Using Analytics to understand Performance and Wellness for a Women's College Soccer Team

Christopher Njunge
cnjunge@callutheran.edu

Paul D. Witman
pwitman@callutheran.edu

California Lutheran University
Thousand Oaks, CA

Patrick Holmberg
Phomey12@gmail.com

Joel Canacoo
jcanacoo@gmail.com

Abstract

This study used analytics to examine the effect of Home Field Advantage (HFA) on the Internal Load (Session Rate of Perceived Exertion (sRPE)), External Load (from GPS trackers) and wellness (based on surveys) for a Division III women's soccer team in Home and Away matches. First the home advantage (HFA) in the Southern California Intercollegiate Athletic Conference (SCIAC) that the team plays in was quantified using conference games only for all nine teams for three seasons. A multiple regression analysis was used to analyze the relationship between Goal Difference and sRPE, External Load variables and Wellness variables at the team, position, and athlete level based on 12 athletes. The results showed the league had an adjusted HFA of 57%. The analysis showed that Defenders were impacted more by away matches based on the medium effect sizes of the difference between their home and away measures for sRPE and sleep quality, and small effect sizes for Distance in high-speed zones and stress. One athlete, a forward, had different mean sRPE in home and away matches. These findings validated the existence of HFA in sports and the findings of differences in the impact of HFA to athletes in specific positions can be used as a guide in analyzing and acting on the performance and well-being of players at position level. The study demonstrated the value of analytics in gaining insights about players performance and well-being.

Keywords: Internal Load, Sports Analytics, Soccer, Home Field Advantage, Performance

1. INTRODUCTION

Home Field Advantage (HFA) and its impact on player performance continue to be an area of interest for fans and practitioners of sports. This paper sought to use analytics to analyze the performance of soccer players in home versus away matches to quantify the differences in their performance and increase the understanding of these differences. Soccer is a sport involving eleven players on each side and featuring one goalkeeper and 10 outfield players. It is played

with a round ball and players' positions are often divided into defenders, midfielders and forwards. Goals are often few, typically less than three over the two 45-minute halves played.

One framework seeking to explain HFA has five major components: game location, game location factors, critical psychological states, critical behavior states and performance outcome (Carron et al.,2005). This framework suggests that multiple factors account for the variation in players' performance in home and away matches. Some of the non-physical factors potentially influencing subjective well-being that Abbott et al. (2018) discussed include match location, quality of match opposition and match result; collectively called situational match variables. Bailey (2013) has proposed regular testing and single-subject design when taking part in athlete monitoring and tracking.

Novelty and Practical Relevance – Use of analytics in analysis of HFA in sports using team, position and individual match variables can extend the understanding of home advantage in sports in general and soccer in particular.

Theoretical Contribution – This paper uses analytics to highlight the differences in the way athletes and players in different positions, even on the same team, are impacted by home field advantage in soccer. These findings can be extended to other sports in order to improve preparation of athletes and teams and improve their performance and well-being.

2. LITERATURE REVIEW

A review of the literature was carried out using the search terms 'Home Field Advantage', 'Internal Load', 'External Load' and 'Wellness'. Out of the 254 articles identified, 76 were selected for further review based on relevance of the title as determined by the authors. Of the 76 selected, further analysis was carried out by reviewing the abstract and high-level reviews of the methodology and findings. Additionally, the articles were categorized according to the following criteria: A further review of the literature was carried out with the following search terms: Sport, Match Play, GPS, Level, Gender and Region. 46 of the 76 articles related to soccer and four of them featured more than one sport. 50 of the 76 articles featured match play and 36 used GPS devices in their study. 30 of the studies were conducted using professional players as research subjects, 21 featured amateurs and 13 identified the subjects as elite. Most of the articles, 61, researched males, eight both male and female athletes, while only five specifically carried out research on female athletes. More than forty of the articles featured research carried out in Europe or the UK and only 11 of the studies were in the US.

