SPORTS AND EXERCISE MEDICINE
Open Journal

Openventio
PUBLISHERS

Research Protocol
Sports Science Data Protocol

Lorena Martin, PhD*

Former Lakers Director of Sports Performance Analytics, Los Angeles, CA 90089, USA

***Corresponding author**
Lorena Martin, PhD
Former Lakers Director of Sports Performance Analytics and Seattle Mariners Director of High Performance, Los Angeles, CA 90089, USA;
E-mail: HealthandSportsAnalytics@gmail.com

▌ ABSTRACT ▌

Purpose
At the most elite level, even an increase of 1% improvement can make the difference between winning and losing. Sports scientists can help athletes gain insights that can be the differentiating factor. The purpose of this protocol is to delineate the process from the identification of key performance indicators to the presentation of the findings for sports scientists. It is designed to provide chronological steps in efforts to mitigate barriers of collecting data and tracking players as well as to help gain athlete buy-in to sports science by helping them maximize their performance.

Study Design
The competitive nature of professional sports leads to players, their agents, and teams to seek the expertise in sports performance areas such as those from athletic trainers, strength and conditioning coaches, physical therapists, nutritionists, and sports psychologists. However, much of the knowledge provided by these entities typically remain in silos. Thus, the whole picture of the athlete's performance enhancement mechanism is not elucidated. This is where the significance of a sports scientist ensues, with the ability to integrate the data from each of the sports performance areas with the objective of obtaining a complete and comprehensive picture of the athlete.

Data Collection
This area of collecting data and monitoring athletes is becoming mainstream. As such professional sports leagues have started to implement privacy rules and regulations on the protection of athlete biometric data. It is important to be aware of the rules related to wearable technology and athlete biometric data as well as how to go about the process of collecting data from professional athletes. That is why it is important that care be taken and a protocol be followed to ensure the integrity of data collection in the field of sports science.

Data Processing and Analyses
Data mining is the extraction of data for the purpose of discovering meaningful patterns, normalization, and the choosing of statistical models that can help in making data driven decisions. The interpretation and presentation of the data can have a large impact on the decisions. As such, a protocol should be adhered to.

Conclusion
The following protocol will serve as a guide on how to collect data efficiently and successfully.
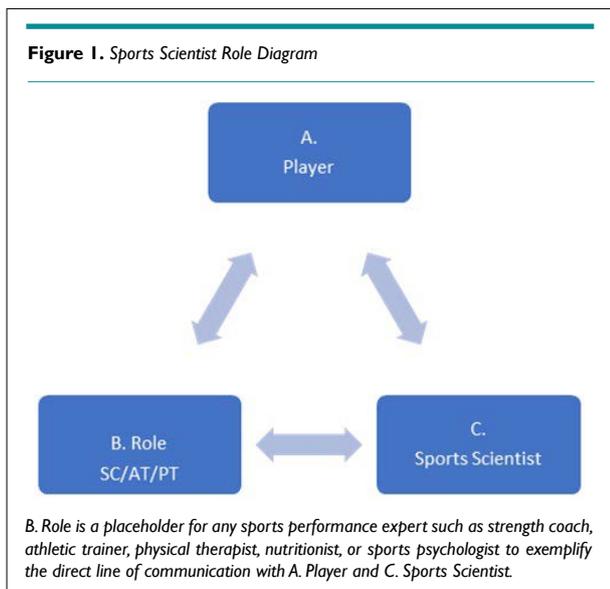
Keywords
Sports science; Data protocol; Sports performance; Sports scientists; Load; Professional athletes; Sports teams.

INTRODUCTION ▐

Professional athletes are always seeking to gain a competitive edge. A slight improvement in any aspect of sports performance can lead to a significant difference in performance outcomes.[1] It has been well-documented in the literature that different areas of sports performance such as athletic training, physical therapy, strength and conditioning, nutrition, and sports psychology function independent of one another without sharing resources and information within professional organizations.[2,3] This is where

the role of the sports scientist comes in, as they can be the connecting link between these distinctive areas of expertise, as their proficiency should be to have a general understanding of all the areas and understand which outcome or dependent variables should be quantified to provide a meaningful metric that is relevant to the athlete's performance.[4] Moreover, the sports scientist should be directly in communication with the professional athletes and the sports performance experts from different areas as shown in Figure 1.

