

Potential risk factors of injuries in professional football using Multivariate Survival Trees: a comparison of female vs. male football players

*Jone Renteria*¹, *Lore Zumeta-Olaskoaga*^{1,2}, *Eder Bikandi*³, *Jon Larruskain*³, *Dae-Jin Lee*⁴

¹{jrenteria,lzumeta}@bcamath.org, Applied Statistics Research Line, Basque Center for Applied Mathematics, Bilbao, Bizkaia, Spain

²Departamento de Matemáticas, Universidad del País Vasco UPV/EHU, Leioa, Bizkaia, Spain

³{e.bikandi,j.larruskain}@athletic-club.eus, Athletic Club, Medical Services, Lezama, Bizkaia, Spain.

⁴ daejin.lee@ie.edu, School of Science and Technology, IE University, Madrid, Spain

Nowadays, the use of information technologies has become a common and necessary practice among professional football clubs. During each match and training session, football teams collect a big amount of data from every movement that a particular player performs during the time he/she is on the field. Moreover, the medical team performs periodical check-ins that include performance physical tests, in which the players are screened for biomechanical movement, endurance, power, speed, and agility, among other characteristics. These screening tests are used to monitor players' functional and strength parameters, assess their physical fitness and thus provide crucial insight into areas that need improvement in order to reduce the likelihood of sports injuries.

In this work, we have gathered, cleaned, and curated all these multidimensional data to identify potential risk factors for sports injuries and better understand the event of injury by means of multivariate survival analysis for high-dimensional data. We have considered a Multivariate Survival Trees approach which can handle many covariates, requires few statistical assumptions, provides a sequence of prognostic rules that are easy to interpret, and performs nicely with the recurrent nature of sports injuries.

For the completion of this work, we have used combined multi-year data from five different seasons (from 17-18 to 21-22), for two female and three male senior teams from a Spanish professional football club. We have compared both sexes generating a different tree for each season. Finally, we have estimated survival curves for each terminal node of the trees, and we have ranked the relative importance of the covariates. Our method has been proven to be valuable for the identification of the potential risk factors involved in female and male non-contact lower limb injuries in football.

Keywords: sports injury, risk factor, multivariate survival tree