2.2 The evolution of sport and technology, and new data

The evolution of sport in recent years has been significant and, as we know, this creates a high level of expectation, has a substantial impact on economics, health and transfer of industry, as well as generating a new area of expertise. We have gone from keeping notes, during a football or basketball game, recording what we are seeing (notational analysis), to obtaining a wide range of data from increasingly sophisticated technologies.

The now-famous vests, which are currently used by many athletes in training and games, have sensors that, when incorporated into the desired articulation, can provide biomedical information. Previously, cross-referencing and coordination of different software were necessary. Now, the tactical, physical and medical aspects are all integrated at once, and all the information can be centralized in a virtual cloud. In the near future, this will probably help to achieve a more multidisciplinary approach across the different professions (i.e. managers, trainers, sports physicians, physiotherapists, physical trainers, rehabilitators, analysts, etc.), whether they are currently involved or will be in the future (that is, the professions that will become involved in sport in the future).

Nowadays, sport is controlled by records that analyze very different environments, ranging from the physiology of exercise, psychology, heat production, to modern digital systems for performance analysis during competition. At the moment, different professional teams are beginning to report information for surveillance or injury control using different devices. But how do athletes and staff perceive and accept these online sports health surveillance systems? According to a recent study, such systems are helping to improve communication between the athlete, coaching staff and medical team (Barboza, Bolling, Nauta, Van Mechelen, and Verhagen, 2017). Furthermore, from the sports physician's point of view, this helps to improve the athlete's knowledge and control, in order to intervene when necessary (Barboza et al., 2017).

One of the main issues that comes with using so much technology is related to measurement errors and with how these types of instruments are validated. Many assessments are carried out in sport using tests or other instruments. However, many times the validity and reliability of these are difficult to guarantee, despite widespread use. Can we ensure that the value we record after an assessment is reliable? And, if it is not, is it reliant on the person responsible for measuring it? We currently make decisions, but are we using data based on records with questionable validity? Do we share the responsibility? These are questions that we must ask ourselves.

At the moment, the challenge of knowledge has not to do with the increase in the type of records that can be generated and used, which will surely increase in parallel with technological progress, but what will come as a result of the selection of the records that are most significant, the creation of indexes that provide increased knowledge and, above all, the creation of statistical models, in terms of the athlete's performance, their adaptation and their injury risks. The sporting community that works closely with the scientific field clearly understands that sports analysis and the statistical and mathematical processing of this huge amount of data being produced is opening up new areas of expertise and business. In summary, we currently have access to more technology, which translates into much more data, but we still need to make better decisions based on adequate statistical models and analyses.

Statistics and Sports

Nowadays, we hear much more about statistics (or analytics) in sports than a few years ago. This coincides with a time when the science of statistics and its profession are booming, and this is considered one of the sexiest professions in the 21st century. Is statistics just a passing fad or a trending topic in the field of sport? History has shown that this is not the case. Within the industry, it is worth noting that the creation of notational analysis by Henry Chadwick took place in 1861, that the first thermal image in sport was produced by Professor James Connors in 1897, that the first statistician hired full-time for a professional team was Mr. Roth in 1947, and it continues until, for example, the famous book by Michael Lewis, "Moneyball: The Art of Winning an Unfair Game", published in 2003, which tells the true story of Billy Beane (played in the film by Brad Pitt), general manager of a modest Californian team that decided to use statistics to inform decision-making.

A very common question is who is really interested in statistics applied to sports. It is surprising to see the number of professionals interested in this area, from sports managers or decision-makers, players, coaches, trainers, sports physicians, rehabilitators, journalists, gambling venues or bookies, scouts or video analysts, to sports scientists, academics, fans, psychologists, epidemiologists, scientists from other fields as well as statisticians.

In the scientific world there has been a growing interest in applied statistics and, above all, in the increased scientific rigor in sports medicine and sports science in general. Statistics, together with other skills or professions, has created different common specializations, such as biostatistics, bioinformatics, geostatistics, econometrics and psychometrics (see Figure 1). The Hollywood production of Moneyball also helped to spark the interest of sports scientists in learning about different sport specializations, such as sabermetrics, Moneyball, sports analysts and sports biostatistics, where the skill of statistics is always present (see Figure 1).

