This is the published version of a paper published in *SORT - Statistics and Operations Research Transactions*.

Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a simulation study

Martí Casals[1,2,3,4], Klaus Langohr[5], Josep Lluís Carrasco[1]
and Lars Rönnegård[6]

Abstract

Generalized linear mixed models are flexible tools for modeling non-normal data and are useful for accommodating overdispersion in Poisson regression models with random effects. Their main difficulty resides in the parameter estimation because there is no analytic solution for the maximization of the marginal likelihood. Many methods have been proposed for this purpose and many of them are implemented in software packages. The purpose of this study is to compare the performance of three different statistical principles – marginal likelihood, extended likelihood, Bayesian analysis – via simulation studies. Real data on contact wrestling are used for illustration.

1. Introduction

One of the methodologies used to study disease incidence in medicine or injuries in sport research is the generalized linear model (GLM). This methodology is able to model counts and proportions besides normally distributed variables (McCullagh and Nelder, 1989). Furthermore, GLMs assume that the observations conditioned on the predictors

[1] Department of Public Health, Universitat de Barcelona, Barcelona, Spain.

[2] Epidemiology Service, Public Health Agency of Barcelona, Barcelona, Spain.

[3] Area of Biostatistics, Universitat Internacional de Catalunya, Barcelona, Spain.

[4] CIBER de Epidemiología y Salud Pública (CIBERESP), Spain.

[5] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya/ BARCELONATECH, Barcelona, Spain.

[6] Statistics Unit, Dalarna University, Falun, Sweden.

are independent and identically distributed. However, these assumptions may be violated in some situations, such as longitudinal studies, where there are repeated measures and, hence, correlated data. Ignoring correlation of data when fitting the model may lead to biased estimates and misinterpretation of results (Bolker et al., 2009).

Generalized linear mixed models (GLMMs) are an extension of GLMs adding random effects in the linear predictor term in a regression setting (Breslow and Clayton, 1993). The GLMM is a more flexible analysis approach for analyzing non-normal data and it is known to be useful for accommodating the overdispersion in Binomial or Poisson regression models, and modelling the dependence structure among outcome variables for longitudinal or repeated measures data (Williams, 1982; Breslow, 1984).

The main difficulty of these models is the estimation of their parameters, as it is often not viable to obtain an analytic solution that allows maximizing the marginal likelihood of the data. Due to this fact, different estimation methods based on approximation or simulation have been developed in recent years. One approximation using numerical integration is the Gauss-Hermite quadrature (GHQ) (McCulloch and Searle, 2001). However, there are alternatives to the marginal likelihood principle including Bayesian statistics and the extended likelihood principle. For example, the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) is a Bayesian implementation and the hierarchical (h-)likelihood is an implementation of the extended likelihood principle (Lee and Nelder, 1996, 2001). It is worth mentioning that the comparison between Bayesian and non-Bayesian methods is difficult to perform given that they are different principles.

Nowadays, GLMMs are implemented in most statistical software packages and several researchers have published and updated different guides and reviews of different software packages for fitting a GLMM. West et al. (2014) introduce the fitting and interpretation of several types of linear mixed models using the statistical software packages SAS, SPSS, Stata, R, and HLM. Dean and Nielsen (2007) review the theoretical background of generalized linear mixed models and the inferential techniques that have been developed for SAS, S-Plus, and contributed R packages. Bolker et al. (2009) describe the use of generalized linear mixed models for ecology and evolution and give information on available functions and packages in SAS or R. For further comparisons of statistical software for GLMMs for binary responses and frailty models, see, for instance, Zhang et al. (2011); Li et al. (2011); Hirsch and Wienke (2012); Kim et al. (2013); or Grilli et al. (2014).

The aim of this work is to compare three different statistical principles – marginal likelihood, extended likelihood, and Bayesian analysis; see, Table 1 – to estimate the parameters of a Poisson Mixed Model in R using both real and simulated data. It is structured as follows: in Section 2, we briefly review the definition of the GLMM and highlight the problem of deriving and maximizing the likelihood. In Sections 3 and 4, we give a theoretical description according to the statistical principle used. Several contributed R packages for the fit of GLMMs are presented in Section 5 and three of them are used in Section 6 for the analysis of the motivating real data set on Leonese Wrestling. These data are then used to define the settings of the simulation study presented in Section 7.

In Sections 8 and 9, the results of the simulation are presented and discussed, and recommendations are given on which statistical principle is, preferably, to be used in each of the settings under study.

2. Generalized linear mixed models

The GLMM extends the GLM by adding normally distributed random effects to the linear predictor. As Bolker et al. (2009) point out, GLMMs combine the properties of linear mixed models (LMMs) and GLMs by using link functions and exponential family distributions such as Binomial or Poisson distributions.

Let $Y_i = (Y_{i1}, \ldots, Y_{im})'$ be a vector of m observations of the response variable of interest corresponding to subject i, $i = 1, \ldots, n$ and u_i, $i = 1, \ldots, n$, be the random effects vector of the same subject. Conditional on u_i, the distribution of Y_i is assumed to be from the exponential family type with density function $f(Y_i|u_i; \cdot)$ and with conditional mean $\mu_i = E(Y_i|u_i)$ and conditional variance $\text{Var}(Y_i|u_i) = \Phi V(\mu_i)$, where Φ is the dispersion parameter and $V(\mu_i)$ is the variance function of the GLMM.

The definition of the GLMM is completed by introducing a monotone and differentiable function $g(\cdot)$ known as the link function (McCullagh and Nelder, 1989) and a linear predictor η as follows:

$$\eta_i = g(\mu_i) = X_i\beta + Z_iu_i, \, i = 1, \ldots, n$$

where X_i (of dimension $m \times k$) and Z_i ($m \times l$) are subject i's design matrices associated with fixed and random effects, respectively. Vector β ($k \times 1$) is the fixed effects vector and u ($l \times 1$) is the random effects vector assumed to follow a multivariate Gaussian distribution with mean vector $\mathbf{0}$ and unknown positive definite covariance matrix Σ. Its density function is denoted by $f(u; \Sigma)$.

Estimation in the GLMM and theoretical description of the likelihood principle

By the local independence assumption, the conditional density of Y given u has the form

$$f(Y|u; \beta) = \prod_{i=1}^{n} f(Y_i|u_i; \beta)$$

and the multivariate density function of u is given by

$$f(u; \Sigma) = \prod_{i=1}^{n} f(u_i; \Sigma).$$

The likelihood principle involves two kinds of objects: observed random variables (the data) and (unknown) fixed parameters. In the case of models with random effects, the estimation is based on the marginal likelihood where the random effects are integrated out (Birnbaum, 1962; Pawitan, 2001). Hence, the following likelihood function needs to be maximized in order to obtain the maximum likelihood (ML) estimates for $\boldsymbol{\beta}$ and the variance components in $\boldsymbol{\Sigma}$:

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = f(\boldsymbol{Y}; \boldsymbol{\beta}) = \int f(\boldsymbol{Y}|\boldsymbol{u}; \boldsymbol{\beta}) f(\boldsymbol{u}; \boldsymbol{\Sigma}) \, d\boldsymbol{u}. \tag{1}$$

The classical method that uses ML estimation and in which $\boldsymbol{u}$ is integrated out does not present problems with linear mixed models. The problem exists with GLMMs because of the more complicated integral (McCulloch and Searle, 2001). For this reason, one of the main interests of the research on the GLMM is to develop more efficient estimation methods for the fixed effects vector and the variance components.

Several ways to solve the integration in (1) and to obtain the marginal likelihood to estimate the parameters of a GLMM have been proposed. The Laplace method for integral approximation is considered to be a possible solution, which can be used to estimate the parameters of interests (Breslow and Clayton, 1993). Alternatives are the GHQ method or pseudo and penalized quasilikelihood methods (Aitkin, 1996). The GHQ method presents better estimation properties than the other methods because the GHQ estimates are maximum likelihood estimates. However, it is not feasible for analyses with more than two or three random effects because the speed of the GHQ decreases rapidly when increasing the number of random effects (Bolker et al., 2009).

3. Theoretical description of the extended likelihood principle

Lee and Nelder (1996, 2001) extended generalized linear models to include random effects by using their hierarchical (h-)likelihood method. This method is based on the extended likelihood principle (Bjørnstad, 1996) and is an implementation of the extended likelihood restricted by a weak canonical link for the random effects (Lee et al., 2006).

The h-likelihood is given by the log joint likelihood, that is, the extended likelihood L_E:

$$h = \log\left(L_E(y; \beta, v)\right) = \log\left(f(y; \beta|v)\right) + \log\left(f(v)\right)$$

where $\log(f(y; \beta|v)$ denotes the log of the density function with β as parameter and conditional on $v = v(\boldsymbol{u})$, where $\boldsymbol{u}$ is a vector of random effects and $\nu(\cdot)$ is an appropriate link function defining the h-likelihood. Unlike the GLMMs, the random effect is not restricted to be normal and can follow other distributions (e.g., gamma, beta, or inverse gamma).

