Science and Medicine in Football

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rsmf20

Data analytics practices and reporting strategies in senior football: insights into athlete health and performance from over 200 practitioners worldwide

Antonio Dello Iacono, Naomi Datson, Jo Clubb, Mathieu Lacome, Adam Sullivan & Tzlil Shushan

To cite this article: Antonio Dello Iacono, Naomi Datson, Jo Clubb, Mathieu Lacome, Adam Sullivan & Tzlil Shushan (14 Mar 2025): Data analytics practices and reporting strategies in senior football: insights into athlete health and performance from over 200 practitioners worldwide, Science and Medicine in Football, DOI: 10.1080/24733938.2025.2476478

To link to this article: https://doi.org/10.1080/24733938.2025.2476478

SCIENCE AND MEDICINE IN FOOTBALL
https://doi.org/10.1080/24733938.2025.2476478

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Data analytics practices and reporting strategies in senior football: insights into athlete health and performance from over 200 practitioners worldwide

Antonio Dello Iacono [a], Naomi Datson [b], Jo Clubb [c], Mathieu Lacome [d,e], Adam Sullivan [f] and Tzlil Shushan [g]

[a]Sport and Physical Activity Research Institute (SPARI), Division of Sport, Exercise and Health, School of Health and Life Sciences, University of the West of Scotland, Glasgow, UK; [b]Department of Sport and Exercise Sciences, Manchester Metropolitan University Institute of Sport, Manchester, UK; [c]Global Performance Insights Ltd, London, UK; [d]Performance & Analytics Department, Parma Calcio 1913, Parma, Italy; [e]Sport Expertise and Performance Laboratory, French National Institute of Sports (INSEP), Paris, France; [f]Sport and Human Performance Research Centre, Health Research Institute, University of Limerick, Limerick, Ireland; [g]Faculty of Science, Medicine and Health, University of Wollongong, Australia

ABSTRACT
Despite the rise of data generation in football, the expertise of data analytics within the sport is relatively underdeveloped. To further understand the landscape, a cross-sectional, observational study design was used to survey practitioners in senior, professional, or semi-professional football. Areas of interest included the personnel involved (the 'who'), the data collected (the 'what'), and the analytical techniques employed (the 'how'). A total of 206 practitioners completed an online survey, with representation from all six FIFA confederations. Of the 206 respondents, 86% were male, 13% female, and 1% preferred not to disclose their gender. Respondents were categorised as working in either the performance (73%), data (18%), or medical (9%) department. Heterogeneity was observed in responses across all departments regarding training load metrics, outcome metrics, methodological attributes, and measurement properties. Evidence sources used prior to implementing a new metric varied between departments, with performance (63%) and medical (67%) staff relying on professional industry and/or community, while data staff (57%) utilised more in-house projects. The analytical approach used most frequently was exploratory data analysis (90%), with modelling, forecasting, and predicting the least frequent (54%). Respondents reported using a mix of solutions for data storage, aggregating and analysing, and reporting and visualising data. Spreadsheets were cited as a popular solution for data wrangling and reporting tasks. The findings provide an overview of current data ecosystems and information systems in modern football organisations. These results can be used to improve data analytics service provision in football by helping identify areas for development and progression.

ARTICLE HISTORY
Accepted 15 February 2025

KEYWORDS
Analytics; decision-making; football; statistics; training monitoring

Introduction

In recent years, sport has seen a rapid rise in data generation, creating new opportunities for comprehensively quantifying and understanding performance (Robertson 2020). Data analytics are applied to inform and optimise decision-making strategies relating to talent identification and development, athletic performance enhancement, and medical service provision (Bartlett and Drust 2021). As part of this trend, the football industry has become data rich, although a lack of clarity on the common infrastructures, resources, and practices used by practitioners to leverage data into actionable insights for player health and performance management remains.

The integration of data analytics expertise in sport is still in its infancy, having been described as 'embryonic', with limited promotion of data-integrated decision support services (Ward et al. 2019; Gregson et al. 2022). Typically, sports scientists within multidisciplinary teams are tasked with translating data into actionable insights for various stakeholders (Bartlett and Drust 2021; Gregson et al. 2022). Scientific training and knowledge of human performance are deemed essential but are not sufficient by themselves to enable sports scientists to effectively apply business intelligence approaches to transform data into actionable insights (Ward et al. 2019). Accordingly, exploring the approaches and practices employed by data specialists within football analytics could drive advancement and innovation in the field (Robertson 2020; Goes et al. 2021).

The effectiveness of data analytics frameworks in enhancing sports performance depends on rigorous data collection, integration, and storage (Lacome et al. 2018). Several studies have highlighted concerns regarding these processes in the sports data analytics landscape. Gerrard and Alamar (Gerrard and Alamar 2014) surveyed men's professional sports leagues, revealing that over a third of these leagues lacked a dedicated database programmer and a fifth did not have an analyst. This study, published over a decade ago, underscores the state of data analytics frameworks at the time, highlighting significant gaps in expertise and infrastructure. Unfortunately, these concerns remain relevant today. For instance, a recent

CONTACT Antonio Dello Iacono antonio.delloiacono@uws.ac.uk Sport and Physical Activity Research Institute (SPARI), Division of Sport, Exercise and Health, School of Health and Life Sciences, University of the West of Scotland, Hamilton International Technology Park, Stephenson Pl, Blantyre, Glasgow G72 0LH, UK
Supplemental data for this article can be accessed online at https://doi.org/10.1080/24733938.2025.2476478

investigation into football academies highlighted a lack of expertise in deploying data analysis insights, with a reliance on off-the-shelf products rather than bespoke in-house solutions (Gregson et al. 2022). Similarly, a more recent study examining data analytics infrastructures in football still found gaps in expertise, resource limitations, and ineffective knowledge translation to key stakeholders (Lolli et al. 2024). Notably, these studies have focused primarily on men's football, with only one study surveying the load monitoring practices in elite women football (Luteberget et al. 2021). With the rapid evolution of women's football and the growing demand for research in this area (Okholm Kryger et al. 2022), it is imperative to understand the current state of data analytics in women's football further.

Given the evolving landscape of data analytics in football (Jayal et al. 2018; Goes et al. 2021), a comprehensive overview of current data ecosystems and information systems in modern football organisations is needed. Understanding the demographics of the personnel tasked with data analytics services (the 'who'), the data collected (the 'what'), and the analytical techniques employed (the 'how') is of interest to practitioners working in football. Additionally, exploring reporting and data dissemination practices can provide insight into the communication and collaboration dynamics within football organisations. Therefore, the aim of this study was to survey and describe data analytics practices in senior football, with a particular interest in the areas of player health and performance.

Materials and methods

Participants

A convenience sample of 206 practitioners working in football were recruited via email, personal or group messaging applications, and invites promoted on social media (e.g., LinkedIn and X, formerly Twitter) through the professional networks of the research team. Eligibility criteria were ≥18 years old; a football practitioner (e.g., any member of data analytics, medical, performance, or support staff); currently work or have worked with senior (≥18 years old) football players competing at professional or semi-professional level; and currently implement or have implemented data analytics practices in the areas of player health and performance as part of their job remit. Participants provided informed consent, and the study received ethical approval by the University of the West of Scotland Institutional Review Committee, United Kingdom (protocol number: 17301).

Design

A cross-sectional, observational study design was used to survey the data analytics practices in senior football. The study followed the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) reporting guidelines for observational studies (Von Elm et al. 2008) (Supplementary File 1: https://osf.io/yasq5). The survey (https://osf.io/pkcs3.) was designed in English language using an online platform (Qualtrics^{XM}: https://www.qualtrics.com.). The survey was developed by a panel of six co-investigators involving academics and practitioners, each with 10 or more years of experience working in professional football. Pilot surveys ($n = 3$ versions) were tested to achieve agreement among the co-investigators. The final agreed version was then reviewed by a pool of academic peers ($n = 2$), sports scientists working in football ($n = 2$) and professionals working in the football analytics industry ($n = 2$). The piloting and review process resulted in minor amendments of the structure and order of the survey's questions, as well as clarifications of the definitions of conceptual constructs and methodological properties of training load and performance outcome metrics. The final version of the survey was publicly released on 23 February 2024, and data were collected until 24 June 2024.

Survey content

The survey included six domains:

- *'Who'*— An array of drop-down lists, multiple checkbox and free-text questions gathering demographic data and professional characteristics of the participants and their working environment.
- *'Evidence, metrics, attributes and approaches'*— An array of multiple checkbox and ranking questions gathering information about a) sources of evidence informing the implementation of b) actionable metrics in the organisation, c) their methodological attributes and properties considered prior to operationalise health and performance constructs of interest and d) analytical approaches to inform decision-making processes.
- *'Where'*— An array of multiple checkbox and free-text questions gathering details about data storage and modelling (i.e., aggregate, analyse), and deployment (i.e., report and visualise) practices implemented in the organisation.
- *'Content'*— Multiple checkbox questions gathering details of tabular and visual summary contents populating data analytics reports.
- *'Target'*— Drop-down list questions gathering information regarding the frequency of sharing and dissemination practices targeting different stakeholders in the organisation.
- *'Live'*— A single question querying whether live data collection inform decision-making during training.

