# Module 3. Problem-Based Learning (PBL) in the sports analytics phenomenon (II)

Martí Casals, professor and investigator in Sport Performance Analysis Research Group (SPARG) of the Center for Studies in Sports and Physical Activity (CEEAF) at the University of Vic – Central University of Catalonia (UVic-UCC). Sports statistician and collaborator in Barça Innovation Hub – Universitas.

### 3.1. Introduction to the Module

Following the PBL CASE carried out previously, we will continue to learn different concepts that can be useful to have a better approach to the Sports Analytics phenomenon. In this module, we will explain what Big Data is, its correct definition applied in a sport like football and what relationship the research and the academy have. In addition, we will report on the most popular scientific journals that deal with aspects related to Sports Analytics and the importance of asking where the data we analyze comes from, so as to think about such important aspects as validation, accuracy and biases in those data before analyzing them. We will remember the different parts of statistics (descriptive and inferential), that many of the academics teach in the academy and we will always have them in mind; finally we will mention the importance of statistical models and also sports biostatistics. We will realize how important it is to have the critical ability to read and understand these concepts, as well as working in a multidisciplinary environment with other specialists in the quantitative world in order to make precise interpretations and advance in the world of Sports Analytics.

### 3.2. Analysis of the 3rd Paragraph of the CASE

Going back to the text of the previous PBL activity of the CASE presented, in this module, we will continue working on some concepts in the following paragraphs.

After carefully reading the next paragraph, I am going to put in bold the words that may help me ask questions in order to consolidate some concepts indirectly related to Sports Analytics.

“Minutes later, the **data scientist** tells me about the importance of being familiar with **research** and how it adapts itself to the **Big Data** and Data Science phenomena. He also explains me that his group meets each Tuesday to discuss some interesting **scientific articles** related to football. He mentions that, at the beginning of the league, he was surprised by the number of yellow cards and, therefore, he would like to know how these could influence in football. I will quickly tell him that I am familiar with a study that was published a few years ago in the **Journal of Quantitative Analysis in Sports** (Anders, A. and Rotthoff, K. W. (2011). Yellow Cards: Do They Matter? Journal of Quantitative Analysis in Sports, 7(1)).

**1. What is the relationship between Research and Big Data? And what is the relationship between Big Data and football?**
**2. What is a scientific article? How are they made up? What is the typology depending on the area of specialization?**
**3. What Sports Analytics journals can I find?**

Have you asked similar questions? Maybe not, but you've probably thought of related questions that make you think about how to answer and obtain more information about it. Let's try to answer to the questions.

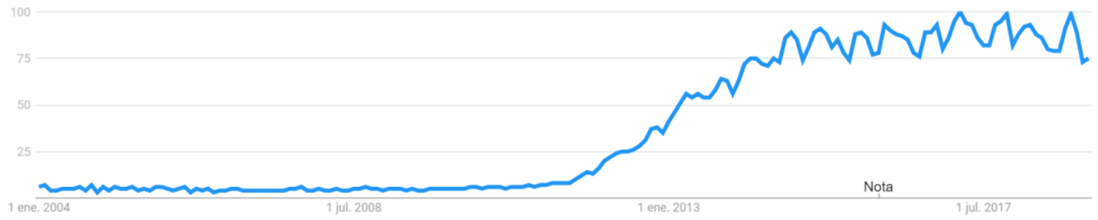
**1. What is the relationship between Research and Big Data? And what is the relationship between Big Data and football?**

As many of us know, the term Big Data has been used a lot recently. One of the points to have in mind is that Big Data is not just about collecting large volumes of data, but it is also about having the ability to know on what types of data to work or not, and which are useful and which are not.

We must first ask ourselves, for example, about the validation of the data and the different types of bias that may exist. In our field of sports analysis, if we don't take into account the internal validation or the external validation, we basically cannot move forward. Before analyzing, making a graph and wasting time misinterpreting from the beginning of the study or problem, one must use critical thinking in the face of the Big Data phenomenon. Therefore, apart from asking good questions in the face of a given problem, it will also be important to ask oneself about the quality of the data and the meaning that the internal and external validation has.

