Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics

Kobe C. Houtmeyers, Arne Jaspers, and Pedro Figueiredo

Elite sport practitioners increasingly use data to support training process decisions related to athletes' health and performance. A careful application of data analytics is essential to gain valuable insights and recommendations that can guide decision making. In business organizations, data analytics are developed based on conceptual data analytics frameworks. The translation of such a framework to elite sport may benefit the use of data to support training process decisions. **Purpose:** The authors aim to present and discuss a conceptual data analytics framework, based on a taxonomy used in business analytics literature to help develop data analytics within elite sport organizations. **Conclusions:** The presented framework consists of 4 analytical steps structured by value and difficulty/complexity. While descriptive (step 1) and diagnostic analytics (step 2) focus on understanding the past training process, predictive (step 3) and prescriptive analytics (step 4) provide more guidance in planning the future. Although descriptive, diagnostic, and predictive analytics generate insights to inform decisions, prescriptive analytics can be used to drive decisions. However, the application of this type of advanced analytics is still challenging in elite sport. Thus, the current use of data in elite sport is more focused on informing decisions rather than driving them. The presented conceptual framework may help practitioners develop their analytical reasoning by providing new insights and guidance and may stimulate future collaborations between practitioners, researchers, and analytics experts.

Keywords: framework, sport science, analysis methods, decision making, prediction

More than ever, practitioners apply various methods and technologies to manage elite athletes' training process.¹ The corresponding data are used to evaluate the training dose and response with the ultimate goal of optimizing athletes' health and performance.² To achieve these goals, practitioners make daily decisions to improve the training process. These decisions are based both on their expertise and on the collected data. In this article, training process data relates to all objective and subjective variables that reflect athlete characteristics, the training dose and response, and training outcomes related to athletes' health and performance.² Nevertheless, the abundance of currently available data, combined with practitioners' limited time, may hamper the valuable application of data to inform decision making.³ Multiple systematic steps should be followed to use data effectively and with confidence. Some of these necessary steps have already been discussed in other publications (eg, the data collection, storage, cleaning, and dissemination steps).^{1,4,5} In this commentary, we specifically focus on data analytics.

Data analytics can be considered the iterative process of discovering insights from data to make better and faster decisions.⁶ In business organizations, conceptual data analytics frameworks are available to guide the organization in developing their strategies.⁶ Similarly, elite sports may benefit from such a framework to streamline the effective use of data. Therefore, we present a conceptual framework that could help develop analytical approaches within the sports organization.

Conceptual Data Analytics Framework

The framework is based on a taxonomy used in business analytics literature.⁶⁻⁹ The presented framework includes 4 steps that are hierarchically structured by value and difficulty/complexity and is based on the perceived developmental needs in data analytics in elite sports (Figure 1).⁶ Ideally, these 4 steps are simultaneously applied, such that insights from one step improve the underlying interacting factors of all the others. It should be emphasized that the use of this (or any other) data analytics framework has the presumption of a quality data set. Therefore, although not the focus of the current article, we highlight the importance of data hygiene (ie, the practices undertaken to ensure minimal error in the collection and storage of data) and data integrity (ie, the consistency, completeness, and accuracy of data) (for more information and guidelines, see Varley et al⁵).

The first step is descriptive analytics. This step attempts to answer what happened in the training process and can be used to compare the actual with the planned process. For example, a practitioner who aims to improve his/her athlete's endurance performance may describe physical test results to monitor progress toward the training goal. However, descriptive analytics do not advance our understanding of why things have happened in the process. The examination of the relations between training process data belongs to the second step, diagnostic analytics. Insights from this step can aid in planning the training process. For example, understanding the relationship between training intensity distribution and improvement in endurance performance can guide a practitioner in planning training intensity. Predictive analytics can then be applied to gain insight into the probability that the outcome of the training process will happen. For example, the practitioner could forecast the probability of improving endurance performance based on the planned training intensity and historical training process data. The descriptive, diagnostic, and predictive

Houtmeyers and Jaspers are with the Faculty of Movement and Rehabilitation Sciences, KU Leuven, Leuven, Belgium. Figueiredo is with the Portugal Football School, Portuguese Football Federation, Oeiras, Portugal, and the Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University Institute of Maia, ISMAI, Maia, Portugal. Houtmeyers (kobe.houtmeyers@kuleuven.be) is corresponding author.