Figure 1: Common specializations in the field of statistics

Box 1 Most common specialisations in the field of statistics

Statistics and Epidemiology, and Public Health | Medicine ≈ *Biostatistics*
Statistics and Computer Science, and Biology and Genetics ≈ *Bioinformatics*
Statistics and Geography ≈ *Geostatistics*
Statistics and Psychology ≈ *Psychometrics*
Statistics and Economics ≈ *Econometrics*

Box 2 Statistics and sports science specialisations

Statistics and Baseball ≈ *Sabermetrics*
Statistics and Sports Science, and Economics and Computer Science ≈ *Moneyball*
Statistics and Sports Science, and Video Analyst and Computer Science ≈ *Sports Analyst*
Statistics and Epidemiology, and Public Health | Medicine and Sports Science ≈ *Sports Biostatistician*

Source: Casals and Finch, 2017

What is statistics and how can it help us?

Many of us studied statistics as part of our university degrees, but we no longer remember what it was about or how useful it was. Statistics is no longer defined as a branch of mathematics, but rather as the science of learning from data, where uncertainty is measured, controlled and communicated. It is a young science that some people still know nothing about. Proof of this is the frequent confusion regarding the distinction between a statistician and a state or government official; some people consider a statistician to be a kind of computer scientist who calculates means, medians and who makes numerical calculations. Statistics quantifies uncertainty and provides information on how to collect data so that said data provides the maximum information 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 major 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. When reading scientific articles or carrying out research projects, it is important to understand different basic statistical terms:

- How to differentiate between population (target of our study and all the data we want to know) and sample (subset of the population and the data with which we can work). Then, we must know how the sample is selected and, for that, different sampling techniques (such as random, stratified) and the possible selection or information biases.

- Knowing how to differentiate the nature, the measurement level or scale and the type of variables (qualitative or quantitative) or characteristics of what we want to study is key to performing appropriate analyses.

- In order to use statistics correctly, it is very important to have a clear understanding of what we want to find out. Questions like what happens (descriptive statistics), what happened (diagnosis), what will happen (prediction) or what we should do (prescription) help us identify possible statistical tools that can be used. Another key distinction to take into account is that studies can explore observed data (descriptive statistics) or use data observed from a sample to infer the study population (inferential statistics). Descriptive statistics shows us the data we have at a given moment through, for example, the use of frequency tables or graphs, according to the variables we are working with. Different measures of central tendency, dispersion, position and form are also used for quantitative variables. For example, at the University of North Carolina, in Chapel Hill, there was a presumption that geography students obtained a higher average salary when they graduated, when compared to students from other faculties. What hadn't been taken into account was that there was one particular student who had graduated in Geography via a sports scholarship. This student was Michael Jordan, whose salary, as you may guess, was different to that of other students who graduated. In this case, we should not calculate the average wage, but instead need to use another central tendency, such as the median. As the infographics expert Alberto Cairo says, "statistics do not lie, the person who manipulates them lies". It would also be important to accompany this measure of central tendency with some measure of dispersion, to see the variability of the data. When describing quantitative variables it is useful to use box plots, where several measures are shown, such as the minimum, maximum, median, and the first and third quartiles, which makes it possible to discover possible outliers. Imagine that the sports biostatistician talks to the club's coordinating sports physician and tells him that, currently, the 3rd percentile (P3) of shoulder injuries in that club is 7 injuries. As experts in sports medicine, we can interpret that 3% of the club's athletes have shoulder injuries, equal to or less than 7 injuries. It will always be relevant to know how to communicate and interpret these types of basic descriptive measures.

