

Can we really predict injuries in team sports?

Dae-Jin Lee

Applied Statistics Research Line

BCAM-Basque Center for Applied Mathematics

ORCID: 0000-0002-8995-8535

dlee@bcamath.org

Lore Zumeta-Olaskoaga

Applied Statistics Research Line

BCAM-Basque Center for Applied Mathematics

Departamento de Matemáticas

Universidad del País Vasco (UPV/EHU)

ORCID: 0000-0001-6141-1469

lzumeta@bcamath.org

Abstract

In the last decade several works have emerged in which statistical and machine learning methods have been proposed for the prediction of sports injuries. The field of medicine and sports science has included in its area multidisciplinary profiles with expertise in data analysis, injury epidemiology or artificial intelligence. However, injury phenomena are very complex and multifactorial. Understanding the mechanisms that produce an injury remains extremely complex and requires expert knowledge. This paper aims to illustrate from a statistical perspective what challenges need to be addressed from data collection, analysis of athlete performance and scientific reflection on questions of interest for knowledge-based decision making in data analysis in sport.

Keywords: sports injury data, athlete's performance, statistical modelling, survival analysis.

MSC Subject classifications: 62J02, 62J07, 62N01, 62M10.

1. Introduction

Injuries are common in professional sports and can have significant physical, psychological and financial consequences on a team performance and considerable impact in athletes' careers. Understanding

injury risk factors and their interplay is thereby a key component of preventing future injuries in sport (Bahr and Krosshaug, 2005; Finch, 2006). During the last decade, thanks to the professionalisation of the specialists involved in sports teams and the use of new technologies (e.g. computer vision, thermal cameras, Global Positioning Systems, etc ...), the interest in the modelling and prediction of injuries in professional sports through machine learning and artificial intelligence algorithms has dramatically grown (see Fiscutean, 2021; Jauhiainen et al., 2021; Ley et al., 2022 for a detailed review). Hence, the field of sports medicine and sports science has become an interesting field of research for data scientist, statisticians and computer scientists, such that a new role of sports biostatistician, with knowledge in statistics, epidemiology, sports medicine and communication skills is more and more required in professional team sports (Casals and Finch, 2017).

Here, we will address some of the most important challenges facings sports science and medicine research, from our (probably biased) perspective and based on our recent experience in collaborations with a professional football team.

2. Some modelling challenges in sports injury

In this section, we focus on sports injury data modelling from the perspective of:

1. **Sports injury data.** From a descriptive analysis, to injury incidence and burden and graphical representations, exploratory data analysis all this is crucial to pose the right questions related to sports injuries epidemiology from a team sports perspective (e.g. is my supported team more impacted by injuries than the others?, which type of injuries were most frequent? And most burdensome? or how does injury affect on the performance of the team in terms of the final classification?).
2. **The analysis of training (internal and external) loads.** Internal Load represents an individual athlete's response to training, and can be quantified by the intensity and duration of the physiological stress imposed on the athlete. The internal load is better explained with the external load that consists of what can be measured by GPS and accelerometers (i.e. distance in different speed zones, total distance covered, etc ...).
3. **Self-report wellness.** Self-report wellness questionnaires are a relatively simple and inexpensive means for determining an athlete's training load and their subsequent responses to that training. In fact, this is the most common method for monitoring athlete fatigue and recovery. A substantial amount of research has been conducted which confirms that wellness questionnaires, can indicate changes in training load/stress in elite team sport athletes.
4. **Modelling injury risk.** Based on the question of interest (either epidemiological or an individual athlete's performance or conditioning) the modelling approach may differ. We consider a time-to-event analysis approach that is a useful statistical tool to analyze the influence of changing exposures on injury risk. Time-to-event modelling allow change in training load to be included as a time-varying exposure for sport injury and modelling recurrent events.

There are other many aspects that are related to sports injury that are not included in the previous classification that are of great interest in the sports injury field.

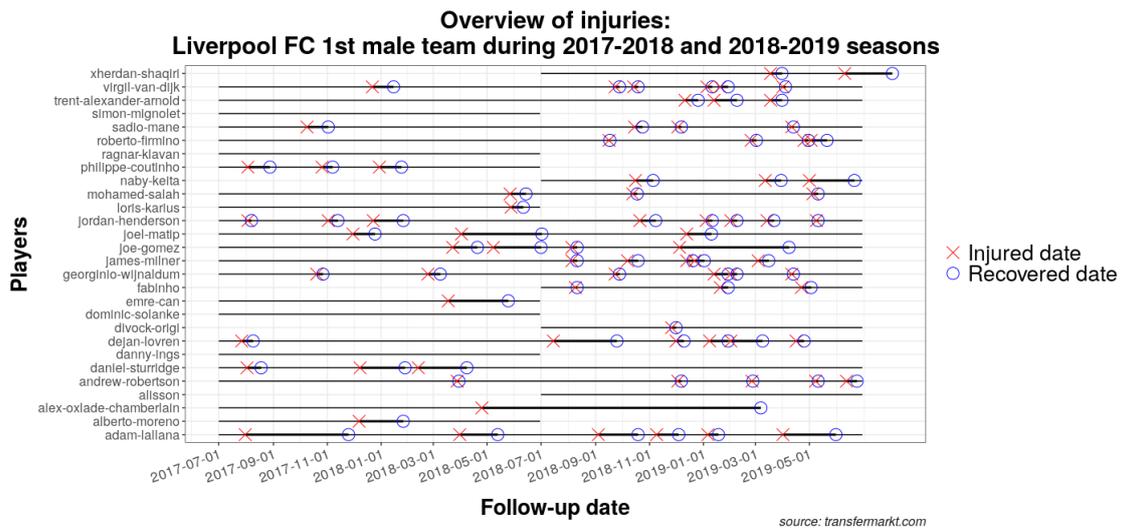


Fig. 1: Representation of Liverpool FC injuries on seasons 2017/18 and 2018/19.

