

BACKGROUND

In this study, play-by-play data from a sample of 123 games from the English Premier League's (EPL) 2010-2011 campaign were obtained from StatDNA. Ball possession was measured by expected probability of pass completion. Via logistic regression, the type of pass and various in-game situations were predictor variables for the probability of a successful pass completion. Multilevel logistic regression was used to identify which teams are most successful in terms of pass completion when controlling for various in-game situations. The play-by-play data were summarized into game-by-game data, and Poisson regression was used to summarize the relationship between pass completion and margin of victory.

THE DATA

- ▶ Play-by-play data obtained from StatDNA from the EPL's 2010-2011 campaign
 - ▶ 123 games
 - ▶ 120,302 observations
 - ▶ 428 players
- ▶ Game Factors
 - ▶ **Event:** Clearance, Cross, Goalie Punt, Goalie Throw, Head Clearance, Throw in, Pass Air (attempted without the head and travels at some point higher than the shoulder), Pass Ground (attempted without the head and does not travel higher than the shoulder), Pass Head, Throw in
 - ▶ **Pressure:** Challenge, Closing (within 10 to 15 yards of player in possession), Marked (within 5 yards of player in possession), No Pressure
 - ▶ **Body Position (with respect to the goal):** Back, Front, Sideways
 - ▶ **Line:** Line 0, Line 1, Line 2, Line 3 (see Figure 1 below for explanation of field location relative to defense)

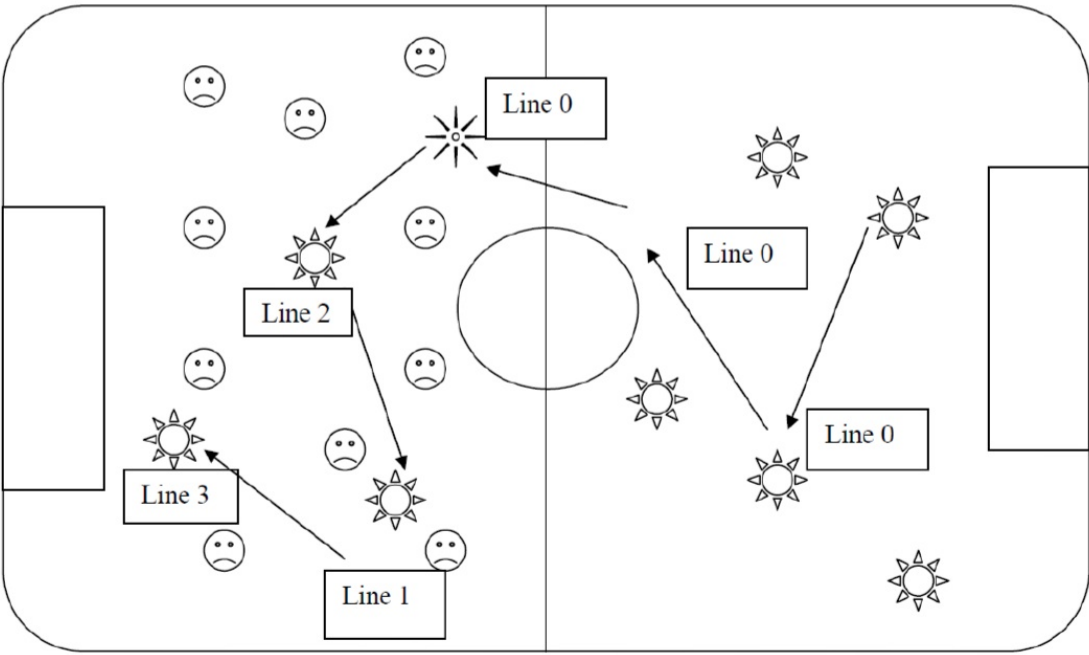


Figure: Line Types in EPL Data Set – Sun is Attacking Player (courtesy of StatDNA)

- ▶ **One Touch:** Pass was one timed, Pass was not one timed
- ▶ **Distance:** Distance of pass or kick measured in yards

ACKNOWLEDGEMENTS

- ▶ Thanks to Jaeson Rosenfeld and the research team at StatDNA for providing the play-by-play data for this study.
- ▶ Thanks to the reviewers of StatDNA's Soccer Analytics Research Competition for their helpful insights and comments.

