Optimal substitution times in soccer

Manuscript for the Journal of Quantitative Analysis in Sports
[DATE OF SUBMISSION]

Abstract

Bret Myers has proposed a decision rule for soccer substitution timings that optimizes a team's likelihood to improve its goal differential when trailing: use the first, second and third substitutions before the 58th, 73rd and 79th minutes, respectively. A replication of Myers' analysis with a larger dataset yielded highly similar decision points, 65-72-78. Myers reported that following the rule led to goal differential improvement 42% of times versus 20% for 'defiers', but did not account for any other characteristics that could affect a team's likelihood to follow the rule or improve its situation regardless. A comparison of goal differential improvement between 509 follower-defier pairs matched on relevant pre- and in-game characteristics suggests a more modest, yet robust results: teams who followed the 65-72-78 rule improved their goal differential 32% of times versus only 21% success otherwise, plus-minus 3%. Coaches would thus benefit from following the 65-72-78 decision rule.

Introduction and relevant literature

Substitutions are the most direct intervention soccer coaches can make to alter the course of a match, particularly needed when their team is behind. Since during all matches but those ending in 0-0 at least one team finds itself trailing, it is a common situation for coaches to try using their substitutions to improve their team's position and outscore their opponent. Since only three substitutions are allowed in competitive matches (except during over-time in certain competitions), it is thus of great interest to coaches to take advantage of their substitutions in a way that optimizes their team's likelihood of improving its situation.

While various academic investigations have been made on the factors influencing substitution timings and tactics (del Corral, Barros & Prieto-Rodriguez 2008; Rey, Lago-Ballesteros and Padrón-Cabo 2015) and their impact on in-game aspects such as possession and shots (Gomez, Lago-Peñas & Owen, 2016), the primary endeavor to derive optimal timings to improve their result by the end of the match was initiated by Myers (2012), partially replicated by Rey, Lago-Ballesteros and Padrón-Cabo (2015) and further developed by Silva and Swartz (2016a).

Myers used regression trees to find optimal timings for the first, second and third substitutions when a team is either winning, tied or trailing. He found significant results only for the latter, and proposed a simple decision rule based on his findings: when behind, use the first, second and third substitutions before the 58th, 73rd and 79th minute, respectively. He tested the effectivity of the rule by comparing teams who followed the rule ($n = 440$) and teams who had the opportunity but did not follow ($n = 843$) on their ability to improve their goal differential (outscore their opponents) from the first moment of following or 'defying' the rule. His results indicated that those who followed managed to outscore their opponent 42.27% of the times versus only 20.52% for defiers, concluding that teams should follow the rule.

Rey, Lago-Ballesteros and Padrón-Cabo (2015) have followed the same procedure as Myers but used a smaller number of substitutions taken from the 2013-2014 UEFA Champions League matches. They found optimal splits at the 53rd, 72nd and 79th minutes, but they have not tested the rule's impact similarly to Myers. Due to their smaller sample size and more limited analysis (optimal timings was not the focus of their study), the focus of this paper is on Myers' derivation and evaluation of the rule.

Silva and Swartz (2016a) have analyzed Myers' rule and identified several scenarios in which the rule misbehaves because it fails to properly consider substitutions a team has made prior to falling behind. Consequently, they have proposed a more complicated version of the rule that considers nine combinations of prior substitutions times and current score, examined it similarly to Myers and found extremely similar results: 40% success for followers versus 17.9% for defiers. They proceeded by performing a different analysis to derive optimal timings using Bayesian logistic regression with a different outcome variable (whether the trailing team will score next) and reported that there was no discernible time in which substitutions were found more beneficial. Both Myers (2016) and Silvea and Swartz (2016b) have defended their respective methods and results in later publications.

This study attempts to derive an easily interpretable decision rule by replicating Myers'

procedure but examine its effect on goal differential more robustly. This is done by only comparing teams who followed or 'defied' the rule that were similar on relevant characteristics such as team strength and match score, using a Genetic Matching algorithm. Myers' original formulation is followed for it is easy to understand and apply, and hence more accessible to practitioners than Silva and Swartz' models. However, they have rightly pointed out difficulties in the simple comparison he performed to examine the rule that question the credibility of its reported impact. Examining only similar match situations yields a more reliable quantitative estimation of the rule's effect on a team's ability to improve its goal differential when behind, while maintaining the rule's straightforward form.

The rest of the paper is organized as follows. In the next section I present Myers' derivation of his original decision rule in greater detail and replicate it with a larger dataset to find a very similar 'revised' decision rule. Then, I analyze Myers' evaluation of the rule and its limitations and propose a more robust evaluation method using genetic matching. Finally, I employ genetic matching to estimate the effect of the revised decision rule and discuss the results.