Figure 1. *Sports Scientist Role Diagram*

B. Role is a placeholder for any sports performance expert such as strength coach, athletic trainer, physical therapist, nutritionist, or sports psychologist to exemplify the direct line of communication with A. Player and C. Sports Scientist.

In many instances, due to teams being short-staffed or the lack of the role of a sports scientist, teams assign one of the aforementioned sports performance area experts to try and carry out the duties of a sports scientist.[5] Ideally, there should be a sports scientist staff with expertise to have the ability to be innovative, ask pertinent research questions, and identify key performance indicators (KPIs) or metrics that can help maximize athletic performance.[6] The role of the sports scientist comprises of knowledge on how to collect, analyze, interpret, present, and secure the data.[7] Furthermore, sports scientists in professional sports teams should have the ability to produce distinct reports to players, coaches, and front office management (general manager, assistant general manager, analytics staff).[8]

The sports scientist's first initiative should be to ask questions,[9] for instance: What are the KPIs that would help the athlete's performance? What meaningful implication will this data have? How to go about collecting data? What data is valuable and simple to measure? How to go about choosing a technology to measure the variables? How long will it take to collect this data? How to go about requesting consent from the athlete? Where is the data going to be stored? Who will have access to the data? Is the data stored in a secure location?[10]

IDENTIFICATION OF DATA-KEY PERFORMANCE INDICATORS

After answering these questions, the focus should shift to the KPIs and the type of load to be quantified. The terms load and training load (TL) are typically applied interchangeably and sub-classified as either internal load or external load. Internal load represents psychological variables such as confidence and anxiety and physiological KPIs such as heart rate variability, lactate threshold, glucose and insulin levels, etc.[11] External load is characterized by physical, biomechanical, behavioral and environmental KPIs. It is important to distinguish between internal load and external load and how they should be quantified. For instance, many of the KPIs of external loads such as power output, acceleration, and speed are derived from wearable technology such as accelerometers, gyroscopes, magnetometers, and global positioning systems (GPS). Whereas, KPIs of internal loads such as lactate threshold and rate of perceived exertion (RPE) are obtained through either biomarker assessments or self-reported questionnaires.[12]

The practical implication of measuring TL is to help the sports scientist and training staff better help the athlete by establishing appropriate TL thresholds.[13] This will also allow the sports scientist to gain insights into strengths and weaknesses of the athlete in efforts to reduce the risk of injury, examine what works for the athlete, and help them to continue to improve their performance.

Table 1. *A Measurement Model for Sports*

Physiological	Physical	Psychological	Behavioral	Environmental
Blood pressure	Agility	Anxiety	Nutrition	Built environment
Glucose and insulin	Anaerobic power	Competitiveness	Sleep	Social support groups (Coaches, Parents, Peers)
Heart rate variability	Balance	Confidence	Substance use	Socioeconomic status
Lactate threshold	Body composition	Depression		
Methylome	Cardiorespiratory endurance	Impulsiveness		
Previous injuries	Coordination ability	Intellect of sport		
Respiratory rate	Flexibility	Motivation		
Resting heart rate	Muscular endurance	Narcissism		
Telomere length	Muscular power	Perfectionism		
Vision	Muscular strength	Resiliency		
VO$_2$max	Reaction time	Self-efficacy		
	Sport-specific skills	Self-esteem		
		Vigor		

Source: Martin L[6]

Based on the sport, position, and player body composition some variables and KPIs may be more relevant than others. See the table below for a list of variables and KPIs related to sports performance (Table 1).