A fundamental difference compared to classical marginal likelihood theory is that estimation and inference based on the h-likelihood includes random effects, whereas in classical likelihood theory the random effects are integrated out and a marginal likelihood is used. Hence, the use of the h-likelihood avoids the integration required for a classical marginal likelihood.

To estimate parameters (β, ν), the fixed and random effects are estimated from the score functions of the h-likelihood:

$$\frac{\partial h}{\partial \beta} = 0, \quad \frac{\partial h}{\partial \nu} = 0.$$

The variance components are estimated by maximizing the adjusted profile h-likelihood defined as

$$p_{\beta,u} = \left(h + \frac{1}{2} \log(2\pi \boldsymbol{H}^{-1}) \right) \big|_{\beta = \hat{\beta}, \boldsymbol{u} = \hat{\boldsymbol{u}}},$$

where $\boldsymbol{H}$ is a Hessian matrix of the h-likelihood.

The estimates can be obtained by using iterative weighted least squares (IRWLS) as implemented in the `hglm` package (Rönnegård et al., 2010). The variance components are then estimated iteratively by applying a gamma GLM to the estimated deviances and with an intercept term included in the linear predictor and appropriate weights (Lee et al., 2006).

4. Theoretical description of the Bayesian principle

The Bayesian methods differ from the likelihood and extended likelihood principles in their philosophy as well as in the specific procedures used.

In order to implement a Bayesian principle, prior distributions are required for all parameters in the model, since under the Bayesian paradigm, all parameters are treated as random variables rather than fixed unknowns.

The Bayesian principle is based on assigning prior distributions to the parameters of the model. Thus, following the model defined in Section 2, the following prior distributions must be specified: $f(\boldsymbol{\beta}|\cdot)$, $f(\boldsymbol{u}|\cdot)$, and $f(\boldsymbol{\Sigma}|\cdot)$. These prior distributions express the beliefs on the parameters and these beliefs are modified by the data to obtain the posterior distribution of the parameters, $f(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\Sigma}|\boldsymbol{Y})$, which is defined as to be proportional to the product of the prior distributions and the likelihood of the data. The posterior distribution is therefore used for inference purposes.

Here, a non-informative normal distribution is assumed as prior distribution for $\boldsymbol{\beta}$, that is, a normal distribution with a huge variance. Let $\boldsymbol{\gamma} = (\boldsymbol{u}, \boldsymbol{\beta})^{\mathsf{T}}$ denote the $G \times 1$ vector of Gaussian parameters. Concerning the random effects, we assume $\boldsymbol{u}$ to follow a multivariate normal distribution, $\boldsymbol{u}|\boldsymbol{\Gamma} \sim \mathcal{N}(0, \boldsymbol{\Gamma}^{-1})$, where the precision matrix $\boldsymbol{\Gamma} =$

$\mathbf{\Gamma}(\boldsymbol{\phi})$ depends on parameters $\boldsymbol{\phi}$. Let $\boldsymbol{\phi}$ also be the vector of the variance components for which the prior $\Pi(\boldsymbol{\phi})$ is assigned. However, often it is not possible to obtain an explicit expression of the posterior distribution and algorithms such as Markov chain Monte Carlo (MCMC) methods are used to generate the posterior distribution by simulation.

The Bayesian principle is attractive because it offers several advantages over the likelihood principle (e.g., it can increase the stability in small samples or in clustered binary data), but it has the difficulty of specifying prior distributions with variance components (Fong et al., 2010).

The use of MCMC methods for GLMMs is the most popular approach, but has problems in terms of convergence and computational time. These problems with Bayesian estimation have been greatly improved by Integrated Nested Laplace Approximations (Rue et al., 2009).

Integrated Nested Laplace Approximation (INLA)

INLA is a new tool for Bayesian inference based on latent Gaussian models introduced by Rue et al. (2009). The method combines Laplace approximations and numerical integration in a very efficient manner. For the GLMM described in Section 2 and using $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ as defined in the previous paragraphs, the posterior density is given by

$$\pi(\boldsymbol{\gamma}, \boldsymbol{\phi}|\boldsymbol{Y}) \propto \pi(\boldsymbol{\gamma}|\boldsymbol{\phi})\pi(\boldsymbol{\phi})\prod_{i=1}^{m} p(\boldsymbol{Y}_i|\boldsymbol{\gamma}, \boldsymbol{\phi}).$$

It is computed via numerical integration as

$$\pi(\boldsymbol{\gamma}|\boldsymbol{Y}) = \int \pi(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{Y})\pi(\boldsymbol{\phi}|\boldsymbol{Y})d\boldsymbol{\phi},$$

where Laplace approximation is applied to carry out the integrations required for the evaluation of $\pi(\boldsymbol{\gamma}|\boldsymbol{\phi}, \boldsymbol{Y})$. For more details we refer the readers to Rue et al. (2009).

5. Contributed R packages for GLMMs

For the likelihood principle, there exist different packages in R such as `glmmML` (Broström and Holmberg, 2013), `lme4` (Bates et al., 2015), or the function `glmmPQL` in the `MASS` package (Venables and Ripley, 2002). One of the most popular and stable functions for fitting GLMMs is called `glmer` and is found within the package `lme4`. This package implements the GHQ to approximate the log-likelihood using numerical integration. By default, it uses the Laplace approximation with one quadrature point.

For the extended likelihood principle, two packages are available for fitting Hierarchical Generalized Linear Models with random effects: `hglm` (Rönnegård et al., 2010) and `HGLMMM` (Molas and Lesaffre, 2011).

Concerning packages for performing Bayesian inference on GLMM, general packages such as `glmmBUGS` (Brown and Zhou, 2010) and `R2WinBugs` (Sturtz et al., 2005), and specialized packages such as `glmmAK` (Komárek and Lesaffre, 2008), `MCMCglmm` (Hadfield, 2010), and `INLA` (Lindgren and Rue, 2015) exist. INLA substitutes MCMC simulations with accuracy and the quality of such approximations is extremely high.

For both the analysis of the wrestling data (Section 6) and the simulation (Section 7), the R packages `lme4`, `hglm`, and `INLA` were used. Note that the `lme4` package does not report standard errors for variance components. The reason is provided by the developer of the package in his book (Bates, 2010) stating that the sampling distribution of the variance is highly skewed, which makes the standard error nonsensical (Li et al., 2011). Regarding the GHQ method, we used 5 quadrature points since it was indicated that this method can give a poor approximation to the integrated likelihood when the number of quadrature points is low (Lesaffre and Spiessens, 2001). This method can be made arbitrarily accurate by increasing the number of quadrature points. We adopted a strategy of increasing the number of quadrature points until there was a negligible difference in the values of the estimators (Ormerod and Wand, 2012).

Recent changes of the packages used

The three packages studied in this paper have implemented some new features in their latest versions.

Regarding the `lme4` package, the authors of the package have been discussing new features and new versions of the package through the forum "R-sig-mixed models" (`https://stat.ethz.ch/mailman/listinfo/r-sig-mixed-models`) since its version 0.99 (better known and used as lme4.0). Currently, the present version is 1.1-7, which offers some advantages with respect to the version 0.99, especially in terms of convergence and optimization. These developments can be found at the following link: `https://github.com/lme4/lme4`.

Concerning the `hglm` package, since version 2.0 it is possible to fit several random effects from different distributions (e.g., gamma or Gaussian), to fit a linear predictor for the dispersion of the random effects, to fit spatial conditional autoregressive (CAR) and spatial autoregressive (SAR) models for the random effects, and to perform a likelihood-ratio test for the dispersion parameter of the random effects (Alam et al., 2014). The method options have also been extended to include the EQL1 method which is a "HL(1,1) correction" (Lee and Lee, 2012; Noh and Lee, 2007) applied on the default EQL method. These developments can be found at the following link: `http://cran.r-project.org/web/packages/hglm/vignettes/hglm.pdf`.

Regarding the INLA package, this analytical framework includes normally distributed latent variables and thus allows for hierarchical data structure, but it is targeted towards complex applications involving temporal and spatial smoothing where MCMC estimation may be too difficult to apply (requiring specialized MCMC samplers) or prohibitively slow. On the one hand, now there is an increase of articles using this package in several application fields such as fishering (Cosandey-Godin et al., 2014) or ecology (Quiroz et al., 2014). Few applications of the INLA methodology in injury epidemiology or sport science have been published (Cervone et al., 2014), and it would be of interest to study in more detail the medical impact in terms of understanding sport injuries with this methodology. On the other hand, one of the most recent papers of Martins et al. (2014) shows a new perspective on the selection of default priors.