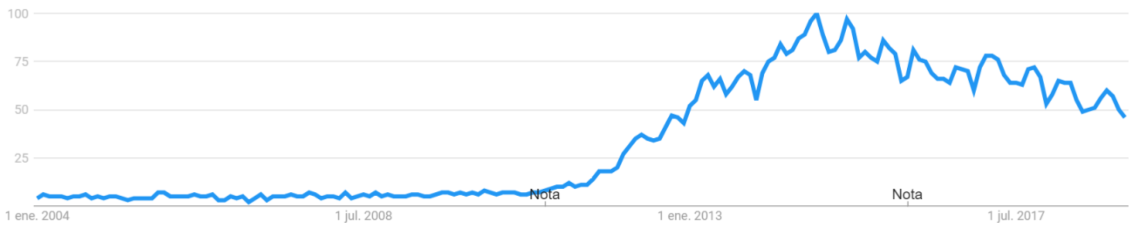
Big Data has been a term and a phenomenon of great interest, but lately it seems that this interest has been decreasing (at least in the United States) gradually, as we see in the following figure from googletrends.

**Figure 1: Google Trends results of the interest evolution, in general (all the countries), of the Big Data phenomenon, All the world, 2004-present.**




Source: own elaboration based on googletrends.

**Figure 2: Big data only in the United States, 2004-present**




Source: own elaboration based on googletrends.

By looking at both pictures, we can say that the moniker Big Data is dropping.

According to some specialized bibliography in the world of research, authors like Zhang or Malhouse reaffirm and remind us that "The big data era doesn't just mean large amounts of data; it also means increased ease and ability to collect data of all types, in all walks of life".

### 1.1. Big Data and Football

Subsection written by José María Fernández Ponce, professor of the University of Seville, member of the Department of Statistics and Operational Research and researcher of the group "*Métodos Cuantitativos en Investigación, FQM-328*" of the Andalusian Regional Government. He also holds a football coach degree (Level I) of the Andalusian Football Federation. At present, he is an external adviser of the football academy of *Cádiz Club de Fútbol*, where he is developing a project to create a database for the selection, detection and development of sports talent.

Big Data is a term widely used in every area of our society, yet it sometimes leads to confusion and misinterpretations due to the intrinsic complexity of what the concept implies. In professional football and in high performance sports in general, Big Data technology cannot be ignored as a tool for helping in decision making, since the budgets that clubs manage are several million euros, that is, companies whose economic management is high or even astronomical, if we talk about the top teams in La Liga. The question is whether Big Data in football is simply a fad that will sooner or later come to nothing or it is an imperative need to achieve modernization and, at the same time, optimize the sports performance of the teams. In this section, I will try to explain the idea of what Big Data can provide to football in its sports dimension, so as to justify its implementation in professional teams.

What is Big Data? The translation of the term, or at least its translation to Spanish, is *Datos Masivos* (massive data). Even though it is true that the definition provided by the Royal Spanish Academy is quite clear and concise, it may not be so for someone who is not a computer or mathematics expert. The first thing that comes to your mind is: "massive data means a lot of data", but how much data is a lot of data? Let's think about a desktop computer. One of the characteristics we look at when buying it is the hard disk memory, i.e. the data storage capacity of the computer. Normally, this memory ranges from 500 Gigabytes (GB) to 2 Terabytes (TB). In order to understand what these units mean, let's look at the following case. In 1GB, you can store about 300 songs in average mp3 size (about three and a half minutes long and with an average quality). Then, on a 500GB PC, you could store about 150,000 mp3 songs. By having in mind that 1 TB is 1000 GB and by cross-multiplying, we could assert that, in a computer with a 2 TB hard disk memory, we could store approximately 600,000 songs in mp3. If each one of them has an average duration of three minutes and a half, it will take us almost four years to hear them all without stopping. Another very clear example is the game FIFA19 for PC, which occupies 41 GB in memory; therefore, we could store about 48 games of the same capacity in a 2 TB computer.

With the use of the Internet, the flow of data has increased tremendously not only by its volume (the memory that would be needed to store and process it) but also by its variety. For example, nowadays we use many different types of social media: Facebook, Twitter, Whatsapp, among others. In them, any kind of information can be exchanged: messages, photos, videos, gifs and so on. This variety, together with the volume of data, can cause a Big Data problem. We could have so much data (massive in volume and/or variety) that it would be impossible to treat them in a conventional PC, mainly because they would exceed the capacity of the TB of the desktop computer. A real Big Data case is Amazon. This company is interesting in having all its clients satisfied so as to get greater benefits. To do this, it classifies them according to their tastes and purchase history, so as to send them individualized advertisement to their electronic devices from which they have purchased or simple viewed their products. How many customers does