Gaps

A review of the literature shows a gap in the study of soccer athletes in the US and particularly female athletes. The use of GPS devices in quantifying external load was quite prevalent as was the use of self-reported RPE to identify internal player load and wellness surveys to track the wellbeing of players. Only one study aimed to use machine learning algorithms to predict the internal load based on external load measures, and this was for professional soccer players in Europe. The authors are unaware of any study where analysis of internal load, external load and wellness of soccer players is carried out at a US Division-III college level (D-III is an indication of relatively small school size and does not permit athletic scholarships). This study seeks to begin to close those gaps while answering new research questions.

3. RESEARCH QUESTIONS

This paper seeks to investigate the following research questions:

- Does homefield advantage exist for a DIII women's soccer team and league?
- Are changes in players wellness and external load measures reflected in the outcome of a match at the team and position level?
- Do the players internal load, external load and wellness measures differ for home versus away matches?

4. METHODS

In selecting the athletes to include in the study several factors were considered. First, three seasons in which the GPS tracking, internal load and wellness data was available for the entire season were selected. During those seasons 50 players had appeared for the team. Given that only 10 outfield players are used in a match, there was significant variation in the data available for the 50 players. The players were therefore ranked based on the following variables: External Load from GPS Trackers (Load, Duration and Distance) as well as Total Games Played. Their ranking on these measures was then added together and divided by the number of measures to find the

average ranking. The top 12 players were then selected, as they had all played more than 20 matches in the time frame.

For the internal and external load analysis, three seasons worth of data were collected for the 12 athletes in the study. The categories of data collected and their respective sources were: External Load (PlayerTek - https://www.playertek.com/us/), sRPE/Internal Load (FitFor90 - https://www.fitfor90.com/) and Wellness (FitFor90). The external load data in Playertek included metrics captured by Catapult GPS trackers during matches showing players total distance covered, duration, high speed running, accelerations and decelerations. The Internal Load data in FitFor90 included responses to surveys that the players filled out after matches indicating the difficulty of the session (rating), and the duration. Session rating is multiplied by duration to arrive at RPE (Rate of Perceived Exertion). The wellness data in FitFor90 included responses to surveys that players filled out daily related to their wellbeing. Measures include: Fatigue, Mood, Soreness, Stress, Sleep Quality and Sleep Hours. A total of 27 different variables were captured, of which a small subset was later found to explain variation in the regression model.

Multiple regression analysis using the enter method was used to analyze the relationship between the Goal Difference (dependent variable) and the sRPE (internal load), external load variables, and wellness variables.

The data was pulled by exporting CSV files with player-level data from the FitFor90 and Playertek sites, as entered by the players. The data was then cleaned, aggregated, and analyzed using R-Studio. Once imported and cleaned, the data was joined to enable the multiple regression analysis to be carried out. sRPE was calculated by multiplying the athlete rating of a session (game), by the duration of the session.

5. RESULTS

HFA was quantified per team, per season, using the widely accepted approach of defining the number of points gained at home as a percentage of all points earned (Pollard, 1986). However, this method has been shown to be impacted by the team's ability and the overall aggregated home advantage of the league during that season. As such, both of these factors need to be accounted for when comparing home advantage from different seasons. Pollard and Gomez (2009) have described a process where team ability and league HFA for a season can be calculated. Maybe add a sentence here to describe that process?

The resulting HFA for the league was 60%, 54% and 58% for the 2017, 2018 and 2019 seasons respectively and 57% over the three seasons, indicating the existence of HFA in the league. The research subject team's HFA for the 2017, 2018 and 2019 seasons was 54%, 57% and 62% respectively and 58% on average for the three seasons. Only regular season matches (40 in total) for the three seasons were included in the HFA analysis in order to have a comparable number of home and away matches.

In response to research question one, the results show that home field advantage does exist both for the team and the league.

Regression Analysis Results

A multiple linear regression analysis was carried out at the team and position level, and results are shown in Table 1. At the team level, the regression results showed an R Square value of 80.4% and an Adjusted R Square value of 49.2%, indicating that the independent variables (sRPE, Distance in Speed Zone 5, Distance in Deceleration Zone 1, Stress, and Sleep Quality) indeed explain nearly half of the variation in the dependent variable of Goal Difference and thus the outcome of a match. The 27 independent variables consisted of several categories as used in similar analyses (Jaspers, 2018) and as shown in Table 3.