6. Real data example: folk wrestling data

Leonese Wrestling (LW) or Aluche is a traditional and popular sport of the province of León, in Northwestern Spain. It is registered with and recognized by three international associations: *Fédération Internationale des Luttes Associées* (FILA), *Asociacion Española de luchas tradicionales* (AELT), and International Belt Wrestling Association (IBWA), respectively. Like with all styles of wrestling, the risk of injury is always present.

The main variable of interest in epidemiological investigation of sports injuries is the incidence of injury, which is generally expressed as the proportion of injuries per fight (Ayán et al., 2010; Hägglund et al., 2010). There are few studies in the international scientific literature about the incidence of injury in combat sports and its associated factors (Klügl et al., 2010; Hewett et al., 2005). Nonetheless, in published papers, it has been found that the incidence of injury in these sports is higher than in other sports activities (Hägglund et al., 2005; Junge et al., 2009).

Concerning factors associated with the incidence of injury, it is known that this incidence is higher during wrestling matches than in training. However, there is not much information on the frequency of injuries, their incidence, and their causes to carry out prevention and control programs in this sport. This lack of knowledge has motivated this analysis of the impact and risk factors of injuries.

Data on matches and injuries of the LW summer seasons were available for 213 wrestlers during the summer seasons from 2005 through 2010. The response variable of interest was the frequency of injuries which was assumed to follow a Poisson distribution. The study design was unbalanced with different numbers of repeated measures given that not all wrestlers participated in official competitions in each of the six years from 2005 to 2010. The possible risk factors for injuries considered were: i) Winner: This variable is defined as a function of the falls during a match. It is set to 'Yes' if the wrestler had more falls in his favor than against him; otherwise, the value of Winner is set to 'No'; and ii) Weight category: a categorical variable with levels Light, Medium

(chosen as reference), Semi-heavy, and Heavy. It has been taken into account that these variables could change from one season to another.

6.1. The model under study

Let Y_i, $i = 1, \ldots, n$, be the vector of the number of injuries per season of wrestler i. The length of Y_i depends on the number of seasons the wrestler took part in official competitions. It is assumed that the distribution of Y_i follows a Poisson distribution: $Y_i \sim Po(\mu_i)$. Usually, the counts are considered in relation to some differential or offset (λ) in order to obtain rates. Here, the offset is the number of the wrestler's matches per season. The Poisson generalized linear mixed model used to analyze the data links the mean of Y_i with both covariates X_i of interest —Winner and Weight category— by means of the following equation

$$\log(\mu_i) = \log(\lambda_i) + X_i\beta + u_i, \, i = 1, \ldots, n, \tag{2}$$

where the vector β contains the fixed effects parameters and u_i stands for the random effect intercept for wrestler i. Random effects are assumed to be independent and normally distributed with mean $\mathbf{0}$ and variance σ^2. Random effects for the slope parameters were not considered in order to keep the model simpler. In addition, posterior model fits including such random effects did not improve the model fit significantly at a 0.05 significance level.

The model's marginal variance (over subjects) can be expressed as

$$\text{Var}(Y_i) = \Phi \cdot \mu_i, \, i = 1, \ldots, n,$$

where Φ is the dispersion parameter. The Poisson distribution assumes $\Phi = 1$ and if $\Phi > 1$, overdispersion is present. In that case, the data have larger variance than expected under the assumption of a Poisson distribution.

The dispersion parameter can be estimated based on the χ^2 approximation of the residual deviance or Pearson residuals. The dispersion parameter is estimated by dividing the χ^2 statistic by the residual degrees of freedom, $n - r$:

$$\hat{\Phi} = \frac{1}{n-r} \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}. \tag{3}$$

If there is high overdispersion, the negative binomial distribution is an alternative to the Poisson distribution for count data because the negative binomial distribution allows for a variance greater than mean. Apparent overdispersion is normally due to missing covariates or interactions, outliers in the response variable, non-linear effects of covariates entered as linear terms in the systematic part of the model, or the choice of

a wrong link function. These are mainly model misspecifications. Real overdispersion exists when none of the previous causes can be identified. The reason for this might be that the variation in the data is, actually, larger than the mean. Or, there may be many zero observations, clustering of observations, or correlation between observations (Hardin and Hilbe, 2007; Zuur et al., 2009).

6.2. Results

We fitted Model (2) with packages `lme4`, `hglm`, and `INLA`, and for the sake of comparison, we also analysed the data as if they were not correlated. That is, we fitted a GLM using function `glm` of the R package `stats`. All analyses were carried out with R, version 3.1.1, and the estimates obtained, together with the 95% confidence and (in the case of the `INLA` package) credible intervals, are presented in Table 2. The confidence intervals of the `lme4` package were calculated using the function `confint.merMod`, which computes a likelihood profile and finds the appropriate cutoffs based on the likelihood ratio test (by choosing `method = "profile"`).

We calculated the dispersion term based on the Pearson residuals using Equation (3) for function `glm` and packages `lme4` and `hglm`. In addition, we checked the possible overdispersion using an individual-level random variable (translating to a lognormal-Poisson model, which is qualitatively similar to a negative binomial (Elston et al., 2001)).

According to the results obtained, the estimates of the coefficients of the linear predictor are quite similar with only slight differences in the second decimal digit. The same is true for the confidence and credible intervals obtained with packages `lme4`, `hglm`, and `INLA`, whereas function `glm` provides smaller confidence intervals because it treats the data as if they were independent observations. Moreover, it can be seen that the estimate of the random effect variance and the dispersion term differ a little from each other.

Concerning both variables of interest, the positive signs of the parameter estimates corresponding to the weight category indicate a higher risk for injuries among all three weight categories as compared with the medium weight category. Nonetheless, all 95% confidence and credible intervals include 0, that is, the differences are not statistically significant at a 0.05 significance level. In the case of variable Winner, the model indicates a lower injury risk $-\hat{\beta} < 0$ and the 95% confidence and credible intervals do not include 0 – among wrestlers with more falls in their favor.

7. Simulation study

In this section, we present the simulation study designed to assess the performance of the three statistical principles using different scenarios based on the wrestling data. The aim is to compare the three methods in different settings defined by overdispersion or sample size with respect to measures of model accuracy, precision, empirical bias, and

empirical coverage of the estimators as well as computation time and possible problems of convergence. A total of 40 different simulation settings were used that can be classified into two main simulation scenarios.

7.1. Simulation scenario 1

For the first simulation scenario, we used the structure of the real data set introduced in Section 6. The values of the response variable – number of injuries – were generated as a function of the observed values of the independent variables and the number of matches of each wrestler in each of the years under study following the model expression in (2). The aim of this scenario was to closely represent the structure of this real data set.

We simulated the number of injuries using the parameter values given in Table 2 obtained with the lme4 package ($\beta_1 = 0.24, \beta_2 = 0.1, \beta_3 = 0.4, \beta_4 = -0.46$). Concerning the model intercept (β_0) and the variance of the random effects (σ_u^2), we used combinations of both parameters that lead to three different values of overdispersion ($\Phi \in \{1.5, 3, 10\}$) that we identify as low, moderate, and high overdispersion settings; see McCulloch and Searle (2001) for technical details. Furthermore, we added the value of $\Phi = 1$, i.e., no overdispersion, to also assess how the GLMM behaves in this situation. In addition, the values of β_0 were chosen such that two different marginal means of the injury numbers were obtained ($\mu = 1$ and $\mu = 10$). As can be seen in the R code of the first simulation study in the Supplemental Material, the values of β_0 ranged from -4.8 to -1.7, those of σ_u^2 from 0 to $2.1^2 = 4.41$.

In total, with four different values of dispersion and two of the marginal mean, the number of simulation settings for the Simulation scenario 1 was $4 \cdot 2 = 8$.

7.2. Simulation scenario 2

The second simulation scenario was motivated by the goal to study the effect of different sample sizes on the parameter estimation. For this purpose, we considered two different sample sizes of $n = 30$ and $n = 100$ wrestlers and for each wrestler, a random number of seasons was generated using a discrete uniform distribution ranging from 1 through 6. In the sequel, the match numbers for each wrestler and season were generated using a Poisson distribution with parameters 60 and 100, respectively. These were the offset terms λ_i in Model (2). The remaining parameters were chosen similar to the first simulation scenario resulting in four dispersion parameters ($\Phi \in \{1, 1.5, 3, 10\}$) and two marginal means for the number of injuries ($\mu = 1$ and $\mu = 10$) so that the number of simulation settings was $2 \cdot 2 \cdot 4 \cdot 2 = 32$ for Simulation scenario 2.

The values of both independent variables – weight category and winner – were generated using equal probabilities for all categories and, as in Simulation scenario 1, the injury numbers were generated using the expression of the model in (2).