2.1. Sports injury data analysis: the R package *injurytools*

The R package *injurytools* (L. Zumeta-Olaskoaga and Lee, 2022) facilitates the data analysis workflow by providing convenience functions and handy tools for sports injury data¹. To illustrate some capabilities of the package, it includes injury data from top European teams in the four leagues: La Liga (Spain), Bundesliga (Germany), Premier League (England) and Serie A (Italy). The package includes several functions that can be classified into (sports injury) data preparation, descriptive analyses and data visualisation routines.

The aim of the package is: 1) to provide a consistent way and general routines to analyse sports injury data, in R, including functions to perform informative visualisations and functions to facilitate the estimation of injury summary statistics, following the standards established in the consensus statement on injuries; 2) to help automate the descriptive reports that are routinely performed for sports injury surveillance. The statistical modelling of sports injuries is for the moment beyond the scope of *injurytools*, but the data structures are suitable for further analyses with other R packages and methods.

To illustrate some examples, we consider data scrapped from the German webpage Transfermarkt (<https://www.transfermarkt.com/>). Figure 1 shows a descriptive visualization of the injuries of Liverpool FC male team during 2017-2018 and 2018-2019 seasons. The horizontal axis represent the time line and the vertical line the Liverpool FC players. For each player, the black line represents the time the player was enrolled to the team with symbols × and ○ to denote the date of the injury and the date of recovery and player's availability to train and play matches respectively.

The extent of the sports injury problem is often described by injury incidence and by indicators of the severity of sports injuries. Sports injury incidence should preferably be expressed as the number of sports injuries per exposure time (e.g. per 1000 hours of sports participation, i.e. training sessions

¹ The *injurytools* package is under construction and can be accessed at <https://lzumeta.github.io/injurytools>.

Team	N players	N Injuries	N Injuries per season (IQR)	Incidence (95% CI)	N days lost	Median days lost per injury (IQR)	Burden (95% CI)
Barcelona	56	227	51.2 (50.5-53.2)	8.65 (7.7-9.6)	5261	12 (6-24)	200.48 (195.9-205.1)
Borussia Dortmund	55	385	93.2 (84.8-99.5)	17.39 (15.9-18.8)	5931	7 (3-17)	267.64 (262.1-273.6)
Liverpool	54	168	39 (28.5-45.5)	6.96 (6.1-7.8)	5275	14 (7-28)	218.41 (213.5-223.4)
Roma	57	224	54.2 (50.5-58.2)	10.25 (9.1-11.4)	4036	10 (5-20)	184.61 (179.8-189.4)

Tab. 1: Numerical comparisons among four European teams (Barcelona, Borussia Dortmund, Liverpool and Roma). Results are reported as 100 player-match exposure; CI: Confidence Interval; IQR: Interquartile Range. Incidence: number of injuries per unit of player-exposure time (frequency). Burden: number of days lost per unit of player-exposure time (severity and frequency).

and matches) in order to facilitate the comparability of research results (WW.Van Mechelen, Hlobil, and Kemper, 2012).

Thus, when attempting to describe the distribution of injuries it is necessary to relate this to the population at risk over a specified time period. This is why the fundamental unit of measurement is a rate. A rate is a measure that consists of a denominator and a numerator over a period of time. Denominator data can be a number of different things (e.g. number of minutes trained/played, number of matches played). As such, it reflects the speed at which new “injury-related” events occurs. There are two important definitions to consider:

Definition 1 (Injury incidence rate). *Injury incidence rate is the number of new injury cases (I) per unit of player-exposure time, i.e.*

$$I_r = \frac{I}{\Delta T} \quad (1)$$

Definition 2 (Injury burden rate). *Injury burden rate is the number of days lost (n_d) per unit of player-exposure time, i.e.*

$$I_{br} = \frac{n_d}{\Delta T} \quad (2)$$

where ΔT is the total time under risk of the study population.

Note that, either injury incidence (I_r) nor injury burden (I_{br}) are ratios, and they are not interpreted as a probability; they are rates and their unit (person-time)⁻¹ (e.g. per 1000h of player-exposure, per player-season etc ...).

In Table 1, exposure time unit is match minutes, hence injury incidence and injury burden are calculated per 100 player-matches of exposure (90 minutes times 100). Indeed, a correct exposure time should include training minutes for the total exposure time. However, Transfermarkt webpage do not collect the training minutes per team or per player.

Figure 2 shows the injury incidence and burden evolution of four European teams from season 2008-2009 to 2018-19. This plot is merely a descriptive and forecasting for the future seasons does not make any sense for such a short time series. The trend in the incidence of all type injuries has increased in Borussia Dortmund, for the rest of the teams the trend is not clear. However the injury burden has not a clear trend in any of the teams analyzed. Overall the most impacted team by

