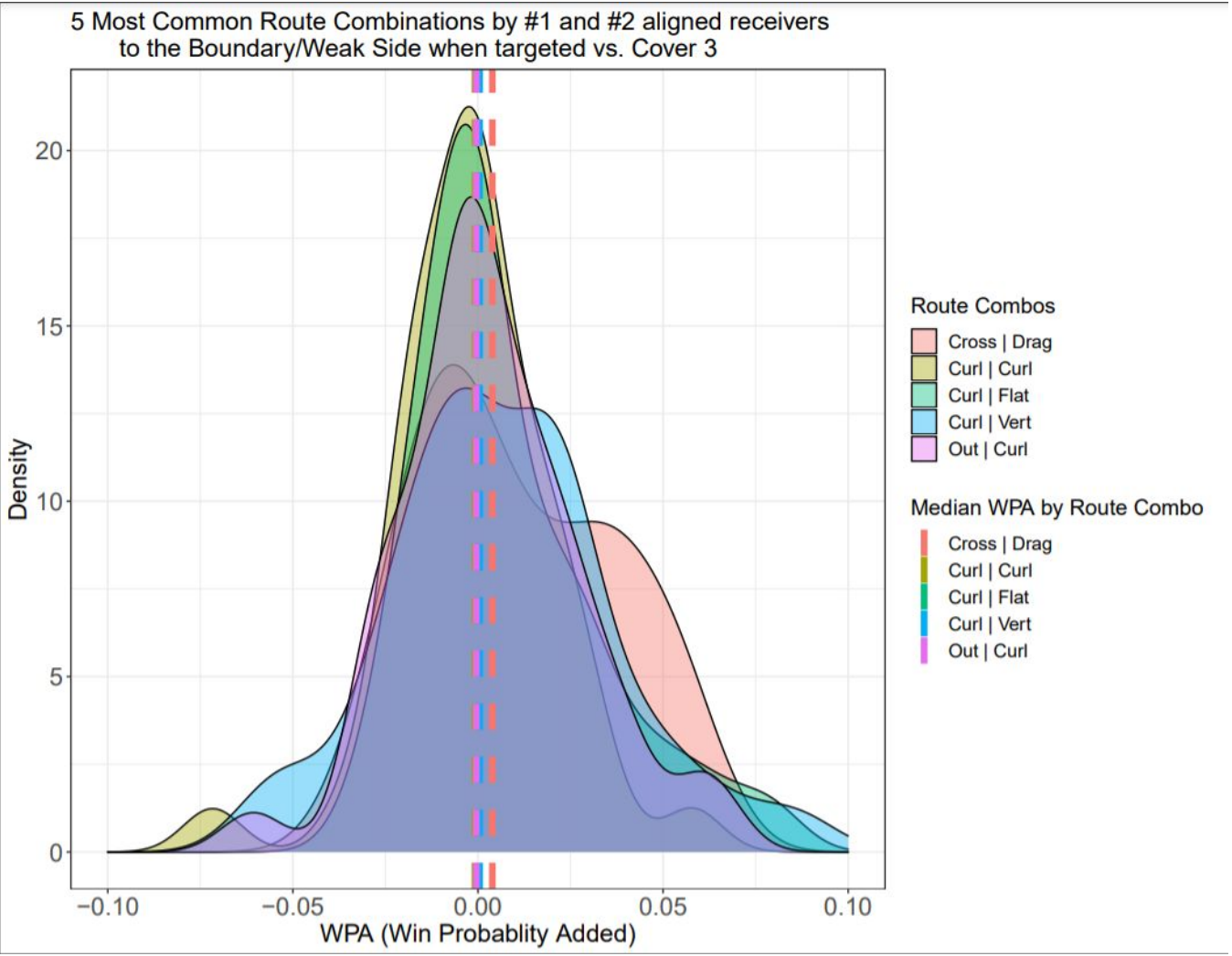
None of the most common routes against Cover 4 to the strong side of the field reveal any significant edge.



