

# Grau en Estadística

---

**Títol:** Qualitat metodològica i informe dels models lineals generalitzats mixtes (GLMM) en ciències de l'esport: una revisió d'abast

**Autor:** Arnau Sánchez Alaix

**Director:** Martí Casals Toquero i Daniel Fernández Martínez

**Departament:** DEIO

**Convocatòria:** Juliol 2023





## RESUM I PARAULES CLAU

Els models lineals generalitzats mixtes (GLMM) estimen efectes fixos i aleatoris i són especialment útils quan la variable dependent és binària, ordinal, de comptatge o quantitativa però no distribuïda normalment i també quan la variable dependent es mesura en diverses ocasions pel mateix subjecte. L'interès per l'estadística esportiva ha crescut en els darrers anys, així doncs també ho ha fet l'ús i l'aplicació d'aquests models, especialment adequats pel tipus de dades que es troben en ciències de l'esport. Aquest estudi pretenia presentar una revisió d'abast seguint la guia PRISMA per determinar com i amb quina freqüència s'utilitzen els GLMM en l'esport i resumir com es presenta la informació sobre ells en articles publicats. Es van cercar articles publicats en grans bases de dades com *PubMed* i *Web of Science*, a més a més de revistes científiques reconegudes en l'àmbit esportiu. Es van seleccionar un total de 35 articles per a la revisió. Els resultats van mostrar que l'ús de GLMM està més present en el futbol (25,7%) i el bàsquet (17,1%). La distribució de variable resposta més observada van ser la *Poisson* i la Binomial (25,7% cadascuna) i la sobredispersió no es va avaluar en un 85,7% dels articles seleccionats. El *software* estadístic més utilitzat va ser R (51,4%), però només es va reportar el paquet estadístic utilitzat en un 34,3% dels articles. S'ha compartit un repositori de dades i/o codi en 6 articles (17,1%). La majoria de la informació important sobre els GLMM no es va indicar a la majoria dels articles i per tant cal millorar la qualitat dels informes d'acord amb les recomanacions actuals per a l'ús de GLMM.

**Paraules clau:** *models lineals generalitzats mixtes, revisió d'abast, esports*

## RESUMEN Y PALABRAS CLAVE

Los modelos lineales generalizados mixtos (GLMM) estiman efectos fijos y aleatorios y son especialmente útiles cuando la variable dependiente es binaria, ordinal, de conteo o cuantitativa pero no distribuida normalmente y también cuando la variable dependiente se mide en varias ocasiones para el mismo sujeto. El interés por la estadística deportiva ha crecido en los últimos años, así pues también lo ha hecho el uso y aplicación de estos modelos, especialmente adecuados para el tipo de datos que se encuentran en ciencias del deporte. Este estudio pretendía presentar una revisión de alcance siguiendo la guía PRISMA para determinar cómo y con qué frecuencia se utilizan los GLMM en el deporte y resumir cómo se presenta la información sobre ellos en artículos publicados. Se buscaron artículos publicados en grandes bases de datos como *PubMed* y *Web of Science*, además de revistas científicas reconocidas en el ámbito deportivo. Se seleccionaron un total de 35 artículos para su revisión. Los resultados mostraron que el uso de GLMM está más presente en el fútbol (25,7%) y el baloncesto (17,1%). La distribución de variable respuesta más observada fueron la Poisson y la Binomial (25,7% cada una) y la sobredispersión no se evaluó en un 85,7% de los artículos seleccionados. El software estadístico más utilizado fue R (51,4%) pero sólo se reportó el paquete estadístico utilizado en un 34,3% de los artículos. Se ha compartido un repositorio de datos y/o código en 6 artículos (17,1%). La mayoría de la información importante sobre los GLMM no se indicó en la mayoría de los artículos y por lo tanto es necesario mejorar la calidad de los informes de acuerdo con las recomendaciones actuales para el uso de GLMM.

**Palabras clave:** *modelos lineales generalizados mixtos, revisión de alcance, deportes*

## ABSTRACT AND KEYWORDS

Generalized Linear Mixed Models (GLMM) estimate fixed and random effects and are especially useful when the dependent variable is binary, ordinal, count, or quantitative but not normally distributed, also when the dependent variable is measured multiple times for the same subject. Interest in sports statistics has grown in recent years, as has the use and application of these models, especially suitable for the type of data found in sports science. This study aimed to present a scoping review following the PRISMA guideline to determine how and how often GLMMs are used in sport and to summarize which information about them is presented in published articles. Articles were searched in large databases such as PubMed and Web of Science, as well as recognized scientific journals in sports field. A total of 35 articles were selected for the review. The results showed that the use of GLMM is more present in soccer (25.7%) and basketball (17.1%). The most observed response variable distribution was Poisson and Binomial (25.7% each) and overdispersion was not evaluated in 85.7% of the selected articles. The most used statistical software was R (51.4%), but only the statistical package used was reported in 34.3% of the articles. A data and/or code repository has been shared in 6 articles (17.1%). Most of the important information about GLMMs was not reported in most articles and therefore there is a need to improve the quality of reporting in line with current recommendations for the use of GLMMs.

**Keywords:** *generalized linear mixed models, scoping review, sports*

## CLASSIFICACIÓ AMS

### ***STATISTICS; Linear inference, regression***

62J05 Linear regression; mixed models

62J12 Generalized linear models

# ÍNDEX DE CONTINGUTS

<b>1. Introducció</b> .....	<b>12</b>
<b>2. Metodologia</b> .....	<b>14</b>
<b>2.1. Models lineals generalitzats mixtes (GLMM)</b> .....	<b>14</b>
2.1.1. L'especificació del model .....	14
2.1.2. Estimació dels paràmetres.....	15
2.1.3. Procediments inferencials .....	16
2.1.4. Aspectes computacionals .....	18
<b>2.2. Tipus de revisions</b> .....	<b>18</b>
2.2.1. <i>Systematic review</i> o <i>scoping review</i> ? .....	20
<b>3. Cos del treball</b> .....	<b>21</b>
<b>3.1. Criteris de la revisió d'abast</b> .....	<b>21</b>
<b>4. Resultats</b> .....	<b>26</b>
<b>4.1. Resultats de la revisió d'abast</b> .....	<b>26</b>
4.1.1. Característiques generals dels articles seleccionats .....	30
4.1.2. Característiques generals de l'esport.....	31
4.1.3. Característiques generals del GLMM ajustat.....	32
<b>5. Discussió</b> .....	<b>34</b>
<b>6. Referències</b> .....	<b>38</b>
<b>7. Annex</b> .....	<b>41</b>