- Statistical inference aims to draw conclusions about a population based on a sample. It is important to understand two concepts: statistically significant differences and clinically relevant differences. Which of the two concepts is more important? Probably the second, but, to answer that, we need the first. In today's world we are constantly making comparisons. One of the points that allows us to compare, relate, test or estimate is what is known as a contrast hypothesis or hypothesis testing. There is also the concept of statistical modeling. The basis of many published scientific conclusions is the concept of statistical significance, which is usually obtained by means of an indicator called the p-value. A p-value is the probability, under a specified statistical model, that a statistic that synthesizes some feature of the data (for example, the difference in the mean when comparing two groups) is equal to or more extreme than the observed value. However, although the p-value may be a useful statistical measure, it is often misused, overused, and misinterpreted. This has led to some scientific journals not recommending its use. In this context, the American Statistical Association (ASA) has made a formal declaration for the scientific community, in which it clarifies some principles that are widely accepted and that are implicit in the correct use and interpretation of the p-value (The American Statistical Association, 2016). A single indicator should not replace scientific reasoning. Some alternatives to the p-value have also been proposed (such as the confidence interval [CI], Cohen's d, the measures of effect, etc.) and it is worth keeping them in mind.

- When we relate two quantitative variables that have a linear relationship, the correlation between the two variables is specified. This concept (linear relationship) is often confused with causality, a very different concept for which the research design and other more complex aspects must be taken into account.

- Target or dependent variables are usually compared with another characteristic of interest or independent variable. Even so, in sports medicine, as happens in other fields, the target (for example: injury [yes/no]) is associated with more than one covariate (e.g.: previous injury, age, playing surface, moment of the season) and multiple factors must be taken into account, some variables are latent and especially the dynamic variables that can influence the target. The specification and validation of statistical regression models, and taking into consideration the confounding or modifying variables, are aspects that must be verified in order to understand the etiology of injuries or the factors associated with them, and also for the use of predictive models.

- Most sports medicine questions which a clinician seeks to answer are: diagnosis, prognosis and identification of risk factors for diseases or injuries. These may be variable factors (for example, whether an athlete smokes, the risk of suffering from a certain disease) and non-modifiable factors (such as age). In sports medicine, it is also common to consider intrinsic factors (e.g., age, ethnicity, sex, genetic predisposition, previous injury history) or extrinsic factors (e.g., nutrition, psychological factors, the ratio of muscle strength, load, fatigue, flexibility). The sports physician tries to take all of this into consideration, for which there is nothing better than a statistical model that identifies those factors. Along with this, in the last decade we have been working on biostatistics applied to personalized medicine, where a set dose is identified to give to a patient according to their unique characteristics, given that treatment is not exactly the same for all patients who suffer from the same disease or injury (goo.gl/gwCUGD). In any case, often knowing the risk factors or identifying subpopulations that increase or decrease the risk of injury is not enough, and we must try to identify how and why this injury or illness have developed.

Research methodology and epidemiological training

Health science professionals must keep in mind the importance of research and sports medicine is no exception in this sense, since it is a science where knowledge is acquired using scientific methods. To conduct research, it is important to know how to formulate questions and use scientific method. In this context intuitive knowledge or unfounded assumptions are not enough. The scientific method is a sequence of steps (observation, induction, hypothesis, test of the hypothesis by experimentation, confirmation or refutation of the hypothesis, and thesis or scientific theory) that must be followed to discover new knowledge and to test unknown hypotheses. Research and each step of the scientific method are linked to the main sections found in a scientific article (see Table 1) (Mabrouki, and Bosch, 2007).

Table 1: Steps of the Scientific Method

STEPS OF THE SCIENTIFIC METHOD	MAIN SECTIONS OF A SCIENTIFIC ARTICLE
Understand the problem to be studied	Introduction
Develop a hypothesis	Material and methods
Collect data	Results

Source: Prepared by the author

A key consideration when investigating is to know how to formulate good questions and, from there, implement the scientific method, being familiar with the different epidemiological designs, the epidemiological measures (frequency, association, impact) and, above all, following the consensus guidelines (e.g.: Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), Consolidated Standards of Reporting Trials (CONSORT), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), WHO Injury Surveillance Guidelines). Epidemiology is the science that studies the distribution and etiology of diseases or injuries in a population. In the world of sports, the International Olympic Committee (IOC) has been working with specialized epidemiology and research groups all over the world for years to improve the health of athletes.

In sports medicine there are different profiles that are related, while having their own skill sets and concerns. All the profiles described below are necessary for a team or a club, in addition to curiosity, enthusiasm and passion for sports medicine as a profession.