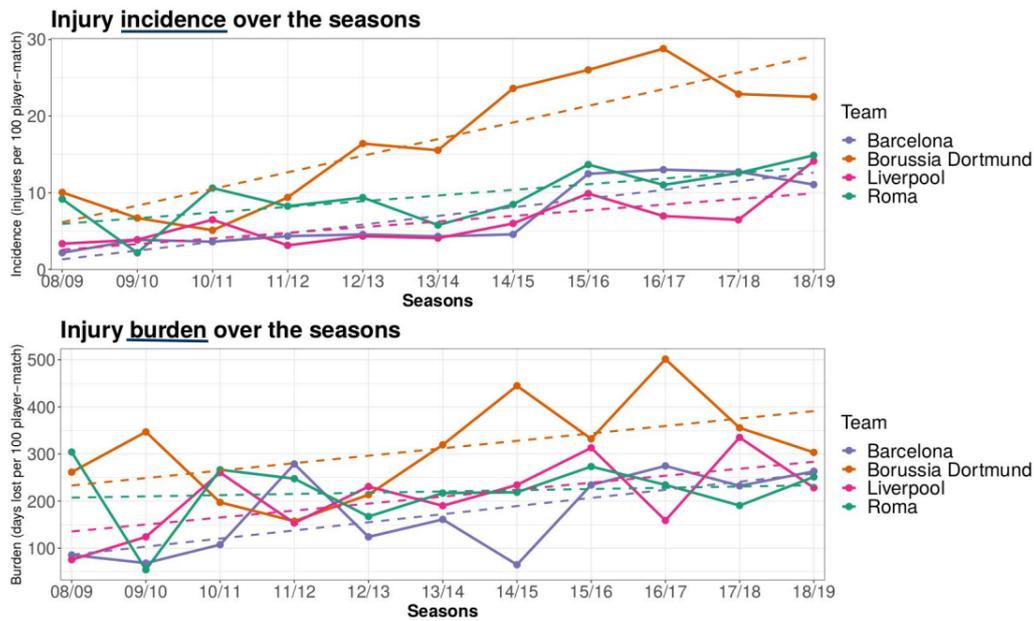


Fig. 2: Comparison of linear trends among four European teams (Barcelona, Borussia Dortmund, Liverpool and Roma). Incidence: number of injuries per unit of player-exposure time (frequency). Burden: number of days lost per unit of player-exposure time (severity and frequency).

injuries was Borussia Dortmund. Liverpool was the team with lowest injury incidence. In terms of the type of injuries (classified in Transfermarkt as muscular, ligament, concussion, bone and unknown), in all seasons and teams, most frequent injuries were muscle injuries. Ligament injuries were by far the most burdensome in Liverpool 2015/16, Roma 2016/17 and Borussia Dortmund 2017/18 (results not shown).

Another way to visualize sports injury data is the so-called risk matrix of injuries in Figure 3. For season 2017/18 it shows the relationship between the severity (consequence) and incidence (likelihood) of the most common injuries (Bahr, Clarsen, and Ekstrand, 2018; Fuller, 2018). The main advantages of using risk matrices, and the reasons for their attractiveness, are the minimal inputs required, the convenience of understanding the visual information presentation, the transparent nature of the assessment standards and the simplicity with which the conclusions can be communicated to stakeholders. Injury burden is most often used for risk evaluation that motivate the lost of days of training and matches, ranking the importance of injuries risk factors and prioritising injury prevention plans.

2.2. Athletes' performance: strength, conditioning and wellness

Strength and conditioning professionals aim to maximize athletic performance and reduce the associated injury risk. Therefore, understanding the relationships between different physical capacities and performance metrics, as well as the acute and long-term effects of distinct training interventions

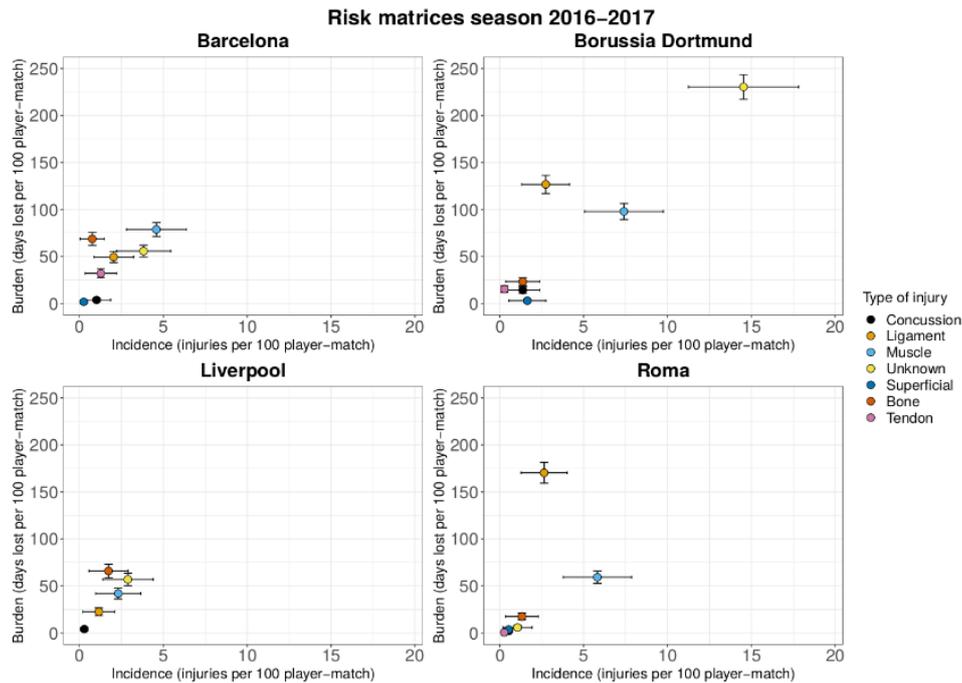


Fig. 3: Risk matrices for Barcelona, Borussia Dortmund, Liverpool and Roma for season 2016-2017.

on athletic populations is crucial for coaches and practitioners. Now, we will first define the internal and the external load.