Analysis:

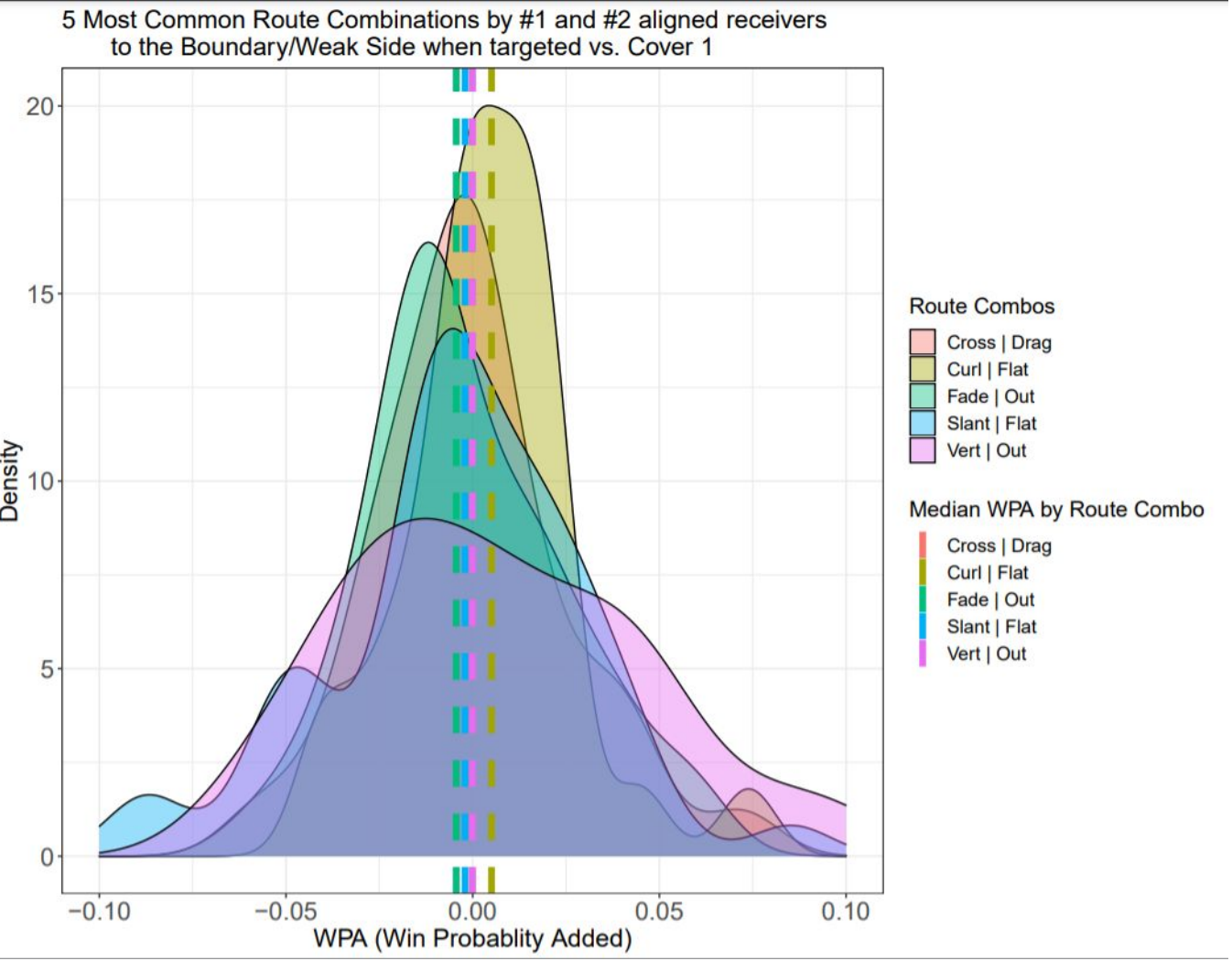
None of the most common routes against Cover 2 to the strong side of the field reveal any significant edge.

When considering targets to the strong side, it seems offenses can find opportunities better opportunities to succeed against one-high coverages (Cover 1 and 3) rather than two-high coverages (Cover 2 and 4).