THE MODELS

Model #1: To study the relationship between pass completion and the in-game factors, a logistic regression model was used.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \sum_{k=1}^{18} \beta_k X_{ki},$$

- ▶ π_i – probability of successful pass completion on play i , $i = 1, \dots, 120,302$;
- ▶ α – intercept; and $\beta_1, \dots, \beta_{18}$ – coefficients for type of event, pressure, body position, line/defense, one-timed pass, and distance

Model #2: To estimate the pass completion rates at a team and player level after accounting for in-game situations, a logistic regression model (similar to Model #1) with varying intercepts and slope coefficients for each team was used.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha_j + \sum_{k=1}^{18} \beta_{kj} X_{ki},$$

- ▶ $\alpha_j \sim N(0, \sigma_\alpha^2)$, for $j = 1, \dots, 20$ and $\beta_{kj} \sim N(0, \sigma_{\beta_j}^2)$, for $k = 1, \dots, 18$, $j = 1, \dots, 20$
- ▶ $\sigma_\alpha \sim \text{Uniform}(0, 100)$ and $\sigma_{\beta_k} \sim \text{Uniform}(0, 100)$, for $k = 1, \dots, 18$

Model #3: To investigate the impact of the difference between a team's offensive pass completion and their opponent's total pass completion on that team's margin of victory while controlling for home field advantage, a Poisson regression used.

$$\log(\mu_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i,$$

- ▶ μ_i – average margin of victory for winning team in game i ; α – intercept;
- ▶ β_1 – home-field coefficient ($x_{1i} = 1$ if winning team team was at home);
- ▶ and β_2 – coefficient for the difference in offensive and defensive pass completion rates
- ▶ $\epsilon_i \sim \text{Poisson}(\mu_i)$
- ▶ $i = 1, \dots, 151 - 123$ games plus the 14 draws counted twice

RESULTS

Coefficients	Estimate	p-value
Intercept	0.4349	$< 2 \times 10^{-16}$
Event (Reference Category: Clearance)		
Cross	-0.4390	$< 2 \times 10^{-16}$
Goalie Punt	-0.1901	0.0157
Goalie Throw	3.3415	$< 2 \times 10^{-16}$
Head Clearance	0.1302	0.0015
Pass Air	-0.3925	$< 2 \times 10^{-16}$
Pass Ground	1.8476	$< 2 \times 10^{-16}$
Pass Head	-0.0100	0.7674
Throw in	1.3971	$< 2 \times 10^{-16}$
Pressure (Reference Category: Challenge)		
Closing	0.8969	$< 2 \times 10^{-16}$
Marked	0.4549	$< 2 \times 10^{-16}$
No Pressure	0.7007	$< 2 \times 10^{-16}$
Body Position (Reference Category: Back)		
Front	-0.8083	$< 2 \times 10^{-16}$
Sideways	-0.2907	$< 2 \times 10^{-16}$
Line (Reference Category: Line 0)		
Line 1	-0.2757	$< 2 \times 10^{-16}$
Line 2	-0.8256	$< 2 \times 10^{-16}$
Line 3	-1.9802	0.0009
One Touch	0.1256	6.09×10^{-8}
Distance	-0.0024	1.76×10^{-8}

AIC = 115,298 McFadden's $R^2 = 0.1970$
Dependent Variable: Successful Pass Completion