The internal load

The Borg scale also known as rate of perceived exertion (or RPE) is an instrument that was created for the purpose of measuring effort in training, it measures, as its name suggests, the perception of effort, intensity and volume of physical activity, so it is a good alternative to assess the level of demand in each workout. The session rating of perceived exertion (sRPE) proposed by Foster, 1998 considers the overall effort of the training session (i.e. the product between RPE and the total time of the training/match session, which is also generally referred to as Training Load (TL)). Two different RPE scales are used in sports: (i) CR-10 where the RPE values are ranged between 0 (no exertion at all) and 10 (maximal exertion), and (ii) 6–20 scale where the values are ranged between 6 (no exertion at all) and 20 (maximal exertion). The TL is widely used in sports as an easy index describing the athletes' internal workload. Another important feature describing the internal workload is heart rate (HR). Even if HR is an important objective index of internal load, the use of heart rate monitoring in team sports is not a standardized procedure due to the fact that the chest strap is uncomfortable while performing contact sports.

The external load

External workloads are defined as the training features that describe the effort performed during training or match sessions. Global Position System (GPS) commonly records such features. The use of GPS ‘wearable technology’ in high-performance sport is becoming increasingly popular (Cummins et al., 2013; Colby et al., 2014; Chambers et al., 2015). The type of variables collected from the devices are:

- “*Kinematic variables*”. Measures athlete’s overall movement during a training session, e.g., total distance and high-speed running distance (Distance in meters covered above 5.5 m/s);
- “*Metabolic variables*”. Measures the energy expenditure of an athlete’s overall movement during a training session, e.g., high metabolic load distance (distance in meters covered by a player with a Metabolic Power is above 25.5 W/Kg);
- “*Mechanical variables*”. Describes athlete’s overall musculo-skeletal load during a training session, e.g., explosive distance (Distance in meters covered above 25.5 W/Kg and below 19.8 Km/h), and the number of accelerations and decelerations above 2 and 3 m/s^2 .

These features are the most used to evaluate external workloads and to predict the risk of injury (Rossi, Pappalardo, et al., 2018).

Self-reported wellness

Perceived wellness has been linked with both internal and external stressors, as well as muscle damage biomarker. Several questionnaires are used in sports to evaluate players’ well-being, the most general one consist of a 5-point Likert scale of 5 items (i.e., fatigue, sleep quality, soreness, stress, and mood), where 1 and 5 indicated the highest and lowest values of wellness for each item. See Table 2 (McLean et al., 2010).

Wellness data is not standardized between individuals, and equivalent scores may not indicate equivalent levels of fatigue and/or wellness (Thornton et al., 2016). The data must be considered within the individual context of each player and, thus, it’s necessary to use relative change within each player when interpreting longitudinal trends amongst groups.

These forms generally consist of 5-12 items using 1-to-5 or 1-to-10 point Likert scales, or modification of existing questionnaires by placing greater emphasis on ratings of muscle soreness, physical fatigue and general wellness.

In the past decade, significant efforts have been made to understand injury risk in sport using subjective (i.e. rating of perceived exertion) and objective (i.e. accelerometers, gyroscopes and magnetometers) player monitoring strategies.

3. Modelling sports injury risks

Modelling sports injury data encompasses the complex time-varying and recurrent nature of injuries: an athlete’s injury susceptibility may change over time, and moreover, an athlete can sustain more than one injury, as subsequent injuries are often influenced by previous ones (Hägglund, Waldén,

	5	4	3	2	1
Fatigue	Very fresh	Fresh	Normal	More tired than normal	Always tired
Sleep quality	Very restful	Good	Difficult falling asleep	Restless sleep	Insomnia
Soreness	Feeling great	Feeling good	Normal	Increase in soreness/lightness	Very sore
Stress	Very relaxed	Relaxed	Normal	Feeling stressed	Highly stressed
Mood	Very positive mood	A generally good mood	Less interested in other and/or activities than usual	Snappiness at teammates, family and co-workers	Highly annoyed/ irritable/down

Tab. 2: Wellness questionnaire test.

and Ekstrand, 2006). Models for recurrent events are appealing for sports injuries prevention (Ullah, Gabbett, and Finch, 2014; Rasmus Oestergaard Nielsen et al., 2019; R. O. Nielsen et al., 2019).

A non-exhaustive list of methods and algorithms in the literature are:

- Generalized linear/additive models, regression trees and random forests.
- Survival analysis and time-to-event data analysis.
- Mixed-effects models (longitudinal modelling).
- Multivariate times series for classification (injury/non-injury).
- Variable selection and dimension reduction.

Figure 4 presents the Kaplan-Meier curves for the four European teams analyzed in Section 2.1 for the time to the first injury of the season (for minutes per match played until the first injury of the season 2017/18). The Kaplan-Meier estimates is to be used to measure the fraction of football players available for training and matches for a certain amount of time. For recurrent events a gap time approach can be considered (Ullah, Gabbett, and Finch, 2014). Lore Zumeta-Olaskoaga et al., 2021 consider the gap time approach for predicting sports injuries with regularized cox regression models with frailty including covariates from functional screening tests and anthropometric measurements of female players during one regular season. A major challenge in sports injury data is usually the small sample size and the few number of injuries.

When internal and external load is considered in the analysis, the most commonly used measure is the acute:chronic workload ratio (ACWR), that comprised an athlete's 'fitness' and 'fatigue', and can be calculated using very basic time series analysis methods such as the rolling average (RA) model or the exponentially weighted moving average (EWMA) model. The actual value computed by the ACWR has different implications, and can assist fitness coaches in understanding the readiness of an athlete, the relative injury risk of an athlete from day-to-day, and therefore, with carefully planned intervention, can help to prevent injury. This ratio is usually considered as a flagging value for injury risk.