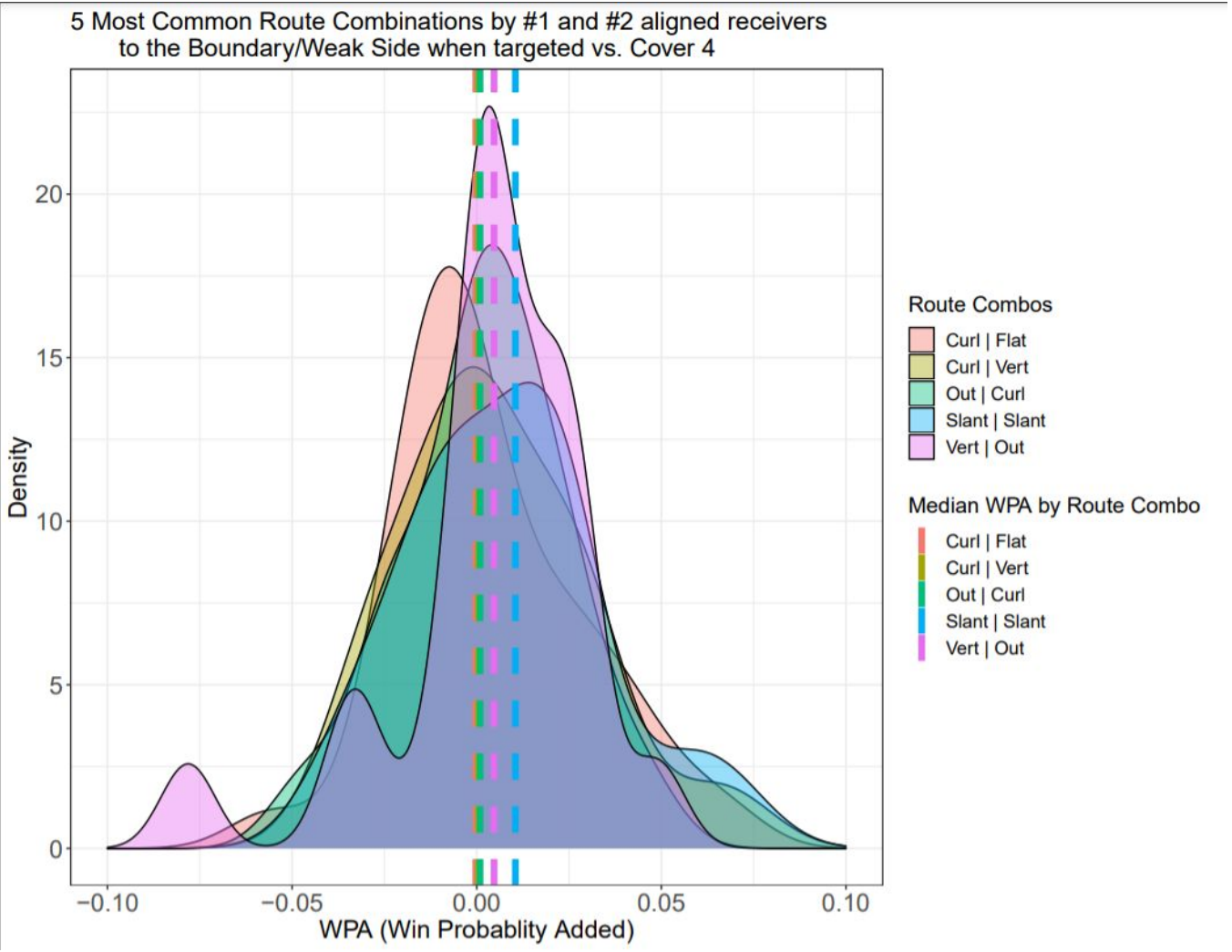
Analysis:

No route combination exhibits much of a significant edge over another.