It is essential that an adequate background literature search that includes becoming knowledgeable in the sport of interest and the metrics to be quantified be performed. This process includes searching through journals, books, as well as inquiring about previous projects that may have been conducted by the team, organization, or league.[14] This will help in the development of a well-defined purpose for the KPIs that will be measured, along with the hypotheses, the purpose of collecting the data on these KPIs and how the findings will be translatable to the athlete's performance and overall team's success.[10]

CHOOSING THE WEARABLE/TECHNOLOGY AND VALIDATION

After choosing the KPIs of interest, the sports scientist should choose the appropriate measurement, assessments, and form of evaluation. Therefore, it is fundamental for sports scientists to be acquainted with and able to distinguish between measurement, assessment, and evaluation. Sports scientists should know that measurement is the assignment of numbers to quantify a characteristic being assessed, while assessment or test is a tool to make the particular measurement, and evaluation is the judgment on the quality of the assessment of the measurement.[10] Furthermore, it is a given that sports scientists know how to validate an instrument of measurement as well as establish reliability. Validity is the term used to describe whether the technology or instrument measures what it is set out to measure.[11] In the pro sports industry, there are many products of wearable technology being marketed to the training staff.[15] It is critical for the sports scientist to assess whether the technology is valid and measures what it is purported to measure or not.[16] As such, it is important to be able to evaluate how valid these technologies and instruments are by validating the technology or instrument against a gold-standard, thus establishing criterion validity. Once the instrument or technology has been chosen and validated, the focus should shift from the instrument or technology to the data output. It is ideal to obtain the raw data files as it can also help the sports scientist understand the underlying algorithm of the formulated commercialized metrics being offered by the equipment manufacturers.

ESTABLISHING INTRA-RATER AND INTER-RATER RELIABILITY TESTING

After choosing and validating a technology or instrument, it is important to conduct reliability assessments.[17] Reliability is the consistency and reproducibility of the measurement.[18] The most commonly agreed-upon types of reliability that should be assessed include absolute and relative reliability.[19] A simple way to distinguish between the two is to think about whether the assessment is cross-sectional, which would pertain to relative reliability typically assessed by intraclass correlation coefficients (ICC), or if it is a longitudinal assessment which would be indicative of absolute

reliability typically assessed through the standard error of measurement (SEM). Other methods include the limit of agreement and Bland Altman plots.

It is strongly recommended that the sports scientists collecting data, establish both intra-rater and inter-rater reliability.[11] Caution is recommended when choosing reliability assessments, as there are several iterations of ICCs (Table 2) based on the number of raters or sports scientists as well as the number of athletes that will be assessed and whether absolute or relative reliability will be examined.

Table 2. *Definition of ICCs and Computation Equations*

Designation	Model	Definition	Computation Formula
ICC (1, 1)	One-way random-effect	$\sigma S^2 / \sigma S^2 + \sigma^2 e$	MSB-MSW/MSB+(k-1)MSW
ICC (1, k)		$\sigma S^2 / \sigma S^2 + \sigma^2 e / k$	MSB-MSW/MSB
ICC (2, 1)	Two-way random-effect	$\sigma S^2 / \sigma S^2 + \sigma T^2 + \sigma^2 e$	MSB-MSE/MSB+(k-1)MSE+k(MST-MSE)/n
ICC (2, k)		$\sigma S^2 / \sigma S^2 + (\sigma T^2 + \sigma^2 e)/k$	MSB-MSE/MSB+(MST-MSE)/n
ICC (3, 1)	Two-way mixed-effect	$\sigma S^2 / \sigma S^2 + \sigma^2 e$	MSB-MSE/MSB+(k-1)MSE
ICC (3, k)		$\sigma S^2 / \sigma S^2 + \sigma^2 e / k$	MSB-MSE/MSB

Adapted from Li et al.[20]

The applied recommendations include practicing setting up the technology to gather data on yourself, peers and other volunteers. The purpose of this practice is three-fold, 1) to become efficient and confident in implementing the technology, 2) establish intra- and inter-rater reliability and 3) be aware of any barriers or hardware difficulties that may come up with the equipment or instrument that will be used to collect the data.