## ÍNDIX DE FIGURES

<b>Figura 1:</b> Llibres de GLMM on es mostren 3 filosofies estadístiques per estimar paràmetres; la versemblança marginal clàssica, la jeràrquica i la filosofia bayesiana. Fonts: (22), (23), (24). .....	16
<b>Figura 2:</b> PRISMA: Diagrama de flux de la revisió d'abast de l'ajust de models GLMM en articles originals en el camp de les ciències de l'esport. ....	23



## ÍNDIX DE GRÀFICS

<b>Gràfic 1:</b> Nombre d'articles publicats per any .....	30
<b>Gràfic 2:</b> Nombre de participants de cada categoria segons la classificació de l'estudi .....	31
<b>Gràfic 3:</b> Softwares estadístics utilitzats en els 35 articles seleccionats. ....	32

## ÍNDIX DE TAULES

<b>Taula 1:</b> Tècniques per l'estimació de paràmetres GLMM, els seus avantatges i desavantatges i els paquets de software que les implementen. Font: (16) .....	15
<b>Taula 2:</b> Taula descriptiva de les diferents classes de revisions de la literatura científica mencionades. Font: (16).....	19
<b>Taula 3:</b> Taula descriptiva de les característiques generals dels articles seleccionats.....	24
<b>Taula 4:</b> Taula descriptiva de les característiques generals de l'esport.....	24
<b>Taula 5:</b> Taula descriptiva de les característiques dels GLMM ajustats en els articles .....	24
<b>Taula 6:</b> Freqüències de les característiques generals dels articles.....	26
<b>Taula 7:</b> Freqüències de les característiques generals de l'esport analitzat en els articles. ....	27
<b>Taula 8:</b> Freqüències de les característiques dels GLMM ajustats en els articles.....	28



# 1. Introducció

L'estadística esportiva és una disciplina de la ciència que ha anat agafant força amb el pas del temps. Durant les *Joint Statistical Meetings (JSM)* de l'any 1992 celebrades a Boston, es va fundar la *Section on Statistics in Sports (SIS)* dins de l'*American Statistical Association (ASA)*, amb l'objectiu de respondre a la necessitat de fomentar el desenvolupament de l'estadística i les seves aplicacions a l'esport (1). Des de llavors, l'ús de dades i anàlisis estadístiques en el món de l'esport s'ha convertit en una pràctica habitual i creixent, com bé s'ha pogut observar en el nombre d'articles publicats en revistes d'estadística reconegudes per la pròpia ASA com el *Journal of Quantitative Analysis in Sports*, *Chance* o *The American Statistician*, les conferències obertes com les impartides al JSM, o la creació de nous departaments de *sports analytics* en universitats (ex: *Harvard*, *Simon Fraser*, *Carnegie Mellon*, *California, Berkeley*, ...).

També han aparegut pel·lícules (ex: *Moneyball* (2), *Concussion* (3)), llibres (ex: *Sports Analytics* (4), *Mathletics* (5)) i recentment podcasts (ex: *Measurables*, *Counterpoints*, ...) que han ajudat i ajuden a entendre la importància de l'estadística en ciències de l'esport i sobretot a divulgar-la, provocant així l'augment d'interès de l'estadística esportiva entre la població.

Un altre factor clau a l'hora d'entendre aquest creixement es troba en els avenços tecnològics. Actualment, la tecnologia ha permès als sistemes d'informació relacionats amb l'esport recollir i analitzar una quantitat cada vegada més gran de dades, com bé s'ha vist amb el *tracking*, una tècnica que recull totes les accions i ubicacions dels jugadors en el terreny de joc en cada instant de temps (6). Això ha creat una gran oportunitat per als professionals que ajuden a les organitzacions esportives a aprofitar al màxim aquestes dades, és a dir, els *data scientists* i els estadístics esportius (7).

Així doncs, la demanda d'aquests professionals està creixent a mesura que les organitzacions esportives busquen aprofitar el gran volum de dades disponibles per obtenir una avantatge competitiva. Això no només afecta les organitzacions professionals, sinó també les acadèmies esportives (7).

De la mateixa manera que en altres àmbits, en ciències de l'esport normalment s'afronten problemes amb un objectiu que pot ser exploratori, predictiu, o causal (8). L'objectiu exploratori es refereix a la descripció i comprensió de dades existents, el predictiu es concentra en predir resultats o tendències futures, i el causal en comprendre les relacions de causa i efecte entre diferents factors i com influeixen en el rendiment o les actuacions dels jugadors o equips (8).

Actualment, la complexitat dels dissenys per poder contrastar adequadament les hipòtesis d'investigació d'interès ha provocat que les dades habitualment presentin una estructura jeràrquica o multinivell. En la pràctica i, concretament en ciències de l'esport, és habitual trobar-se amb jugadors o atletes que són observats dins d'unitats més grans, com equips, partits, competicions, temporades. Aquesta estructura de dades apareix també en els estudis longitudinals, on les mesures estan agrupades dins dels propis subjectes.

Els *Generalized Linear Mixed Models* (GLMM) són una extensió dels GLM que incorpora efectes aleatoris en el predictor lineal (9). Els GLMM proporcionen un procediment d'anàlisi més flexible quan la variable resposta no és continua, i on no es compleix l'assumpció d'independència donat que permet modelar dades agrupades mitjançant els efectes aleatoris. A més a més, els GLMM són útils per modelar la sobredispersió (10) i autocorrelació en models amb distribució de Poisson o Binomial. Els GLMM també es coneixen a la literatura com a models lineals generalitzats jeràrquics (HGLM) i models lineals generalitzats multinivell (MGLM), depenent de l'àmbit de recerca on s'apliqui.

En el món de l'esport ja s'han ajustat i aplicat aquests models, per exemple per identificar les variables que més inflüen en l'anotació de diversos jugadors de la NBA, on es va realitzar un GLMM assumint que la variable resposta (punts anotats) seguia una distribució *Poisson* (11). També s'han utilitzat per analitzar si el tipus de competició (a nivell de clubs, estat o internacional) afectava a la intensitat de partits de rugbi femení (12) mitjançant diversos models entre els quals es va ajustar un on la variable resposta (comptatge de faltes realitzades) seguia també una distribució de *Poisson*, o també s'ha fet un estudi en el qual s'han desenvolupat diversos GLMMs amb dades de futbol americà i bàsquet per ajustar respostes normals o *Poisson* amb prediccions binàries (13).

Tot i l'estructura jeràrquica o multinivell i també l'ús de models amb distribució exponencial encara fins al moment no existeixen revisions sobre l'ús i la qualitat de la informació reportada pels GLMM en ciències de l'esport, com bé ja s'ha fet fins al moment en l'àmbit de la medicina (14) i de la psicologia (15).

