Turner Schwiebert

Seminar in Data Analytics

Dr. Sarah Supp

September 23, 2022

Domain Review: A Synthesis of Soccer Research

As with all sports analytics, soccer analytics is a very crowded space. Along with being the world's most popular/watched sport, it is also the most researched (Kirkendall, 2020). In fact, much research has already been done on the financial, physical, and performance aspects of the sport throughout the past 20 years (Bush, 2015; Di Salvo, 2007). In addition to published literature, professional teams also conduct their own analyses in an attempt to improve their performance. Thus, there is much variability in both the form and source of research conducted in this space. Soccer research is also very old. Some sources date soccer research back to the creation and organization of the sport in the late 1800s, although it started to become more and more common in published literature by the 1960s (Kirkendall, 2020). Research related to injury, health benefits, training, and performance were all very common in this time period, but research related to finance, wages, and transfers did not become common until the professional level saw rapid financial growth in the 2000s (Morrow, 2014). This is also around the same time when professional teams funded their own research/analytics teams in hopes to improve performance on the pitch. It has been shown that, since investing in data analytics efforts, teams have been able to rapidly improve play year on year, making it more and more difficult to break into the sport's elite clubs both financially and tactically (Bradley, 2016). With this increased focus on the professional game in the past 20 years, published literature has followed suit. In fact, the vast majority of research covers topics related to the modern professional landscape of the sport due

to its increased international coverage (Kirkendall, 2020). In fact, about 95% of published literature on the professional game has been published since 1998 (Kirkendall, 2020). This goes to show the rapid (and recent) rise in soccer analytics. Furthermore, these analytical efforts of professional play have mainly focused on the performance and financial side of the sport. On the other hand, recreational research has focused on the sport's health benefits for non-athletes and amateur athletes alike.

Thus, there are differences in published research across the various levels of play in soccer. As previously stated, a large volume of research has been done on various subjects across the game (Bush, 2015; Di Salvo, 2007). It is also worth noting that the overwhelming majority of published literature focuses solely on professional players (Kirkendall, 2020). This makes sense given the astronomical popularity and financial might the professional game holds. However, there are many differences in how the game is studied across levels of play. For example, amateur play is mainly studied in the context of health benefits or injury prevention. This is evident in one study that analyzes injuries in amateur play in youth and adult levels (Pfirrmann, 2016). This is a common theme across research on amateur-level play. Looking at the professional context, however, there is still much research being done on injury prevention. For example, one study analyzes injury trends across various time periods of a professional player's yearly cycle, such as the off-season or mid-season (Lu, 2022). As previously mentioned, the sphere of professional soccer analytics dominates soccer analytics as a whole. If a topic is researched at the amateur level, it is likely that it is also being researched at the professional level at a much higher volume. This is likely down to the convenience of analyzing topics in a professional context that produces mountains of readily accessible data. On the other hand, amateur-level research likely involves the researcher collecting their own data. This is much

more costly overall. Furthermore, analyzing topics at the professional level is more likely to draw attention due to the popularity of the professional game. One example of this is analyzing performance data of popular players.

As with the previously discussed research topics, analysis of performance data varies widely across published literature in the soccer analytics space. Not only does published literature vary, but unpublished findings generated by professional teams also vary widely. There is a huge amount of topics related to performance data, and more and more are being introduced year on year. One example of the most popular uses of performance data is chance creation. Another is a player's value when they have the ball. The recent development of expected goals per shot is another example of cutting edge analytics in this space (Rathke, 2017). Thus, performance data has a wide range of uses. It is worth noting that extensive performance data is usually unavailable outside of the professional level, so this type of analysis is often limited to professional players only. Nevertheless, descriptive analytics are useful when evaluating how well (or poorly) professional players perform on average. The findings generated from these kinds of studies are interesting to soccer experts and casual fans alike. Thus, cutting edge performance analysis can bring lots of attention from those interested in the sport (of which there are many).