Data handling

Data from questions with preset answers (i.e., multiple choices or ranks) were converted into standardised codes using a designated Microsoft Excel spreadsheet; all automated responses were checked for veracity, and absence of duplicates from the same club or national team was confirmed. The remaining data (i.e., free-text answers) were analysed independently by two authors (Dello Iacono and Shushan) using the same standardised codes. Relevant information was recoded or discarded through a discussion between the same two authors, while two other authors (Lacome and Sullivan) acted as moderators in the case of a disagreement. Upon a consensus achieved among the co-investigators, part of the data from the *'Who'* domain was re-coded to facilitate subgroup analyses, contextualization, and interpretation of the main findings. Specifically, the categorical responses pertaining to the participants' department were re-coded and grouped into three categories such as *data*, *medical,* and *performance* depending on the close affinity of the original responses with these grouping categories. Analysing departmental differences sought to understand whether distinct strategies, approaches, and practices are adopted within each department. Moreover, within the working environment subdomain, responses indicating the organisation's league tier either as third, fourth, or fifth tier, were re-coded and pooled into a single category named *third tier and lower*. Differences between competition tiers reflect the availability and variability of financial, human, and technological resources within football organisations, which data analytics practices largely depend upon. Finally, one objective of the subgroup analyses was to explore differences in data analytics practices between practitioners working with male and female players. However, we observed that only a minority of respondents worked exclusively with female players ($n = 27$; 13%). Given this imbalance, the subgroup analysis could have been biased and misrepresentative of the study findings. Therefore, we decided against conducting this subgroup analysis. The full raw data set is available as Supplementary File 2: https://osf.io/fvzn5.

Statistical analysis

Due to the cross-sectional and observational study design, data are presented using a variety of descriptive statistics (Amrhein et al. 2019). Frequency analysis was conducted for participant characteristics, working environment, multiple-choice, checkboxes, ranking, and rating-scale questions, with the results presented as counts (absolute frequency) and percentages (relative frequency) of respondents. For the subgroup analyses, we calculated the relative frequencies as the number of responses divided by the total number of possible responses per category (e.g., data, medical, or performance). We described results using qualitative terms assigned to determine the magnitude of the observed frequencies as follows: All = 100% of respondents; Most = ≥75%; Majority = 55% to 74%; Approximately half = ~50%; Approximately a third = ~30%; Minority = <30% (Starling and Lambert 2018). Responses involving a numerical answer in single questions (i.e., count data or ranking) were presented as mean and standard deviation (SD) or median and interquartile range (IQR). All statistical analyses were performed using R (Version 2024.04.2, R Foundation for Statistical Computing). The script including the data wrangling, analysis, and plots coding is available as Supplementary File 3: https://osf.io/j6c8v.

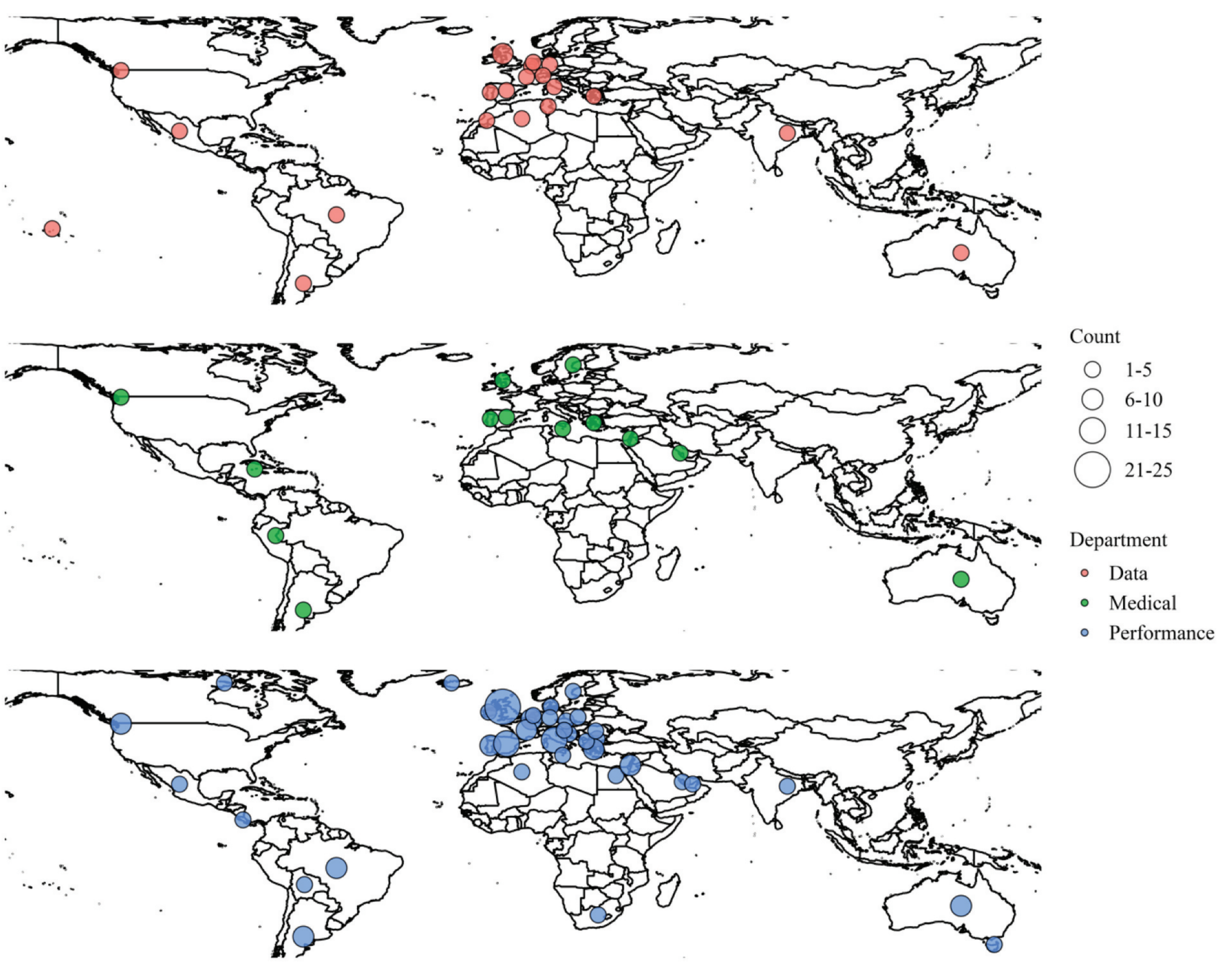

Figure 1. Geographical location of respondents grouped by department.

Results

'The who'

A total of 206 surveys were included in the final analysis. Of these, 176 (86%) were male practitioners, 26 (13%) were female, and 2 (1%) preferred not to disclose their gender. Respondents represented all six of FIFA's confederations (Figure 1), with the majority (68%) working with football organisations based in Europe (UEFA). This was followed by respondents from the Confederation of North, Central America, and Caribbean Association Football (CONCACAF; 10%), Asia (AFC; 9%), South America (CONMEBOL; 8%), Africa (CAF; 4%), and Oceania (OFC; 1%). The distribution of respondents across departments included those working in performance (*n* = 150; 73%), data (*n* = 38; 18%), and medical (*n* = 18; 9%).

Across the three departments, there was a relative balanced distribution regarding competition level (tier), national team representation, and the sex of the players (Figure 2). Of the total respondents, 127 (62%) indicated they work exclusively with male players, 27 (13%) worked solely with female players, and 52 (25%) reported working with both within their organisations. We provide an interactive version of Figure 2 as a supplementary file (Supplementary File 4: https://osf.io/uch6n) to enable readers to conduct a comprehensive exploratory analysis by hovering over the nested domains.

'Evidence, metrics, attributes and approaches'

Ranking responses regarding evidence sources before implementing a new metric are presented in Figures 3 and 4. Notably, performance (63%) and medical staff (67%) frequently relied on the professional industry and/or community, while data staff (57%) leaned more on in-house projects for guidance (Figure 3). The frequency of rankings across tiers was identical (*n* = 43; 57%, *n* = 16; 57%, and *n* = 12; 57% for top, second and third tier, and lower, respectively) with most respondents ranking the professional industry/community first. In national teams, both professional industry/community and in-house projects (*n* = 9; 43% for both) were ranked equally (Figure 4).

Figures 5 present the top two ranked operationalised metrics of training load constructs across respondents' departments according to common use. Across the three departments, locomotive variables (e.g., speed-zone-based

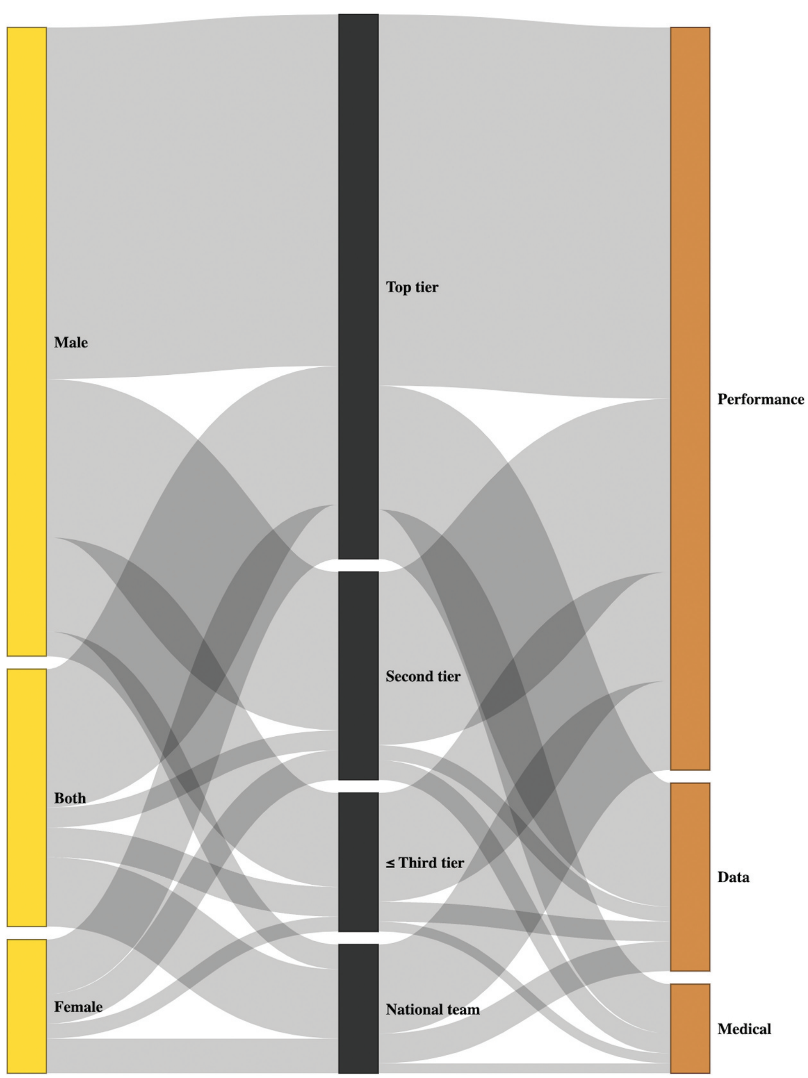

Figure 2. Sankey plot of respondents' departments across tiers and players' sex.