It is important to make it an easy and successful trial the first time around otherwise it may be difficult to obtain player buy-in. Therefore, it is recommended that the wearable or player tracking device be initially implemented on professional athletes that buy-in and are likely to be compliant. The reality is that with high profile athletes one needs to be cautious as there may not be a second opportunity. Therefore, it is important to be confident and efficient when approaching them. If the professional team or league has a B team, a second division, a minor league, or a G league (the NBA's official minor league), this may be an ideal starting place to see what technology the athletes gravitate towards and which end up being impractical.[21] It would be ideal to have this scenario as a pilot trial. If that is not an option or there is a strong appeal from the majors or 1st division teams, then be particular as to which technologies to present, while avoiding overwhelming the players.[22]

PRESENTATION TO THE PLAYERS

A major barrier for collecting data can be a lack of buy-in from the athletes. In order to obtain player buy-in, the following are recommended: give a general presentation to the group of players, have a one on one meeting with each player if possible, and provide the

rationale for collecting the athlete's data. Remember professional athletes are inundated daily with media, autograph requests, and their own training and performance. Typically, it is rare to find a professional athlete to volunteer to be assessed, although there are a few exceptions. Many are skeptical because they don't know what is being collected, why, or how their data is going to be used.[23] Therefore, these concerns need to be addressed upfront.

To address their concerns, it is recommended that the sports scientists be familiar with the respective team's, league's, or organization's rules and regulations on player health and performance data. As each league and each organization may differ on who owns the data, how it will be used and who will have access to the data.[22] For instance, if the coaches and front office are going to have access to the data, be transparent with the athlete as they have a right to know. Ideally, it would be optimal if the team, league, or organization has granted permission to allow the player to choose whom he wants to show the data to.

Explain what will be collected and how it will relate to their performance. Emphasize that the main purpose of collecting and analyzing data is to prevent injuries and maximize sports performance. Be honest about how the data is going to be used and who will have access. This leads us to the next section.

INFORMED CONSENT FORM

Informed consent forms are basically a contract between the player and the sports scientists detailing how the data will be collected, what type of data will be collected, what is the intended use of the data, and who will have access to the data. It is a written statement granting permission to the sports scientist to obtain the player's data for the specified purposes on a voluntary basis.[23] If the team, league or organization supply a form, implement the form provided, otherwise, draft up a consent form that provides at least the following: rationale for collecting data, assessments that will be conducted along with a short description of each and how the outcomes translate to performance. Also, on the informed consent, there should be a space for the athlete's printed name, signature, date, and sports scientist's initials that is providing the informed consent. There should also be an area that describes who the athlete is willing to grant access to the data.[24] Of course, check that this informed consent is compliant with the professional league's, team's or sports organization's policy.

COLLECTION OF DATA

If all the prior steps have been completed, this part becomes simple. Set up a date and time that works well for the players and coaches.[25] Make sure that the technology or equipment that is going to be implemented has been tested and retested. Refine the processes as necessary. Try and keep everything similar, from the sports scientist's demeanor during the data collection process to the time of data collection, and situation in regard to whether the data collection will take place during practice or games.[10] Finally, keep in mind the end goal of data collection, which is to help make data-driven decisions that will increase the probability of maintain-

ing players healthy and maximizing their performance.[22]

DATA PROCESSING AND ANALYSES

Ideally there should already be a platform in place that has an application programming interfaces (APIs) integration for the technologies that will be collecting data and servers that can hold the data securely.[26] A repository of excel built-in from the organization or a commercially available database platform or athlete management system should also be up and running. For example, currently, many of the professional leagues subscribe to an athlete management system platform that provides trends and aesthetically pleasing dashboards.[26] The next step is to make sure to have access to the raw data form and work off a copy of the original data, and aggregate to a large database. In simple terms combine the datasets from different technologies to a master dataset. This phase involves data cleaning, processing, and the handling of missing data. Running the statistical analyses will depend on the type of variables collected and the questions of interest, whether comparisons of the athlete to their prior performance or comparisons to other players will be performed.[27] Statistical modeling is a lot more complex and is not the focus of this paper, for detailed descriptions of types of analyses recommended based on variables to be collected and research questions, refer to Sports Performance Measurement and Analytics.[6]

PRESENTATION OF RESULTS

Data visualization is one of the most important aspects of this process. If the data is not presented in a simple, palatable manner then the message and content will be automatically discarded. Although it is critical to collect meaningful metrics that are relevant to the athlete's performance, it is just as important to know the audience.[28] The presentation of the data should be consistent, easy to understand and structured slightly different for each of the following; players, coaches, and front office management.[8] For example, some may prefer a more detailed report, while others may be expecting a simple chart or graph.