Performance data is also used by both teams and independent researchers to generate valuable insights into tactics in a variety of ways. In other words, some research on performance data is used to optimize performance through tactical decisions. The tactical decisions being made from these insights could result in lineups, positional matchups, situational decision-making, and formational nuances. In other words, it could help coaches determine who to play and who not to play against specific opponents. They could also affect a player's decision

making in certain moments, such as where to pass the ball or how to go about breaking down a defense by analyzing its weak points through their defenders' historical performance data. Outside of actually having an impact on games, tactical analyses can generate interesting conclusions about past seasons/games or predict future outcomes, all of which holds intellectual merit for fans and teams alike. Performance data falls into the "Big data" umbrella since there is a huge volume of historical performance data available to researchers interested in generating these kinds of insights about professional players and teams (Rein, 2016). Even though there has been much research done on the sport, analytics have only played a role in tactical evolution for about 15 years. This means that there is much more to be optimized when it comes to soccer analytics. As previously mentioned, professional teams have been able to improve their play year on year, and much of this is down to this consistent evolution of tactics (Bradley, 2016). This constant evolution means that it is important to constantly revisit older, more traditional styles of play and innovate on them with modern research methods. Professional teams have been doing this kind of work to get ahead of opponents, and much of the cutting edge research is not being published for this exact reason. Professional teams are investing lots of time and money into getting ahead, and it is therefore important to keep these innovations to themselves.

One way data analytics and modern coaching innovation have reshaped the sport is in modern positional play. This area, which is much more relevant to my research project, is at the forefront of soccer analytics. One example of an innovative new insight in this area is that of the wing back: a more versatile version of the more traditional full back (Konefal, 2015). A full back is usually a wide defender who will push up the field slightly to assist in attack, but they will generally stay back as to not leave their team vulnerable. Wing backs, however, run from box to box and will commit to attacking even if they leave space behind them. Teams will use players in

this new positional role to gain the upper hand in various situations. This is the exact kind of tactical innovation that improves teams year on year, and many more are occurring in other positions as well, like the striker. Conventional strikers typically make runs behind the opponent's back line in an attempt to break through and score. Some strikers, however, will try to draw defenders away from the backline and actually run away from the goal, leaving space for other players. Each professional team uses each position somewhat differently, so the amount of nuance from position to position is complex (Bush, 2015). This innovation is not as new, but it serves as another example of tactical innovation in a traditional position. Thus, performance analysis provides insights into future tactical innovations in other areas of the field. While much of this work has been done before, the fact the game is constantly evolving means that finding the current positional roles being used through the professional soccer world is and will always be a knowledge gap in soccer analytics.

Soccer analytics is a space filled with various groups, researchers, and teams providing insights into a wide range of topics including finance, health, injuries, performance, tactics, positions, and even customer satisfaction. Thus, there is a vast amount of published literature out there. When it comes to generating insights into performance, tactics, and positions, not all research is even published. This is due to teams wanting to keep innovations secret to get ahead of their opposition. However, generating insights into the modern game is interesting to fans, and similar methods can be used in other sports. Thus, it is important to understand the current literature that is relevant to tactical and positional evolution.

Works Cited:

Bradley, P. S., Archer, D. T., Hogg, B., Schuth, G., Bush, M., Carling, C., & Barnes, C. (2016). Tier-specific evolution of match performance characteristics in the English Premier League: it's getting tougher at the top. *Journal of sports sciences*, *34*(10), 980–987. https://doi.org/10.1080/02640414.2015.1082614

Bush, M., Barnes, C., Archer, D. T., Hogg, B., & Bradley, P. S. (2015). Evolution of match performance parameters for various playing positions in the English Premier League. *Human movement science*, *39*, 1–11. https://doi.org/10.1016/j.humov.2014.10.003

Di Salvo, V., Baron, R., Tschan, H., Montero, F. C., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International journal of sports medicine*, *28*(03), 222-227.

Kirkendall, D. T. (2020). Evolution of soccer as a research topic. *Progress in Cardiovascular Diseases*, *63*(6), 723-729. https://doi.org/10.1016/j.pcad.2020.06.011

Konefal, M., Chmura, P., Andrzejewski, M., Pukszta, D., & Chmura, J. (2015). Analysis of match performance of full-backs from selected European soccer leagues. *Central European Journal of Sport Sciences and Medicine*, *3*(3).

Lu, D., McCall, A., Jones, M., & Duffield, R. (2022). The relationship between team-level and league-level injury rate, type and location in a professional football league. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2022.03.017

Morrow, S. (2014). Football finances. In *Handbook on the economics of professional football*. Edward Elgar Publishing. https://doi.org/10.4337/9781781003176.00012

Pfirrmann, D., Herbst, M., Ingelfinger, P., Simon, P., & Tug, S. (2016). Analysis of injury incidences in male professional adult and elite youth soccer players: a systematic review. *Journal of athletic training*, *51*(5), 410-424. https://doi.org/10.4085/1062-6050-51.6.03

Rathke, A. (2017). An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise*, *12*(2), 514-529.

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, *5*(1), 1-13.