The table shows the five independent variables that had a positive correlation with Goal Difference and were significant at the 95% level. No variables were significant for midfielders only, and the share of variability of the dependent variable Goal Difference explained by the independent variables for forwards, was very low (.027). These results show that defenders' independent variables have the strongest correlation with the dependent variable, Goal Difference.

In response to the second research question 'are changes in players wellness and external load measures reflected in the outcome of a match at the team and position level?', the results show that changes to these measures are statistically significant at a position level but not at a team level. See Figure 2 for the differences in mean player load by position.

T-Test Comparing Variables and Matched Pairs

The variables that were significant were therefore selected for further analysis in comparing them for home and away matches. This was carried out to establish whether their means were statistically different during home matches versus away matches. None of the five significant variables had p-values of less than the alpha value of 0.05, indicating that the means were not statistically different for home and away matches based on the sample. However, the effect sizes were still included. A paired samples t-test was used, and the Cohen D statistic was used to measure the effect size of those differences as shown in Table 2 below. These results show that at the team level the mean of the five variables that correlated with Goal Difference, sRPE, Distance in Speed Zone 5, Distance in Deceleration Zone 1, Stress and Sleep Quality were not statistically different in home versus away matches as reflected by their low p-values. Despite the lack of significant difference in this data set sRPE and Sleep Quality had medium effect sizes and warrant further investigation.

Similar to the team-level analysis, the means of the five variables selected based on the regression analysis were not statistically different based on the matched pair t-test. However, their Cohen's D effect sizes are included in Figure 1. Additionally, a player-level analysis was carried out to establish whether the mean of the five variables above was statistically different in their home versus away matches. For one of the players, whose position was forward, the sRPE (internal load) had a p-value of .043 indicating that the mean of their sRPE is statistically different between home and away matches. This suggests that there are likely differences in the way individual players experience and are impacted by home and away matches. A follow up study such as that proposed by Baily (2013) that studies single subjects may shed more light on the impact of home and away matches on individual athletes based on additional measures.

In response to the third research question, "do the players internal load, external load and wellness measures differ for home versus away matches?", the results show that these measures are not statistically different for home versus away matches for players on this team over the three seasons. The differences in the means of the variables that were statistically significant in the regression analysis were not statistically significant at a team or positional level. However, for one of the players, a forward, the mean of sRPE in home versus away matches was statistically different.

6. DISCUSSION

This study used analytics to evaluate performance and wellness of female college soccer athletes on one team over three seasons and found that HFA exists in the SCIAC conference at 57%, and the team whose athletes were selected for analysis had a home advantage of 58%. Five variables, sRPE, Distance in Speed Zone 5, Distance in Deceleration Zone 1, Stress, and Sleep Quality were found to be significantly correlated with Goal Difference. It is notable that of the five variables, one related to internal load, two related to external load and two to wellness. Although the team-level and position-level means of these variables was found to not be statistically different given the sample size, one of the athletes had significantly different mean sRPE between home and away matches. Given that only 12 out of 51 athletes were selected for the analysis (limited by their tenure with the team across all three seasons of study), it is likely more players have internal load, external load and wellness variables that are different between their home and away matches. If coaches and trainers can identify the players with significant differences in these variables for home and away matches, they can intervene or modify their performance to improve individual, position, and team performance.

Two of the three research questions were validated through the research. The study showed that home field advantage does exist at the league and team level over the three years. The study also showed that changes to players external, internal and wellness load measures impact the outcome of a game, as measure by goal difference, particularly at the position level. For the third research question, the study showed that the internal, external and wellness measures were not statistically different in home versus away matches.

Limitations and Future Work – Given the duration of the study was three years, it limited the number of athletes that could be included in the study, which in turn limits the generalizability of the results. Broader (more seasons) and deeper (more athletes) analyses can be carried out to validate the findings of the study. Additional match variables such as wellness or injuries can be added to the study to better understand impacts to players beyond the day of the match. Also, providing insights to athletes and coaches in a short period of time, during or soon after the

game, can help teams make adjustments within or between matches and improve their performance.