Per tant, l'objectiu principal d'aquest estudi és realitzar per primera vegada una revisió d'abast (RA) de l'aplicació dels GLMM i avaluar la qualitat dels resultats i de la informació reportada en articles originals en el camp de les ciències de l'esport a través d'una guia coneguda com PRISMA.

## 2. Metodologia

### 2.1. Models lineals generalitzats mixtes (GLMM)

Com ja s'ha mencionat, els models generalitzats mixtes (GLMM) són una extensió dels models lineals generalitzats (GLM) que afegeixen efectes aleatoris distribuïts normalment en el predictor lineal. Com bé definia Ben Bolker (16), els GLMM combinen les propietats dels models lineals mixtes (LMMs) i dels GLMs mitjançant l'ús de funcions d'enllaç i distribucions de la família exponencial com la Binomial o la Poisson (17).

#### 2.1.1. L'especificació del model

Sigui  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})'$  un vector de  $m$  observacions de la variable resposta corresponent al subjecte  $i$ ,  $i = 1, \dots, n$ , i  $\mathbf{u}_i$ ,  $i = 1, \dots, n$  el vector d'efectes aleatoris del mateix subjecte  $i$ . S'assumeix que  $(\mathbf{Y}_i|\mathbf{u}_i)$  segueix una distribució dins de la família exponencial amb funció de densitat  $f(\mathbf{Y}_i|\mathbf{u}_i; \cdot)$ , amb una mitjana i variància condicionals definides com  $\boldsymbol{\mu}_i = E(\mathbf{Y}_i|\mathbf{u}_i)$  i  $\text{Var}(\mathbf{Y}_i|\mathbf{u}_i) = \Phi \text{Var}(\boldsymbol{\mu}_i)$  respectivament, on  $\Phi$  és el paràmetre de dispersió i  $\text{Var}(\boldsymbol{\mu}_i)$  la funció de variància del GLMM.

La definició del GLMM es completa amb la introducció d'una funció d'enllaç monòtona i diferenciable  $g(\cdot)$  i amb un predictor lineal  $\boldsymbol{\eta}$ , com es pot veure a continuació:

$$\boldsymbol{\eta}_i = g(E(\mathbf{Y}_i)) = g(\boldsymbol{\mu}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i \quad (2.1)$$

On  $\mathbf{X}_i$  (de dimensió  $m \times k$ ) i  $\mathbf{Z}_i$  ( $m \times l$ ) són matrius de disseny del subjecte  $i$  associades amb efectes fixos i aleatoris, respectivament.  $\boldsymbol{\beta}$  ( $k \times 1$ ) és el vector d'efectes fixos i  $\mathbf{u}$  ( $l \times 1$ ) és el vector d'efectes aleatoris que s'assumeix que segueix una distribució Gaussiana amb un vector mitjana  $\mathbf{0}$  i una matriu de covariàncies definida positiva desconeguda  $\boldsymbol{\Sigma}$ :  $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . La seva funció de densitat és definida com  $f(\mathbf{u}; \boldsymbol{\Sigma})$ .

La funció de densitat condicionada de  $\mathbf{Y}$  donat  $\mathbf{u}$  té la següent forma:

$$f(\mathbf{Y}|\mathbf{u}; \boldsymbol{\beta}) = \prod_{i=1}^n f(\mathbf{Y}_i|\mathbf{u}_i; \boldsymbol{\beta}) \quad (2.2)$$

I la funció de densitat multivariant de  $\mathbf{u}$  s'expressa com:

$$f(\mathbf{u}; \boldsymbol{\Sigma}) = \prod_{i=1}^n f(\mathbf{u}_i; \boldsymbol{\Sigma}) \quad (2.3)$$

### 2.1.2. Estimació dels paràmetres

El principi de versemblança implica dos objectes: les variables aleatòries observades (les dades) i els paràmetres fixes (desconeguts). Per obtenir les estimacions de màxima versemblança (ML) per a  $\beta$  i els components de la variància en  $\Sigma$ , s'ha de maximitzar la següent funció de versemblança:

$$l(\beta, \Sigma|Y) = f(Y; \beta) = \int f(Y|u; \beta)f(u; \Sigma) du \quad (2.4)$$

L'estimació dels paràmetres en els GLMM no és una qüestió trivial, perquè la derivació de la versemblança sovint no es pot aconseguir d'una manera analítica. Donat que els efectes aleatoris no són observats directament, apareixen dificultats pel que fa a l'ús de la funció de versemblança conjunta.

Per a anàlisis senzills on les variables resposta són normals, tots els tractaments tenen mides mostrals iguals (disseny balancejat) i tots els efectes aleatoris són efectes niats, el clàssic mètode ANOVA basat en calcular les diferències de suma de quadrats dona els mateixos resultats que la màxima versemblança (ML). En canvi, aquesta equivalència ja no es compleix per models GLMM més complexos: per trobar les estimacions de màxima versemblança, cal integrar probabilitats sobre tots els valors possibles dels efectes aleatoris, un procés molt lent o inclús computacionalment inviable (16).

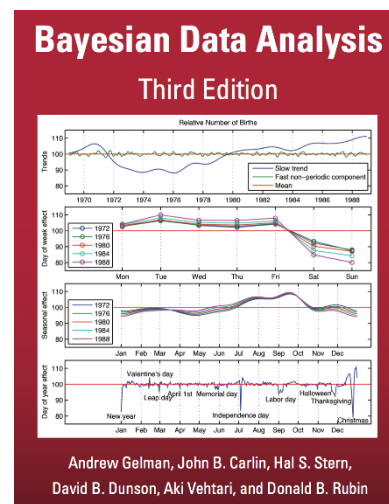
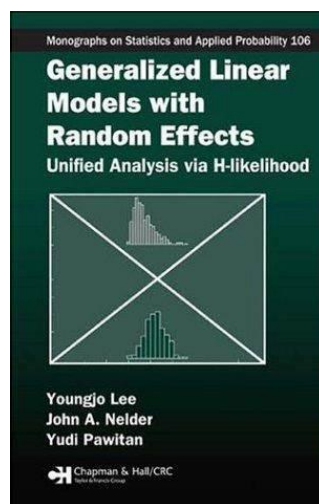
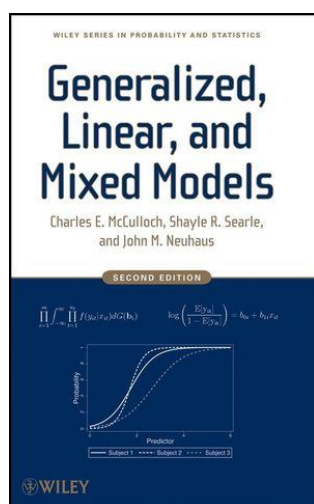
Així doncs, s'han proposat diferents mètodes per estimar els paràmetres del GLMM, com el mètode *Penalized quasi-likelihood* (PQL) (9), el mètode de *Laplace*, *Gauss-Hermite quadrature* (GHQ) (18); (19) i procediments jeràrquics (20) i procediments bayesians basats en les tècniques de *Markov Chain Monte Carlo* (MCMC) (21).