Figure 1 — A conceptual data analytics framework for elite sport (adapted from Delen and Ram⁶).

steps inform decisions by providing insight into the past, which can be applied to the future. However, these steps do not directly offer insight into what precisely needs to be planned to optimize the path to the goal, that is, what is the optimal decision regarding the training process for that athlete at that particular moment in a given context. For example, several strategies can be used to improve endurance performance. Prescriptive analytics could be applied to try to identify the optimal, most efficient, and controllable strategy by balancing the probability of different training process outcomes, such as performance improvement and injury risk. In the following sections, we provide some comments on the different steps to show the framework's expressiveness and usability.

Descriptive Analytics

Before starting with descriptive analytics, a consistent data collection based on valid and reliable data needs to be organized and encouraged to the various stakeholders such as athletes, coaches, and the management board.¹⁰ It is then essential that the collected data are efficiently centralized and summarized to provide a clear overview of the data, such as interactive dashboards.¹⁰ In general, practitioners develop these dashboards using data visualization tools or commercially available athlete management software.

To effectively describe data, it is first essential to select the variables that will be described carefully. Nowadays, considerable amounts of training process data are collected, such as data from the global navigation satellite systems, heart rate monitors, athlete questionnaires, physical tests, and data related to injury epidemiology.² Besides selecting variables based on scientific evidence, it is recommended to reduce the number of variables without losing unique information. The latter can be achieved by applying data reduction methods, such as principal component analysis.¹¹ Such methods allow one to describe correlated data more efficiently. Second, one is advised to consider the uncertainty around the collected data for a correct interpretation. Almost all data are subject to various noise sources such as measurement error and biological variation. Therefore, the inclusion of uncertainty measures, such as confidence intervals, is recommended.¹² Finally, appropriate visualization methods are recommended to correctly inform and disseminate the message to the other stakeholders,

which implies using appropriate chart types, colors, axis ranges, and so forth.¹³ Accurate descriptive analytics are critical to proceed toward diagnostic analytics, which in turn aid the improvement of descriptive analytics by updating variables based on new insights (for more examples on descriptive analytics in elite sports, see Thornton et al¹⁰).

Diagnostic Analytics

Measures such as correlation coefficients are often used in elite sports to analyze the association between training process data.¹⁴ Although this may help understand related factors, it is essential to note that strong associations do not provide evidence for causation.¹⁵ Establishing a causal relation requires a causal experimental design, which is often not attainable in observational sports science research.¹⁶

Identifying strong and consistent associations between training process data is challenging. This can be explained by the validity, reliability, and sensitivity of variables that are commonly analyzed in elite sports.¹⁷ Typical training process data, such as athlete-reported outcome measures or global navigation satellite system-derived variables, frequently lack careful validation and often reflect indirect measurements of the causal mechanism of interest (eg, distance covered at high speed as a proxy measure of muscular strain and stress).^{18,19} Moreover, the system's complexity, which refers to the athlete in his/her environment, can largely contribute to the lack of consistent associations.²⁰ Many variables interact in a dynamic and nonlinear manner. In addition, the association between variables is strongly influenced by other variables. This stresses the need for a multifaceted approach. To unravel complex data, the application of advanced statistics and machine learning techniques can be used. Advanced statistics includes, for example, hierarchical and mixture models, while machine learning techniques such as neural networks and association rule learning can be applied.^{21,22} The gathering of a quality data set (ie, based on valid, reliable, and sensitive data that are related to causal mechanisms) and the use of appropriate analytical techniques may help to determine better possible related factors, which is useful for applying predictive analytics (for more examples on diagnostic analytics in elite sports, see Bittencourt et al²⁰ and Fox et al²³).

Predictive Analytics

Predictive analytics is a collective term for all methods that aim to predict new or future (training process) data accurately, such as temporal forecasting.^{24,25} Again, given the complexity of the training process, advanced statistics and machine learning techniques such as regression and decision trees can be recommended for this purpose.^{21,22,25–28} To evaluate predictive accuracy, appropriate statistical measures such as sensitivity, specificity, and mean absolute error should be used.^{14,22} Also, it is advised to distinguish the model's in-sample and out-of-sample performance when interpreting predictive accuracy.²¹ While in-sample performance is based on the collected data, the out-of-sample performance is based on the ability to predict future “unseen” data. In machine learning, in-sample performance is often overestimated in relation to out-of-sample performance, known as overfitting. Different validation methods such as split-sample validation (ie, dividing the collected data into a training and testing data set) should be used to address this issue.^{21,22}