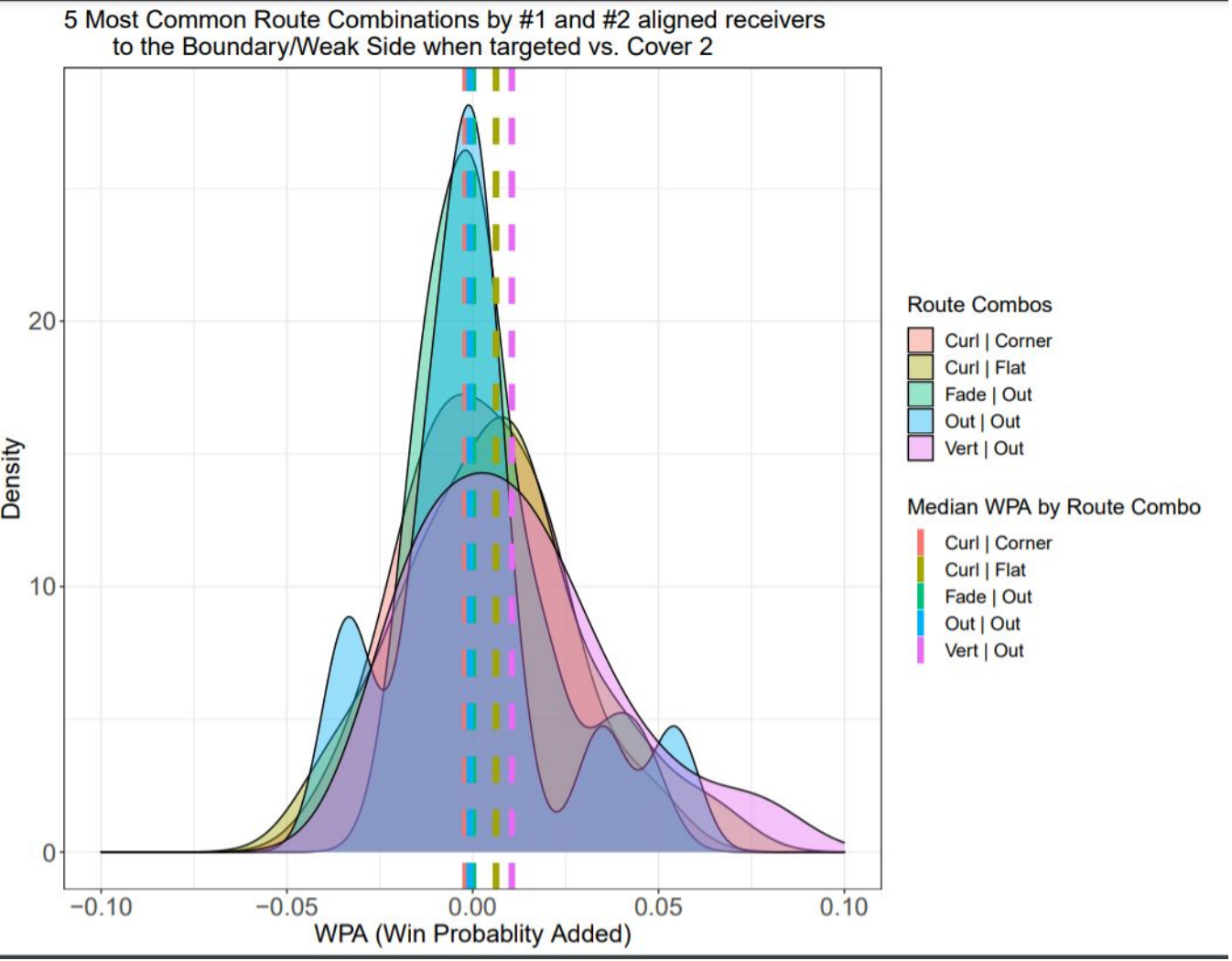


Analysis:

The “Curl | Flat” concept is the only combination that offers a median win probability added greater than 0. Unlike when targeting the strong side against one high coverages, targeting the weak side against one high does not seem to offer nearly as much opportunity for the offense.



Analysis:
This chart suggests that offenses are likely to find success when targeting “Slant | Slant” and “Vert | Out” combinations compared to other route combos against Cover 4 to the boundary.



Analysis:

“Vert | Out” and “Curl | Flat” combinations imply slight advantages for the offense when targeting to the weak side against Cover 2.

Interestingly, it appears offenses have more success against two high coverages when targeting the weak side than when targeting the strong side.

Data Augmentation/Feature Engineering

- After observing the most common route combinations per coverage, binary categorical variables were made that identify each 2-man route combination.
- Binary categorical features were also created to indicate the coverage being played on a specific play
- Our newly created dataframe, generated from our data preparation stage, was then partitioned to represent segments of the field, which can be seen below:
    - Backed Up = Own 1-20 yard line
    - Own 21 = Own 21-39 yard line
    - Midfield = Own 40- Opponent 40 yard line
    - Shot Area = Opponent 39-21 yard line
    - High Redzone = Opponent 20-11 yard line
    - Low Redzone = Opponent 10-1 yard line
- Unlike the previous graphs, we decided that dividing our data by field and boundary would only confine the amount of data to a greater extent than desired, considering different segments of the field were observed.

Modeling Process

- Following the data augmentation/feature engineering step, we decided to create ordinary least squares (OLS) regression models to predict EPA based on the coverage played and the 2-man route combinations identified in the previous graphs.
- For the models, features such as Drop Type, Yards After Catch, Targeted Inside (checks if receiver targeted was at #2), Throw Depth, Categorical Route Combination Features and Categorical Coverage Scheme Type.
- These features were used in conjunction with the partitioned field data to run OLS models that pertain to EPA based on the coverage played and the 2-man route combinations present.
- Route Combinations are listed in order from the outside most receiver (#1) to the inside most (#2).
- An adjusted R-squared is representative of the amount of variance in EPA that is explained by the predictor variables, while the coefficient of each feature is indicative of the change in EPA if EPA were equal to 0 while all other features remain constant.