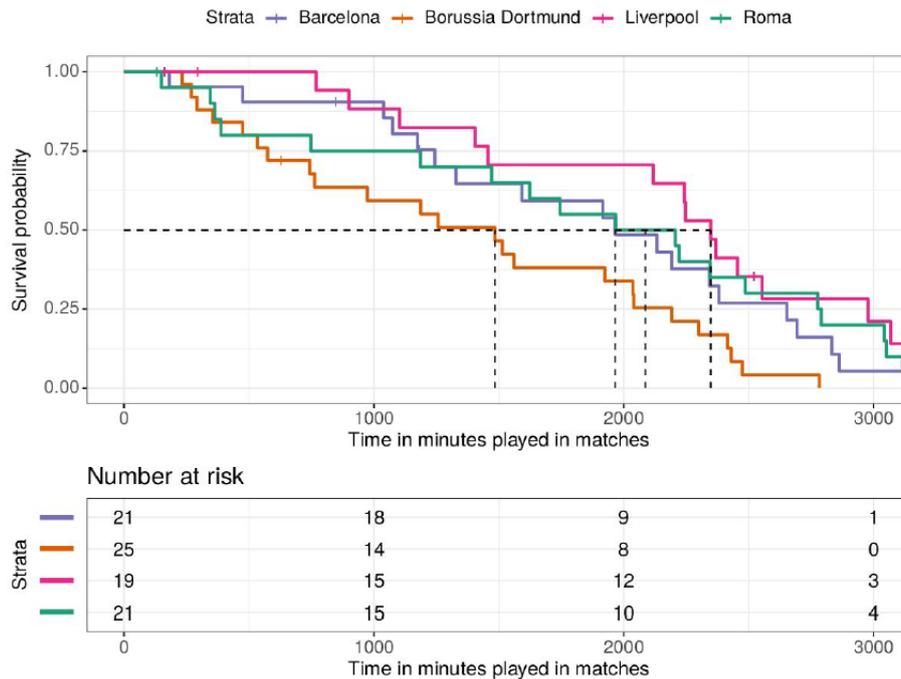


Fig. 4: Comparison of Kaplan-Meier curves for four European teams (Barcelona, Borussia Dortmund, Liverpool and Roma).

Definition 3 (Acute workload). *Typically, this is the workload performed by an athlete in 1-week (7 days). This value contains both training-and match-load information over this 7-day period. The acute workload represents the ‘fatigue’ aspect of the ACWR.*

Definition 4 (Chronic workload). *The chronic workload is typically the 4-week (28 day) average acute workload. This value is important as it provides a clear indication of what an athlete has done leading up to the present training or match day. Therefore, it is commonly viewed as an indication of an athlete’s ‘fitness’.*

Several studies suggested that large increases in acute workload with respect to the chronic workload (i.e. the average training workload of the previous month) are associated with an increased injury risk (Hulin et al., 2014). In particular, they showed that players with a high ratio between acute and chronic workload are more likely to become injured compared to those with a lower ratio. Traditional calculations of ACWR are ‘mathematically coupled’, as the most recent week is included in estimates of both the acute and chronic workloads. The uncoupled version consists of using the ACWR where the acute load is not part of the chronic load instead.

The R package ACWR (Fernandez-Santos, 2022), allows for computing the ACWR using three different methods: exponentially weighted moving average (EWMA), rolling average coupled (RAC) and rolling averaged uncoupled (RAU) in Williams et al., 2017; Windt and Gabbett, 2019.

Figure 5 illustrates the daily training load of an athlete (sPRE) through a regular season. The vertical lines represent the sRPE per type of session (match or training) and the grey shaded areas are the time period the athlete was injured. Similar plots can be obtained from other external and

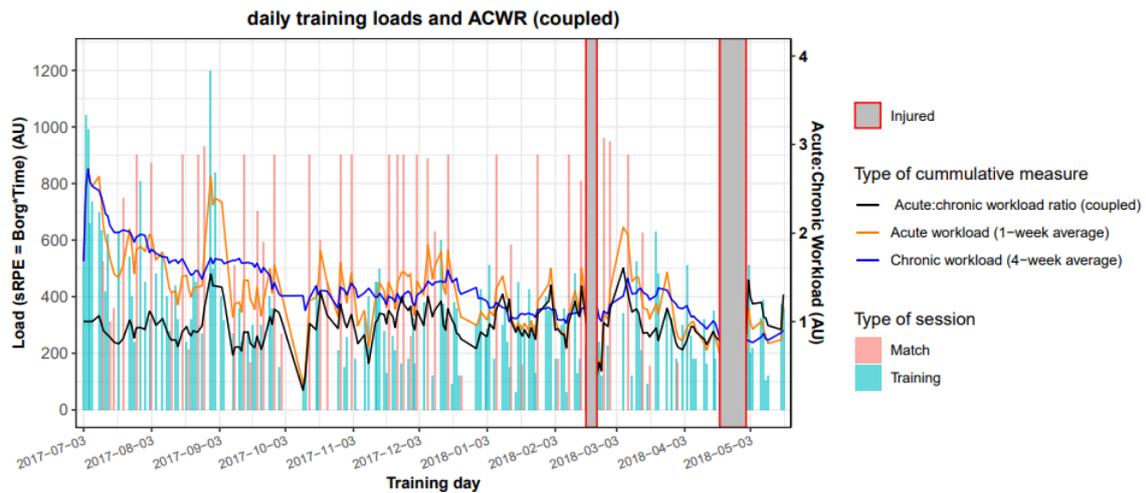


Fig. 5: Simulated daily training loads (sRPE) of an athlete over a season. There are training and match sessions and the external load measure consist of ACWR (coupled version). Additionally, acute and chronic workload are show. Grey shaded areas show the days the athlete was injured.

internal loads (i.e. kinematic, metabolic and mechanical variables, training loads and wellness tests) and a multivariate approach for forecasting injuries in soccer for evaluating and interpreting the complex relations between injury risk and training performance (Rossi, Pappalardo, et al., 2018; Rossi, Perri, et al., 2022).