Currently, predictive analytics is not generally adopted in elite sports. Some studies have provided a first insight into machine learning methods to predict training process outcomes, such as injuries.^{29–31} However, to date, there is no strong evidence on the accurate prediction of training process data. The system's complexity could again explain this and also the (sometimes limited) validity, reliability, and sensitivity of (sometimes inconsistently) collected data, which can be a pitfall and is occasionally quoted as “garbage in, garbage out.”^{17–20,26} In addition, the amount of training process data that are collected in elite sports is relatively small compared with other “big data” domains, such as finance using the historical data of stock prices or social sciences using social media data. Therefore, we advise being careful in dealing with commercial assertions that claim accurate training process data predictions (for more examples on predictive analytics in elite sports, see Van Eetvelde et al²⁷ and Richter et al²⁶).

Prescriptive Analytics

In elite sports, practitioners' expertise is the primary driver of decision making. The training process is often planned based on training principles, which are continuously updated based on descriptive, diagnostic, or predictive analytics insights.³² However, expert knowledge can be prone to cognitive biases, such as confirmation bias, and previous studies, both outside and inside sport, have suggested the superiority of data-driven decision making over human judgment.^{22,33,34} Therefore, prescriptive analytics is considered useful as it provides a more objective data-driven assessment of the decision options. Prescriptive analytics takes large amounts of data and possible situations and provides a series of possible outcomes and the paths to these and, especially, to the “best” possible outcome. Thus, it relates to “decision support systems,” supporting the organizational decision making focused on optimization. Prescriptive analytics can be built on probabilistic predictive analytics using data-driven algorithms or expert-driven systems such as mathematical optimization and constraint programming.^{3,7} In contrast to predictive analytics, prescriptive analytics suggests actions that will benefit from the predictions and provides direct insight into the consequences of each decision option.

Prescriptive analytics is seldomly reported in the context of training process data in elite sports. To perform prescriptive analytics, advanced methods such as logic-based models are

required.⁷ Given the current shortcomings (see aforementioned steps), it is challenging to apply such methods effectively in elite sports. However, prescriptive analytics is already applied in other disciplines such as operation research/management science.^{7,8} Prescriptive analytics is also used in well-known examples such as Google's self-driving car or health insurance companies to look at scenarios of reimbursement costs considering associated comorbidities of the customers. In the context of fast technological innovation, it may not be surprising that prescriptive analytics finds its way into elite sports in the near future. The provided conceptual framework may help guide these future developments.

Adopting prescriptive analytics in elite sports does not only depend on accuracy; other criteria, such as feasibility, complexity, and usability, also determine the acceptance of prescriptive analytics.^{3,22} This stresses the importance of a balanced trade-off between analytics and practitioners' expertise and preferences. First, practitioners need to be aware of expert bias's influence on their decision making and need to understand the benefit of prescriptive analytics to drive their decisions. Second, the practitioner needs to be accustomed to effectively using such analytics in terms of interpretation, dissemination, and application to the training process (for more information and guidelines on prescriptive analytics in elite sports, see Schelling and Robertson³).

Practical Applications and Conclusions

This commentary provided a conceptual data analytics framework that can be used to develop analytical reasoning within elite sports organizations. Nowadays, elite sports organizations are exploring the use of data to inform training process decisions. However, this exploration comes with many challenges in using the data effectively. Where challenges around data collection, storage, cleaning, and dissemination are already discussed elsewhere, this commentary focused on providing a framework that could help develop data analytics in sports.^{1,4,5}

Finding its origin in business analytics, the framework may stimulate elite sports practitioners and researchers to translate the analytical knowledge and skills from other domains into their practices. We focused on the effective use of data, which can be improved in both a simple context, such as dashboarding, and in a more advanced context, such as decision-support systems. However, and in a more advanced context, reporting data in clear dashboards but primarily based on presenting the collected data (ie, descriptive analytics) runs the danger of being cherry-picked for confirmation of beliefs and may therefore have limited value. Adding (simultaneous) applications of the 3 other analytical steps may improve analytical reasoning by providing new insights and guidance. In this respect, the framework can be used to streamline collaborations between practitioners, researchers, and other data scientists, by encouraging accurate and systematic data collection, questioning, and implementation of findings into everyday practice.