DATA SAFEKEEPING

Ensuring that the athlete's data is secure is fundamental to this protocol. Part of the security and confidentiality is included in the legislation of The Health Insurance Portability and Accountability Act of 1996 (HIPAA).[22] It requires health care providers and organizations to protect the confidentiality of protected health information (PHI), which can also include some if not all of the athlete biometric data. In addition, it covers how the data is handled, transferred and shared.[29] Furthermore, this phase typically involves being knowledgeable about the respective league, team, or organization's rules and regulations as well as collective bargaining agreements (CBAs) on player health data.[22] Depending on the league, for example, the NBA's CBA has a provision on wearable data compared to other leagues that do not. Finally, the last step is to confirm that the data is stored in a safe and secure location.[30]

CONCLUSION

The sports science data protocol is aimed at guiding the sports scientists through a logical progression of steps towards collecting data from professional athletes in a team, league, or organizational setting (Figure 2). The formulation of this protocol is based on a combination of experiences working for different professional sports organizations, input from professional athletes, training staff, coaches, and evidence-based research. The goal was to combine purposeful research with the applicability in professional sports settings. In theory, there would be larger sample sizes, athletes would grant their permission to have sports scientists collect their health and performance data, privacy would not be an issue, and the players would be completely compliant. This serves as a guide to understanding barriers to the collection of data from professional athletes as well as to be able to successfully overcome such obstacles. Overall, the end goal is to reduce the risk of injury and maximizing sports performance through the process delineated in the sports science data protocol.

Figure 2. *Sports Science Data Protocol Steps*
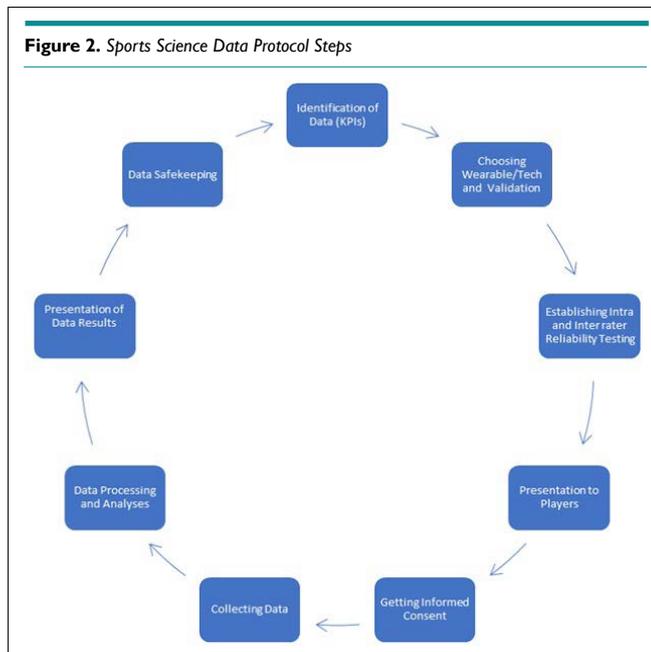
REFERENCES

1. Nevill A, Atkinson G, Hughes M. Twenty-five years of sport performance research in the Journal of Sports Sciences. *J Sports Sci.* 2008; 26(4): 413-426. doi: 10.1080/02640410701714589

2. Barnes M, Cousens L, MacLean J. From silos to synergies: A network perspective of the Canadian sport system. International *Journal of Sport Management and Marketing.* 2007; 2(5-6): 555-571. doi: 10.1504/IJSMM.2007.013967

3. McDuff DR, Garvin M. Working with sports organizations and teams. *Int Rev Psychiatry.* 2006; 28(6): 595-605. doi: 10.1080/09540261.2016.1212820

4. Reade I, Rodgers W, Hall N. Knowledge transfer: How do high performance coaches access the knowledge of sport scientists?. *Int J Sports Sci Coaching.* 2008; 3(3): 319-334. doi: 10.1260/174795408786238470

5. Miles KH, Clark B, Fowler PM, Miller J, Pumpa KL. Sleep practices implemented by team sport coaches and sports science support staff: A potential avenue to improve athlete sleep?. *J Sci Med Sport.* 2019; 22(7): 748-752. doi: 10.1016/j.jsams.2019.01.008

6. Martin L. *Sports Performance Measurement and Analytics: The Science of Assessing Performance, Predicting Future Outcomes, Interpreting Statistical Models, and Evaluating the Market Value of Athletes.* New Jersey, USA: FT Press; 2016.