7. CONCLUSION

Given the increased importance of sports in society in general and of college sports in particular, understanding team, position, and individual athlete experiences of matches both at home and away is crucial. Since a minority of the athletes are likely to become professional players, understanding how they are impacted by the sport and keeping track of various match variables in home and away matches can improve their experience as students, workers, and family members and improve their individual well-being as well as that of society.

8. ACKNOWLEDGEMENTS

I would like to acknowledge my co-authors for their tireless work on the paper, and thank the reviewers and conference chairs for their constructive feedback and support.

9. REFERENCES

Abbott, W., Brownlee, T., Harper, L., Naughton, R. and Clifford, T. (2018) The independent effects of match location, match result and the quality of opposition on subjective wellbeing in under 23 soccer players: a case study. Research in Sports Medicine. ISSN 1543-8627

Armatas, V. & Pollard, R. (2014) Home advantage in Greek football, European Journal of Sports Science, 14(2): 116-122

Bailey, C. Longitudinal Monitoring of Athletes: Statistical Issues and Best Practices. J. of SCI. IN SPORT AND EXERCISE 1, 217–227 (2019). https://doi.org/10.1007/s42978-019-00042-4

Carron, A.V., Todd M Loughhead & Steven R Bray (2005) The home advantage in sport competitions: Courneya and Carron (1992) conceptual framework a decade later, Journal of Sports Sciences, 23:4, 395-407, DOI: 10.1080/02640410400021542

Davenport, T. H. 2014. "Analytics in Sports: The New Science of Winning," International Institute for Analytics (2014:2), pp.1-28.

Fowler, P, Duffield, T., Waterson, A., & Vaile, J. (2015) Effects of Regular Away Travel on Training Loads, Recovery, and Injury Rates in Professional Australian Soccer Players, International Journal of Sports Physiology and Performance, 2015, 10, 546 -552, DOI: https://doi.org/10.1123/ijspp.2014-0266

Guillaume, R., Urs, G., Boullosa, D., Hackney, A. & Hassane, Z. (2020) How to Use Global Positioning Systems (GPS) Data to Monitor Training Load in the "Real World" of Elite Soccer, Frontiers in Physiology https://www.frontiersin.org/article/10.3389/fphys.2020.00944 11 ,1664-042X

Jaspers, A., Op De Beéck, T., Brink, M.S., Frencken, W., Staes, F., Davis, J., and Helsen, W. (2018) Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning?, International Journal of Sports Physiology and Performance, 2018, 13, 625-630, DOI: https://doi.org/10.1123/ijspp.2017-0299

Kinugasa, T. (2013) The Application of Single-Case Research Designs to Study Elite Athletes' Conditioning: An Update, Journal of Applied Sport Psychology, 25:1, 157-166, DOI: 10.1080/10413200.2012.709578

Pollard, R. (1986) Home advantage in soccer: A retrospective analysis, Journal of Sports Sciences, 4(3): 237-248.

Pollard, R. & Pollard, G. (2005) Long-term trends in home advantage in professional team sports in North America and England (1876 – 2003), Journal of Sports Sciences, 23(4): 337-350.

Sams, M., Wagle, J., Sato, K., DeWeese, B., Sayers, A., & Stone, M. (2020) Using the Session Rating of Perceived Exertion to Quantify Training Load in a Men's College Soccer Team. J Strength Cond Res. 2020 Oct;34(10):2793-2799. doi: 10.1519/JSC.0000000000003793. PMID: 32868677.

Scheck, A. Evaluation of the Validity of the Fit For 90 Subjective Training Load and Wellness Measures. (2017)

Appendices

Category	Variable	Team (Adj R^2 = .421) Beta	Sig.	Defenders (Adj R^2 = .455) Beta	Sig.	Midfielders (Adj R^2 = .252) Beta	Sig.	Forwards (Adj R^2 = .027) Beta	Sig.
Internal Load	sRPE	-.602	.026	-.762	.013	-.261	.406	-1.49	.030
External Load	Distance in Speed Zone 5	.491	.042	.834	.020	.138	.601	-.057	.855
External Load	Distance in Deceleration Zone 1	2.161	.233	.462	.851	2.39	.198	5.55	.026
Wellness	Stress	.398	.011	.500	.004	.093	.647	.101	.712
Wellness	Sleep Quality	.516	.032	.398	.016	-.172	548	.185	.612