Amazon have worldwide? In the United States, there are around 93 million of prime customers; therefore, the volume of all these clients, together with their data, cannot be managed from a normal PC. The problem in question is not very exhaustive, since the reality is somewhat more complex, but it does allow us to understand the principle of the concept of Big Data technology as the set of hardware (that is, electronic components and computers) and software (computer programs) that allow us to treat and analyze massive data. It is not enough to have only one machine, it is necessary to have many of them (computers connected in a network) containing all the information. Moreover, in order to manipulate the data, it is also essential to have software that facilitates the distribution of that information and the obtaining of results (what is called, in computer jargon, the parallel and/or distributed computing). But it is even more complicated. The emergence of this computer technology for mass data processing has created the need for a new approach to statistics. Paraphrasing Professor Daniel Peña of the University Carlos III of Madrid, the appearance of Big Data has meant, to classical statistics (that which deals with small samples), something equivalent to what Einstein's Theory of Relativity meant to Newton's classical physics. In short, the Big Data has forced mathematicians/statisticians into posing hitherto unknown problems and providing solutions to this new paradigm.

**Figure 1: Professional players of football wearing Wimu.**

Source: https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub_231195_102.html

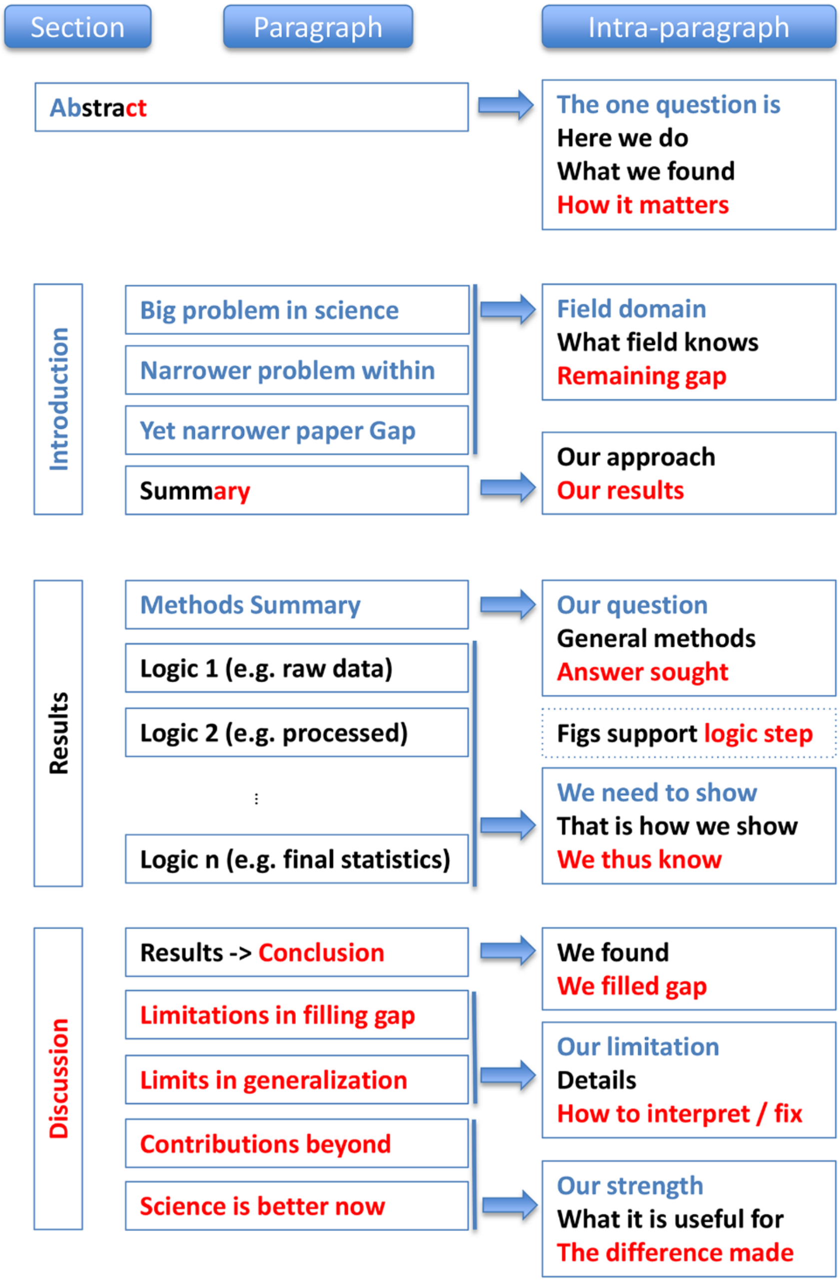
And in football, is there so much data that it is impossible to manipulate them on a PC? The advance on the Information Technologies has caused us to currently live in a technological revolution of football. For example, since the 2015-16 season, the International Football Association Board has allowed teams to wear GPS in official games, and it was not until this last season that these data has been allowed to be sent to team benches in real time via electronic devices (tablets, mobile phones or others). The GPS worn by the players can be clearly observed underneath the shirts, on the back just between the shoulder blades, at shoulder height (see figure 1). There are many companies that commercialize this product, but among all of them we can highlight the Andalusian Real Track Systems, which works with its WIMU PRO system. This company that applies GPS to the Sciences of the Sport is quite widespread at an international level and, naturally, among LFP teams. This device generates data every second per player in more than 250 recorded variables (heart rate, oxygen saturation, distance covered, maximum speed, accelerations/minute, decelerations/minutes, sprints, tactical information, among many others). It is also important to highlight that La Liga has developed its own software (Mediacoach) for data collection during games, which can be complemented with those of Wimu Pro. The analysis of these data can be really helpful depending on the information that needs to be extracted. For example, this can help detect possible injuries of players by programming algorithms that notifies of such risks and, therefore, it could be decided to replace a player or reduce their load during training sessions. It could also be really useful to analyze the tactical pattern recognition from the data provided by these technologies. But, what is the problem in reality? In order to explain the problem in reality, let's assume first that a team has 23 players and they perform 200 sessions of 90 minutes during the season. Moreover, let's suppose that each datum occupies the smaller volume possible in memory (6 bytes). Therefore, the problem is that, with this technology, we would have a data volume of 149 in total, that is to say we would be dealing with a much larger data volume than a normal PC. If we also consider that during a season there are about 40 games, it is clear that in professional football there is a need to work with Big Data technology and with professional data analysts from the field of mathematics/statistics. As a curiosity, we could add that Liverpool FC has three data analysts with this scientific profile in their Performance Analysis Department. Moreover, all English Premier League teams have at least one specialist in this field, not to mention those in American professional sports (NBA, MLB and NFL).