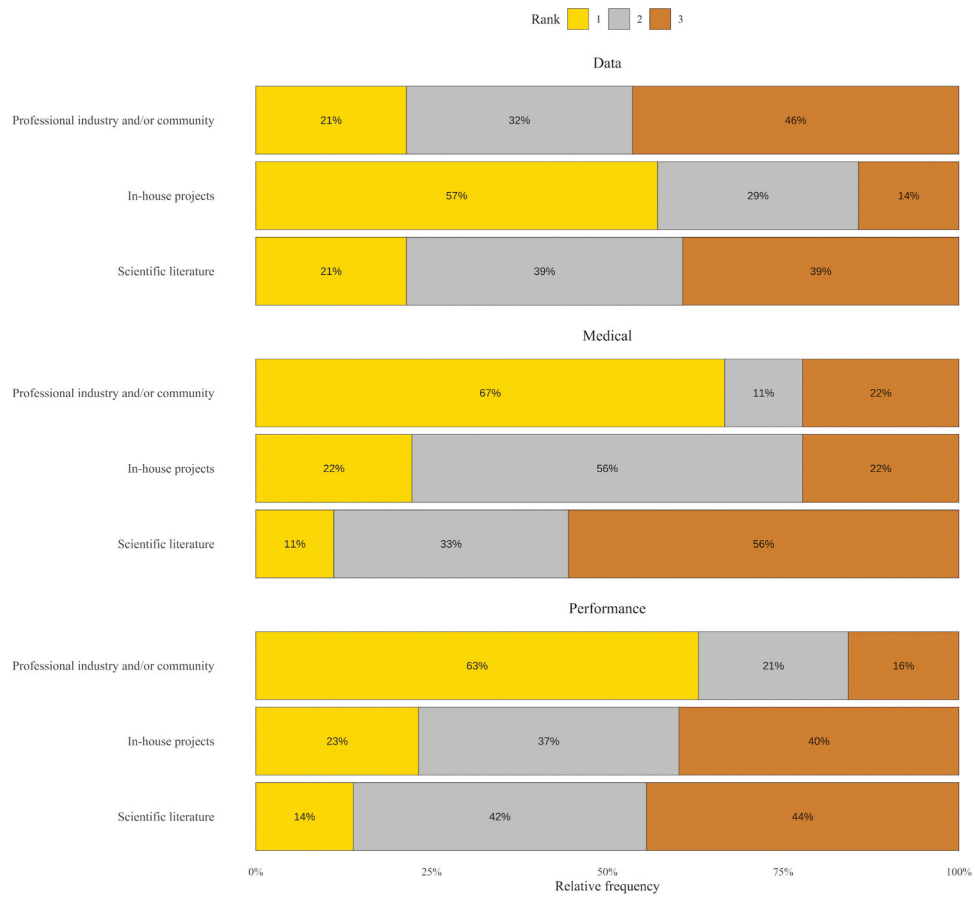

Figure 3. Source ranking grouped by department.

distances or counts) were the most commonly used (69% to 72%, dark grey), while perceptual measures (e.g., ratings of perceived exertion) were the second most used among medical and performance staff (50% and 45%, respectively, light grey). Data staff indicated more evenly distributed responses for their second-ranked metric (26% to 32%, light grey).

For performance outcome metrics, field-based and gym-based testing procedures were ranked first (58% to 83%, dark grey) and second (53% to 72%, light grey) according to common use, respectively, with similar trends observed across the three departments (Figure 6).

Respondents provided heterogeneous responses regarding their focus on the methodological attributes of a new metric before implementation, with no clear trend favouring any particular measurement property (Figure 7).

Regarding analytical approaches, respondents indicated a greater use of exploratory data analysis ($n = 192$; 90%), while the least implemented approaches included modelling, forecast, and predictions ($n = 141$; 54%). These trends were consistent across the three departments (Figure 8). A left skewed distribution was presented when respondents were asked about the importance of analytical approaches in decision-making, with exploratory data analysis showing a trend toward being rated higher in importance compared to other analytical methods (Figure 9).

'Where and content'

Reflecting the subgroup analysis investigating the availability of technological resources, Figure 10 displays the frequency of storage services used, stratified by player sex and tier. The majority of respondents ($n = 158$; 73%) reported to use more than a single storage service with cloud-based systems combined to local/in-house storage ($n = 113$; 55%) or analytical companies' services ($n = 100$; 48%) representing the majority and approximately half, respectively. We provide an interactive version of Figure 10 as a supplementary file (Supplementary File 5: https://osf.io/qdpsk) to enable readers to conduct a comprehensive exploratory analysis by hovering over the nested domains.

The choices across cloud-based systems and analytical company services are further illustrated in word cloud visualisations, presented as Supplementary File 6 (https://osf.io/7xcpd.) and Supplementary File 7 (https://osf.io/t56ch).

Supplementary File 8 (https://osf.io/u7crw.) and Supplementary File 9 (https://osf.io/n52ku.) display the frequency of software used for aggregating & analysing and reporting & visualisation tasks, with spreadsheets representing the choice used by most ($n = 157$; 76%) and the majority ($n = 128$; 62%) of respondents, respectively.

Finally, Figure 11 illustrates the frequency of contents used in data analytics reports and visualisation strategies, with summary tables representing the most common choice ($n = 189$; 92%) used

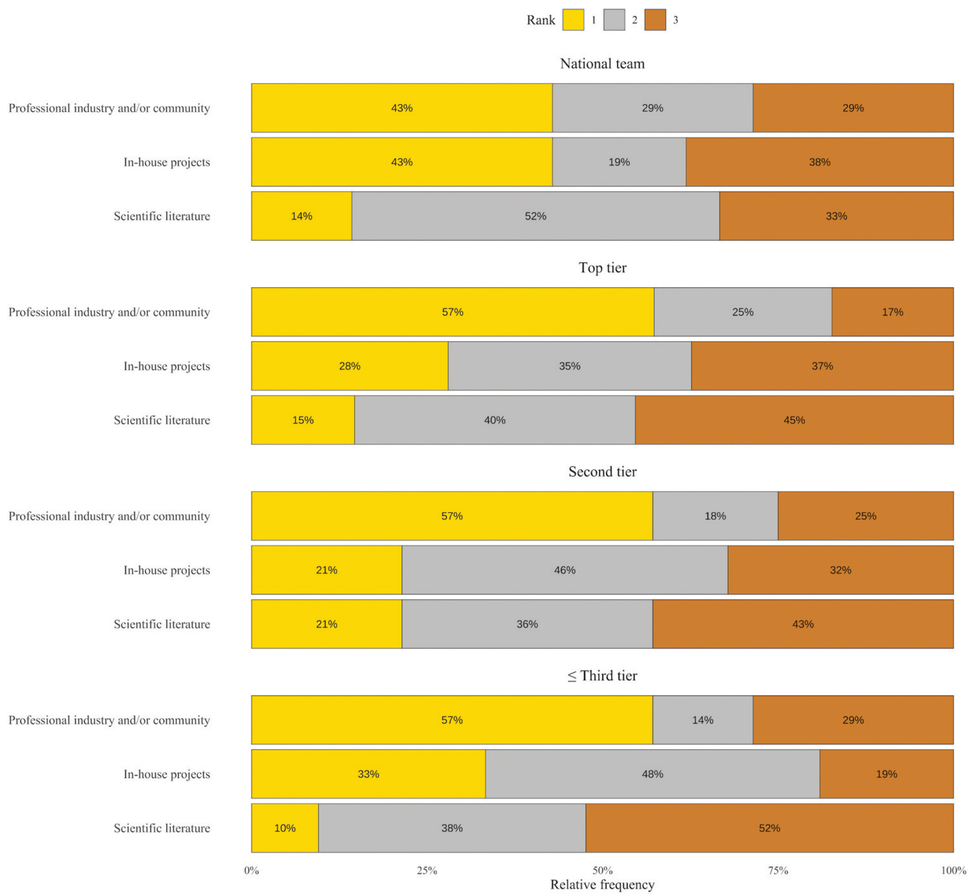

Figure 4. Source ranking grouped by tier.