7.3. Evaluation criteria

For each of the $8 + 32 = 40$ simulation settings, we simulated 1000 data sets of the model under study and used the three methods to estimate the model parameters (Marginal Likelihood, Extended Likelihood, and Bayesian Analysis). In addition, we used R function `glm` treating the data as if we dealt with a GLM. Measures for the comparison of the different estimation methods were the empirical mean squared error (MSE) as a measure of model accuracy, the ratio of precision, the empirical bias, and the empirical coverage of the confidence and credible intervals, respectively. Moreover, we recorded the computation times and studied possible problems of convergence.

For each simulation setting and estimation method, the empirical bias was calculated as the mean bias over the 1000 data sets and its squared value was used together with the empirical variance to compute the empirical MSE. The rate of precision was computed as the ratio between the estimator's empirical variance and the mean of the squared standard errors. In order to calculate and compare the empirical bias and the empirical MSE in the case of the `INLA` package, the distribution of the parameters provided by `INLA` were reduced to only one value (the posterior mean).

For the likelihood and the extended likelihood principles, we used the 95% confidence interval and for the Bayesian principle, we used the 95% credible interval of parameter given by the 0.025 and 0.975 sample quantiles of the posterior parameter distributions. In the case of the `lme4` package, we used 5 quadrature points for the GHQ method and non-informative priors were assumed for the Bayesian analysis. Moreover, the random intercepts were assumed to have a normal distribution. Regarding the prior distribution for the precision, a half-normal distribution with mean 0 and precision 0.0001 was assigned to the standard deviations (Gelman et al., 2006).

The comparison was done for the two main parameters of interest: the parameter β_4, which corresponds to the variable Winner, and the variance of the random effects (σ_u^2). The former was chosen, since the analysis of the wrestling data showed a statistically significant association between the number of injuries and this variable. Whereas the value of σ_u^2 varied across the simulation settings, the value of β_4 was kept constant in all settings.

8. Results of the simulation study

For the simulation scenarios, results are presented only for the intercept (β_0), the slope (parameter β_4, corresponding to the covariate Winner), and variance of the random effect (σ_u^2). The performance of the estimation methods in terms of empirical bias, empirical MSE, precision ratio, and empirical coverage of $\hat{\beta}_4$ and $\hat{\sigma}_u^2$ is summarized in Figures 1a, 1b, 2a, and 2b. The corresponding figures for $\hat{\beta}_0$ are provided in the Supplemental Material.

In order not to mix up statistical principles, methods, and algorithms as presented in Table 1, following, we only refer to these by the names of the R packages used.

Note that the results are yielded by package INLA with the prior distribution selected, a half-normal distribution with mean 0 and precision 0.0001.

Concerning the percentage of convergence of the estimation methods, a model was considered as "not convergent" if either the estimation process did not converge or if the estimate or its standard error was not provided. For example, in some cases the parameter can be estimated but the estimation process may be unable to provide a positive definite variance-covariance matrix of the parameters (for problems with the Hessian matrix), mainly due to the instability of the model. Convergence was checked and obtained using the criteria offered in each software package. In the case of the first simulation scenario, the convergence percentages were always equal to 100% with only one exception: package hglm achieved convergence in 99.2% of all data sets in the case of $\mu = 1$ and $\Phi = 10$. The results for Simulation scenario 2 are shown in Table 3. The rate of convergence of all estimation methods was close to 100% for most of the settings. However, in the case of packages hglm and lme4, the percentage of convergence slightly decreased in some settings with $\mu = 1$ and $n = 30$ as overdispersion increased.

Regarding the empirical bias of the slope, all packages provided mostly unbiased and similar estimates. In terms of accuracy, the highest empirical MSE for the slope for all GLMM packages is given when $\Phi = 10$, $\mu = 1$, and $n = 30$. In this case, function glm is the one that presents highest values. The empirical MSE value obtained with INLA is higher than with hglm and lme4, which are both similar; see Figure 1a.

In terms of precision (upper panel of Figure 1b), we calculated the ratio of the estimator's empirical variance and the mean of the squared standard errors as a precision measure. In general, we found that almost all estimation methods presented an underestimation in the case of $\Phi = 1$ and $\Phi = 1.5$ together with $\mu = 10$. More differences between the packages were observed with sample size equal to 30. In the case of the lme4 package, the ratio was slightly larger than 1 (equivalent to 100%) especially with moderate and high overdispersion. By contrast, the values of the hglm and INLA packages were close to 100% for that sample size independently of the offset, the marginal mean, and the dispersion term. The function glm, in general, showed values far larger than 100% (and, hence, out of the range of the corresponding plots) from $\Phi = 1.5$.

The empirical coverage of the confidence intervals for the GLMM packages were close to 95% in all settings; see the lower panel of Figure 1b. The GLM appeared to have bad coverage, only acceptable for those combinations with low overdispersion. It suffered from substantial undercoverage (down to 75%) when $\Phi = 3$ and $\Phi = 10$. This result may be expected for the GLM since it does not include random effects and therefore can not assume any overdispersion. As overdispersion increased, the empirical coverage behavior became worse.