Table 1 - Regression Analysis Results

	Home		**Away**		**Difference**		**Effects**	
Variable	Mean	SD	Mean	SD	Mean	SD	Cohen's D	Effect Size
sRPE	486.41	131.10	523.42	95.58	(37.01)	35.52	(0.30)	M
Distance in Speed Zone 5	0.01	0.01	0.01	0.01	0.00	0.00	0.12	L
Distance in Deceleration Zone 1	0.53	0.18	0.48	0.13	0.04	0.04	N/A	N/A
Stress	1.17	0.55	1.23	0.46	(0.07)	0.10	(0.01)	L
Sleep Quality	1.57	0.41	1.50	0.56	0.07	(0.15)	0.31	M

Table 2 - T-Test Values and Effect Sizes

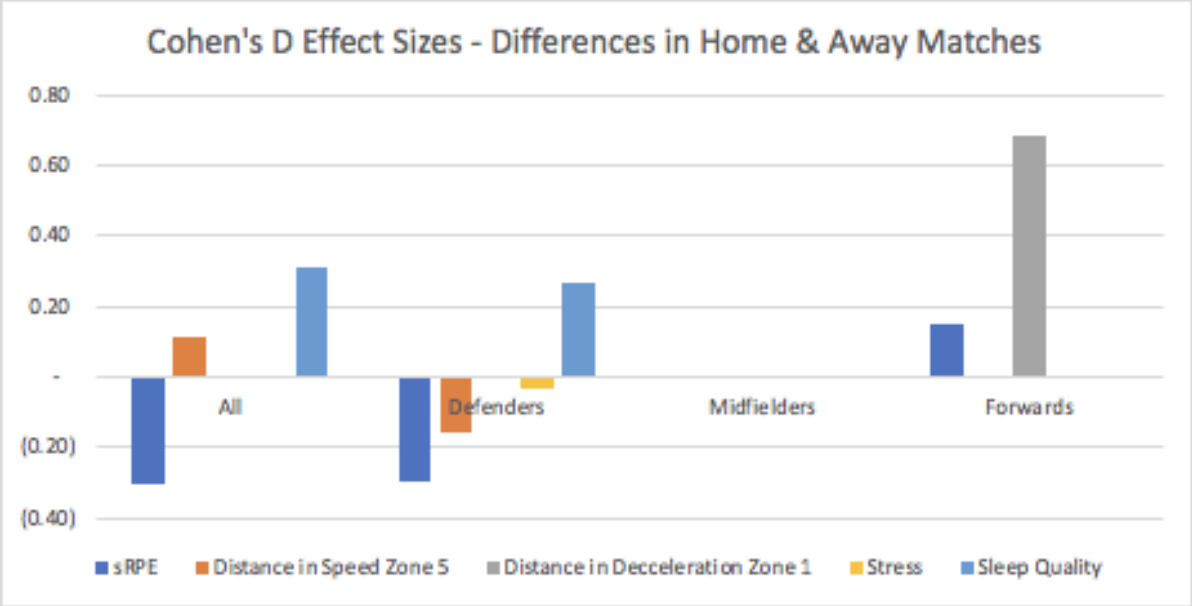

Figure 1 - Effect Sizes of Differences in Means at Position Level

Figure 2 – Player Load by position

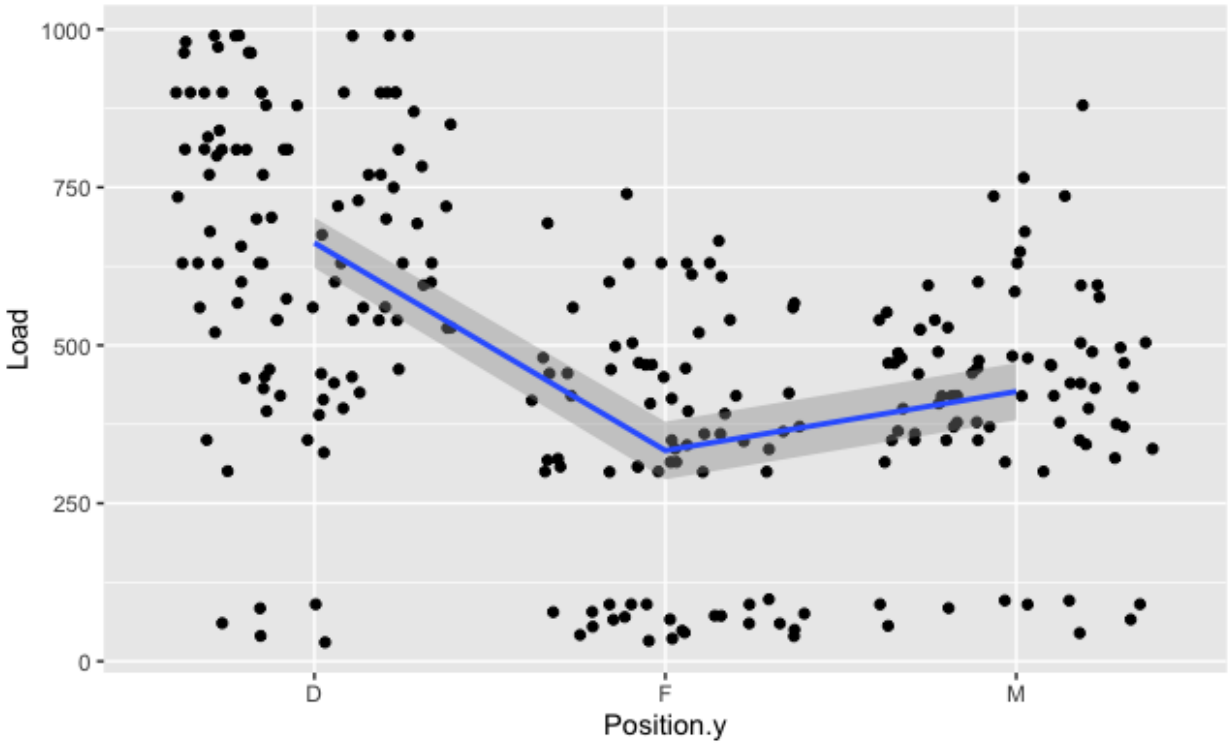

Table 3 – List of Independent Variables

No.	Variable	Category
1	Load (sRPE)	Internal Load
2	Duration	External Load
3	Distance_(miles)	External Load
4	Sprint_Distance_(yards)	External Load
5	Top_Speed_(mph)	External Load
6	Distance_Per_Min_(yd/min)	External Load
7	Distance_in_Speed_Zone_1__(miles)	External Load
8	Distance_in_Speed_Zone_2__(miles)	External Load