**Taula 1:** Tècniques per l'estimació de paràmetres GLMM, els seus avantatges i desavantatges i els paquets de software que les implementen. Font: (16)

Tècnica	Avantatge	Desavantatge	Paquet (software)
<b><i>Penalized quasilielihood</i></b>	Flexible, àmpliament implementat	La inferència de probabilitat pot ser inadequada; esbiaixat per a grans variàncies o petites mitjanes	PROC GLIMMIX (SAS), GLMM (Genstat), glmmPQL (R), lme4 (R)
<b><i>Laplace</i></b>	Més precís que el PQL	Més lent i menys flexible que PQL	PROC GLIMMIX, lme4 (R), glmm. admB (R), AD Model Builder, HLM
<b><i>Gauss-Hermite</i></b>	Més precís que <i>Laplace</i>	Més lent que <i>Laplace</i> ;	PROC GLIMMIX, PROC

<b>quadrature</b>		limitat a 2-3 efectes aleatoris	NLMIXED (SAS), lme4 (R), glmmML (R)
<b>Markov Chain Monte Carlo</b>	Nombre d'efectes aleatoris molt flexible i arbitrari; precís	Lent, tècnicament difícil, marc bayesià	WinBUGS, JAGS, MCMCpack, (R), AD Model Builder

Figura 1: Llibres de GLMM on es mostren 3 filosofies estadístiques per estimar paràmetres; la versemblança marginal clàssica, la jeràrquica i la filosofia bayesiana. Fonts: (22), (23), (24).



### 2.1.3. Procediments inferencials

Després d'estimar els valors dels paràmetres per als GLMM, el següent pas és la inferència estadística, és a dir, extreure conclusions de les dades examinant les estimacions i els seus intervals de confiança, provant hipòtesis, seleccionant el millor o millors models i avaluant les diferències de bondat d'ajust entre els models. En general parlem de tres tipus d'inferència: prova d'hipòtesis, comparació de models i enfocaments bayesians (16).

Els tests d'hipòtesi freqüentistes comparen els estadístics de prova (exemple: estadístic F en l'ANOVA) amb les seves distribucions esperades sota la hipòtesi nul·la, estimant un p valor per determinar si es pot rebutjar la hipòtesi nul·la. En la selecció de models, en canvi, es comparen els ajustaments dels models candidats. Els mètodes bayesians tenen el mateix abast general que els enfocaments freqüentistes, però difereixen tant en els seus fonaments filosòfics com en els procediments específics utilitzats (16).

#### Proves d'hipòtesi

Els tests de Wald Z,  $\chi^2$ , t i F per a GLMM comproven la hipòtesi nul·la que no hi ha efecte mitjançant



l'escala d'estimacions o combinacions de paràmetres pels seus errors estàndard estimats i comparant l'estadístic de prova resultant amb zero. Wald Z i  $\chi^2$  només són apropiats per a GLMM sense sobredispersió, mentre que les proves de Wald t i F expliquen la incertesa en les estimacions de sobredispersió. Això depèn del nombre de graus residuals de llibertat, que pot ser molt difícil de calcular perquè el nombre de paràmetres utilitzats per un efecte aleatori es troba entre 1 (un únic paràmetre de desviació estàndard) i N-1 (un paràmetre per a cada nivell addicional de l'efecte aleatori) (16).

La *likelihood ratio* (LR) determina la contribució d'un sol factor (aleatori o fix) mitjançant la comparació de l'ajust (mesurat com la *deviance*, és a dir, 2 vegades el logaritme de la LR) per als models amb i sense el factor, anomenats models niats. Encara que és molt utilitzat a tot arreu, la prova LR no es recomana per a la prova d'efectes fixes en GLMM, perquè no és fiable per a petites mides mostrals. La prova LR només és adequada per provar efectes fixos quan tant la relació entre la mida total de la mostra i el nombre de nivells d'efecte fix que s'estan provant i el nombre de nivells d'efecte aleatori (blocs) són grans (16).

### *Selecció del model*

Les proves LR poden avaluar la importància de factors particulars o, de manera equivalent, escollir el millor d'entre dos models niats, però alguns investigadors han criticat la selecció de models mitjançant aquestes comparacions com un abús de la prova d'hipòtesis. En canvi, existeixen mètodes d'informació teòrica que permeten la comparació de múltiples models que no tenen per què ser niats. L'*Akaike Information Criteria* (AIC) utilitza la *deviance* com una mesura d'ajust, afegint un terme per penalitzar models més complexos, és a dir, models amb un nombre més gran de paràmetres. Hi ha variants de l'AIC que són útils per exemple quan les mides mostrals són petites (AIC<sub>c</sub>), quan les dades pateixen sobredispersió (quasi-AIC, QAIC) o quan es vol identificar el nombre de paràmetres en un model "vertader" (*Bayesian or Schwarz information criterion*, BIC) (16).

### *Enfocament Bayesià*

Els enfocaments bayesians de la inferència GLMM ofereixen diversos avantatges respecte a la freqüentista i els mètodes d'informació teòrica. En primer lloc, el MCMC proporciona intervals de confiança pels paràmetres GLMM (i, per tant, proves de si aquests paràmetres podrien ser plausiblement iguals a zero) de manera que s'eviten moltes de les difícils aproximacions utilitzades en les proves d'hipòtesis freqüentistes (16).

En segon lloc, les tècniques bayesianes defineixen probabilitats de model posterior que penalitzen automàticament models més complexos, proporcionant una manera de seleccionar models. Com que aquestes probabilitats poden ser molt difícils de calcular, les anàlisis bayesianes solen utilitzar el

*Bayesian information criteria* (BIC), un criteri similar a l'AIC, que també requereix una estimació del nombre de paràmetres (16).