Elite sports organizations can compare their actual approaches with the presented steps. This can help organizations in aligning their analytical expectations to the type of analytics they currently perform. In the context of increased reliance on data, a critical reflection of analytics is helpful to manage expectations and may provide an increased understanding of the current state of evidence and possibilities. For example, when reflecting on elite sports, the current application of this framework confirms that we are in a reality of data-informed decision making instead of data-driven decision making.³⁵

Finally, we acknowledge that we only discussed a snapshot of the available knowledge on the different analytical steps. Therefore,

we hope that this commentary encourages research to provide more detailed explanations and recommendations in the near future.

Acknowledgments

The authors would like to thank Michel Brink, Werner Helsen, Pieter Robberechts, and Jos Vanrenterghem for their advice and feedback on the manuscript. K.C.H. and A.J. are involved in TopSportsLab, an athlete management software company. The company does not provide predictive or prescriptive analytics in their software.

References

- Windt J, MacDonald K, Taylor D, Zumbo BD, Sporer BC, Martin DT. "To tech or not to tech?" A critical decision-making framework for implementing technology in sport. *J Athl Train*. 2020;55(9):902–910. PubMed ID: [32991702](#) doi:[10.4085/1062-6050-0540.19](#)
- Impellizzeri FM, Marcora SM, Coutts AJ. Internal and external training load: 15 years on. *Int J Sports Physiol Perform*. 2019;14(2):270–273. PubMed ID: [30614348](#) doi:[10.1123/ijspp.2018-0935](#)
- Schelling X, Robertson S. A development framework for decision support systems in high-performance sport. *Int J Comput Sci Sport*. 2020;19(1):1–23. doi:[10.2478/ijcss-2020-0001](#)
- Robertson S, Bartlett JD, Gastin PB. Red, Amber, or green? Athlete monitoring in team sport: the need for decision-support systems. *Int J Sports Physiol Perform*. 2017;12(suppl 2):S273–S279. doi:[10.1123/ijspp.2016-0541](#)
- Varley MC, Lovell R, Carey D. Data hygiene. In: French DN, Ronda LT, eds. *NSCA's Essentials of Sport Science*. Champaign, IL: Human Kinetics; 2021.
- Delen D, Ram S. Research challenges and opportunities in business analytics. *J Bus Anal*. 2018;1(1):2–12. doi:[10.1080/2573234X.2018.1507324](#)
- Lepeniotti K, Bousdekis A, Apostolou D, Mentzas G. Prescriptive analytics: literature review and research challenges. *Int J Inf Manage*. 2020;50:57–70. doi:[10.1016/j.ijinfomgt.2019.04.003](#)
- Mortenson MJ, Doherty NF, Robinson S. Operational research from Taylorism to Terabytes: a research agenda for the analytics age. *Eur J Oper Res*. 2015;241(3):583–595. doi:[10.1016/j.ejor.2014.08.029](#)
- Delen D, Demirkan H. Data, information and analytics as services. *Decis Support Syst*. 2013;55(1):359–363. doi:[10.1016/j.dss.2012.05.044](#)
- Thornton HR, Delaney JA, Duthie GM, Dascombe BJ. Developing athlete monitoring systems in team sports: data analysis and visualization. *Int J Sports Physiol Perform*. 2019;14(6):698–705. doi:[10.1123/ijspp.2018-0169](#)
- Ryan S, Kempton T, Coutts AJ. Data reduction approaches to athlete monitoring in professional Australian football. *Int J Sports Physiol Perform*. 2020;16(1):59–65. doi:[10.1123/ijspp.2020-0083](#)
- de Koning JJ, Noordhof DA. Embrace uncertainty. *Int J Sports Physiol Perform*. 2019;14(6):697. PubMed ID: [31185772](#) doi:[10.1123/ijspp.2019-0419](#)
- Nimphius S, Jordan MJ. Show me the data, Jerry! Data visualization and transparency. *Int J Sports Physiol Perform*. 2020;15(10):1353–1355. PubMed ID: [33129199](#) doi:[10.1123/ijspp.2020-0813](#)
- McCall A, Fanchini M, Coutts AJ. Prediction: the modern-day sport-science and sports-medicine "Quest for the Holy Grail." *Int J Sports Physiol Perform*. 2017;12(5):704–706. PubMed ID: [28488907](#) doi:[10.1123/ijspp.2017-0137](#)
- Altman N, Krzywinski M. Association, correlation and causation. *Nat Methods*. 2015;12(10):899–900. doi:[10.1038/nmeth.3587](#)