Replicating Myers' decision rule

Brett Myers (2012) proposed a decision rule for optimal substitution timings when a team is behind. According to the rule, teams trailing should use their first, second and third substitutions before the 58th, 73rd and 79th minute, respectively. If the team has equalized, it can continue substituting at will, but if it falls behind again prior to the next decision point it shall resume following. A team is considered to have followed the rule in a match when it fell behind and had the opportunity to substitute according to the rule at one or more of the decision points (65, 72, 78) and has done so at all decision points during which it was still behind.

The rule was derived by separately analyzing first, second and third substitutions from 485 substitution patterns from matches played in the English Premier League, Italian Serie A and Spanish La Liga (the specific season was not mentioned). Only substitutions done while the team was behind were considered, but their exact numbers out of the original 484, 464 and 403 first,

second, and third substitutions in the data were not reported. For each substitution, a binary outcome was defined as 1 (success) if the team outscored its opponents from the substitution to the end of the match, and 0 otherwise. For each category, a tree algorithm was used to find the minute that splits the data such that it optimally predicts success rates before and after the split. 58, 73 and 79 were reported as the optimal splits, although differences in success rates before and after all three points were not significant by a standard p-value of 0.05.

I followed Myers' procedure to replicate the decision rule using data from matches played between the 2011/2012 and 2016/2017 seasons in five major European leagues: English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A, and French Ligue 1 (Secareanu, 2017). The main difference from Myers' procedure was the exclusion of matches with substitutions due to injuries and matches with red cards; both suggest forced substitutions while only those from tactical reasons are of interest. In the original paper, these matches were excluded only when testing the rule but not during its derivation. After these exclusions, substitutions from 4346 matches remained. Substitutions from half of these matches during which the team was behind ($n = 4546$) were used to derive the decision rule while the other 2173 matches were used to estimate its impact.

Results from Myers' trees and the replication trees for first, second and third substitutions timings are shown in Table 1. The second and third decision points were remarkably similar to Myers' 73-79 and the 71-79 timings found by Rey, Lago-Ballesteros and Padrón-Cabo. The timing of the first substitution was not as similar to previous analyses and slightly less robust in the present one (see Table 1 and Appendix). Also, the mean success rates (proportion of teams improving their scores after substitutions) were generally lower in the replication, and so did the differences between the before- and after-the-split groups which were about 10% versus about 20% in Myers' analysis. This is perhaps due to the substantially larger sample used for the replication, which also grants greater robustness to the smaller observed differences. Thus, the evaluation of the impact of using the first, second and third substitutions before the 65th, 72nd and 78th minutes (the 'revised' rule) was tested, as presented after discussing Myers' evaluation.

Table 1. Proposed Substitution Decision Rules by Myers and its Replication

	Myers (2012)			*Replication*		
	Sub #1	**Sub #2**	**Sub #3**	**Sub #1**	**Sub #2**	**Sub #3**
Mean success rate	0.31	0.24	0.16	0.24	0.19	0.15
Split point	**57.5**	**72.5**	**78.5**	**64.5***	**71.5**	**77.5**
Mean success before	0.41	0.3	0.24	0.26	0.23	0.21
Mean success after	0.18	0.06	0.07	0.18	0.13	0.11
Sample size	180**	172**	150**	1509	1596	1441
p-value	0.052	0.057	0.134	0.0001	< 0.00001	< 0.00001

* The algorithm did not find a split for the first substitution. Further analysis of possible splits by before-after differences indicated lower differences than other categories and yielded this split (see Appendix A for an illustrative figure of this analysis)

** Not reported, estimated for magnitude using the proportion of replication substitutions.

Limitations of the original rule's evaluation

To examine the effect of using the rule, Myers compared the mean success rate of instances of teams following ($n = 440$) and 'defying' ($n = 843$) the rule. Defiance consisted of being behind during at least one decision point but not completing one, two or three substitutions as prescribed by the rule; success for these cases was improving goal differential (outscoring the opponent) from the first decision point of defiance to the end of the match. Myers reported that 42.27% of 'followers' succeeded in outscoring their opponents while versus 20.52% success for 'defiers', concluding the rule is highly effective. Others in the growing soccer analytics literature have since endorsed these findings (for example, Anderson & Sally, 2013).

While the simplicity and interpretability were advantageous when deriving the rule, its evaluation did not consider relevant characteristics that could determine the likelihood of following the rule or succeed in improving goal differential and thus confound the results. If most followers were stronger than their opponents, they had better chances of success regardless

of the rule; if a team conceded two minutes before a decision point, it is unreasonable to expect it could follow it; if more defiers than followers were only presented a chance to follow on the third point (for example by conceding on the 75th minute) they had less time to outscore their opponent and thus lower success chances. The latter point also indicates a broader issue: there are many patterns that 'follow' and 'defy' the rule with various combinations of trailing at different decision points.