#### 2.1.4. Aspectes computacionals

Actualment els GLMM estan disponibles en diferents *softwares* estadístics. A nivell computacional, la funció GLIMMIX de SAS (1992) va ser la primera que es va disposar per ajustar GLMM amb el mètode d'estimació PQL, el qual produiria estimacions esbiaixades en els paràmetres (22). De totes maneres, la versió 9.2 de GLIMMIX seria la que permetria ajustar els GLMM amb les aproximacions de mètodes d'estimació de *Laplace* i GHQ de manera eficient (23,24). En l'actualitat, hi ha diferents *softwares* disponibles per ajustar els GLMM, però fins l'any 2000 no es van implementar la seva majoria, com ara ASReml (25), R (paquet lme4), ADMB, MLwiN, SAS (PROC GLIMMIX va esdevenir el procediment estàndard amb la versió 9.2, 2008), entre d'altres (24)(27). Els *softwares* disponibles poden ajustar diferents variables resposta de la família exponencial, com la distribució de *Poisson*, Binomial, Gamma, Inversa Gaussiana, encara que les més utilitzades i que la majoria té per defecte són la distribució *Poisson* i la Binomial (28).

Darrerament, s'estan realitzant diferents estudis que comparen els paquets estadístics disponibles i mètodes d'estimació (29)(11), i també es disposa de diferents llibres acadèmics sobre els GLMM, la seva aplicació, amb diferents filosofies i paquets de R (30). Així doncs, aquests estudis, llibres, els diferents tutorials dels diferents paquets i la web <http://glmm.wikidot.com/faq> revisada per Ben Bolker constitueixen una sèrie de recursos amb els que es pot obtenir informació actualitzada sobre el *software* estadístic disponible dels GLMM (28).

## 2.2. Tipus de revisions

Amb el pas de les dècades, els avenços tecnològics han anat augmentant i en conseqüència també ho han fet el nombre de publicacions científiques que volen compartir i actualitzar tota aquesta nova informació. Òbviament quantes més fonts existeixin, més facilitats tindrà la població d'adquirir i sintetitzar coneixements. En canvi, però, aquest augment de referències pot tendir a arribar a conclusions errònies o esbiaixades (31). És aquí on apareixen i cobren importància les revisions. Les revisions permeten sintetitzar de forma lògica i objectiva la informació, justificar la realització d'una investigació, explorar-ne la metodologia i analitzar de forma crítica els resultats del tema d'interès.

Per fer una revisió és important saber quines són les necessitats per les que es planteja realitzar un estudi ja que existeixen diferents tipus de revisions de la literatura científica. S'ha de tenir en compte el tipus de pregunta de la revisió, el tipus i la quantitat d'estudis disponibles i la seva finalitat, entre d'altres factors. Algunes de les diferents classes de revisions de la literatura científica són la *narrative review*, la *umbrella review*, la *scoping review*, la *rapid review* i la *systematic review* (Taula 2).

*Taula 2: Taula descriptiva de les diferents classes de revisions de la literatura científica mencionades.  
Font: (16)*

<b>Tipus</b>	<b>Descripció</b>	<b>Avantatges</b>	<b>Inconvenients</b>
<b><i>Narrative review</i></b>	Es tracta de material publicat que proporciona una revisió de la literatura recent o actual	Reuneix el que s'ha aconseguit sense repeticions i identificant omissions	No hi ha un mètode establert que s'asseguri que es consideri tota la literatura d'un tema
<b><i>Umbrella review</i></b>	Es refereix específicament a la revisió de l'evidència de múltiples vistes en un document accessible i utilitzable	Compila els resultats de múltiples revisions per a respondre a una pregunta específica. Crea un equilibri entre les opinions generals i les opinions que estan fragmentades a causa de la seva especificitat	Depèn del fet que ja hi hagi una revisió de components més restringida
<b><i>Scoping review</i></b>	Té com a objectiu identificar la naturalesa i l'abast de l'evidència de recerca	S'utilitza per a determinar si serà necessària una revisió sistemàtica completa per a arribar a una conclusió	No és un producte financer i corre un major risc de ser esbiaixat
<b><i>Rapid review</i></b>	És una avaluació del que ja se sap sobre una qüestió política o pràctica, mitjançant l'ús de mètodes de revisió sistemàtica per a buscar i avaluar críticament la recerca existent	Està dissenyat per a fer-ho ràpidament utilitzant estratègies de cerca menys sofisticades, mirant altres revisions, sense incloure la matèria grisa, i fent avaluacions de qualitat limitades	La reducció del temps d'avaluació de la qualitat augmenta el risc d'utilitzar estudis esbiaixats o de mala qualitat
<b><i>Systematic review</i></b>	Busca sistemàticament la cerca, avaluació i síntesi de proves de recerca	Cerca incloure tots els coneixements sobre un tema	És restrictiu per a enfocar un determinat mètode utilitzat en els estudis

### 2.2.1. *Systematic review o scoping review?*

Una revisió sistemàtica o *systematic review* es pot definir com un tipus de recerca duta a terme per identificar i recuperar evidències internacionals que són rellevants per a una pregunta o preguntes concretes i per avaluar i sintetitzar els resultats d'aquesta cerca. Les revisions sistemàtiques segueixen una estructura i un procés predefinit que requereix mètodes rigorosos per garantir que els resultats siguin fiables i significatius (32).

Es pot dur a terme una revisió sistemàtica per confirmar o refutar si la pràctica actual es basa o no en evidències rellevants, per establir la qualitat d'aquesta evidència i per abordar qualsevol incertesa o variació en la pràctica que es pugui produir. Aquestes variacions en la pràctica poden ser degudes a proves contradictòries i la realització d'una revisió sistemàtica hauria de resoldre aquests conflictes. La realització d'una revisió sistemàtica també pot identificar llacunes, deficiències i tendències en l'evidència actual i pot ajudar a sustentar i informar futures investigacions en l'àrea (32). A grans trets, les raons per les quals pot escaure la realització d'una revisió sistemàtica són les següents:

1. Descobrir l'evidència internacional
2. Confirmar la pràctica actual / abordar qualsevol variació / identificar noves pràctiques
3. Identificar i informar àrees per a futures investigacions
4. Identificar i investigar resultats conflictius
5. Elaborar declaracions per guiar la presa de decisions

Malgrat la utilitat de les revisions sistemàtiques per abordar aquestes situacions, hi ha casos en què les revisions sistemàtiques no poden satisfer els objectius o requisits necessaris o en què una activitat de cerca preliminar estructurada i metodològicament robusta és necessària per informar la realització de les revisions sistemàtiques. Com a tal, les *scoping reviews* o revisions d'abast han sorgit com un enfocament vàlid amb indicacions força diferents a les de les revisions sistemàtiques: els autors no sempre volen fer preguntes tan simples o precises, i potser estan més interessats en la identificació de determinades característiques/conceptes en articles o estudis, i en el mapeig, informe o discussió d'aquestes característiques/conceptes. En aquests casos, una *scoping review* és la millor opció (32).