Regarding the empirical bias of the variance component (upper panel of Figure 2a), the three packages performed similarly for $\Phi = 1$ and $\Phi = 1.5$. For $\Phi = 3$, the INLA package showed the largest empirical bias with $n = 30$, whereas with $n = 100$, the differences among the packages were small for $\Phi = 3$. For $\Phi = 10$ and $n = 30$, package lme4 had the smallest empirical bias in terms of the absolute value, INLA the largest
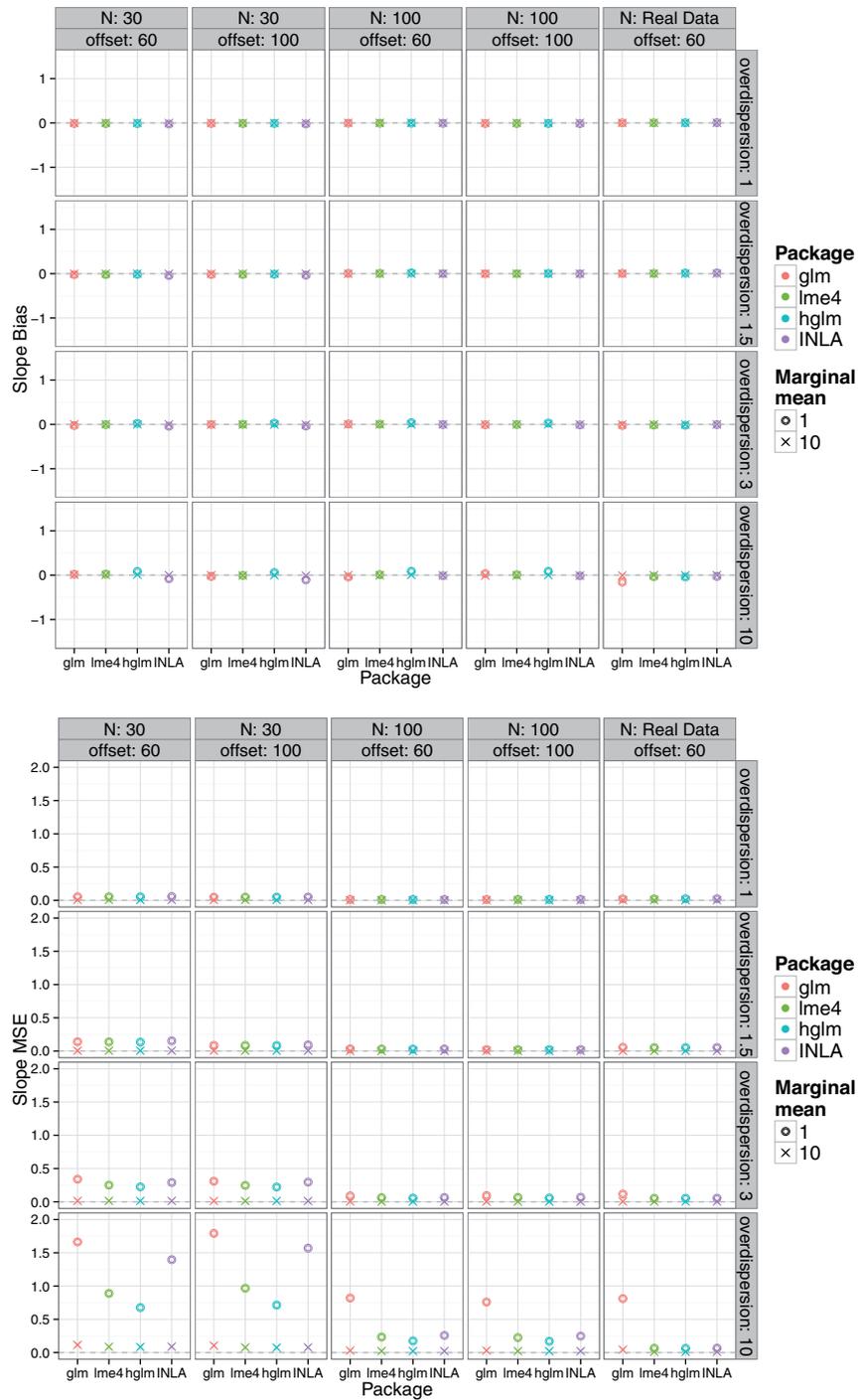
Figure 1a: *Empirical bias (upper panel) and empirical MSE of the slope estimate ($\hat{\beta}_4$) as a function of overdispersion, marginal mean, offset, and sample size.*
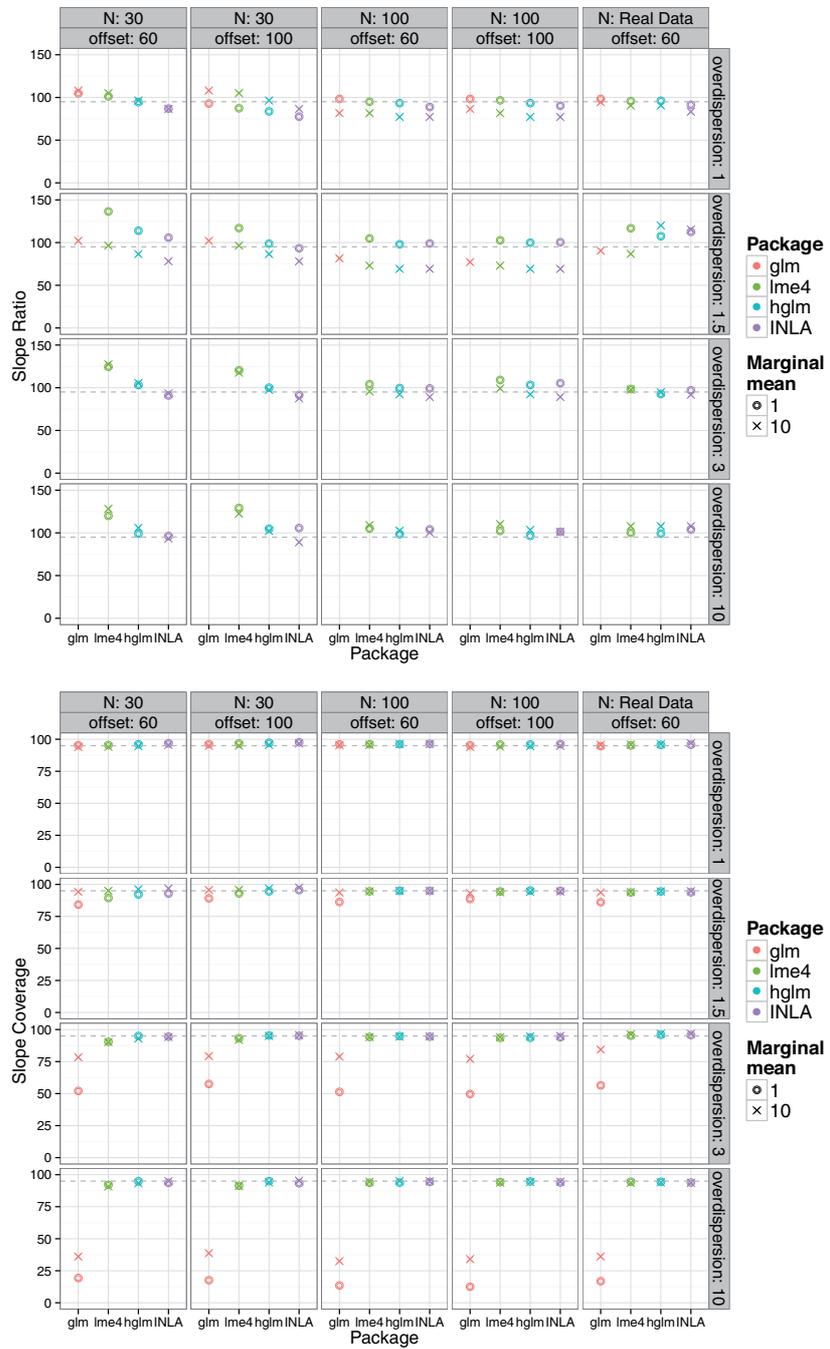
Figure 1b: *Precision (upper panel) and empirical coverage of the slope estimate ($\hat{\beta}_4$) as a function of overdispersion, marginal mean, offset, and sample size. Precision is measured as the ratio of the estimator's empirical variance divided by the average of the squared standard errors.ope estimate ($\hat{\beta}_4$) as a function of overdispersion, marginal mean, offset, and sample size.*
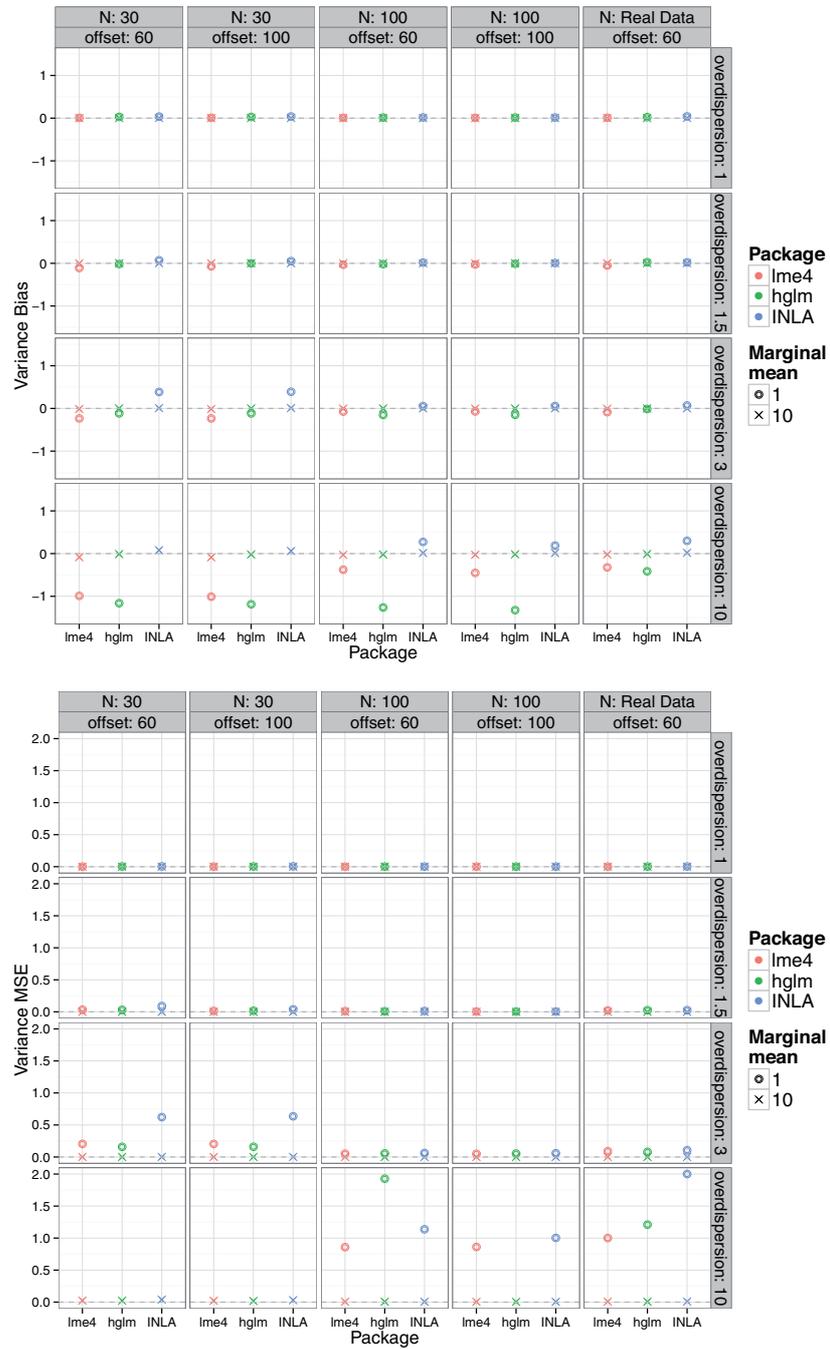
Figure 2a: *Empirical bias (upper panel) and empirical MSE of the variance component estimate ($\hat{\sigma}_u^2$) as a function of overdispersion, marginal mean, offset, and sample size.*
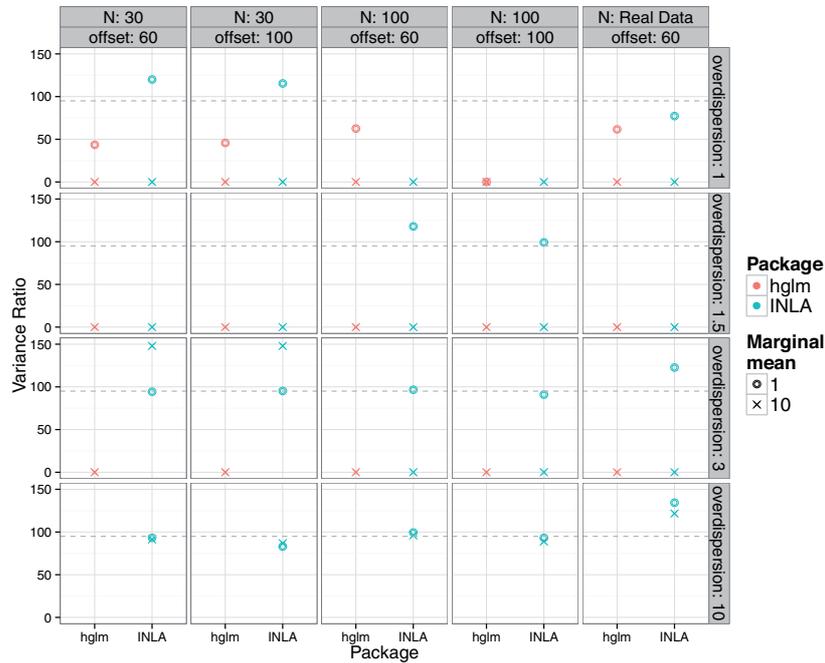
Figure 2b: *Precision of the variance component estimate* $((\hat{\sigma}_u^2))$ *as a function of overdispersion, marginal mean, offset, and sample size. Precision is measured as the ratio of the estimator's empirical variance divided by the average of the squared standard errors.*