In general, we can say that Big Data in football is not a fantastic machine, but rather it is a lot of machines doing the work that humans could not do. Obviously, Big Data is not an infallible system, even less in football, which is probably the most hazardous of all sports, although it seems the opposite. Nevertheless, it is an essential tool in professional football teams for decision making, where the human factor based on experience and knowledge should intervene. Big Data is already present in Spanish football and, sooner or later, professional teams will incline towards having specialists in mathematics, statistics and computers (some already have them) in their technical secretariats. The reason for this is that even if data privacy is guaranteed, would we hire a sports analysis company for our football club that also works for our opponent?

2. **What is a scientific article? What parts does it normally have? What is the typology depending on the area of specialization?**

When we enter the scientific world, it seems that many concepts or skills are obvious, but this is not the case. It is really important to know the information and what type of sections can be found (for example: introduction, methods, results, discussions and conclusions) in order to better interpret the content of the article in question.

**Figure 2: Summary of a paper's structural elements at three spatial scales: Across sections, across paragraphs, and within paragraphs.**




Source: Mensh, B. & Kording, K. (2017).

### 3. What Sports Analytics journals can I find?

With society's growing fascination with sport and the progressive availability of sport-related data, there are great opportunities for sports analytics research. In the article published by Swartz (2018), the author explains some relevant problems that quantitative scientists can find when publishing in the field of sports analytics. He also elaborates on some experiences and opinions that make us aware of journals that make reference to sports statistics, sports technology and computing, sports analysis, and sports science and sports medicine. The two most popular academic journals in sport analysis are from the United States: *Journal of Quantitative Analysis in Sports (JQAS)* (https://www.degruyter.com/view/j/jqas?lang=en) and *Journal of Sports Analytics (JSA)* (https://www.journalofsportsanalytics.com/).