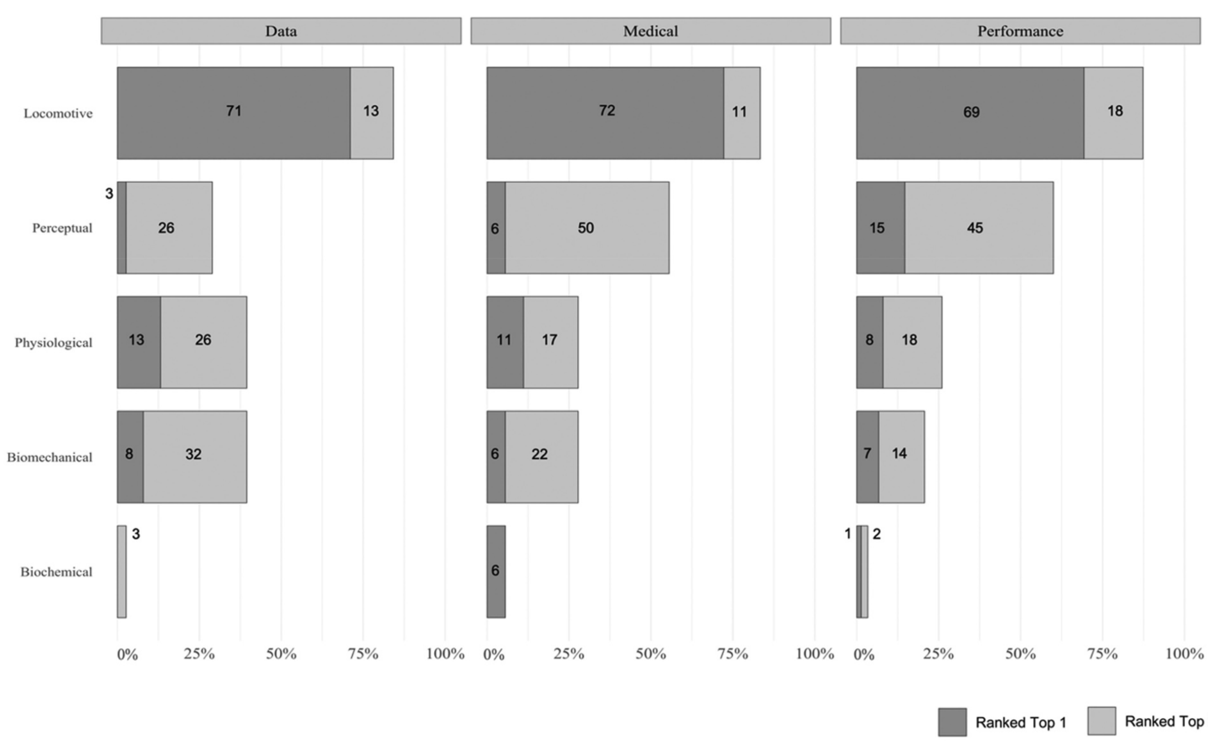

Figure 5. Top two ranked (according to common use) metrics to gauge insights on training load grouped by department.

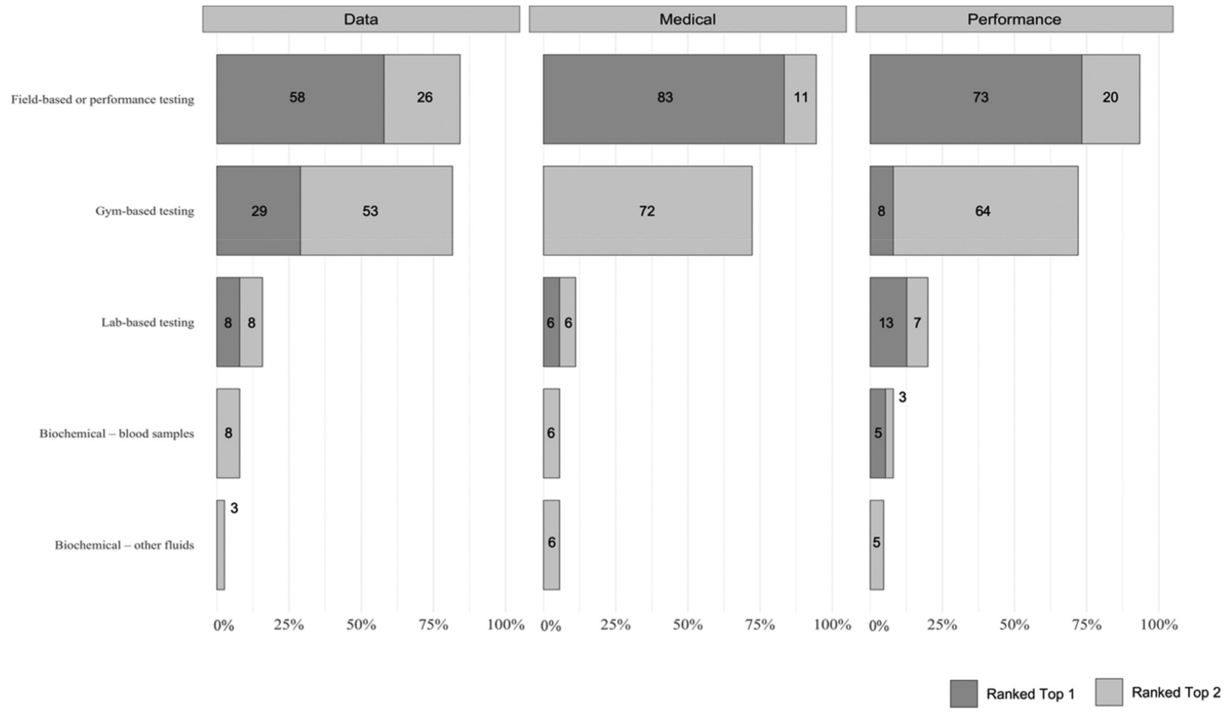

Figure 6. Top two ranked (according to common use) metrics to gauge insights on performance outcomes grouped by department.

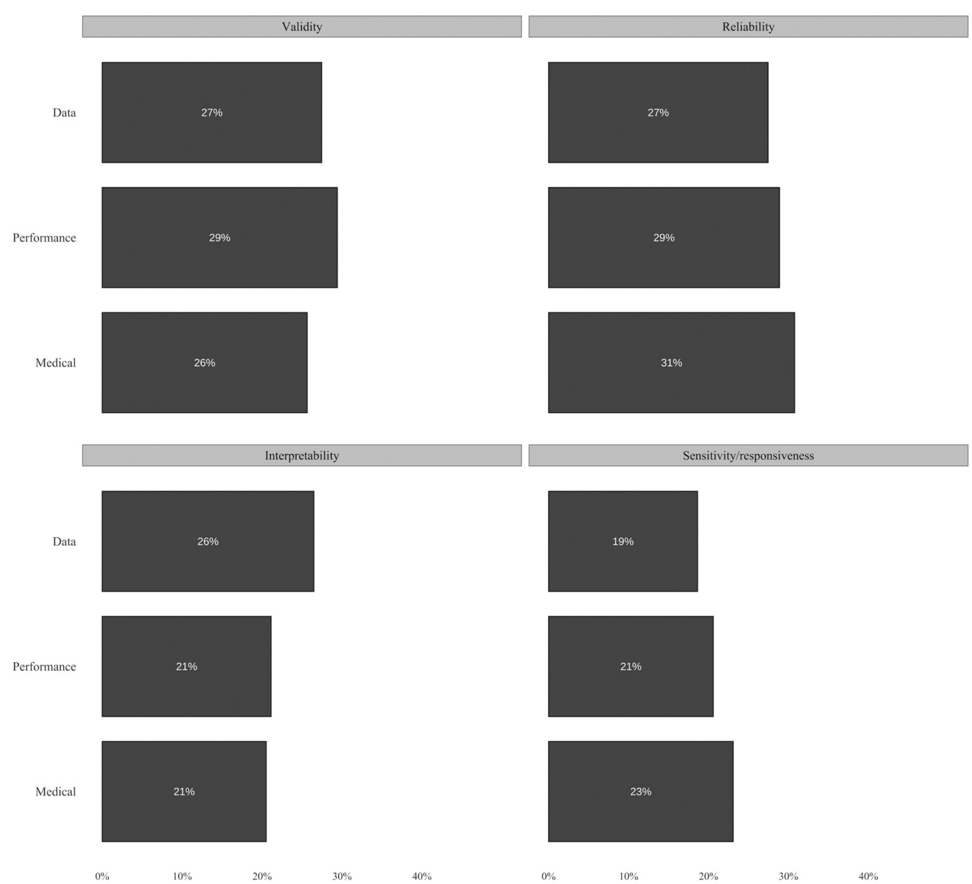

Figure 7. Frequency of the methodological attributes grouped by department.