(with values out of the range of the plot). By contrast, for $\Phi = 10$ and $n = 100$ the absolute values of the empirical bias of `lme4` and `INLA` were roughly the same. It was largest for `hglm` in that setting. Values for function `glm` do not appear in that figure since a GLM does not consider any random effects.

In terms of the accuracy of the variance component (lower panel of Figure 2a), the empirical MSEs were very similar except when $\Phi > 1.5$ and $\mu = 1$. The `INLA` package had the largest empirical MSE when $\Phi = 3$, $\mu = 1$, and $n = 30$. By contrast, in this same setting, the `lme4` and `hglm` packages had very similar values. None of the packages showed a satisfactory behavior when $\Phi = 10$, $\mu = 1$, and $n = 30$: all empirical MSEs are excessively high and, hence, out of the range of the corresponding plot. With $n = 100$, the empirical MSE of package `hglm` was still out of the plot's range, whereas that of `lme4` and `INLA` were close to 1.

Concerning the precision of the estimation of the variance component (Figure 2b), the ratio obtained with `hglm` and `INLA` was close to zero when $\Phi \neq 10$ and $\mu = 10$. For $n = 30$, in `INLA`, the values were close to 150 when $\mu = 10$, and they were close to 100 when $\Phi > 1.5$ and $\mu = 1$. For `hglm` and when $\Phi > 1$ and $\mu = 1$, the values were excessively high in most of the simulation settings. It could not be computed with `lme4` since this package does not provide the standard error of the estimation of the random effect's variance.

Contrary to overdispersion, sample size, and marginal mean, the choice of the offset did not seem to have any effect on the estimators' performance.

Finally, we also compared the computational times measured with R function `system.time`. Among the three packages that consider random effects, the average computing times of packages `lme4` and `hglm` were very similar in each setting. On average, they were four times faster than package `INLA`.

In summary, the approaches involving a random effect (`lme4`, `INLA`, and `hglm`) showed good performance on estimating the model parameters except for the estimation of the random effect's variance in the case of combinations of huge overdispersion, a small marginal mean, and a small sample size. Given that the empirical MSE and the empirical bias of the `lme4` package are close to zero for most of the simulation settings, it seems that this package, generally, outperforms the other packages, even though often only slightly.

9. Discussion

An overview of statistical principles for GLMMs is presented in this paper. The problem of selecting the best approach for estimation and inference within Poisson Mixed Models is very complex and too difficult to solve analytically. For this reason, we have carried out a simulation study that has evaluated the impact of overdispersion, marginal mean, offset, and sample size assuming Poisson Mixed Models using different statistical principles.

The fact that the bias and mean square error improve with larger sample size can be interpreted as that the estimators are consistent when overdispersion is taken into account in the model by means of random effects. By contrast, the empirical bias and MSE were larger with the GLM since this type of model does not include random effects and, hence, ignores overdispersion (Bolker et al., 2009; Milanzi et al., 2012). The results for `INLA`, in general, did hardly differ with the values obtained with both `lme4` and `hglm`; at least not with the prior distribution used.

We have found that for small sample sizes, the random effects variances are difficult to estimate, which has also been described in other studies (Li et al., 2011). In addition, the results are worse in the case of a moderate dispersion term ($\Phi = 3$) and especially with high overdispersion ($\Phi = 10$), in which none of the methods provide satisfactory results. As Zuur et al. (2009) explain, with a dispersion term up to 1.5 there are no important overdispersion problems. To solve overdispersion problems, other methods and distributions may be used, for instance, the Poisson-lognormal distribution, GEE models, or quasi-Poisson distributions (Booth et al., 2003; Bolker et al., 2009). In the case of high overdispersion with real data, it would be more reasonable to change the Poisson distribution for another distribution that does not have the restriction of the variance equalling the mean, for instance, the Negative Binomial distribution (Czado et al., 2007). Recently, Aregay et al. (2013) suggested a Hierarchical Poisson-Normal overdispersed

model (HPNOD) as an alternative using the Bayesian principle. The HPNOD performs better than a Hierarchical Poisson-Normal model for data with low, moderate, and high overdispersion.

In the simulation study, for most combinations, we observed similar performance in terms of the empirical bias and the empirical MSE whatever was the package applied. On the other hand, differences arise with respect to the precision of estimates. This fact may indicate a problem of underestimation because the methods do not capture well all the variability present in data. A bootstrap approach may be a solution to solve the standard error problem.

Regarding computational time and convergence, the `glm` function requires less time because it does not capture the presence of the random effect. The `hglm` and `lme4` packages need similar computational times, whereas the `INLA` package takes more time. Nonetheless, at least for the simulation settings under study, computation time was not excessive and results were obtained in less than 5 seconds in most of the cases. In some combinations of the simulation with small sample size ($n = 30$), huge overdispersion ($\Phi = 10$), and small marginal mean ($\mu = 1$), we found problems of convergence in the `hglm` package. To solve a convergence problem we recommend specifying other starting values.

Several studies carried out until now have compared estimation methods only for the Bayesian principle, the marginal likelihood principle or both in the GLMMs (Zhang et al., 2011; Ormerod and Wand, 2012; Li et al., 2011; Kim et al., 2013). For example, Li et al. (2011) recommend the use of the `lme4` package. The authors highlight that in case a Bayesian package is chosen, the parameter estimates might be influenced by the priors for the variances of the random effects. They also mention that when the data set is small, the random effects' variances are difficult to estimate with both frequentist and Bayesian methods. Kim et al. (2013) recommend the use of the GHQ method given that it performs well in terms of accuracy, precision, convergence rates, and computing speed. According to the authors, this is also valid with small sample sizes and for longitudinal studies with a few time points. On the other hand, there are some studies that compare the extended likelihood approach with the Bayesian principle (Pan and Thompson, 2007; Collins, 2008). However, they do not take into account the `INLA` method, which is a recently proposed approximate Bayesian approach for latent Gaussian models. According to Collins (2008), in some case studies both Bayesian and extended likelihood approach estimators of the variance component were positively biased, whereas GHQ method based estimators were not.

Our work is different from these previous studies given that we have focused our work on different estimation methods, principles, and more commonly used free software packages. Regarding the real data example, most GLMMs are used in applications of ecology, epidemiology, genetics, clinical medicine, and other applications but these models are now gaining attention in sports sciences, too (Avalos et al., 2003; Bullock and Hopkins, 2009; Sampaio et al., 2010; Casals and Martínez, 2013; Casals et al., 2014). However, there are only a few studies in epidemiology of sports injuries.