4. So can we really predict injuries in team sports?

In the last decade, the number of studies about machine learning algorithms applied to sports, e.g., injury forecasting and athlete performance prediction, have rapidly increased. However, a world where we can prevent sports injuries before they happen is impossible, sports injuries occur and will continue to occur. However, it is entirely possible to accurately assess your risk level in terms of physical activity and injury. From lifestyle to biological constitution or genetic characteristics, there are many factors that influence an athlete's level of sports injury risk. In this paper, we have presented some challenges in team sports injury risk modelling, from the type of data collected, the concepts of performance and strength of the internal/external training loads and self-report wellness questionnaires. However, the leading approaches in machine learning are notoriously data-hungry. Unfortunately, in teams sports injury field there is no large number of injury data because acquiring data involves a process that is expensive or time-consuming.

However, the most important aspect in sports injury data modelling comes from a sports science and medical staff perspective. It is important to effectively use evidence-based knowledge to develop decision-making processes that reduce injury risk and optimize athlete performance (Drew, Raysmith, and Charlton, 2017; Meyer, 2017; Nassis, 2017). From our perspective, as statisticians, statistical modelling plays an important role in bridging the gap for understanding and quantifying the risk of team sports injuries where awareness about relevant concepts such as causality, association and

complexity are crucial rather prediction of an athletes' injury itself (Meeuwisse, 1994; Ruddy et al., 2019; Fonseca et al., 2020). An evidence-based injury risk assessment can help prevent future injuries and increase your potential for better performance.

Unfortunately, we can never predict injuries with complete certainty because we certainly can't predict the future. However, there is a way to determine injury risk, which in part can help predict or even prevent sports injuries.

Based on scientific research and the sciences of biomechanics, kinesiology, and ergonomics, the sports and medical communities have identified certain risk factors that can lead to sports injuries. Of course, risking any or all of these factors doesn't necessarily mean you'll end up hurting. However, knowing that you are at risk will help prevent many types of sports injuries in the future.

Acknowledgments

This research was funded by projects PID2020-115882RB-I00 funded by Agencia Estatal de Investigación and acronym "S3M1P4R", by the Basque Government (BERC 2022-2025 program) and by the Spanish Ministry of Science, Innovation, and Universities (BCAM Severo Ochoa accreditation SEV-2017-0718). This project has been also funded by the Provincial Council of Bizkaia within the Technology Transfer Programme 2022 and is co-financed by the European Regional Development Fund (ERDF) through the project "MATH4SPORTS - Modelización matemática para la industria deportiva: salud y rendimiento." Provincial Council of Bizkaia 6/12/TT/2022/00006 (BFA/DFB).

Acerca de los autores



Dae-Jin Lee Dae-Jin is a researcher at the Basque Center for Applied Mathematics (BCAM) and research line leader of the Applied Statistics research line (Data Science and Artificial Intelligence Area). His primary research interests are non-parametric smoothing methods in general with applications in Biostatistics, Environmental Modelling, Epidemiology, Mortality and Times Series Analysis. He is also working on Health-Related Quality of Life data modeling, complex data modeling, computational statistics and sports analytics. He is also Data Science coordinator of the Knowledge Transfer Unit at BCAM. The aim of BCAM Knowledge Transfer Unit is to develop mathematical solutions for scientific challenges based on real-life applications.



Lore Zumeta-Olaskoaga She is a PhD student at the BCAM - Basque Center for Applied Mathematics in the Applied Statistics Group under the supervision of Dr. Dae-Jin Lee. She is interested in statistical modelling research, software development and applications of statistics, in particular, to the fields of Sports Medicine, Epidemiology and Biomedicine. Her PhD project aims to develop statistical methods and software for the estimation of injury risk and incidence in the context of professional sport (football), focusing on the framework of recurrent time-to-event data analysis. The project is supported by the Severo Ochoa predoctoral fellowship (MICINN, Spain).