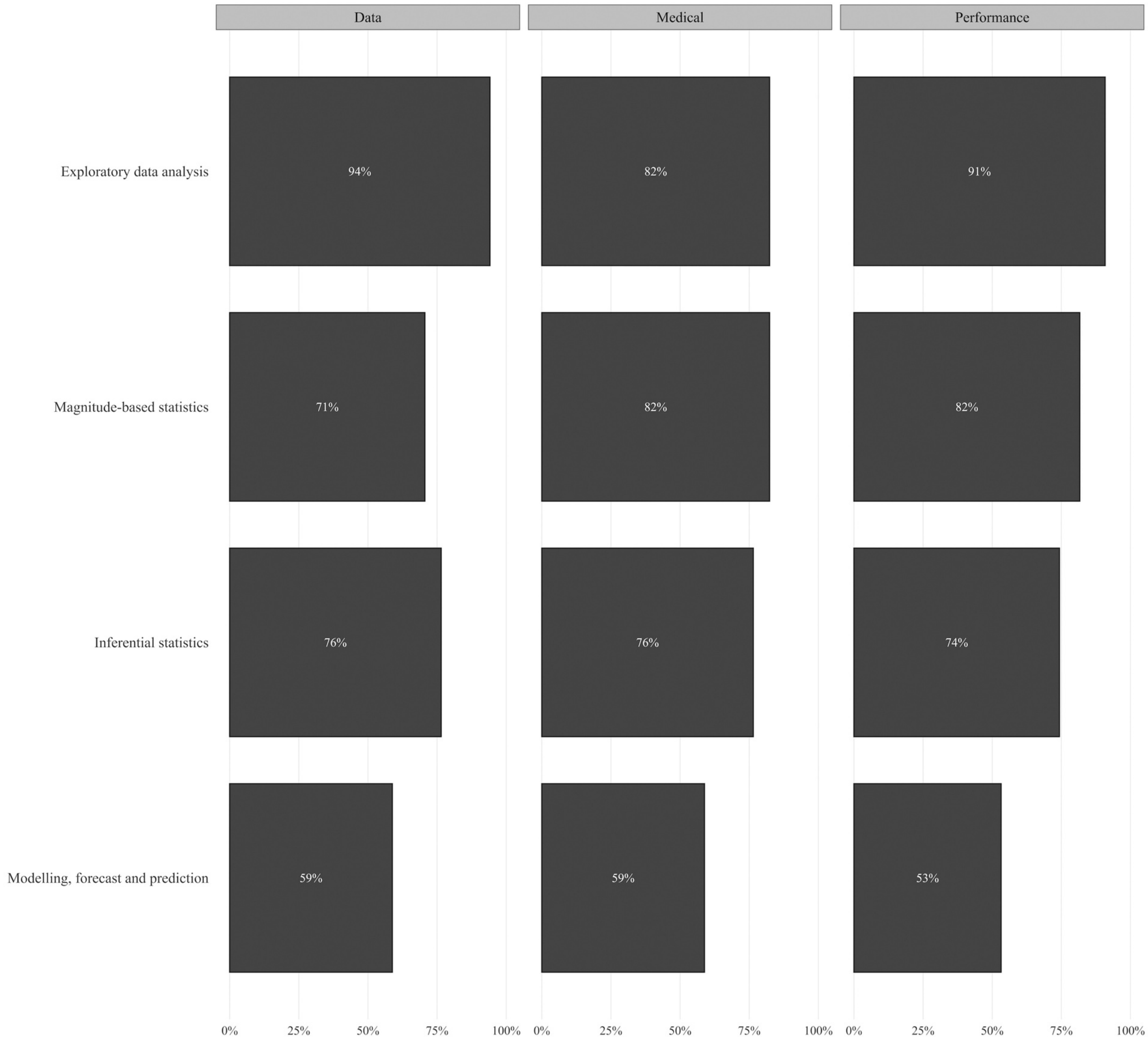

Figure 8. Frequency of the analytics approached used in the organisation grouped by department.

by the respondents. We provide a bar plot version of Figure 11 as a supplementary file (Supplementary File 10: https://osf.io/4xtq6.) to enable readers to identify the frequencies associated with each data analytics report and visualisation content.

'Target'

Table 1 presents a heat map illustrating the frequency of reporting and dissemination to different target staff groups (coaching, performance, and medical), across the different tiers and national teams. Daily and weekly reporting communications were the most common timeframes for reporting across all tiers, national teams, and target staff groups.

'Live'

The majority of respondents working in first or second tiers, or national teams (76% to 79%), indicated regularly using live data to inform immediate training modifications at both the squad and individual levels. In contrast, approximately half (54%) of respondents in third or lower-tier levels reported using live data (Figure 12).

Discussion

The aim of this study was to survey and describe data analytics practices in senior football, focusing on the areas of player health and performance. A total of 206 practitioners, representing all six FIFA confederations, participated in the survey. The majority of respondents were male and held roles within the performance department. The study revealed heterogeneity in the use of training load and performance outcome metrics, as well as their methodological attributes and measurement properties. Practitioners relied on different sources of evidence when adopting new analytics metrics. Exploratory data analysis emerged as the most commonly used analytical approach, with modelling, forecasting, and prediction being the least utilised. Participants reported using a range of tools and solutions for data storage, aggregation, analysis, reporting, and visualization.

Figure 9. Importance attributed to analytics approached used in the organisation. Note: the dots represent individual datapoints; the thick vertical lines and coloured areas in the boxes represent median values and interquartile ranges, respectively.

Reporting and dissemination practices varied by organisational level.

The 'who'

The breadth, roles, levels of expertise, and professional environments within the surveyed sample contribute valuable knowledge and insightful perspectives for practitioners interested in data analytics practices in football. To elaborate, most of the respondents held a postgraduate ($n = 124$; 60%) or higher degree ($n = 49$; 24%) in a field relevant to player health, data, and performance, coupled with an average of more than 10 years of professional experience. Notably, the surveyed sample represents clubs from 61 distinct national leagues, with the majority from top-tier ($n = 110$; 53%) and second-tier ($n = 42$, 20%) levels (Figure 2). Of these, approximately half ($n = 82$; 46%) have competed in international competitions over the past 3 years. Additionally, approximately half of the remaining 54 respondents were employed by national teams. Therefore, we can reasonably assume that our findings are representative of data analytics practices in top-level football worldwide.

We initially intended to conduct a subgroup analysis to explore differences in data analytics practices between men's and women's football, which proved unfeasible because only a minority of respondents worked exclusively with female players (Figure 2). However, with more than a third of respondents employed in women's football environments, our findings can be considered somewhat representative of women's football. Notably, our study expands upon the work of Luteberget and colleagues (Luteberget et al. 2021), who examined load monitoring practices in elite women's football. With a larger sample size that includes representatives from all FIFA confederations and national teams, spanning roles beyond the performance department, our study provides deeper insights into data analysis, reporting, and visualization practices. Conceptually, establishing data analytics frameworks within

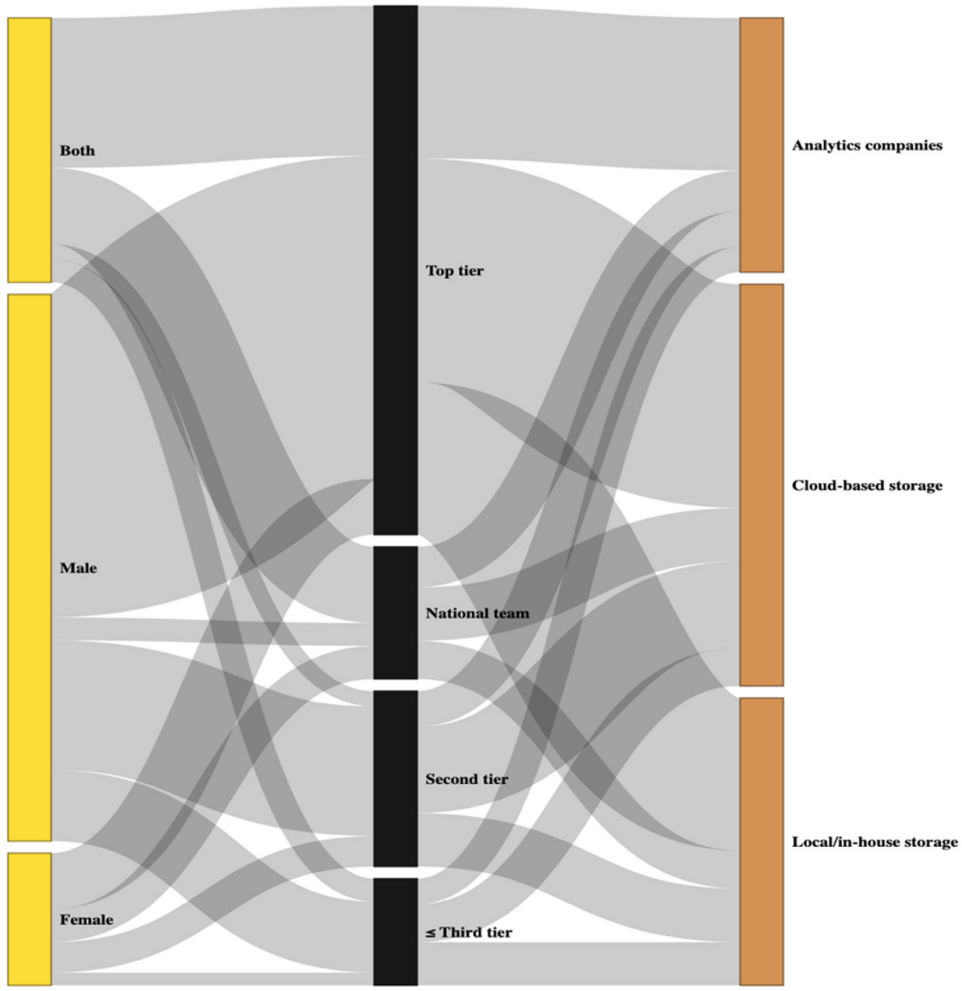

Figure 10. Sankey plot of storage services across tiers and players' sex.