- Clinical: seeing and knowing how to diagnose patients on a daily basis keeps them close to clinical practice and the reality of sports medicine.
- Epidemiologist: reporting injuries, studying their distribution and possible causes helps them to prevent, monitor and care for athletes.
- Scientific: writing and understanding scientific papers following clinical practice guidelines keeps them up-to-date with developments in sports medicine.
- Video analysts: watching games and training sessions in person or on TV, and trying to understand the evolution of sports will help them to formulate better scientific and clinical questions.
- Statistical/analyst profile: asking questions and trying to understand the patterns and the probability of events will help them develop quantitative skills.

Current discussions about whether a sports physician is more of a clinical doctor or epidemiologist are useless because in the professional world both skills are necessary. Often there is still a divide between clinical practice and research, and between practitioners and academics and researchers. Although not all of us conduct research, we must be active consumers of research literature, to incorporate more knowledge and rigor to our actions. Even so, this divide is increasingly less noticeable, thanks to the requirement for more multidisciplinary work and the ability to interpret data that is being recorded. Currently, the British Journal of Sports Medicine (BJSM) is publishing several educational editorials that emphasize epidemiology and statistics. This allows professionals who work with sports teams to connect with the more academic world (Nielsen, et al., 2017a, Nielsen, et al., 2017b). These editorials are helping professionals to be able to understand scientific articles, and helping coaches and sports physicians to distinguish between measures such as injury prevalence and incidence. In any case, scientists, as in other disciplines, are trying to improve different aspects. John Ioannidis (2005), one of the pioneers of so-called metascience, a discipline that analyzes the work of other scientists and checks whether the fundamental rules that define good science are being followed, discovered that there is significant margin for improvement in the majority of scientific articles. In this regard, two key aspects are being worked on: the reproducibility and the replicability of data, to respond to this scientific crisis.

Sports Biostatistician: a new profession that contributes to injury prevention

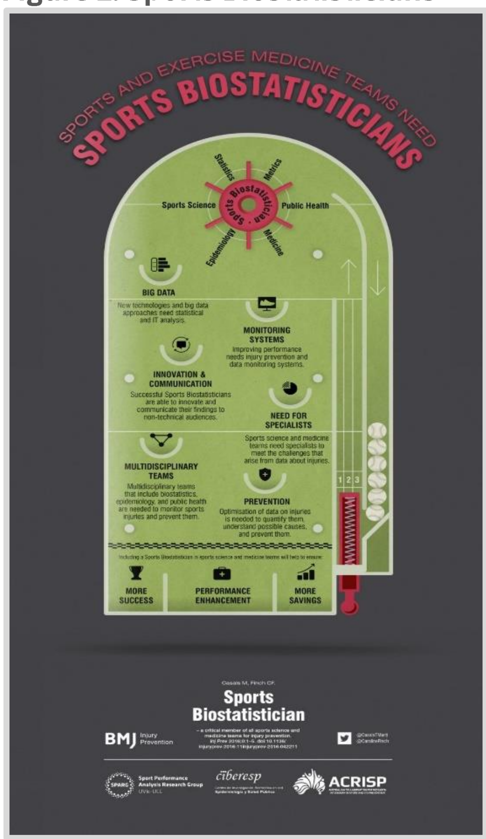
"The new professional specialization, sports biostatistician, can help in the optimization of injury data to quantify it, comprehend the possible causes and thus prevent injuries" ("La Biestadística Deportiva pide cancha para evitar lesiones", 2017, https://goo.gl/B633Kg), according to a new study published in Injury Prevention (Casals and Finch, 2016). In this investigation, Martí Casals, professor and researcher at Sport Performance Analysis Research Group (SPARG) of the University of Vic-Central University of Catalonia (UVic-UCC), current FC Barcelona biostatistician and with experience working with an NBA team, together with Caroline Finch, from the Australian Collaboration for Research into Sports and its Prevention (Federation University Australia), which is one of the nine research centers recognized by the IOC for injury prevention and the protection of athletes' health, describe the emerging field of sports biostatistics (Casals and Finch, 2016).