References

- Bahr, R., B. Clarsen, and J. Ekstrand (Aug. 2018). «Why we should focus on the burden of injuries and illnesses, not just their incidence». In: *Br J Sports Med* 52 (16), pp. 1018–1021. ISSN: 14730480. DOI: [10.1136/bjsports-2017-098160](https://doi.org/10.1136/bjsports-2017-098160).
- Bahr, R. and T. Krosshaug (June 2005). «Understanding injury mechanisms: a key component of preventing injuries in sport». In: *Br J Sports Med* 39 (6), pp. 324–329. ISSN: 03063674. DOI: [10.1136/bjism.2005.018341](https://doi.org/10.1136/bjism.2005.018341).
- Casals, M. and C. F. Finch (Dec. 2017). «Sports Biostatistician: A critical member of all sports science and medicine teams for injury prevention». In: *Injury Prevention* 23 (6), pp. 423–427. ISSN: 14755785. DOI: [10.1136/injuryprev-2016-042211](https://doi.org/10.1136/injuryprev-2016-042211).
- Chambers, R., T. J. Gabbett, M. H. Cole, and A. Beard (July 2015). «The Use of Wearable Microsensors to Quantify Sport-Specific Movements». In: *Sports Medicine* 45 (7), pp. 1065–1081. ISSN: 11792035. DOI: [10.1007/s40279-015-0332-9](https://doi.org/10.1007/s40279-015-0332-9).
- Colby, M. J., B. Dawson, J. Heasman, B. Rogalski, and T. J. Gabbett (2014). «Accelerometer and GPS-derived running loads and injury risk in elite Australian footballers». In: *Journal of Strength and Conditioning Research* 28 (8), pp. 2244–2252. ISSN: 15334295. DOI: [10.1519/JSC.0000000000000362](https://doi.org/10.1519/JSC.0000000000000362).
- Cummins, C., R. Orr, H. O'Connor, and C. West (Oct. 2013). «Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review». In: *Sports Medicine* 43 (10), pp. 1025–1042. ISSN: 01121642. DOI: [10.1007/s40279-013-0069-2](https://doi.org/10.1007/s40279-013-0069-2).
- Drew, M. K., B. P. Raysmith, and P. C. Charlton (Aug. 2017). «Injuries impair the chance of successful performance by sportspeople: A systematic review». In: *British Journal of Sports Medicine* 51 (16), pp. 1209–1214. ISSN: 14730480. DOI: [10.1136/bjsports-2016-096731](https://doi.org/10.1136/bjsports-2016-096731).
- Fernandez-Santos, J. R (2022). *ACWR: Acute Chronic Workload Ratio Calculation*. R package version 0.1.0. URL: <https://CRAN.R-project.org/package=ACWR>.
- Finch, C. F. (2006). «A new framework for research leading to sports injury prevention». In: *Journal of Science and Medicine in Sport* 9 (1-2), pp. 3–9. DOI: [10.1016/j.jsams.2006.02.009](https://doi.org/10.1016/j.jsams.2006.02.009).
- Fiscutean, A. (Apr. 2021). «Data scientists are predicting sports injuries with an algorithm». In: *Nature* 592 (S10-S11). DOI: [10.1038/d41586-021-00818-1](https://doi.org/10.1038/d41586-021-00818-1).
- Fonseca, S. T., T. R. Souza, E. Verhagen, R. van Emmerik, N.F.N. Bittencourt, L.D.M. Mendonça, A. G.P. Andrade, R. A. Resende, and J. M. Ocarino (2020). «Sports Injury Forecasting and Complexity: A Synergetic Approach». In: *Sports Medicine* (0123456789). ISSN: 11792035. DOI: [10.1007/s40279-020-01326-4](https://doi.org/10.1007/s40279-020-01326-4).

- Foster, C. (July 1998). «Monitoring training in athletes with reference to overtraining syndrome». In: *Med Sci Sports Exerc.* 30 (7), pp. 1164–1168. ISSN: 01959131. DOI: [10.1097/00005768-199807000-00023](https://doi.org/10.1097/00005768-199807000-00023).
- Fuller, C. W. (July 2018). «Injury Risk (Burden), Risk Matrices and Risk Contours in Team Sports: A Review of Principles, Practices and Problems». In: *Sports Medicine* 48 (7), pp. 1597–1606. ISSN: 11792035. DOI: [10.1007/s40279-018-0913-5](https://doi.org/10.1007/s40279-018-0913-5).
- Häggglund, M., M. Waldén, and J. Ekstrand (Sept. 2006). «Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons». In: *British journal of sports medicine* 40 (9), pp. 767–772. ISSN: 1473-0480. DOI: [10.1136/BJSM.2006.026609](https://doi.org/10.1136/BJSM.2006.026609).
- Hulin, B. T., T. J. Gabbett, P. Blanch, P. Chapman, D. Bailey, and J. W. Orchard (2014). «The acute-chronic workload ratio-injury figure and its ‘sweet spot’ are flawed». In: *British Journal of Sports Medicine* 48 (8), pp. 708–712. ISSN: 14730480. DOI: [10.1136/bjsports-2013-092524](https://doi.org/10.1136/bjsports-2013-092524).
- Jauhiainen, S., J. P. Kauppi, M. Leppänen, K. Pasanen, J. Parkkari, T. Vasankari, P. Kannus, and S. Ayramo (Feb. 2021). «New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes». In: *International journal of sports medicine* 42 (2), pp. 175–182. ISSN: 1439-3964. DOI: [10.1055/A-1231-5304](https://doi.org/10.1055/A-1231-5304).
- Ley, C., R. K. Martin, A. Pareek, A. Groll, R. Seil, and T. Tischer (2022). «Machine learning and conventional statistics: making sense of the differences». In: *Knee Surgery, Sports Traumatology, Arthroscopy* 30 (3), pp. 753–757. ISSN: 14337347. DOI: [10.1007/s00167-022-06896-6](https://doi.org/10.1007/s00167-022-06896-6).
- McLean, B. D., A.J. Coutts, V. Kelly, M. R. McGuigan, and S. J. Cormack (Sept. 2010). «Neuromuscular, Endocrine, and Perceptual Fatigue Responses During Different Length Between-Match Microcycles in Professional Rugby League Players». In: *International Journal of Sports Physiology and Performance* 5 (3), pp. 367–383. ISSN: 1555-0273. DOI: [10.1123/IJSPP.5.3.367](https://doi.org/10.1123/IJSPP.5.3.367).
- Meeuwisse, W.H. M.D. (1994). «Assessing causation in sport injury: A multifactorial model». In: *Clinical Journal of Sport Medicine* 4 (3), pp. 166–170. ISSN: 1050642X. DOI: [10.1097/00042752-199407000-00004](https://doi.org/10.1097/00042752-199407000-00004).
- Meyer, T. (May 2017). «How much scientific diagnostics for high-performance football?» In: *Science and Medicine in Football* 1 (2), p. 95. ISSN: 24734446. DOI: [10.1080/24733938.2017.1342327](https://doi.org/10.1080/24733938.2017.1342327).
- Nassis, G. P. (Sept. 2017). «Leadership in science and medicine: can you see the gap?» In: *Science and Medicine in Football* 1 (3), pp. 195–196. ISSN: 24734446. DOI: [10.1080/24733938.2017.1377845](https://doi.org/10.1080/24733938.2017.1377845).
- Nielsen, R. O., M. L. Bertelsen, D. Ramskov, M. Møller, Adam H., D. Theisen, C. F. Finch, L. V. Fortington, M. A. Mansournia, and E. T. Parner (Jan. 2019). «Time-to-event analysis for sports injury research part 2: Time-varying outcomes». In: *British Journal of Sports Medicine* 53 (1), pp. 70–78. ISSN: 14730480. DOI: [10.1136/BJSPORTS-2018-100000](https://doi.org/10.1136/BJSPORTS-2018-100000).
- Nielsen, Rasmus Oestergaard, Michael Lejbach Bertelsen, Daniel Ramskov, Merete Møller, Adam Hulme, Daniel Theisen, Caroline F. Finch, Lauren Victoria Fortington, Mohammad Ali Mansournia, and Erik Thorlund Parner (Jan. 2019). «Time-to-event analysis for sports injury research part 1: time-varying exposures». In: *British Journal of Sports Medicine* 53 (1), pp. 61–68. ISSN: 0306-3674. DOI: [10.1136/BJSPORTS-2018-099408](https://doi.org/10.1136/BJSPORTS-2018-099408).
- Rossi, A., L. Pappalardo, P. Cintia, F. M. Iaia, J. Fernández, and D. Medina (July 2018). «Effective injury forecasting in soccer with GPS training data and machine learning». In: *PLOS ONE* 13 (7). Ed. by Jaime Sampaio, e0201264. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0201264](https://doi.org/10.1371/journal.pone.0201264).
- Rossi, A., E. Perri, L.Pappalardo, P. Cintia, G. Alberti, D. Norman, and F. M. Iaia (June 2022). «Wellness Forecasting by External and Internal Workloads in Elite Soccer Players: A Machine Learning Approach». In: *Frontiers in Physiology* 13. ISSN: 1664042X. DOI: [10.3389/FPHYS.2022.896928/FULL](https://doi.org/10.3389/FPHYS.2022.896928/FULL).
-