Concerning the Bayesian principle, when the sample size is small, the posterior distribution may be more influenced by the choice of prior distributions than when the sample size is moderate or large. Note that the posterior means depend on the choice of the non-informative prior of the variance component (Li et al., 2011). Bayesian algorithms such as MCMC offer different advantages over frequentist algorithms but they have problems in terms of convergence and computational time. These aspects have improved with INLA. However, for some combinations, the default initial value is not adequate for the data because of the use of very weak priors. For example, it is known that the inverse Gamma(0.001, 0.001) prior is not always a good choice (Fong et al., 2010). Frühwirth-Schnatter and Wagner (2010b) and Frühwirth-Schnatter and Wagner (2010a) demonstrate overfitting due to Gamma-priors and suggest using a (half) Gaussian prior for the standard deviation to overcome this problem, as suggested by Gelman et al. (2006). Another interesting possibility that could be considered in the future is the use of Penalized Complexity (PC) priors that have heavier tails than the half Gaussian, but lighter tails than the half-Cauchy (Martins et al., 2014). According to the developers of the INLA package, the use of PC priors works pretty much identically to the half-Cauchy in practice.

There are some limitations of the present work. First, in this study we have worked with a Poisson mixed model with a random intercept, but we have not considered models with random slopes. The reasons for this decision were twofold: On one hand, the inclusion of random slopes in the real data example did not significantly improve the model fit at a 0.05 significance level. On the other hand, to study the impact of overdispersion, marginal mean, sample size, and offset, we preferred to analyze a Poisson mixed model with a random intercept due to its frequency in sports medicine research. Given the importance of more complex mixed models, future research should investigate the performances of these principles in mixed Poisson models including random slopes, cross-classified random effects and multiple membership structures that may be analyzed in future simulation studies. Second, we only examined three packages corresponding to the three estimation methods and principles. There are other R packages such as glmmML and MCMCglmm as well as the function glmmPQL of the MASS package that could be included in further simulation studies. In addition, such simulation studies could include other software packages such as SAS, STATA, or SPSS, which were not considered for the present work. We decided to focus our work on R because of its great popularity among statisticians (Muenchen, 2015) and the constant development of new packages and functions to deal with GLMMs.

In addition to the two limitations mentioned, the simulation study could have also studied other parameters of interest such as the cluster size, which, in the real data example in Section 6, is equivalent to the number of seasons of a wrestler. Indeed, several simulation studies point out that for binary responses, the performance of the estimators is influenced by the cluster size, e.g., that clusters of size two usually entail problems (Breslow and Clayton, 1993; Diaz, 2007; Kim et al., 2013; Grilli et al., 2014). Following the suggestion of one of the reviewers, we decided to assess the role of the cluster size

on empirical bias and MSE with a small additional simulation study with eight different settings defined by two values of cluster size ($m \in \{2, 5\}$), two values of overdispersion ($\Phi \in \{1.5, 10\}$), and two different marginal means ($\mu \in \{1, 10\}$). The offset ($\lambda = 60$), the sample size ($n = 100$), and the slope parameter ($\beta_4 = -0.46$) were kept constant. Regarding the results of the empirical bias and the MSE, which are shown in Tables 4 and 5 in the Supplemental Material, there were hardly any differences between cluster sizes two and five. However, the estimates obtained by the three packages had larger empirical bias and MSE when $\Phi = 10$ and $\mu = 1$. Under this scenario, the values of package `hglm` were larger than those of the `lme4` and `INLA` packages. Moreover, with respect to computing time under the settings of this additional simulation study, the `hglm` package was somewhat faster than `lme4`. The computing times of the `INLA` package were, again, much larger. Problems of convergence were not detected.

It is important to highlight that the concepts of estimation and standard error are different in the three statistical principles used. In the case of the marginal likelihood, estimation is the value that maximizes the likelihood and the standard error reflects the sample variation of the estimator. In Bayesian analysis, estimation is a summary of the posterior distribution and the standard error is a measure of the variability of this distribution. As for the classical likelihood, the extended likelihood advocates the use of standard errors computed from the Fisher information matrix of the marginal likelihood. In practice, the standard errors are computed from the matrix of second derivatives for the adjusted profile h-likelihood, which is an approximation of the marginal likelihood (Lee et al., 2006).

Parameter estimation of GLMM is nowadays possible using different statistical principles, though estimation methods as well as statistical packages are still under development (Bolker et al., 2009). The problem of selecting the most adequate approach for the estimation and inference within GLMM is very complex. In addition, the software implementations can differ considerably in flexibility, computation time and usability (Austin, 2010; Li et al., 2011). A strategy could be to carry out a simulation study that emulated the data design and to apply the different estimation methods. Although one may think that this strategy is not very practical, it would be indeed worse to use an estimation method that could provide biased and inefficient estimates.

We have shown through simulations that ignoring overdispersion in Poisson Mixed Models can have serious consequences on the parameter estimates. Available R packages can handle this problem very satisfactorily; however, care must be taken in situations with small sample size, large overdispersion, and small marginal mean. In such situations, the `lme4` package seems to have a slightly better performance than packages `hglm` and the `INLA`, which also depends on the choice of the prior (Grilli et al., 2014), especially concerning the estimation of the random effect's variance (Figure 2a). This observation coincides with the recommendation of Kim et al. (2013) to use the GHQ method under such settings and the active discussions of the package's authors (Douglas Bates, Martin Maechler, and Ben Bolker) and members of the R sig-mixed-models mailing list (`https://stat.ethz.ch/mailman/listinfo/r-sig-mixed-models`). All pack-

ages under study have recently been improved in terms of convergence and optimization.

Appendix

Table 1: *Overview of statistical principles.*

Principle	Method	Algorithms
Marginal Likelihood	Maximum likelihood	Newton-Raphson (N-R), Fisher scoring, Penalized iteratively reweighted least squares (PIRLS) Adaptative Gauss Hermite Quadrature (GHQ)
Extended likelihood	h-likelihood	N-R, Iterative weighted least squares (IRWLS)
Bayesian	Posterior mean	MCMC, Integrated Nested Laplace Approximations (INLA)

Table 2: *Results from the Poisson mixed model applied to the folk wrestling data. CI stands for confidence interval and credible interval (in the case of INLA), respectively.*

	Function `glm`		Package `lme4`		Package `hglm`		Package INLA	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Intercept	−4.34	[−4.73, −4.0]	−4.37	[−4.82, −3.99]	−4.37	[−4.79, −3.95]	−4.41	[−4.85, −4.01]
Weight category[1]								
Light	0.25	[−0.2, 0.71]	0.24	[−0.26, 0.76]	0.25	[−0.28, 0.78]	0.25	[−0.26, 0.76]
Semiheavy	0.1	[−0.36, 0.57]	0.1	[−0.41, 0.63]	0.11	[−0.43, 0.65]	0.12	[−0.4, 0.64]
Heavy	0.39	[−0.1, 0.87]	0.4	[−0.14, 0.96]	0.4	[−0.16, 0.97]	0.41	[−0.14, 0.97]
Winner	−0.48	[−0.82, −0.15]	−0.46	[−0.82, −0.07]	−0.46	[−0.85, −0.07]	−0.44	[−0.82, −0.06]
σ_u^2		–	0.08	[0.0, 0.39]	0.08	[0.0, 0.22]	0.12	[0.01, 0.31]
Dispersion (Φ)[2]	1.45		1.29		1.35		–	

[1] The reference category is Medium

[2] Obtained by means of equation (3)

Table 3: *Percentages of convergence in Simulation scenario 2 as a function of the marginal mean (μ), the average match number per season (Offset), overdispersion (OD), and the sample size.*

	Offset	OD	glm $n = 30$	glm $n = 100$	lme4 $n = 30$	lme4 $n = 100$	hglm $n = 30$	hglm $n = 100$	INLA $n = 30$	INLA $n = 100$
$\mu = 1$	60	$\Phi = 1$	100	100	99.7	100	99.7	100	100	100
		$\Phi = 1.5$	100	100	99.4	100	99.7	100	100	100
		$\Phi = 3$	100	100	99.4	100	99.6	100	100	100
		$\Phi = 10$	100	100	98.1	99.9	98	99.8	100	100
	100	$\Phi = 1$	100	100	100	100	100	100	100	100
		$\Phi = 1.5$	100	100	99.8	99.9	99.9	100	100	100
		$\Phi = 3$	100	100	99.4	100	99.7	100	100	100
		$\Phi = 10$	100	100	97.2	100	97.9	99.9	100	100
$\mu = 10$	60	$\Phi = 1$	100	100	100	100	100	100	100	100
		$\Phi = 1.5$	100	100	100	100	100	100	100	100
		$\Phi = 3$	100	100	100	100	100	100	100	100
		$\Phi = 10$	100	100	100	100	100	100	100	100
	100	$\Phi = 1$	100	100	100	100	100	100	100	100
		$\Phi = 1.5$	100	100	100	100	100	100	100	100
		$\Phi = 3$	100	100	100	100	100	100	100	100
		$\Phi = 10$	100	100	100	100	100	100	100	100

Acknowledgements

This article contains supplementary material that can be consulted at the web page: www.idescat.cat/sort/artpublished.html

References

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6, 251–262.