Building comparable followers-defiers groups with genetic matching

To account for these limitations, teams identified as followers or defiers out of the 2173 matches not used for deriving the rule were assigned additional in-game variables: the first decision point from which they could (defiers) or did follow, the score at that point, and the time the team was behind before the decision point. The difference in average betting odds (positive if the team had higher odds, negative otherwise) was added as a covariate for each team to account for both season-long strength and per-match variation such as recent form and injuries. In addition, only teams who could follow from the first decision point were analyzed to minimize some of the diversity in 'rule following' and creating more distinct followers and defiers.[1] Moreover, only teams who were behind for more than seven minutes were considered to allow time for warm-up and substitution. This left a total of 539 followers and 800 defiers.

A 'genetic matching' algorithm was used to identify followers and defiers who are similar on all the aforementioned characteristics, as well as their interactions and squared terms. The genetic matching algorithm finds the optimal weights to assign each variable used for matching such that it maximizes the minimal p-value across all difference-in-means t-tests and KS-test comparisons between the groups (more on genetic matching in Diamond & Sekhon 2013).

Table 2 shows the characteristics used for matching, the means of both groups and a p-value indicating the level of balance (the algorithm aims to make the smallest value in this column as

[1] To better account for remaining variation in rule-following options, further analysis could compare followers only with 'perfect defiers' who failed to substitute in all three points or those who substituted the furthest away from them. If the rule is effective, A higher positive effect for followers is expected.

large as possible) for followers and defiers before and after matching. Before matching, most characteristics differ between the groups. For example, followers tended to have higher betting odds, suggesting they are usually stronger, and were behind for longer, indicating a greater motivation to substitute. After matching, the two groups are extremely similar in all characteristics with the lowest p-value being 0.59. In the process, 30 followers for which a good match could not be found were excluded.[2]

Table 2. The balance between followers and defiers before/after genetic matching

	Before genetic matching			*After genetic matching*		
	Mean followers	**Mean defiers**	**p-value***	**Mean followers**	**Mean defiers**	**p-value***
Score at chance	-1.7	-1.6	0.05	-1.53	-1.53	1
Time behind	35.4	38.2	0.0003	34.62	34.6	0.67
Odds	3.32	4.19	0.07	1.89	1.9	0.88
Odds^2	73.81	83.25	0.02	17.79	17.52	0.59
Side (home-1/away-2)	1.59	1.6	0.76	1.57	1.58	0.6
Score at chance^2	3.85	3.31	0.046	2.89	2.89	1
Time behind^2	1492.1	1797.9	0.0003	1442.7	1441.6	0.79
Score at chance*time behind	-65.22	-66.72	0.63	-56.78	-56.77	0.98
odds*time behind	128.63	166.15	0.06	69.2	68.8	0.81
Score at chance*odds	-8.85	-8.5	0.82	-3.62	-3.66	0.62

***** p-value is the smaller of a t-test and KS-tests produced in R.

A 0.45 caliper was used for the first four variables; adequate balance was not obtained otherwise.

Estimating the effect of following the 65-72-78 decision rule

[2] The mean of Odds^2 changed the most before-after, suggesting matches were extremely unbalanced, which directly affects the likelihood of improvement after substitution. However, matched units included teams with odds differences as high as 21.8 for the team and 13.2 against, so there is no loss of generality.

The final 509 followers and 509 defiers similar on pre- and in-game characteristics were then suitable for a difference-in-means comparison to estimate the effect of substituting according to the rule on improving goal differential until the end of the match. Followers succeeded in outscored their opponents 32% of the time while defiers did so only 20.9% of the times, an average percentage point difference of 11.1% with a standard error of 3%. Therefore, *coaches following the 65-72-78 rule when their team is behind are expected to improve their situation in 32% of the time versus only 21% for those who do not substitute according to the rule, plus-minus 3%.* This is a lower effect than the 20 percentage points difference found in Myers' study, but it is more reliable because it was obtained from more comparable cases of teams who did and did not follow. It is also meaningful in light of the generally low chances of improvement found. Running a similar procedure as outlined thus far with the original 58-73-79 rule on the entire dataset used in this study (since none of its matches were used to derive it) yielded 764 eligible followers and defiers for which the average success was 28.4% and 23% respectively, a lower impact than the revised rule.

Conclusion

When a soccer team is trailing, using substitutions to maximize the chance of a comeback is of great interest to coaches. Myers' analysis of improvement after substitutions yielded a simple decision rule: if behind, use the first, second and third substitutions before the 58th, 73rd and 79th minutes. A replication of the analysis with a larger dataset and only tactically-driven substitutions found slightly different optimal decision points: 65th, 72nd and 78th minutes, still in close alignment with the original rule. Following the original rule was reported as leading to success in goal differential improvement about 42% of the time versus 20% for defiers, but it did not consider other characteristics that could significantly impact a team's ability to follow the rule, such as timing of falling behind, or improving its situation, such as its relative strength. Success from following or defying the 65-72-78 decision rule was thus compared between 509 pairs of teams who had the chance to follow from the first decision point and were similar on pre-game odds of winning, score during the decision point and the time the team was trailing before it. The results suggest that coaches who follow the 65-72-78 rule would improve their

goal differential 32% of the times versus only 21% chances of improvement for those who do not follow the rule, plus-minus 3%. Soccer coaches would thus benefit from following the 65-72-78 decision rule.