## 3.4. Analysis of the 4th paragraph of the CASE

If we consider the following paragraph:

“"The data scientist is in a bit of a hurry and introduces me to one of the FCB's tactical or game analysts from the first team. This provides me with a **database** (**validated** by the IT team) of the last 15 league games of all teams (**they couldn't get the first 10 games**), where there are different **characteristics** of the teams (example: Home/Away Stadium), initial training (example: 4-3-3, 4-4-2), possession percentage, number shot left, number shot middle, number shot right, shot conversion rate, number of corners, number of fouls, final result) so I can **see some patterns and describe behaviors** of some of them. The analyst tells me that he **doesn't have a specific question** to ask me **right now**, but he wanted to **know if there's a connection** between scoring before the opponent and the final result of the match”.

We could ask ourselves questions of the following kind:

**4. What is data validation? What are biases really and what types are there?**
**5. What is the difference between sample data and population data? What kind of variables can we find?**
**6. What do we get from descriptive statistics?**
**7. And what do we get from inferential statistics?**

In order to answer most questions (4-7), it is important and useful to refresh and learn concepts from the world of statistics education. If the person who is taking the course comes from the world of sports, there are different books that have been explained in the academy where some of these concepts are exemplified through sport and statistics (O'Donoghue, P. and Holmes, L., 2014). Other related concepts can also be explored in detail in bibliography that is more applied and modern or more focused on advanced statistics and sport (O'Donoghue, P., 2013).

Although it is not possible to cover or summarize the knowledge of statistics and of the most advanced sports statistics in a minicourse, we will mention below some academic aspects that cannot be overlooked in this field before starting to analyze in front of a computer and going deeper into the historical and initial statistical thinking (Albert, J. and Koning, R. H. Eds., 2007).

**A Journey through some Basic Concepts of the Statistics World**

Many of us have had a statistics course in our university's program of studies, but we no longer remember what it was about or how useful it was. Some have been lucky enough to have a more motivating teacher, who may have shown them a video like this: https://www.youtube.com/watch?v=cQx_yP0pPpk so as to start this course full of interesting knowledge. Statistics is no longer defined as one of the branches of mathematics, as we have already seen, but it is the science of learning from data, uncertainty is measured, controlled and communicated. This is a young science, which some part of society does not know yet. One proof of this is that the statistician is usually mistaken for the statesman (State profession) or the statistician is thought to be a kind of computer scientist who calculates means, medians and makes some numerical calculations. Statistics quantifies uncertainty and advises on how to collect the data so that they provide as much information as possible. In medicine, biostatistics has been studied, which is the science that refers to the application and development of statistics in life sciences and, more particularly, in health sciences. The big difference between statistics and biostatistics is that, in the latter, you have to be familiar with other disciplines, such as epidemiology, genetics, demography, public health, etc. The same happens in Sports Analytics. The big difference between the latter and statistics is that one must also know about other disciplines so as to learn more about this phenomenon. When we read scientific articles or we want to carry out research projects, it is important to understand different basic statistical terms:

- How to differentiate between population (objective of our study and all the data we want to know) and sample (subset of population and the data we can work with). From here, it is important to know how the sample is selected and, for this purpose, different sampling techniques (for example: random, stratified) and their possible selection or information biases (Elfil, M. and Negida, A., 2017).

- Knowing how to differentiate the nature, the level of measurement or scale and the type of variables (qualitative or quantitative) or characteristics on which we work is essential so as to carrying out proper analysis on these. An important point of the correct use of statistics is knowing what we really want to answer. Questions like what is going on? (Descriptive statistics), what happened? (Diagnosis), what is going to happen? (Prediction) or what should we do? (Prescription) point out possible statistical tools that can be used. Another key distinction to keep in mind is that studies can either explore observed data (descriptive statistics) or use observed data from a sample to infer the study population (inferential statistics). In fact, at present, according to one of the most popular scientists and statisticians in the world of health, there are three very different fields to study that answer three different questions and can get us closer to the goal of analysis quickly: Descriptive statistics, prediction and causal inference (Hernán, M. A., Hsu, J. and Healy, B., 2019).