Table: Results of Logistic Regression of Pass Completion on Game Factors

RESULTS

Team	2.5%	50%	97.5%	Points 2010-2011
Chelsea	0.7580	0.7790	0.7941	71
Manchester United	0.7619	0.7707	0.7807	80
Arsenal	0.7610	0.7705	0.7797	68
Manchester City	0.7584	0.7677	0.7771	71
Tottenham	0.7274	0.7385	0.7482	62
WBA	0.7268	0.7355	0.7452	47
Liverpool	0.7183	0.7280	0.7376	58
Wigan Athletic	0.7020	0.7110	0.7202	42
Blackpool	0.6947	0.7043	0.7159	39
Fulham	0.6892	0.7010	0.7128	49
West Ham Utd	0.6900	0.7005	0.7118	33
Newcastle	0.6823	0.6938	0.7040	46
Everton	0.6793	0.6901	0.7008	54
Aston Villa	0.6748	0.6873	0.6994	48
Sunderland	0.6755	0.6868	0.6975	47
Wolves	0.6734	0.6837	0.6937	40
Birmingham	0.6656	0.6769	0.6888	39
Bolton	0.6294	0.6412	0.6530	46
Blackburn	0.6230	0.6335	0.6439	43
Stoke City	0.5614	0.5761	0.5901	46

Table: Estimated Probability of Pass Completion for 2010-2011 EPL Teams

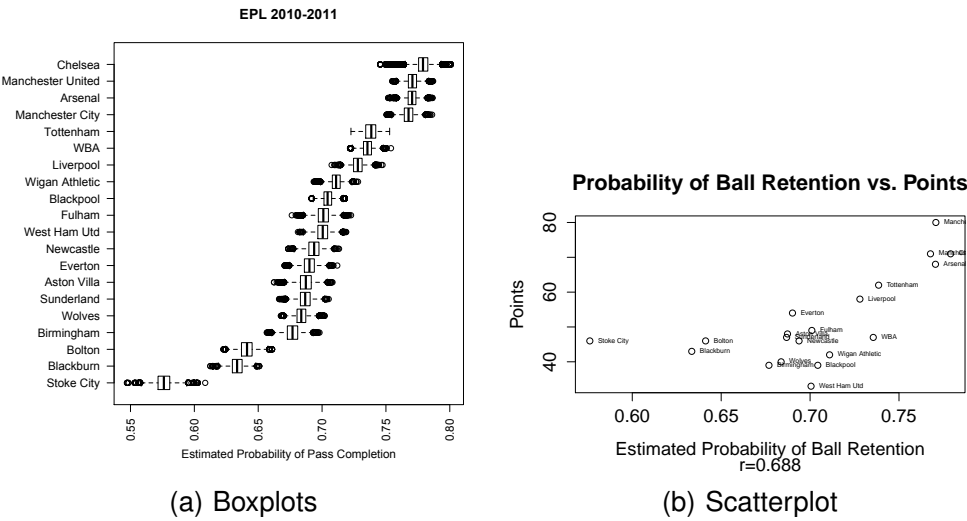


Figure: Estimated Probability of Pass Completion for 2010-2011 EPL Teams

	df	Deviance	p-value	Coefficient	Std. Err.
Intercept				0.0738	0.1429
Home Field	1	0.7043	0.4014	0.1246	0.1828
Difference	1	0.5563	0.4558	0.6321	0.8389

Table: Analysis of Deviance and Results of Poisson Regression of Margin of Victory on Home Field Advantage and Difference of Raw Offensive & Raw Defensive Pass Completion

CONCLUSIONS & FUTURE WORK

- ▶ Easier types of passes (goalie throws, ground passes, throw ins, and head clearances) and pressure (closing, marked, and no pressure) all lead to significantly higher pass rates, controlling for other in-game factors.
- ▶ There is a significant positive relationship between pass completion and team's winning percentage in the EPL (p-value of 0.0008 from simple linear regression).
- ▶ The difference in raw offensive pass completion minus raw defensive pass completion had nothing to do with a team's margin of victory.
- ▶ Future work: Investigate whether choice of pass type truly separates the good teams from the bad.