Així doncs, es pot concloure que si l'investigador té una pregunta sobre la viabilitat, l'adequació, la significació o l'eficàcia d'un determinat tractament o pràctica, llavors una revisió sistemàtica és probablement l'enfocament més adequat. En canvi, quan l'objectiu de la revisió és identificar llacunes de coneixement, abordar un conjunt de literatura, aclarir conceptes, investigar la conducta de la investigació o informar una revisió sistemàtica, serà preferible la realització d'una *scoping review* (32).

Per tant, com que aquest projecte tracta d'avaluar la qualitat d'un tractament en articles científics mitjançant la cerca de paraules clau però sense fer un anàlisi rigorós ni profund, caldrà dur a terme una revisió d'abast.

## 3. Cos del treball

### 3.1. Criteris de la revisió d'abast

#### Disseny

S'ha dut a terme una revisió d'abast de l'aplicació de models lineals generalitzats mixtes (GLMM) en el camp de les ciències de l'esport seguint les recomanacions de la declaració PRISMA (33).

#### Estratègia de cerca

La cerca d'articles per a la revisió d'abast es va realitzar el 14 de març de 2023 en les bases de dades *PubMed Central (PMC)* i *Web of Science (WoS)*, així com també en aquelles revistes més reconegudes de contingut esportiu citades (34) que no es trobaven ni a PMC ni a WoS, com ara bé *Journal of Quantitative Analysis in Sports (JAQS)* i *South African Statistical Journal (SASJ)*. Tant en les bases de dades com en les revistes es va fer una recerca utilitzant els mateixos operadors booleans i paraules clau:

```
("sport*") AND ("mixe*" OR "multilevel*" OR "hierarch*" OR "generali*" OR "GLMM*" OR "HGLM*" OR "MGLM*") AND ("ordinal" OR "binary" OR "count" OR "nominal" OR "categorical" OR "dichot*" OR "polyto*")
```

#### Selecció dels estudis

Com a criteris d'elegibilitat dels estudis, es van definir els següents criteris d'inclusió i exclusió:

- Criteris d'inclusió

Els articles inclosos en la revisió d'abast havien de ser articles originals, escrits en anglès i que ajustessin algun model lineal generalitzat mixte.

- Criteris d'exclusió

Es van excloure de la RS els treballs de tesi, conferència i articles no originals (*systematic review, editorial, ...*), així com també tots aquells articles escrits en un idioma diferent a l'anglès.

Per al procés de selecció es van seguir els següents passos: en primer lloc, es van agrupar tots els articles trobats i es van eliminar tots els duplicats. A continuació es van filtrar els estudis en funció del títol i del resum. Finalment, en els casos que no es podia prendre una decisió a partir de la informació proporcionada pel títol i l'*abstract*, es van revisar els texts complets dels articles per tal d'avaluar si aquests complien els criteris d'elegibilitat.

### Identificació dels estudis

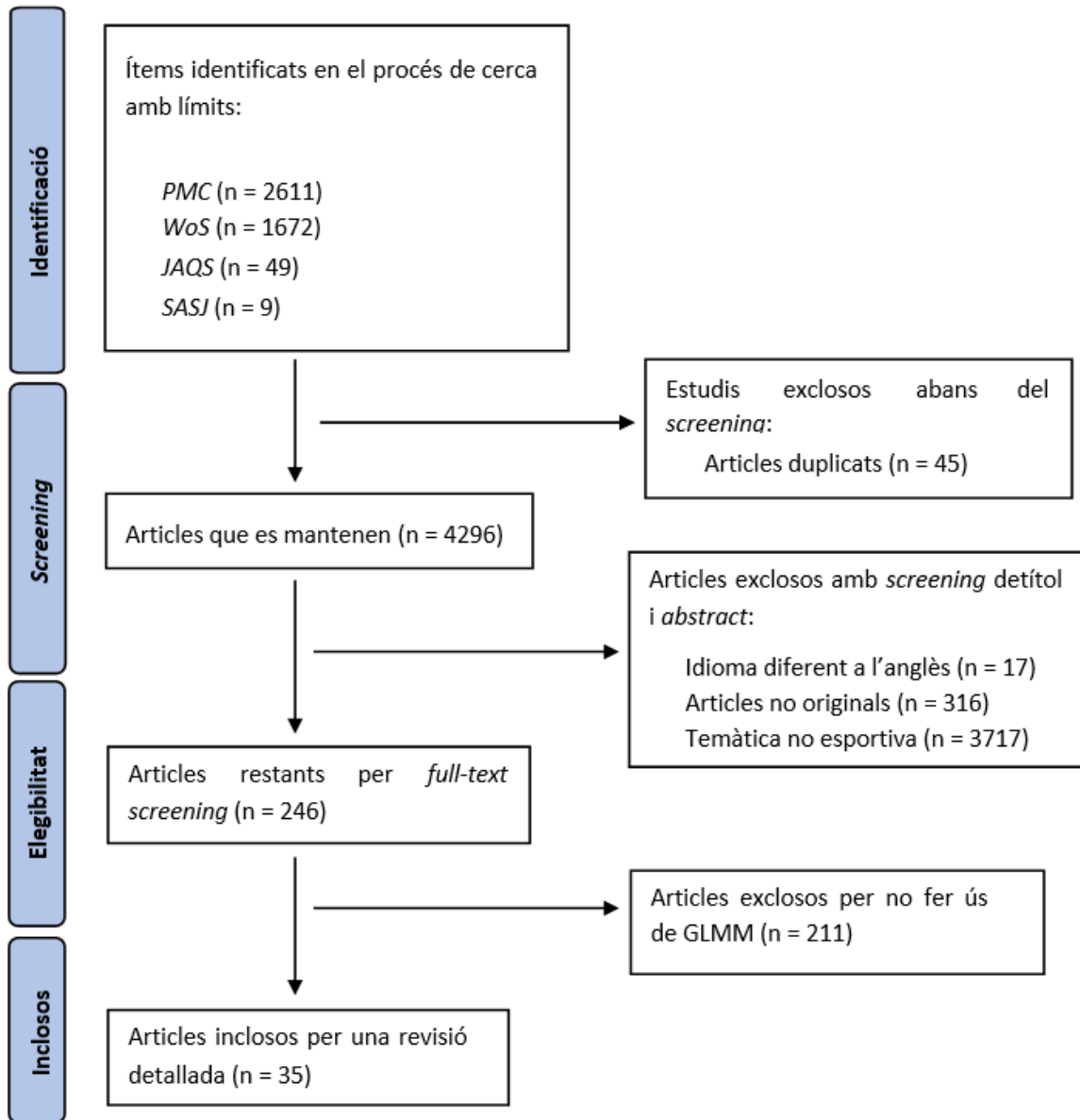
La Figura 2 mostra el diagrama de flux PRISMA per resumir totes les etapes del procés de selecció. En el primer procés de cerca es van recollir un total de 2611 ítems a *PubMed*, 1672 a *WoS* i un total de 58 entre totes les dues revistes esportives seleccionades. A l'agrupar tots aquests articles mitjançant *EndNote*, es van excloure 45 duplicats.