> "To discuss performance is also to indirectly discuss injury prevention and injury prevention surveillance systems... Sports science and medicine need specialists to solve the challenges that arise with injury data. One of the specialty professions that can help carry out these functions is the sports biostatistician, which is more widely known in the United States or Australia." (Casals and Finch, as cited in "La estadística también ayuda a prevenir lesiones", 2016, https://goo.gl/UtpUHd).

> "This new role requires understanding the etiology of injuries, being highly skilled in statistics, epidemiology, and computer programming with data, as well as excellent communication skills, since the biostatistician needs to convey their conclusions to the diverse range of people who make up the sports community, from parents to coaches, to players, sports physicians, clinicians, physical trainers, physiotherapists, sports scientists, epidemiologists and managers who make decisions in clubs." (Pichel Andrés, 2017).

The large clubs in the most important sports in our environment have begun to include (in addition to sports analyst) sports biostatisticians. The following infographic, published in the British Journal of Sports Medicine, shows a summary of the specialty's characteristics (Casals, Bekker and Finch, 2017).

Figure 2: Sports Biostatisticians



Source: Casals et al., 2017.

References

Barboza, S. D., Bolling, C. S., Nauta, J., Van Mechelen, M., and Verhagen, E. (2017). Acceptability and perceptions of end-users towards an online sports-health surveillance system. *BMJ Open Sport & Exercise Medicine, 3*(1).

Casals, M., and Finch, C. F. (2016). Sports Biostatisticians – a critical member of all sports science and medicine teams for injury prevention. *Injury Prevention*, *23*(6), 423-427.

Casals, M., Bekker, S., and Finch, C. F. (2017). Infographic: Sports Biostatisticians as a critical member of all sports science and medical teams for injury prevention. *Inj Prev. 2017 Dec;23(6):423-427.*

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8).

La biestadística deportiva pide cancha para evitar lesiones. (Tuesday, January 03, 2017). Retrieved from http://www.immedicohospitalario.es/noticia/10183/la-biestadistica-deportiva-pide-cancha-para-evitar-lesiones

La estadística también ayuda a prevenir lesiones. (Thursday, January 05, 2017). *Sinc, la ciencia es noticia* [Digital version]. Retrieved from http://www.agenciasinc.es/Noticias/La-estadistica-tambien-ayuda-a-prevenir-lesiones

Mabrouki, K., and Bosch, F. (2007). Redacción científica en biomedicina. Lo que hay que saber. Barcelona, SP: Prous Science.

Nielsen, R. O., Debes-Kristensen, K., Hulme, A., Bertelsen, M. L., Møller, M., Parner, E. T., and Mansournia, M. A. (2017a). Are prevalence measures better than incidence measures in sports injury research?. *Br J Sports Med. 2017 Oct 13.*

Nielsen, R. O., Chapman, C. M., Louis, W. R., Stovitz, S. D., Mansournia, M. A., Windt, J., Møller, M., Thorlund Parner, E., Hulme, A., Lejbach Bertelsen, M., Finch, C. F., Casals, M., and Verhagen, E. (2017b). Seven sins when interpreting statistics in sports injury science. *Br J Sports Med. 2017 Dec 20.*

Nielsen, R. O., Bertelsen, M. L., Verhagen, E., Mansournia, M. A., Hulme, A., Møller, M., and Casals, M. (2017). When is a study result important for athletes, clinicians and team coaches/staff? *British Journal of Sports Medicine*, *51,* 1454-1455.

Pichel Andrés, J.(January 16, 2017). Bioestadística y “big data” para prevenir las lesiones deportivas. Retrieved from https://www.bez.es/882750780/bioestadistica-big-data-para-prevenir-lesiones.html

The American Statistical Association (ASA). (March 7, 2016). American Statistical Association Releases Statement On Statistical Significance and *P*-Values. Provides Principles to Improve the Conduct and Interpretation of Quantitative Science. Retrieved from http://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

Verhagen, E., Stovitz, S. D., Mansournia, M. A., Nielsen, R. O., and Shrier, I. (2017). BJSM educational editorials: methods matter. *Br J Sports Med. 2017 Aug 17.*