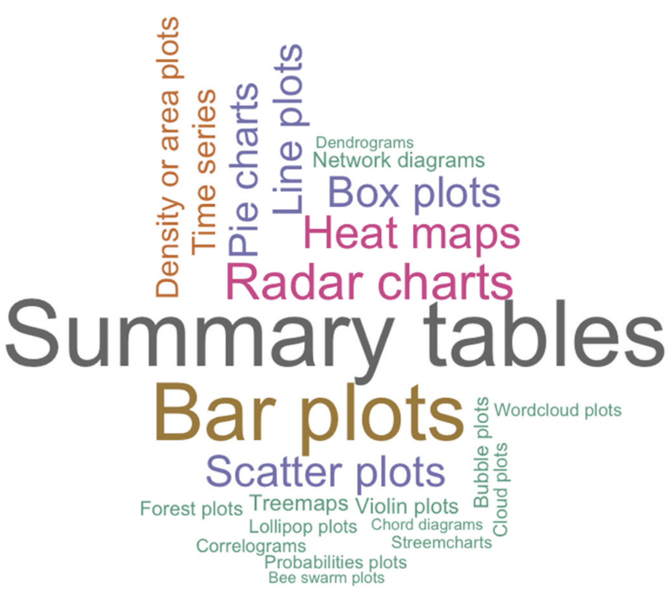

Figure 11. Frequency of data summary and visualisation contents data in the organisation.

a football organisation requires the same infrastructures, resources, expertise, and processes regardless of the players' sex. However, differences likely exist at financial, social, and contextual levels. The type and amount of data collected, combined with the data science knowledge and analytical skills required, significantly increase financial demands related to equipment acquisition, staff time, and the recruitment of specialist expertise. These demands can differ between men's and women's football, and recent evidence suggests that greater financial resources in men's football allow for more extensive data collection compared to women's football (Houtmeyers et al. 2021). While the sex gap disparities are expected to narrow in the future due to substantial investments made by governing bodies in women's football (Beissel et al. 2024), the sport remains a male-dominated industry, with significantly lower funding available in women's football to date and fewer female employees in medical, performance, and data roles compared to their male counterparts (Luteberget et al. 2021; Bryan 2022). Our investigation confirms this gender disparity, as only 13% of the surveyed sample was female, compared to 86% male (1% preferred not to disclose), and only 13% worked solely with female players compared to 62% who worked solely with male players. At a contextual level, while the same conceptual frameworks and constructs, analytical and reporting processes are used to leverage data analytics and inform health and performance services in both men's and women's football (Luteberget et al. 2021), differences may arise from women-specific benchmarks appraisals related to key performance indicators (Harkness-Armstrong et al. 2022), injury risk factors (López-Valenciano et al. 2021; Okholm Kryger et al. 2023), physical performance outcomes, and monitoring female health (Beato et al. 2023). These areas should be prioritised in the research agenda and in data analytics projects focused on women's football, particularly in relation to health and performance. For example, conducting interviews and surveys with players and coaches can help identify their needs and provide insights into where and how data analytics efforts should be directed to address these priorities effectively.

The demographic data revealed a lack of equal representation among the respondents' working departments. The majority were employed in the performance department, while those in data and medical departments represented only a minority within the surveyed sample. Similar findings were observed when stratifying the responses across tiers and national teams, with this disparity becoming more pronounced at lower competitive tiers (i.e., second tier and lower; see Figure 2). This finding aligns with previous studies indicating that performance staff outnumber other staff roles in football multidisciplinary teams. In many cases, sports scientists are tasked with translating data into actionable insights (Gerrard and Alamar 2014; Gregson et al. 2022; Lolli et al. 2024). A logical explanation for this disparity is that the financial resources available to a club largely determine the scale of the data analytics department (Houtmeyers et al. 2021; Gregson et al. 2022). Furthermore, the strategy of the organisation may also influence the existence of a dedicated data analytics department. For instance, fully decentralised organizations may outsource data analytics services and rely on off-the-shelf products rather than bespoke in-house solutions (see 'Where and content' section), with medical and performance staff being primarily confined to the role of service consumers (Ratten 2016; Casals and Finch 2017). Future studies should explore the dynamics of data strategies and action chains that influence decision-making across football organisations. Such research should address the interplay between financial resources, organisational structures, and data operations across departments (Houtmeyers et al. 2021). Additionally, understanding how these factors shape the integration of data-driven processes into routine practice could identify barriers to adopting bespoke, in-house analytics solutions. Finally, investigating how clubs and organisations at different competitive levels overcome resource limitations to implement effective data analytics strategies would also be valuable.

'Evidence, metrics, attributes and approaches'

The analysis of the sources of evidence revealed some interesting findings. Firstly, scientific literature was the least preferred source of guidance, regardless of the respondents' roles, professional context (e.g., clubs or national teams), or competitive level (e.g., tiers). A plausible explanation is that player management processes have advanced more rapidly within applied settings than within the scientific community. Additionally, research conducted in well-controlled lab-based environments (i.e., *efficacious* research, or 'does it work?') does not always translate directly into practice in high-performance settings (i.e., *effective* research, or 'does it work in practice?'), making it difficult to implement (Fullagar et al. 2019). Nevertheless, knowledge from scientific literature may still influence the professional and industry community through indirect dissemination pathways such as workshops, webinars, and other forms of advanced professional training, even if the primary sources (i.e., peer-reviewed papers) are not directly consulted by the surveyed participants. In this way, the scientific literature may function as an indirect source of guidance, as many respondents identified the professional community and their practices as key sources of information (Luteberget et al. 2021; Houtmeyers et al. 2021; Asimakidis et al. 2024). Secondly, while the majority of performance and medical staff reported seeking guidance from the professional or industry community, respondents in data-related roles were more reliant on in-house projects (Figure 3). This finding aligns with the notion that performance and medical staff tend to trust the face validity of accepted beliefs and practices within their peer groups (Fullagar et al. 2019). In contrast, data department staff's reliance on in-house projects is more consistent with the methodological rigor and systematic approach required for the use of metrics in data-informed fields. Implementing new metrics in football depends not only on the quantity and accuracy of the data but also on other pivotal factors such as feasibility, complexity, and usability, all of which must be carefully evaluated before the implementation process (Houtmeyers et al. 2021).

The analysis of the metrics confirmed the expected results. First, locomotive and perceptual measures were the top- and second-ranked common operationalised metrics of training load among the majority of respondents and approximately half, respectively, regardless of their role (Figure 5). This is not surprising, given the consensus around the training load framework and the load management practices consolidated in the football industry over the past 15 years (Akenhead and Nassis 2016; Impellizzeri et al. 2019; Jeffries et al. 2022). Second, field- and gym-based testing was the top- and second-ranked common settings for collecting performance outcome metrics among the majority of respondents, again regardless of their department (Figure 6). This preference likely stems from the fast-paced demands of applied football contexts. Field- and gym-based testing offers pragmatic solutions by providing direct or surrogate (proxy) measures of performance. These methods are also time-efficient and non-invasive compared to lab-based tests or obtrusive procedures that require the collection of blood or other biochemical fluids (Asimakidis et al. 2024). As a result, they facilitate more immediate application to performance evaluation and training programming, making them the preferred practices for collecting performance outcome metrics.

Collectively, we observed considerable heterogeneity in responses across all departments regarding the methodological attributes and measurement properties (Figure 7). A simple, yet reasonable explanation for this finding is that the decision to prioritise specific attributes and their associated measurement properties depends on the theoretical construct being assessed (Impellizzeri and Marcora 2009). To illustrate this point, we present the following contemporary example. Submaximal fitness tests have recently gained interest in football as a practical approach for evaluating physiological state and changes in performance (Shushan et al. 2022). When using a validated protocol with reliable test measures, practitioners may aim to investigate whether changes in test outcomes strongly correlate with changes in the indicators associated with the construct of interest. Alternatively, others might prioritise understanding whether comparisons of test outcomes against normative or baseline data can identify minimal important and detectable changes with high credibility or probability. In these scenarios, the emphasis would be on sensitivity and interpretability, respectively, potentially bypassing the assessment of validity and reliability attributes. Conversely, when developing a new submaximal fitness test, the validity and reliability attributes may be given higher hierarchical importance, underscoring the general principle that methodological attributes and measurement properties assessment are context and goal specific.

The analysis of responses regarding the analytical approaches revealed somewhat contrasting results. Specifically, approximately half of the respondents indicated that they do not use modelling, forecasting, or prediction as analytical approaches. A minority also excluded magnitude-based and inferential statistics, with exploratory data analysis indicated as the most utilised approach (Figure 8). These findings are consistent with two other surveys on load monitoring practices in elite men's and women's football (Luteberget et al. 2021; Houtmeyers et al. 2021), which revealed that only a minority of surveyed practitioners reported using machine learning techniques for data analysis. In contrast, in our study, among those confirming to use these approaches and asked to rate their importance, we observed left-skewed scores distributions, indicating overall high importance for all approaches similarly across departments (Figure 9). Exploratory data analysis was consistently rated the highest for importance, despite the responses being spread across the entire scale. This finding was expected, as exploratory data analysis is a fundamental step in descriptive statistics. It enables an initial understanding of data distributions and offers quantitative summaries that inform subsequent statistical tasks, such as comparing means, measuring associations between variables, and calculating effect sizes. These tasks are considered some of the simplest statistical models (French and Ronda 2021) and were reported as the most common data analysis tasks in recent surveys among practitioners working in elite men's and women's football (Luteberget et al. 2021; Asimakidis et al. 2024).