Després de la inspecció del títol i l'*abstract* dels 4296 articles restants, es van excloure 17 articles no escrits en anglès, 316 articles no originals i 3717 articles que no eren de temàtica esportiva.

Com a darrer pas, es va procedir a revisar el text complet dels 246 articles seleccionats fins al moment per veure si empraven algun model lineal generalitzat mixte. Es va observar que en 211 casos no es complia, generalment corresponents a articles on s'ajusten models de regressió logística, models mixtos o models lineals mixtes.

Per tant, finalment 35 articles van ser inclosos per una revisió més detallada a posteriori. La Figura 2 resumeix el nombre d'articles identificats i les raons d'exclusió en cada fase.

**Figura 2:** PRISMA: Diagrama de flux de la revisió d'abast de l'ajust de models GLMM en articles originals en el camp de les ciències de l'esport.



### Extracció de les dades

La informació recollida en els articles seleccionats es va agrupar en tres taules: característiques generals de l'article, característiques generals de l'esport i característiques del GLMM ajustat. La darrera es va subdividir en dues taules, una centrada en informació referent a inferència i mètodes d'estimació i l'altre a especificació, validació i construcció del model.

A continuació es presenten les variables d'interès de cada taula juntament amb les seves descripcions.

**Taula 3:** Taula descriptiva de les característiques generals dels articles seleccionats.

<b>Variable</b>	<b>Description</b>
<b>General characteristics of the selected articles</b>	
<i>Author (citation)</i>	Citació en format APA dels autors
<i>Country</i>	País d'on provenen les dades de l'article
<i>Publication Year</i>	Any de publicació de l'article
<i>Journal Name</i>	Nom de la revista on està publicat l'article
<i>Type of Journal</i>	Tipus de revista ( <i>Sport, Statistics, Multidisciplinary, ...</i> )
<i>N (participants)</i>	Nombre de participants en l'estudi
<i>Age (participants)</i>	Mitjana o interval d'edat dels participants
<i>Principal Aim</i>	Descripció de l'objectiu principal de l'estudi
<i>Outcome type</i>	Descripció de la variable resposta
<i>Type of design</i>	Tipus de disseny ( <i>Cross-sectional, Longitudinal/repeated measures, Mixed</i> )
<i>Multilevel (nested design)</i>	Multinivell ( <i>Yes, No</i> )

**Taula 4:** Taula descriptiva de les característiques generals de l'esport.

<b>General characteristics of the sport</b>	
<i>Author (citation)</i>	Citació en format APA dels autors
<i>Sport</i>	Esport/s en el/s que es centra l'article
<i>Gender</i>	Sexe dels participants en el que es centra l'article ( <i>Male, Female, Both</i> )
<i>Category participants</i>	Categoria dels participants ( <i>Professional, Amateur, Both</i> )
<i>Name of source data (League, Association, Organization, Federation,...)</i>	Nom de la font de les dades de l'estudi
<i>Category classification</i>	<i>Sports Performance Analysis, Sports Technology, Movement Integration, Health, eSports</i>

**Taula 5:** Taula descriptiva de les característiques dels GLMM ajustats en els articles

<b>General characteristics of the GLMM</b>	
<b>Characteristics of inference and estimation methods reported in the review articles</b>	
<i>Test for fixed effects</i>	Test aplicat per efectes fixes
<i>Test for random effects</i>	Test aplicat per efectes aleatoris
<i>Variance estimates of random effects</i>	Estimació de la variància dels efectes aleatoris ( <i>Yes, No</i> )
<i>Statistical software</i>	<i>Software</i> utilitzat
<i>Estimation method</i>	Mètode d'estimació
<i>Statistical software package or procedure</i>	Paquet utilitzat del <i>software</i>



<i>Statistical software function or macro</i>	Funció utilitzada del <i>software</i>
<i>Data shared</i>	Si l'article comparteix o no les dades utilitzades en l'estudi
<i>Code shared</i>	Si l'article comparteix o no el codi escrit durant l'estudi
<i>Repository of Data or Code shared</i>	Si l'article comparteix un repositori amb les dades o el codi utilitzat
<b><i>Characteristics of the specification, validation and construction of the model for the reviewed articles</i></b>	
<i>Variable response distribution</i>	Distribució de la variable resposta
<i>Overdispersion evaluation</i>	Avaluació de la sobredispersió ( <i>Yes, No</i> )
<i>Overdispersion measurement</i>	Mesura de sobredispersió
<i>Proposed alternative for overdispersion</i>	Proposta alternativa per la sobredispersió
<i>Method of variable selection</i>	Mètode de selecció de variables
<i>Method of model selection</i>	Mètode per comparar i seleccionar models
<i>GLMM validation</i>	Validació del model ajustat ( <i>Yes, No</i> )

## 4. Resultats

### 4.1. Resultats de la revisió d'abast

En aquest apartat, es mostren els resultats de l'anàlisi descriptiu de les variables descrites a les taules de l'apartat anterior referents als diferents tipus de característiques dels 35 articles seleccionats per dur a terme la revisió d'abast. En primer lloc, es presenta una taula de freqüències de les variables més destacades.

*Taula 6: Freqüències de les característiques generals dels articles.*

<b>Variable (N = 35)</b>	<b>Category</b>	<b>n (%)</b>
<b>General characteristics of the selected articles</b>		
<i>Country</i>	<i>USA</i>	10 (28,6%)
	<i>International*</i>	9 (25,7%)
	<i>Spain</i>	4 (11,4%)
	<i>Australia</i>	3 (8,6%)
	<i>UK</i>	2 (5,7%)
	<i>Netherlands</i>	1 (2,9%)
	<i>Argentina</i>	1 (2,9%)
	<i>Finland</i>	1 (2,9%)
	<i>Slovenia</i>	1 (2,9%)
	<i>Germany</i>	1 (2,9%)
	<i>Poland</i>	1 (2,9%)
	<i>NR**</i>	1 (2,9%)
<i>Publication Year</i>	<i>2012</i>	2 (5,7%)
	<i>2013</i>	1 (2,9%)
	<i>2014</i>	1 (2,9%)
	<i>2015</i>	2 (5,7%)
	<i>2016</i>	5 (14,3%)
	<i>2017</i>	2 (5,7%)
	<i>2018</i>	4 (11,4%)
	<i>2019</i>	5 (14,3%)
	<i>2020</i>	2 (5,7%)
	<i>2021</i>	7 (20,0%)
	<i>2022</i>	3 (8,6%)
	<i>2023</i>	1 (2,9%)
<i>Type of Journal</i>	<i>Sports</i>	20 (57,1%)
	<i>Multidisciplinary</i>	13 (37,1%)
	<i>Statistics</i>	2 (5,7%)