- Stovitz SD, Verhagen E, Shrier I. Distinguishing between causal and non-causal associations: implications for sports medicine clinicians. *Br J Sports Med*. 2019;53(7):398–399. PubMed ID: [29162620](#) doi:[10.1136/bjsports-2017-098520](#)
- Thorpe RT, Atkinson G, Drust B, Gregson W. Monitoring fatigue status in elite team-sport athletes: implications for practice. *Int J Sports Physiol Perform*. 2017;12(suppl 2):S227–S234. doi:[10.1123/ijspp.2016-0434](#)
- Verheul JN, Nedergaard NJ, Vanrenterghem J, Robinson MA. Measuring biomechanical loads in team sports: from lab to field. *Sci Med Footb*. 2020;4(3):246–252. doi:[10.1080/24733938.2019.1709654](#)
- Jeffries AC, Wallace L, Coutts AJ, McLaren SJ, McCall A, Impellizzeri FM. Athlete-reported outcome measures for monitoring training responses: a systematic review of risk of bias and measurement property quality according to the COSMIN guidelines. *Int J Sports Physiol Perform*. 2020;15(9):1203–1215. PubMed ID: [32957081](#) doi:[10.1123/ijspp.2020-0386](#)
- Bittencourt NFN, Meeuwisse WH, Mendonca LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. *Br J Sports Med*. 2016;50(21):1309–1314. PubMed ID: [27445362](#) doi:[10.1136/bjsports-2015-095850](#)
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Berlin, Germany: Springer-Verlag; 2009.
- Robertson S. Data mining and nonlinear data analysis. In: French DN, Ronda LT, eds. *NSCA's Essentials of Sport Science*. Champaign, IL: Human Kinetics; 2021.
- Fox JL, Stanton R, Sargent C, Wintour SA, Scanlan AT. The association between training load and performance in team sports: a systematic review. *Sports Med*. 2018;48(12):2743–2774. PubMed ID: [30225537](#) doi:[10.1007/s40279-018-0982-5](#)
- Shmueli G. To explain or to predict? *Stat Sci*. 2010;25(3):289–310. doi:[10.1214/10-STS330](#)
- Fonseca ST, Souza TR, Verhagen E, et al. Sports injury forecasting and complexity: a synergetic approach. *Sports Med*. 2020;50(10):1757–1770. PubMed ID: [32757162](#) doi:[10.1007/s40279-020-01326-4](#)
- Richter C, O'Reilly M, Delahunt E. Machine learning in sports science: challenges and opportunities [published online ahead of print April 20, 2021]. *Sports Biomech*. PubMed ID: [33874846](#) doi:[10.1080/14763141.2021.1910334](#)
- Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: a systematic review. *J Exp Ortop*. 2021;8(1):27. PubMed ID: [33855647](#) doi:[10.1186/s40634-021-00346-x](#)
- Passfield L, Hopker JG. A mine of information: can sports analytics provide wisdom from your data? *Int J Sports Physiol Perform*. 2017;12(7):851–855. PubMed ID: [27967295](#) doi:[10.1123/ijspp.2016-0644](#)
- Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernandez J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*. 2018;13(7):e0201264. PubMed ID: [30044858](#) doi:[10.1371/journal.pone.0201264](#)
- Rommers N, Rössler R, Verhagen E, et al. A machine learning approach to assess injury risk in elite youth football players. *Med Sci Sports Exerc*. 2020;52(8):1745–1751. PubMed ID: [32079917](#) doi:[10.1249/MSS.0000000000002305](#)
- Lövdal SS, Den Hartigh RJR, Azzopardi G. Injury prediction in competitive runners with machine learning. *Int J Sports Physiol Perform*. 2021;16(10):1522–1531. PubMed ID: [33931574](#) doi:[10.1123/ijspp.2020-0518](#)

32. Impellizzeri FM, Menaspa P, Coutts AJ, Kalkhoven J, Menaspa MJ. Training load and its role in injury prevention, part I: back to the future. *J Athl Train*. 2020;55(9):885–892. PubMed ID: [32991701](#) doi:[10.4085/1062-6050-500-19](#)
33. Maymin PZ. The automated general manager: can an algorithmic system for drafts, trades and free agency outperform human front offices? *J Glob Sport Manag*. 2017;2(4):234–249. doi:[10.1080/24704067.2017.1389248](#)
34. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess*. 2000;12(1):19–30. PubMed ID: [10752360](#)
35. Gamble P, Chia L, Allen S. The illogic of being data-driven: reasserting control and restoring balance in our relationship with data and technology in football. *Sci Med Footb*. 2020;4(4):338–341. doi:[10.1080/24733938.2020.1854842](#)