9	Distance_in_Speed_Zone_3__(miles)	External Load
10	Distance_in_Speed_Zone_4__(miles)	External Load
11	Distance_in_Speed_Zone_5__(miles)	External Load
12	Distance_in_Deceleration_Zones:_0_-_1_m/s/s__(miles)	External Load
13	Distance_in_Deceleration_Zones:_1_-_2_m/s/s__(miles)	External Load
14	Distance_in_Deceleration_Zones:_2_-_3_m/s/s__(miles)	External Load
15	Distance_in_Deceleration_Zones:_3_-_4_m/s/s__(miles)	External Load
16	Distance_in_Deceleration_Zones:_>_4_m/s/s__(miles)	External Load
17	Distance_in_Acceleration_Zones:_0_-_1_m/s/s__(miles)	External Load
18	Distance_in_Acceleration_Zones:_1_-_2_m/s/s__(miles)	External Load
19	Distance_in_Acceleration_Zones:_2_-_3_m/s/s__(miles)	External Load
20	Distance_in_Acceleration_Zones:_3_-_4_m/s/s__(miles)	External Load
21	Distance_in_Acceleration_Zones:_>_4_m/s/s__(miles)	External Load
22	Fatigue	Wellness
23	Mood	Wellness
24	Soreness	Wellness
25	Stress	Wellness
26	SleepQuality	Wellness
27	SleepHours	Wellness