References

Anderson, C. & Sally, D. (2013). *The Numbers Game: Why Everything you know about Soccer is Wrong*. Penguin Books: New York.

Diamond, A. & Sekhon, S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. The Review of Economics and Statistics, 932-945.

Del Corral, Julio & Pestana Barros, Carlos & Prieto-Rodriguez, Juan. (2008). The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. *Journal of Sports Economics, 9*, 160-172.

Myers, B. R. (2012). A proposed decision rule for the timing of soccer substitutions. *Journal of Quantitative Analysis in Sports*, *8,* Article 9.

Myers, B. R. (2016). Analysis of substitution times in soccer (Silva and Swartz). *Journal of Quantitative Analysis in Sports, 12*(3), 123-124.

Ezequiel Rey, Joaquín Lago-Ballesteros & Alexis Padrón-Cabo (2015) Timing and tactical analysis of player substitutions in the UEFA Champions League, International Journal of Performance Analysis in Sport, 15:3, 840-850

Miguel-Angel Gomez, Carlos Lago-Peñas & L. Adam Owen (2016). The influence of substitutions on elite soccer teams' performance, International Journal of Performance Analysis in Sport, 16:2, 553-568

M. Silva, Rajitha & Swartz, Tim. (2016a). Analysis of substitution times in soccer. Journal of Quantitative Analysis in Sports. 12. 10.1515/jqas-2015-0114.

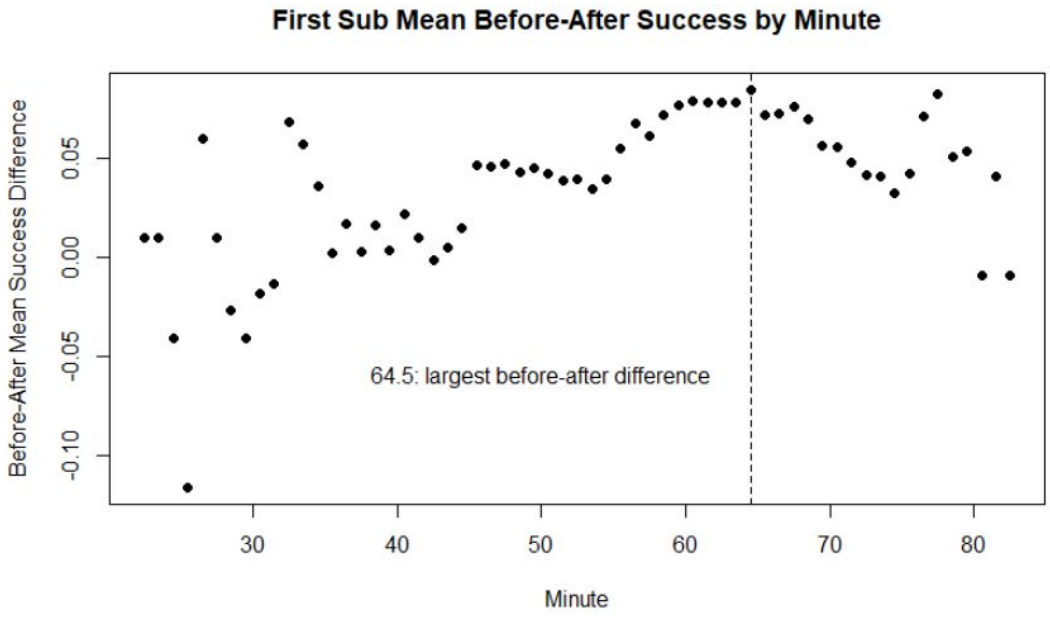
M. Silva, Rajitha & Swartz, Tim. (2016b). Rejoinder to Myers (2016). *Journal of Quantitative Analysis in Sports, 12*(3), 125.

Secareanu, A. (2017). Football events: More than 900,000 events from 9,074 football games

across Europe. In *Kaggle.* Retrieved from: https://www.kaggle.com/secareanualin/football-events

Appendix: Finding a split for first substitutions

The following figure shows the difference in mean success before and after each possible minute for first substitutions, used to select a split in lack of a result from the tree. The dashed vertical line only highlights the optimal point used in the analysis.



The data and the code in R that produced the results, tables and figures can be received upon demand by emailing tom.kremer@minerva.kgi.edu.