7. 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

8. Sands W. *Communicating with coaches: Envisioning data. In: Applied Proceedings: Acrobatics.* Perth, WA, USA: School of Biomedical and Sports Science; 2005: 11-20.

9. Atkinson G, Nevill AM. Selected issues in the design and analysis of sport performance research. *J Sports Sci.* 2001; 19(10): 811-827. doi: 10.1080/026404101317015447

10. Morrow JR. *Measurement and Evaluation in Human Performance*, 5th ed. Champaign, IL, USA: Human Kinetics; 2015.

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

12. Saw AE, Main LC, Gastin PB. Monitoring athletes through self-report: factors influencing implementation. *J Sports Sci Med.* 2015; 14(1): 137-146.

13. Burgess DJ. The research doesn't always apply: Practical solutions to evidence-based training-load monitoring in elite team sports. *Int J Sports Physiol Perform.* 2017; 12(s2): S2136-S2141. doi: 10.1123/ijspp.2016-0608

14. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000; 30(1): 1-15. doi: 10.2165/00007256-200030010-00001

15. Düking P, Hotho A, Holmberg HC, Fuss FK, Sperlich B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Front Physiol.* 2016; 7: 71. doi: 10.3389/fphys.2016.00071

16. Johnston RJ, Watsford ML, Kelly SJ, Pine MJ, Spurrs RW. Validity and interunit reliability of 10 Hz and 15 Hz GPS units for assessing athlete movement demands. *J Strength Cond Res.* 2014; 28(6): 1649-1655. doi: 10.1519/JSC.0000000000000323

Openventio

PUBLISHERS

17. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996; 1: 30-46. doi: 10.1037/1082-989X.1.1.30

18. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979; 86: 420-428. doi: 10.1037//0033-2909.86.2.420

19. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998; 26(4): 217-238. doi: 10.2165/00007256-199826040-00002

20. Li L, Zeng L, Lin ZJ, Cazzell M, Liu H. Tutorial on use of intraclass correlation coefficients for assessing intertest reliability and its application in functional near-infrared spectroscopy–based brain imaging. *J Biomed Opt*. 2015; 20(5): 050801. doi: 10.1117/1.JBO.20.5.050801

21. Zych N. Collection and ownership of minor league athlete activity biometric data by major league baseball franchises. *DePaul J. Sports L*. 2018; 14: 129.

22. Osborne B. Legal and ethical implications of athletes' biometric data collection in professional sport. *Marq. Sports L. Rev*. 2017; 28: 37.

23. Gale K. The sports industry's new power play: Athlete biometric data domination. Who owns it and what may be done with it. *Ariz. St. U. Sports & Ent. LJ*. 2016; 6: 7.

24. Lowens B, Motti VG, Caine K. *Wearable privacy: Skeletons in the data closet*. Paper presente at: *2017 IEEE International Conference on Healthcare Informatics (ICHI);* 2017; Park City, UT, USA; doi: 10.1109/ICHI.2017.29

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

26. 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

27. Stein M, Janetzko H, Seebacher D, et al. How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*. 2017; 2(1): 2.

28. De Beéck TO. Data science approaches for the analysis and interpretation of training load data of athletes. 2019.

29. Karlsson D, Timpka T, Jacobsson J, et al. Electronic data capture on athletes' pre-participation health and in-competition injury and illness at major sports championships: An extended usability study in athletics. *Health Informatics Journal*. 2018; 24(2): 136-145. doi: 10.1177/1460458216661861

30. Durkin TM. Health data privacy and security in the age of wearable tech: Privacy and security concerns for the NFLPA and WHOOP. *J. High Tech. L*. 2018; 19: 279.