The less frequent use of modelling, prediction, and inferential approaches may reflect limited data analytics literacy among part of the surveyed sample, as well as conceptual and methodological constraints inherent to football environments. For instance, while some studies in football have explored the potential of modelling, forecasting, and prediction to anticipate outcomes like injuries, there is limited robust evidence supporting the accurate prediction of training process data (Rossi et al. 2018; Rommers et al. 2020). Additionally, the volume of training data collected in football is relatively small compared to other big data domains, such as finance, thus limiting the efficacy of these approaches only to niche areas. Magnitude-based analytics, which depend on predefined smallest effect size of interest, face challenges in football because defining and estimating these effect sizes is not always easy, clear, or justifiable from a causal perspective. Finally, inferring causation is crucial in player management, as recommending treatments and training interventions without proven causal benefits will not enhance player health or performance. To address this, future studies must be designed according to established guidelines leading to robust and high-quality evidence (Higgins et al. 2019). Unfortunately, such rigorous designs, like randomised controlled trials, are uncommon in applied football settings, where interventional, control-group comparisons are rarely feasible. Consequently, analytics approaches aimed at establishing cause-and-effect relationships may be perceived as unnecessary and are often underutilised. Nevertheless, future efforts should focus on creating clear guidelines and recommendations for implementing advanced analytics modelling and standardising reporting practices, analogous to those commonly employed in medical and clinical fields (Collins et al. 2024). These guidelines should include frameworks for validating models, methodologies for addressing the challenges of small data sets, and best practices for transparently reporting analysis processes and results. Such initiatives would enable practitioners to evaluate case studies critically, replicate findings, and assess the robustness of analytics models within their contexts.

'Where and content'

Responses regarding data storage revealed considerable heterogeneity across professional contexts (e.g., clubs or national teams) and competitive levels (i.e., tiers). Respondents reported using a mix of local storage, cloud-based services, and analytics company services for storing data within their organisations (Figure 10). A similar diversity was observed in the software used for aggregating & analysing (Supplementary File 5: https://osf.io/7xcpd.) and reporting & visualising data (Supplementary File 6: https://osf.io/t56ch). Interestingly, spreadsheets emerged as the most used tool for data wrangling and reporting tasks compared to databases and coding software. This preference can be attributed, in part, to strategic factors (e.g., decentralised outsourcing), financial constraints (e.g., budget availability), and expertise-related issues (e.g., data literacy and analytics skills), as previously discussed (see 'Who' section). Additionally, from a practical standpoint, spreadsheets are often perceived as more user-friendly (Asimakidis et al. 2024). They are cost-free, require minimal setup

Table 1. Frequency (%) of data analytics reports dissemination across target stakeholders and tiers.

Target	Tier	Daily	Weekly	>= Monthly	Never
Coaching staff	National team	12	6	7	2
	Top tier	63	25	12	4
	Second tier	21	16	15	0
	Third tier or lower	13	8	4	2
Medical staff	National team	13	8	3	3
	Top tier	50	35	9	10
	Second tier	20	13	6	3
	Third tier or lower	10	6	6	5
Performance staff	National team	16	7	4	0
	Top tier	67	27	6	4
	Second tier	27	7	2	6
	Third tier or lower	13	7	5	2

Legend	Minority: 0-30%	Approximately a third: 31-49%	Approximately half: 50-54%	Majority: 55-74%

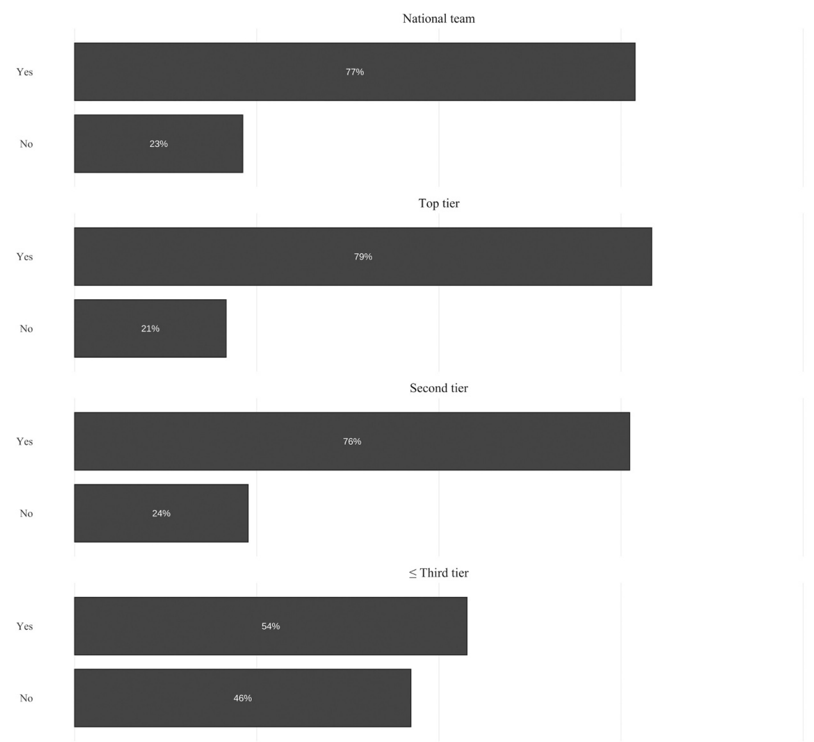

Figure 12. Use of live data to inform ongoing training sessions modification.

time, and are easy to use by multiple staff members. However, it is likely that a combination of methods is often employed for data wrangling and reporting, depending on the analytical approach and the specific goal of the analysis – whether descriptive, predictive, or prescriptive, as justified by causal inference. Supporting this assumption, respondents also reported using a variety of methods and combinations for summarising and visualising data (Figure 11). This may indicate the tendency to create ad-hoc data visualisations based on the target audience. Notably, summary tables and bar plots were the most frequently used, suggesting that descriptive statistics remain the most fundamental and common objective of data analytics practices in player health and human performance within the surveyed sample. Promoting progress in the field, interdisciplinary collaboration between data professionals and practitioners in football should be encouraged to ensure the practical relevance of data analytics applications. Accordingly, education programs designed to enhance data literacy within multidisciplinary teams and the development of user-friendly tools for data reporting, dissemination, and visualisation can further facilitate the adoption and effective implementation of analytics approaches in football settings.

'Target'

Regardless of the role of the dissemination audience, data analytics reports are most commonly disseminated on a daily or weekly basis (Table 1). This suggests that reporting practices primarily support decision-making in training programming – such as planning, monitoring, evaluating, and adjusting training prescription – quantifying acute training effects, and assessing player availability for training and matches. Minor differences were observed among national teams and lower-tier clubs (third tier and lower), where monthly reports are slightly more common. For national teams, this is likely attributed to the nature of service provision, which is less frequent compared to the day-to-day operations of professional clubs (Buchheit and Dupont 2018). In lower-tier clubs, limited financial and human resources likely dictate less frequent reporting.

'Live'

The differences in live monitoring usage between national teams, top- and second-tier clubs, and lower-tier clubs can be attributed to variations in human and technological resources (Figure 12). National teams and higher-tier clubs typically have the necessary staff and technology to make real-time adjustments during training sessions. In contrast, third and lower tier clubs, often semi-professional or developmental in some countries, operate with more limited resources, which likely restricts their ability to fully utilise live monitoring, resulting in approximately half of the respondents from these clubs relying on it.

Limitations

Despite the comprehensive nature of this survey, our study is not without limitations. First, the study utilised a convenience sample which may not be fully representative of the broader population and different roles of football practitioners. Whilst we believe our survey had significant reach, as emphasised by the large sample size ($n = 206$) and the wide geographical spread (all six FIFA confederations represented), it must be recognised that most respondents were male (86%), and the majority were based in Europe (68%) (UEFA). Second, the lack of a clear classification between professional and semi-professional levels may have impacted the accuracy of the respondent groupings and limited the comparisons between these groups. Since the distinction between professional and semi-professional levels is contextually dependent on the country and league, and as even lower-tier clubs may operate professionally in some contexts, using a tier-based approach for classification may not fully capture the nuances of these categories. This potential ambiguity could have increased variability and influenced the interpretation of results, particularly when comparing practices and approaches across competitive levels. Third, the survey only being available in a single language (English) has likely limited the ability to capture a more diverse dataset. These factors potentially limit the generalisability of the findings to wider contexts. Fourthly, the aim of the study was to understand the current landscape of data analytics in men's and women's football. This aim would have been realised via completion of a subgroup analysis to explore differences in data analytics practices. However, this subgroup analysis was not conducted as only a minority (13%) of respondents worked exclusively with female players, while 25% of respondents stated working with both male and female players within their organisation. Nevertheless, these respondents represented a total of 38% working with female players, suggesting that the findings may still be somewhat representative of women's football. Finally, although the survey was rigorously designed and piloted, there may still be limitations related to the clarity and comprehensiveness of the questions. These limitations may have resulted in some data analytics practices not being fully captured.

Conclusion

The present study provides a comprehensive overview of data analytics practices within top-level football worldwide. Our findings highlight that practitioners with a responsibility for data are primarily employed within the performance department, whilst those in the data and medical departments represented only a minority within the surveyed sample. A gender disparity was observed, with only 13% of the survey sample identifying as female.

Significant heterogeneity was observed in responses across all departments regarding training load and performance outcome metrics, methodological attributes, and measurement properties. The analytical approach utilised most frequently was exploratory data analysis, with modelling, forecasting, and predicting the least frequent approach. Interestingly, the use of scientific literature was the least preferred source of evidence, regardless of the respondent's department. This finding suggests the need for greater congruence between research and applied practice. Respondents reported using a mix of solutions for data storage, aggregating & analysing, and reporting & visualising data. Spreadsheets were cited as the most popular solution for data wrangling and reporting tasks which may be a result of strategic factors, financial constraints, and/or expertise-related issues. Data analytics reports were commonly disseminated daily or weekly, emphasising their potential to support decision-making processes.

Collectively, our findings provide an overview of current data ecosystems and information systems in modern football organisations. We anticipate that our results can be used to improve the data analytics service provision in football by helping identify areas for development and progression.

