Sydney Lockwood

# The Impact of Defensive Pressure and Sacks on NFL Win Probability

**Project Overview**

This project examines whether generating consistent quarterback pressure, through sacks and defensive disruptions, significantly influences a team's win probability in the NFL. By comparing defensive pressure metrics to a team's win success, the analysis can help identify the hidden value of defensive efficiency beyond traditional metrics.

**Research Question**

The overarching research question is: Does consistent quarterback pressure, measured by sack rate and pressure rate, significantly impact a team's win probability in the NFL? NFL teams invest heavily in defensive players and design their defensive strategy to disrupt opposing quarterbacks. By understanding how defensive pressure can influence win probability, stakeholders can make more informed decisions on undervalued defensive talent and optimize defensive strategy.

**KPIs**

Two key metrics:

1. Defensive Sack Rate: total sacks divided by opponent passing attempts
2. Pressure Rate: total QB pressure divided by opponent dropbacks

**Unit of Analysis & Scope**

The primary observation level will be team-season level, with some analysis conducted on a game level. The report will focus on the NFL in the regular season only for 2018-2024.

**Data Sources & Access Plan**

The primary dataset will be NFLFastR GitHub, which provides play-by-play data since 1999 and includes sacks, pressures, and wins. Supplemental data will come from NFL Next Gen Stats, which provides advanced tracking metrics such as fastest sacks, average time to pressure, and defensive disruption. Data preparation will involve downloading and merging the data sets by team. I will also need to standardize column names, clean or replace missing values, and calculate the measurable variables. Together, these two datasets provide both traditional and advanced measures of defensive pressure, allowing for a robust analysis.

**Mini Literature Scan**

Paine (2021) analyzed play-by-play data and found that quarterback pressure correlated more strongly with defensive efficiency and team wins than sacks. Next Gen Stats (2023) reported that teams with a high pressure rate experienced significant decreases in opponents' completion percentage and occasionally EPA. This concluded that defensive pressure leads to more negative offensive plays and increases win probability. ESPN (2022) published a league-wide evaluation of pass-rush win rate (PRWR) and found that teams with a higher percentage tend to have stronger defensive EPA. Unlike sacks, the PRWR measures how often

a defensive lineman beats their blocker within 2.5 seconds. This research shows that PRWR may need to be analyzed on top of the other data to make an accurate hypothesis, as it may be a better predictor to answer the research question. Together, these studies and previous research demonstrate that consistent quarterback pressure, rather than sacks alone, is a reliable predictor of team wins.

**Early Analysis Approach**

Exploratory Data Analysis will begin by inspecting each dataset's variables, identifying any missing values, and computing basic descriptive statistics such as mean, median, and standard deviation. Preprocessing Steps will include filtering out non-regular season data, merging the two datasets for one complete working sheet, and creating the measurable KPI calculations.

**Analytical/Modeling Paths**

Performance and analytics will be evaluated using R2 values and significance testing to determine the strength of the relationship between defensive pressure metrics and the team's success. For this analysis, I will use linear regression as my modeling path to predict win probability/percentage using sack rate and pressure rate as independent variables.

**Potential biases and data access issues**

Sack totals can overemphasize a few standout plays while missing the consistent pressure in the pocket. To mitigate this bias, I am including both sack rate and pressure rate. In regards to data access, Next-Gen Stats data may be incomplete or require manual scraping. NFLFastR data is publicly available, so no concerns there.

**Immediate Tasks**

1. Set up GitHub repository
2. Download and review datasets to ensure all key metrics are included
3. Merge both datasets into a single working file, along with the calculated data
4. Research AI leveraging uses and decide if that is something I want to include

**References**

ESPN Analytics. (2022, December 5). *Pass-rush win rate and its impact on defensive performance.* ESPN. https://www.espn.com/nfl/story/_/id/35085050/pass-rush-win-rate-nfl-analysis

Next Gen Stats. (2023). *Impact of pass rush on offensive efficiency.* NFL Next Gen Stats. https://www.nfl.com/news/next-gen-stats-how-defensive-pressure-defined-the-2023-nfl-season

Paine, N. (2021, October 14). *Why pressuring the QB wins games.* FiveThirtyEight. https://fivethirtyeight.com/features/why-pressuring-the-qb-wins-games/