The descriptive statistics show us the data we have at that moment by means of, for example, appropriate frequency tables and graphs, depending on the variables we are working with. Different measures of centralization, dispersion, position and shape are also used for the quantitative variables. At the University of North Carolina at Chapel Hill, for example, they tended to pride themselves on the salary of their geography students who achieved a higher average salary after graduation than students of other schools. What they didn't know was that there was a student who graduated in geography through an athletic scholarship. This student was Michael Jordan, whose salary, as you can guess, was different from the others who graduated. In this case, we would not have to calculate the average wage, but another measure of centralization, such as the median. As the infographer Alberto Cairo claimed: “statistics doesn’t lie, the person who manipulates it does". It would also be important to complement this measure of centralization with some measure of dispersion, in order to always know the variability of its data. When describing quantitative variables, it is useful to use boxplots or other alternatives found in the bibliography, in which several measures are shown, such as the minimum, the maximum, the median and the first and third quartiles, which allow us to discover possible outliers. Imagine that the analyst says that the 3rd percentile of corner goals received from league teams is 2. This will mean that 3 % of the teams receive 2 or less corner goals. Therefore, it will always be interesting to know how to communicate and interpret this type of basic descriptive measures.

Statistical inference is intended to draw conclusions about a population from a sample. It is important to know two concepts, such as statistically significant differences or differences that are clinically, tactically, technically, economically (or other) relevant. Which of the two concepts is more important? Probably the second one, but, in order to answer, we need the first one. We are in a world where we are constantly comparing ourselves. One of the points that allows us to compare, relate, test or estimate is what is known as a contrast or hypothesis test. There is also the concept of modeling. The basis of many published scientific conclusions is the concept of statistical significance, usually evaluated by means of an index called p-value. A p-value is the probability, for a specified statistical model, that a statistic that synthesizes some characteristic of the data (for example, the difference in means when comparing two groups) is equal to or more extreme than the observed value. However, although p-value can be a useful statistical measure, it is often misused, abused and misinterpreted. This has caused some scientific journals to not recommend its use so much. In this context, the American Statistical Association (ASA) provided a formal statement to the scientific community, clarifying some principles that are widely accepted and implicit in the correct use and interpretation of p-value (The American Statistical Association, 2016). A single index should not replace scientific reasoning. Some alternatives to this index (such as the confidence interval [CI], Cohen's d, effect measures, etc.) have also been proposed and are worth considering. As a consequence of the American Statistical Association's Statement on Statistical Significance and P-Values, the New England Journal of Medicine recently published an editorial on New Guidelines for Statistical Reporting in the Journal (Harrington, D. et al., 2019). The editorial begins by stating that perhaps some readers may have noticed a more parsimonious reporting regarding p-value information. It also explains that this is because recently implemented journal guidelines limiting the use of p-values for secondary and other comparisons. They state that they have now clarified, expanded and refined the statistical guidelines for authors to cover both clinical trials and observational studies. You can find the link to the statistical guidelines for authors below: *Statistical reporting guidelines* (https://www.nejm.org/author-center/new-manuscripts).

- When relating two quantitative variables that have a linear relationship, it is specified that these variables are correlated. This concept (linear relationship) is often mistaken for causality, a very different concept for which design and other more complex aspects must be taken into account.

- It is normal to compare our target or dependent variable with another characteristic of interest or independent variable. Even so, in sports medicine, for example, as in other fields, our objective (e.g., to be injured [yes/no]) is often associated with more than one covariate (e.g., previous injury, age, playing surface, time of the season) and we must take into account the multiple factors, some of which are latent and, above all, dynamic, which can have an influence. The specification and validation of statistical regression models, and the taking into account of confounding or modifying variables, are aspects that must be verified in order to know the etiology of injuries or the factors associated with them and also for the use of predictive models.

- Most of the questions in sport analysis that a staff wants answers to are: the diagnosis, the prognosis and the search for risk factors of a particular problem. These can be variable factors (position on the pitch) and non-modifiable factors (such as age). In sports medicine, we also often talk about intrinsic factors (e.g. age, ethnicity, gender, genetic predisposition, previous history of injury) or extrinsic factors (e.g. nutrition, psychological factors, balance of muscle strength, load, fatigue, flexibility). The doctor tries to recognize all this, using statistical models that identify these factors and help them. Along with this, in the last decade, work has been done on biostatistics applied to personalized medicine, in which an attempt is made to answer what dose of treatment should be given to a patient according to their unique characteristics, since the treatment does not have to be the same for all patients suffering from the same disease or injury, see: (https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/). However, knowing the risk factors or identifying subpopulations that increase or decrease the risk of injury is often not enough, and one must try to answer how and why this injury or disease has developed.