Acknowledgements

The authors wish to thank Dr Shaun McLaren, Professor Alistair McRobert, Dr Steve Barrett, Mr Eyal Eliakim, Mr Ori Kobi, and Dr. Adriano Arguerdas-Soley for their feedback during the developmental stages of the survey.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Antonio Dello Iacono http://orcid.org/0000-0003-0204-0957
Naomi Datson http://orcid.org/0000-0002-5507-9540
Jo Clubb http://orcid.org/0000-0002-6509-7531
Mathieu Lacome http://orcid.org/0000-0002-1082-0200
Adam Sullivan http://orcid.org/0000-0003-1370-8385
Tzlil Shushan http://orcid.org/0000-0002-0544-1986

Author contributions

Antonio Dello Iacono: Conceptualisation, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Project administration.

Naomi Datson: Methodology, Writing – original draft, Writing – review & editing.

Jo Clubb: Methodology, Writing – original draft, Writing – review & editing.

Mathieu Lacome: Methodology, Software, Writing – review & editing.

Adam Sullivan: Methodology, Software, Writing – review & editing.

Tzlil Shushan: Conceptualisation, Methodology, Software, Writing – review & editing.

Code availability statement

The script, including data wrangling, analysis, and plot coding, is available as a supplementary file.

Ethics approval statement and informed consent

This study was approved by the University of the West of Scotland Institutional Review Committee, United Kingdom (protocol number: 17301).

Data availability statement

The survey copy and raw dataset are available as supplementary files.

Open Scholarship

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The data and materials are openly accessible at https://osf.io/pkcs3

References

Akenhead R, Nassis GP. 2016. Training load and player monitoring in high-level football: current practice and perceptions. Int J Sports Physiol Perform. 11(5):587–593. doi: 10.1123/ijspp.2015-0331.

Amrhein V, Trafimow D, Greenland S. 2019. Inferential statistics as descriptive statistics: there is No replication crisis if we Don't expect replication. The Am Statistician. 73(sup1):262–270. doi: 10.1080/00031305.2018.1543137.

Asimakidis ND, Bishop CJ, Beato M, Mukandi IN, Kelly AL, Weldon A, Turner AN. 2024. A survey into the current fitness testing practices of elite male soccer practitioners: from assessment to communicating results. Front Physiol. 15:1376047. doi: 10.3389/fphys.2024.1376047.

Bartlett JD, Drust B. 2021. A framework for effective knowledge translation and performance delivery of sport scientists in professional sport. Eur J Sport Sci. 21(11):1579–1587. doi: 10.1080/17461391.2020.1842511.

Beato M, Datson N, Anderson L, Brownlee T, Coates A, Hulton A. 2023. Rationale and practical recommendations for testing protocols in female soccer: a narrative review. J Strength Cond Res. 37(9):1912–1922. doi: 10.1519/JSC.0000000000004509.

Beissel AS, Postlethwaite V, Grainger A, Brice J. 2024. A new hope? FIFA 2.0, FIFA women's football strategy, and event bidding for the 2023 FIFA women's world CupTM. Soccer Soc. 25(1):1–28. doi: 10.1080/14660970.2023.2214512.

Bryan A. 2022. Women in a man's world: an examination of women's leadership work in the 'extremely gendered'Organisation of men's football in England.

Buchheit M, Dupont G. 2018. Elite clubs and national teams: sharing the same party? Sci Med Football. 2(2):83–85. doi: 10.1080/24733938.2018.1470156.

Casals M, Finch CF. 2017. Sports biostatistician: a critical member of all sports science and medicine teams for injury prevention. Inj Prev. 23 (6):423–427. doi: 10.1136/injuryprev-2016-042211.

Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, et al. 2024. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. e078378. doi: 10.1136/bmj-2023-078378.

French D, Ronda LT. 2021. NSCA's essentials of sport science. Champaign, IL, US: Human Kinetics.

Fullagar HHK, McCall A, Impellizzeri FM, Favero T, Coutts AJ. 2019. The translation of sport science research to the field: a current opinion and overview on the perceptions of practitioners, researchers and coaches. Sports Med. 49(12):1817–1824. doi: 10.1007/s40279-019-01139-0.

Gerrard B, Alamar BC. 2014. Sports analytics: a Guide for coaches, managers and other decision makers. Sport Manag Rev. 17(2):240–241. doi: 10.1016/j.smr.2013.06.005.

Goes FR, Meerhoff LA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, Torres RS, et al. 2021. Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. Eur J Sport Sci. 21 (4):481–496. doi: 10.1080/17461391.2020.1747552.

Gregson W, Carling C, Gualtieri A, O'Brien J, Reilly P, Tavares F, Bonanno D, Lopez E, Marques J, Lolli L, et al. 2022. A survey of organizational structure and operational practices of elite youth football academies and national federations from around the world: a performance and medical perspective. Front Sports Act Living. 4:1031721. doi: 10.3389/fspor.2022.1031721.

Harkness-Armstrong A, Till K, Datson N, Myhill N, Emmonds S. 2022. A systematic review of match-play characteristics in women's soccer. PLOS ONE. 17(6):e0268334. doi: 10.1371/journal.pone.0268334.

Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. 2019. Assessing risk of bias in a randomized trial. Cochrane Handb Systematic Rev Intervent. 23:205–228.

Houtmeyers KC, Jaspers A, Figueiredo P. 2021. Managing the training process in elite sports: from descriptive to prescriptive data analytics. Int J Sports Physiol Perform. 16(11):1719–1723. doi: 10.1123/ijspp.2020-0958.

Houtmeyers KC, Vanrenterghem J, Jaspers A, Ruf L, Brink MS, Helsen WF. 2021. Load monitoring practice in European elite football and the impact of club culture and financial resources. Front Sports Act Living. 3:679824. doi: 10.3389/fspor.2021.679824.

Impellizzeri FM, Marcora SM. 2009. Test validation in sport physiology: lessons learned from Clinimetrics. Int J Sports Physiol Perform. 4 (2):269–277. doi: 10.1123/ijspp.4.2.269.

Impellizzeri FM, Marcora SM, Coutts AJ. 2019. Internal and external training load: 15 years on. Int J Sports Physiol Perform. 14(2):270–273. doi: 10.1123/ijspp.2018-0935.

Jayal A, McRobert A, Oatley G, O'Donoghue P. 2018. Sports analytics: analysis, visualisation and decision making in sports performance. Abingdon, UK: Routledge.

Jeffries AC, Marcora SM, Coutts AJ, Wallace L, McCall A, Impellizzeri FM. 2022. Development of a revised conceptual framework of physical training for use in research and practice. Sports Med. 52(4):709–724. doi: 10.1007/s40279-021-01551-5.

Lacome M, Simpson B, Buchheit M. 2018. Monitoring training status with player-tracking technology: still on the road to Rome. Aspetar Sports Med J. 7:54–63.

Lolli L, Bauer P, Irving C, Bonanno D, Höner O, Gregson W, Di Salvo V. 2024. Data analytics in the football industry: a survey investigating operational frameworks and practices in professional clubs and national federations from around the world. Sci Med Football. 2024:1–10. doi: 10.1080/24733938.2024.2341837.

López-Valenciano A, Raya-González J, Garcia-Gómez JA, Aparicio-Sarmiento A, Sainz De Baranda P, De Ste Croix M, Ayala F. 2021. Injury profile in Women's football: a systematic review and meta-analysis. Sports Med. 51(3):423–442. doi: 10.1007/s40279-020-01401-w.

Luteberget LS, Houtmeyers KC, Vanrenterghem J, Jaspers A, Brink MS, Helsen WF. 2021. Load monitoring practice in elite women association football. Front Sports Act Living. 3:715122. doi: 10.3389/fspor.2021.715122.

Okholm Kryger K, Wang A, Mehta R, Impellizzeri F, Massey A, Harrison M, Glendinning R, McCall A. 2023. Can we evidence-base injury prevention and management in women's football? A scoping review. Res Sports Med. 31(5):687–702. doi: 10.1080/15438627.2022.2038161.

Okholm Kryger K, Wang A, Mehta R, Impellizzeri FM, Massey A, McCall A. 2022. Research on women's football: a scoping review. Sci Med Football. 6(5):549–558. doi: 10.1080/24733938.2020.1868560.

Ratten V. 2016. Sport innovation management: towards a research agenda. Innovation. 18(3):238–250. doi: 10.1080/14479338.2016.1244471.

Robertson PS. 2020. Man & machine: adaptive tools for the contemporary performance analyst. J Sports Sci. 38(18):2118–2126. doi: 10.1080/02640414.2020.1774143.

Rommers N, Rössler R, Verhagen E, Vandecasteele F, Verstockt S, Vaeyens R, Lenoir M, D'HONDT E, Witvrouw E. 2020. A machine learning approach to assess injury risk in elite youth football players. Med Sci Sports Exercise. 52(8):1745–1751. doi: 10.1249/MSS.0000000000002305.

Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernàndez J, Medina D, Sampaio J. 2018. Effective injury forecasting in soccer with GPS training data and machine learning. PLOS ONE. 13(7):e0201264. doi: 10.1371/journal.pone.0201264.

Shushan T, McLaren SJ, Buchheit M, Scott TJ, Barrett S, Lovell R. 2022. Submaximal fitness tests in team sports: a theoretical framework for evaluating physiological state. Sports Med. 52(11):2605–2626. doi: 10.1007/s40279-022-01712-0.

Starling LT, Lambert MI. 2018. Monitoring rugby players for fitness and fatigue: what do coaches want? Int J Sports Physiol Perform. 13(6):777–782. doi: 10.1123/ijspp.2017-0416.

Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2008. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 61(4):344–349. doi: 10.1016/j.jclinepi.2007.11.008.

Ward P, Windt J, Kempton T. 2019. Business intelligence: how sport scientists can support organization decision making in professional sport. Int J Sports Physiol Perform. 14(4):544–546. doi: 10.1123/ijspp.2018-0903.