Findings: Cover 1

- Backed Up:
  - Adjusted R-squared: 0.618
  - Curl Flat combination was statistically significant (at a 5% significance level) but had a negative coefficient (-0.2626)
- Own 21:
  - Adjusted R-squared: 0.6261
  - No route combinations were statistically significant
- Midfield:
  - Adjusted R-squared: 0.6295
  - No route combinations were statistically significant
- Shot Area:
  - Adjusted R-squared: 0.660
  - Vert Out combination was statistically significant with a coefficient of 0.2525
- High Red Zone:
  - Adjusted R-squared: 0.6019
  - No route combinations were statistically significant
- Low Red Zone:
  - Cover 1 is not a commonly played coverage in this part of the field

Findings: Cover 2

- Backed Up:
  - Adjusted R-squared: 0.6207
  - Curl Flat combination was statistically significant but had a negative coefficient (-0.2673)
- Own 21:
  - Adjusted R-squared: 0.6256
  - No route combinations were statistically significant
- Midfield:
  - Adjusted R-squared: 0.6284
  - No route combinations were statistically significant
- Shot Area:
  - Adjusted R-squared: 0.6553
  - Vert Out combination was statistically significant with a coefficient of 0.2644
- High Red Zone:
  - Adjusted R-squared: 0.6020
  - No route combinations were statistically significant
- Low Red Zone:
  - Adjusted R-squared: 0.3042
  - Curl Flat combination was statistically significant but had a negative coefficient (-0.7776), while Slant Slant has a positive coefficient of 0.6901

Findings: Cover 3

- Backed Up:
  - Adjusted R-squared: 0.6206
  - Curl Flat and Out Curl were both statistically significant though both had negative coefficients of -0.2622 and -0.6460 respectively
- Own 21:
  - Adjusted R-squared: 0.6206
  - No route combinations were statistically significant
- Midfield:
  - Adjusted R-squared: 0.6285
  - No route combinations were statistically significant
- Shot Area:
  - Adjusted R-squared: 0.6552
  - Vert Out combination was statistically significant with a coefficient of 0.2629
- High Red Zone:
  - Adjusted R-squared: 0.6000
  - No route combinations were statistically significant
- Low Red Zone:
  - Adjusted R-squared: 0.3063
  - Curl Flat combination was statistically significant but had a negative coefficient (-0.7795), while Slant Slant has a positive coefficient of 0.6814

Findings: Cover 4

- Backed Up:
  - Adjusted R-squared: 0.6196
  - Curl Flat and Out Curl combinations were statistically significant but had a negative coefficients of -0.2777 and -0.6664
- Own 21:
  - Adjusted R-squared: 0.6255
  - No route combinations were statistically significant
- Midfield:
  - Adjusted R-squared: 0.6290
  - No route combinations were statistically significant
- Shot Area:
  - Adjusted R-squared: 0.6563
  - Vert Out combination was statistically significant with a coefficient of 0.2632
- High Red Zone:
  - Adjusted R-squared: 0.6166
  - Curl In combination was statistically significant but had a negative coefficient of -1.2580
- Low Red Zone:
  - Adjusted R-squared:
  - Cover 4 is not a commonly played coverage in this part of the field

Findings Summary

- The route combinations that possessed the most statistical significance were the Curl Flat, Out Curl and Vert Out combinations.
- Though these had negative coefficients in predicting EPA, Curl Flat and Out Curl route combinations were statistically significant in condensed parts of the field (i.e. Backed Up and Red Zone areas). Intuitively, this could be the case due to the high number of turnovers generated from these route combinations, especially in the aforementioned parts of the field.
- Vert Out was consistently significant with a positive impact in the prediction of EPA in particular in the Shot Area of the field. This would seem to make sense as vertical plays would generate more EPA if successful.
- Throughout the modeling process, other features that had strong statistical significance in the prediction of EPA included Yards After Catch and Throw Depth, making sense as an increase in these features on a given play would likely increase EPA, if successful.

Limitations and Future Analysis

- In future analyses, we would be interested in studying:
  - Which route combinations offer the most yards after the catch
  - Which route combos are the best for getting targets near the sticks on 3rd downs
  - Investigate play action. Does the middle of the field offer the most opportunity with linebackers coming downhill?
- Build a more robust model that does not need to be spliced into different sections of the field. Obviously more seasons of data could potentially be a huge help in this.
- The use of geospatial data to understand the splits and alignments of players pre-snap would also be helpful, as it would be interesting to observe what kinds of splits are beneficial to offensive EPA, depending on the coverage being played.
- This would be helpful in understanding which route combinations generate the most separation against from defenders in different coverages

Other data sources used

- nflfastR