- In the field of sport analysis, it is also essential to know the statistical models and notions of probability. As far as statistical models are concerned, it is of great importance to bear in mind that, as some of the best coaches remind us, space and time are key when it comes to making decisions. Statistical models that take into account these two concepts (temporal and spatial) are being developed in different specializations of Sports Analytics. Some current reflections can be found on this website: https://barcainnovationhub.com/es/espacios-futbol-perspectiva-cuantitativa/. Although, as George Box said, it's also good to remember that "Essentially, all models are wrong, but some are useful". Finally, regarding the concept of probability, apart from going into concepts such as conditioned probability, Bayes' theorem, it is also interesting to have in mind the ELO qualification system (method for calculating the relative skill of chess players) and different metrics applied to sports. Statistical models and probability lead us to concepts from the world of sports like the different popular metrics described in basketball, such as PER, Usage, +/-, Winscore, PTC, among others. And also in the world of football, they lead us to the already known expected goals (xG) among others. There is an extensive bibliography in: https://wikieducator.org/Sport_Informatics_and_Analytics/Performance_Monitoring/Expected_Goals#Expected_goals_literature and even some R-package created: (https://github.com/Jelagmil).

# References

**Albert, J. & Koning, R.** (Eds.) (2007). *Statistical thinking in sports*. CRC Press.

**Albert, J., Glickman, M., Swartz, T. & Koning, R.** (Eds.) (2017). *Handbook of Statistical Methods and Analyses in Sports*. CRC Press.

**Barca Innovation Hub** (May 9, 2019). *Los espacios en el fútbol desde una perspectiva cuantitativa* [Spaces in Football from a Quantitative Perspective]. Retrieved from: https://barcainnovationhub.com/es/espacios-futbol-perspectiva-cuantitativa/

**Barca Innovation Hub** (August 20, 2018). *La importancia de la bioestadística en las ciencias del deporte y la medicina deportiva* [The Importance of Biostatistics in Sports Science and Sports Medicine]. Retrieved from: https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/

**Elfil, M. & Negida, A.** (2017). Sampling methods in clinical research; an educational review. *Emergency*, 5(1).

**Harrington, D. *et al.***(2019). *New guidelines for statistical reporting in the journal.*

**Hernán, M., Hsu, J. & Healy, B.** (2019). A second chance to get causal inference right: A classification of data science tasks. *Chance*, 32(1), 42-49.

**Horton, M.** (2018). *Algorithms for the Analysis of Spatio-Temporal Data from Team Sports.*

**Jones, I.** (2014). *Research methods for sports studies*. Routledge.

***Journal of Quantitative Analysis in Sports (JQAS).*** Retrieved from: (https://www.degruyter.com/view/j/jqas?lang=en)

***Journal of Sports Analytics (JSA).***
Retrieved from: (https://www.journalofsportsanalytics.com/)

**Kai Zhang** (2018). How big data has created a big crisis in science. 05/01/2019, The Conversation. Retrieved from: https://theconversation.com/how-big-data-has-created-a-big-crisis-in-science-102835

**Lagos, J.** Create your own github profile. Retrieved from: https://github.com/Jelagmil

**Malthouse, E. & Li, H.** (2017). Opportunities for and pitfalls of using big data in advertising research. *Journal of Advertising*, 46(2), 227-235.

**Mensh, B. & Kording, K.** (2017). Ten simple rules for structuring papers. *PLoS computational biology*, 13(9), e1005619.

**Miller, T.** (2016). Sports Analytics and Data Science. *Winning the Game with Methods and Models.*

**Newell, J., Aitchison, T. & Grant, S.** (2014). *Statistics for sports and exercise science: a practical approach*. Routledge.

**O'Donoghue, P.** (2013). *Statistics for sport and exercise studies: An introduction*. Routledge.

**O'Donoghue, P. & Holmes, L.** (2014). *Data analysis in sport*. Routledge.

**Osborne, J.** (July 17, 2017). Understanding sports through Statistics. Retrieved from: https://www.youtube.com/watch?v=cQx_yP0pPpk

**Quintana O.** (March 22, 2019). Real track systems. *El primer ejemplo de éxito del Barca Innovation Hub.* [The first Successful Example of the Barca Innovation Hub.] Retrieved from: https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub_231195_102.html

**Swartz, T.** (2018). Where Should I Publish My Sports Paper? *The American Statistician*, (just-accepted), 1-18.

**Wiki Educator.** Sports Informatic and Analytics/ Performance Monitoring/ Expected Goals. Retrieved from: https://wikieducator.org/Sport_Informatics_and_Analytics/Performance_Monitoring/Expected_Goals#Expected_goals_literature