Alam, M., L. Rönnegård, and X. Shen (2014). Fitting conditional and simultaneous autoregressive spatial models in hglm. Submitted.

Aregay, M., Z. Shkedy, and G. Molenberghs (2013). A hierarchical Bayesian approach for the analysis of longitudinal count data with overdispersion: a simulation study. *Computational Statistics & Data Analysis*, 57, 233–245.

Austin, P. (2010). Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures. *The International Journal of Biostatistics*, 6, Article 16. doi: 10.2202/1557–4679.1195.

Avalos, M., P. Hellard, and J. C. Chatard (2003). Modeling the training-performance relationship using a mixed model in elite swimmers. *Medicine and Science in Sports and Exercise*, 35, 838–846.

Ayán, C., A. J. Molina, H. García, G. González, M. J. Álvarez, T. Fernández, and V. Martín (2010). Rules modification effect's in incidence of injuries in Lucha Leonesa (Leonesa wrestling). *Apunts. Medicina de l'Esport*, 45, 17–22.

Bates, D. (2010, February). lme4: Mixed-effects modeling with R.
URL: `http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf`.

Bates, D., M. Maechler, B. Bolker, and S. Walker (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8.

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association*, 57, 269–306.

Bjørnstad, J. F. (1996). On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association*, 91, 791–806.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J. S. White (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24, 127–135.

Booth, J. G., G. Casella, H. Friedl, and J. P. Hobert (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, 3, 179–191.

Breslow, N. E. (1984). Extra-poisson variation in log-linear models. *Applied Statistics*, 33, 38–44.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.

Broström, G. and H. Holmberg (2013). *glmmML: Generalized linear models with clustering*. R package version 1.0.

Brown, P. E. and L. Zhou (2010). Mcmc for generalized linear mixed models with glmmbugs. *The R Journal*, 2, 13–17.

Bullock, N. and W. G. Hopkins (2009). Methods for tracking athletes' competitive performance in skeleton. *Journal of Sports Sciences*, 27, 937–940.

Casals, M., M. Girabent-Farrés, and J. L. Carrasco (2014). Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000–2012): a systematic review. *PloS one*, 9, e112653.

Casals, M. and J. A. Martínez (2013). Modelling player performance in basketball through mixed models. *International Journal of Performance Analysis in Sports*, 13, 64–82.

Cervone, D., A. D'Amour, L. Bornn, and K. Goldsberry (2014). A multiresolution stochastic process model for predicting basketball possession outcomes. Submitted. arXiv:1408.0777v1.

Collins, D. (2008). *The performance of estimation methods for generalized linear mixed models*. Ph. D. thesis, School of Mathematics & Applied Statistics-Faculty of Informatics, University of Wollongong.

Cosandey-Godin, A., E. Teixeira Krainski, B. Worm, and J. Mills Flemming (2014). Applying Bayesian spatio-temporal models to fisheries bycatch in the canadian arctic. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 1–12.

Czado, C., V. Erhardt, A. Min, and S. Wagner (2007). Zero-inflated generalized poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Statistical Modelling*, 7, 125–153.

Dean, C. B. and J. D. Nielsen (2007). Generalized linear mixed models: a review and some extensions. *Lifetime Data Analysis*, 13, 497–512.

Diaz, R. E. (2007). Comparison of PQL and Laplace 6 estimates of hierarchical linear models when comparing groups of small incident rates in cluster randomised trials. *Computational Statistics & Data Analysis*, 51, 2871–2888.

Elston, D., R. Moss, T. Boulinier, C. Arrowsmith, and X. Lambin (2001). Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology*, 122, 563–569.

Fong, Y., H. Rue, and J. Wakefield (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, 11, 397–412.

Frühwirth-Schnatter, S. and H. Wagner (2010a). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West (Eds.), *Bayesian Statistics*, 9, 1–21. Oxford University Press.

Frühwirth-Schnatter, S. and H. Wagner (2010b). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154, 85–100.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1, 515–534.

Grilli, L., S. Metelli, and C. Rampichini (2014). Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models. *Journal of Statistical Computation and Simulation*, 85, 2718–2726.

Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.

Hardin, J. W. and J. Hilbe (2007). *Generalized Linear Models and Extensions*. Stata Press.

Hewett, T., C. Pasque, R. Heyl, and R. Wroble (2005). Wrestling injuries. *Medicine and Sport Science*, 48, 152–178.

Hägglund, M., M. Waldén, R. Bahr, and J. Ekstrand (2005). Methods for epidemiological study of injuries to professional football players: developing the UEFA model. *British Journal of Sports Medicine*, 39, 340–346.

Hägglund, M., M. Waldén, L. Til, and R. Pruna (2010). The importance of epidemiological research in sports medicine. *Apunts. Medicina de l'Esport*, 45, 57–59.

Hirsch, K. and A. Wienke (2012). Software for semiparametric shared gamma and log-normal frailty models: an overview. *Computer methods and programs in biomedicine*, 107, 582–597.

Junge, A., L. Engebretsen, M. L. Mountjoy, J. M. Alonso, P. A. F. H. Renström, M. J. Aubry, and J. Dvorak (2009). Sports injuries during the Summer Olympic Games 2008. *The American Journal of Sports Medicine*, 37, 2165–2172.

Kim, Y., Y.-K. Choi, and S. Emery (2013). Logistic regression with multiple random effects: a simulation study of estimation methods and statistical packages. *The American Statistician*, 67, 171–182.

Klügl, M., I. Shrier, K. McBain, R. Shultz, W. H. Meeuwisse, D. Garza, and G. O. Matheson (2010). The prevention of sport injury: an analysis of 12 000 published manuscripts. *Clinical Journal of Sport Medicine*, 20, 407.

Komárek, A. and E. Lesaffre (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random-effects distribution. *Computational Statistics and Data Analysis*, 52, 3441–3458.

Lee, W. and Y. Lee (2012). Modifications of REML algorithm for HGLMs. *Statistics and Computing*, 22, 959–966.

Lee, Y., J. Nelder, and Y. Pawitan (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Boca Raton: Chapman & Hall/CRC.

Lee, Y. and J. A. Nelder (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B (Methodological)*, 58, 619–678.

Lee, Y. and J. A. Nelder (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, 88, 987–1006.

Lesaffre, E. and B. Spiessens (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50, 325–335.

Li, B., H. Lingsma, E. Steyerberg, and E. Lesaffre (2011). Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Medical Research Methodology*, 11, 77–87.

Lindgren, F. and H. Rue (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63, 1–25.

Martins, T. G., D. P. Simpson, A. Riebler, H. Rue, and S. H. Sørbye (2014). Penalising model component complexity: a principled, practical approach to constructing priors. Submitted. arXiv:1403.4630v3.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. London: Chapman and Hall.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, Linear and Mixed Models*. New York: John Wiley & Sons.

Milanzi, E., A. Alonso, and G. Molenberghs (2012). Ignoring overdispersion in hierarchical loglinear models: possible problems and solutions. *Statistics in Medicine*, 31, 1475–1482.

Molas, M. and E. Lesaffre (2011). Hierarchical generalized linear models: the R package HGLMMM. *Journal of Statistical Software*, 39, 1–20.

Muenchen, R. A. (2015). The popularity of data analysis software.
URL: http://r4stats.com/articles/popularity/.

Noh, M. and Y. Lee (2007). Reml estimation for binary data in GLMMs. *Journal of Multivariate Analysis*, 98, 896–915.

Ormerod, J. and M. Wand (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 21, 2–17.

Pan, J. and R. Thompson (2007). Quasi-Monte Carlo estimation in generalized linear mixed models. *Computational Statistics & Data Analysis*, 51, 5765–5775.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. OUP Oxford.

Quiroz, Z. C., M. O. Prates, and H. Rue (2014). A Bayesian approach to estimate the biomass of anchovies off the coast of Perú. *Biometrics*, 71, 208–217.

Rönnegård, L., X. Shen, and M. Alam (2010). hglm: A package for fitting hierarchical generalized linear models. *The R Journal*, 2, 20–28.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 71, 319–392.

Sampaio, J., E. J. Drinkwater, and N. M. Leite (2010). Effects of season period, team quality, and playing time on basketball players' game-related statistics. *European Journal of Sport Science*, 10, 141–149.

Sturtz, S., U. Ligges, and A. Gelman (2005). R2winbugs: A package for running WinBUGS from R. *Journal of Statistical Software*, 12, 1–16.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.

West, B. T., K. B. Welch, and A. T. Galecki (2014). *Linear Mixed Models: A Practical Guide Using Statistical Software* (2nd ed.). CRC Press.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, 31, 144–148.

Zhang, H., N. Lu, C. Feng, S. W. Thurston, Y. Xia, L. Zhu, and X. M. Tu (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine*, 30, 2562–2572.

Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.