-
- Ruddy, J. D., S.J. Cormack, R. Whiteley, M. D. Williams, R. G. Timmins, and D.A. Opar (July 2019). «Modeling the Risk of Team Sport Injuries: A Narrative Review of Different Statistical Approaches». In: *Frontiers in Physiology* 10, p. 829. ISSN: 1664-042X. DOI: [10.3389/fphys.2019.00829](https://doi.org/10.3389/fphys.2019.00829).
- Thornton, H. R., J. A. Delaney, G. M. Duthie, B. R. Scott, W. J. Chivers, C. E. Sanctuary, and B. J. Dascombe (May 2016). «Predicting Self-Reported Illness for Professional Team-Sport Athletes». In: *International journal of sports physiology and performance* 11 (4), pp. 543–550. ISSN: 1555-0273. DOI: [10.1123/IJSP.2015-0330](https://doi.org/10.1123/IJSP.2015-0330).
- Ullah, S., T. J. Gabbett, and C. F. Finch (2014). «Statistical modelling for recurrent events: An application to sports injuries». In: *British Journal of Sports Medicine* 48 (17), pp. 1287–1293. DOI: [10.1136/bjsports-2011-090803](https://doi.org/10.1136/bjsports-2011-090803).
- Williams, Sean, Stephen West, Matthew J Cross, and Keith A Stokes (2017). «Better way to determine the acute:chronic workload ratio?» In: *British Journal of Sports Medicine* 51 (3), pp. 209–210. ISSN: 0306-3674. DOI: [10.1136/bjsports-2016-096589](https://doi.org/10.1136/bjsports-2016-096589).
- Windt, J. and T. J. Gabbett (Aug. 2019). «Is it all for naught? What does mathematical coupling mean for acute:chronic workload ratios?» In: *British Journal of Sports Medicine* 53 (16), pp. 988–990. ISSN: 0306-3674. DOI: [10.1136/BJSPORTS-2017-098925](https://doi.org/10.1136/BJSPORTS-2017-098925).
- WW.Van Mechelen, H. Hlobil, and H.C.G. Kemper (Oct. 2012). «Incidence, Severity, Aetiology and Prevention of Sports Injuries». In: *Sports Medicine 1992 14:2* 14 (2), pp. 82–99. ISSN: 1179-2035. DOI: [10.2165/00007256-199214020-00002](https://doi.org/10.2165/00007256-199214020-00002).
- Zumeta-Olaskoaga, L. and D.-J. Lee (2022). *injurytools: A toolkit for Sports Injury Data Analysis*. <https://github.com/lzumeta/injurytools>, <https://lzumeta.github.io/injurytools/>.
- Zumeta-Olaskoaga, Lore, Maximilian Weigert, Jon Larruskain, Eder Bikandi, Igor Setuain, Josean Lekue, Helmut Küchenhoff, and Dae Jin Lee (Nov. 2021). «Prediction of sports injuries in football: a recurrent time-to-event approach using regularized Cox models». In: *AStA Advances in Statistical Analysis*, pp. 1–26. ISSN: 1863818X. DOI: [10.1007/S10182-021-00428-2/FIGURES/6](https://doi.org/10.1007/S10182-021-00428-2/FIGURES/6).
-