A zero-inflated Bayesian modeling of sports injury risk incidences

Oihane Álvarez[1], *Lore Zumeta-Olaskoaga*[1], *Joaquín Martínez-Minaya*[2], *Dae-Jin Lee*[3],

[1]{oalvarez,lzumeta}@bcamath.org, Applied Statistics Research Line, BCAM - Basque Center for Applied Mathematics, Bilbao, Bizkaia, Spain

[2]jmarmin@eio.upv.es, Department of Applied Statistics and Operations Research and Quality, Universitat Politècnica de València (UPV), Valencia, Spain

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

Sports injury risk incidence is a significant concern for athletes, coaches, and medical professionals in any sport or physical activity. Injuries can not only limit the ability to engage in sports, but also have long-term negative effects on health. Therefore, it is crucial to take preventive measures and establish risk management strategies to reduce the incidence of injuries. Hence, accurately modelling injury risk can aid in injury prevention and management strategies. One statistical method used for modeling count data is the Poisson regression model. However, in many cases, the count data may have an excess of zeros, which violates the assumptions of the Poisson regression model. The zero-inflated Poisson (ZIP) model is a commonly used approach to handle such data, accounting for both the overdispersion and excess zeros.

Bayesian methods provide a flexible framework for modeling complex data structures and incorporating prior knowledge. Moreover, Bayesian inference allows for uncertainty quantification, which we found particularly useful in the context of sports injury risk incidence. Recently, some studies have shown a relationship between maturity and growth timing and sports injuries. There exists some factors that may increase the risk of sports injuries in individuals who are still growing and developing. For instance, during periods of rapid growth, bones, muscles, and tendons may be more vulnerable to injury. This is because growth can occur at different rates, causing the bones and muscles to grow at different rates, which can lead to imbalances and increased stress on the body.

In this work, we illustrate a zero-inflated Poisson and Binomial likelihood model, where we allow for a linear predictor in both the zero-inflation and in the mean. We analyzed data from a study conducted on 110 football players from a professional youth academy team over two decades. Multiple predictive factors which allows exploring the possible influence of different variables on injury burden such as maturity or timing of injury were included in both linear predictors: i) the one corresponding to the zero-inflated part, and ii) the one corresponding to the count part. More robust and reliable results than frequentist ZIP models were obtained. Our results show that the proposed method can provide valuable insights for injury prevention strategies in sports.

Keywords: Zero-inflated Poisson, Bayesian inference, Sports injuries risks incidences.