<i>Outcome type</i>	<i>Number of events***</i>	14 (40,0%)
	<i>Presence of injury (Yes/No)</i>	4 (11,4%)
	<i>Outcome of an event****</i>	4 (11,4%)
	<i>Ranking of the team</i>	2 (5,7%)
	<i>Match result (Win/Lose)</i>	2 (5,7%)
	<i>Others*****</i>	8 (22,8%)
	<i>NR**</i>	1 (2,9%)
<i>Type of design</i>	<i>Longitudinal / repeated measures</i>	29 (82,9%)
	<i>Cross-sectional</i>	4 (11,4%)
	<i>NR**</i>	2 (5,7%)
<i>Multilevel (nested design)</i>	<i>Yes</i>	24 (68,6%)
	<i>No</i>	11 (31,4%)

*\*International: estudis realitzats en diversos països.*

*\*\*NR: Not reported.*

*\*\*\*Number of events: nombre de gols marcats (n=2), punts aconseguits (n=3), faltes comeses (n=2), nombre de lesions (n=2), targetes rebudes (n=2), ...*

*\*\*\*\*Outcome of an event: desenllaç d'una jugada en atac (n=3) i desenllaç d'un cop de bat (n=1)*

*\*\*\*\*\*Others: tipus de variable resposta que només s'han observat en un article, com per exemple si l'equip local ha comès o no la falta, la màxima categoria a la que ha arribat un esportista professional, l'opinió sobre la influència del terreny de joc en el resultat final, si un atleta ha superat o no una prova, ...*

*Taula 7: Freqüències de les característiques generals de l'esport analitzat en els articles.*

<b>Variable (N = 35)</b>	<b>Category</b>	<b>n (%)</b>
<b>General characteristics of the sport</b>		
<i>Type of Sport</i>	<i>Soccer</i>	9 (25,7%)
	<i>Basketball</i>	6 (17,1%)
	<i>Multidisciplinary</i>	4 (11,4%)
	<i>Rugby</i>	3 (8,6%)
	<i>Swimming</i>	2 (5,7%)
	<i>Baseball</i>	2 (5,7%)
	<i>Australian Football</i>	2 (5,7%)
	<i>Others*</i>	7 (20,0%)
<i>Gender</i>	<i>Female</i>	1 (2,9%)
	<i>Male</i>	17 (48,6%)
	<i>Both</i>	8 (22,8%)
	<i>NR**</i>	9 (25,7%)

<i>Category Participants</i>	<i>Professional</i>	22 (62,8%)
	<i>Amateur</i>	11 (31,4%)
	<i>Both</i>	1 (2,9%)
	<i>NR**</i>	1 (2,9%)
<i>Category classification</i>	<i>Sport performance analysis</i>	25 (71,4%)
	<i>Health</i>	10 (28,6%)

\*Others: esports que només s'han observat en un article: American football, fishing, Formula 1, Leonese wrestling, field hockey, volleyball, ice hockey.

\*\*NR: Not reported.

**Taula 8:** Freqüències de les característiques dels GLMM ajustats en els articles.

<b>General characteristics of the GLMM</b>		
<b>Characteristics of inference and estimation methods reported in the review articles</b>		
<i>Test for fixed effects</i>	<i>Wald test</i>	5 (14,3%)
	<i>Z test</i>	2 (5,7%)
	<i>Chi-square test</i>	1 (2,9%)
	<i>NR*</i>	27 (77,1%)
<i>Test for random effects</i>	<i>NR*</i>	35 (100%)
<i>Variance estimates of random effects</i>	<i>Yes</i>	12 (34,3%)
	<i>NR*</i>	23 (65,7%)
<i>Statistical software</i>	<i>R</i>	18 (51,4%)
	<i>SAS</i>	5 (14,3%)
	<i>SPSS</i>	5 (14,3%)
	<i>Others**</i>	2 (5,7%)
	<i>NR*</i>	5 (14,3%)
<i>Estimation method</i>	<i>Gauss-Hermite Quadrature (GHQ)</i>	3 (8,6%)
	<i>2 methods: PQL, Laplace approximation</i>	1 (2,9%)
	<i>Monte Carlo Markov Chain (MCMC)</i>	1 (2,9%)
	<i>NR*</i>	30 (85,7%)
<i>Statistical software package or procedure</i>	<i>lme4</i>	8 (22,8%)
	<i>glmmLasso</i>	1 (2,9%)
	<i>brms</i>	1 (2,9%)
	<i>ordinal</i>	1 (2,9%)
	<i>3 packages: lme4, hglm, INLA</i>	1 (2,9%)
	<i>NR*</i>	23 (65,7%)

<i>Statistical software function or macro</i>	<i>PROC GLIMMIX</i>	3 (8,6%)
	<i>glmer</i>	2 (5,7%)
	<i>glmmLasso</i>	1 (2,9%)
	<i>NR*</i>	29 (82,8%)
<i>Data shared</i>	<i>Yes</i>	5 (14,3%)
	<i>No</i>	30 (85,7%)
<i>Code shared</i>	<i>Yes</i>	5 (14,3%)
	<i>No</i>	30 (85,7%)
<i>Repository of data or code shared</i>	<i>Yes</i>	6 (17,1%)
	<i>No</i>	29 (82,9%)

\*NR: Not reported.

\*\*Others: softwares utilitzats només en un article: MLwiN i Statistica

#### **Characteristics of the specification, validation and construction of the model for the reviewed articles**

<i>Variable response distribution</i>	<i>Poisson</i>	9 (25,7%)
	<i>Binomial</i>	9 (25,7%)
	<i>Ordinal</i>	5 (14,3%)
	<i>Multinomial</i>	3 (8,6%)
	<i>Negative-Binomial</i>	2 (5,7%)
	<i>2 distributions: Poisson, Negative Binomial</i>	1 (2,9%)
	<i>NR*</i>	6 (17,1%)
<i>Overdispersion evaluation</i>	<i>Yes</i>	5 (14,3%)
	<i>NR*</i>	30 (85,7%)
<i>Overdispersion measurement</i>	<i>Pearson's dispersion parameter</i>	2 (5,7%)
	<i>Pearson residuals</i>	1 (2,9%)
	<i>NR*</i>	32 (91,4%)
<i>Proposed alternative for overdispersion</i>	<i>Negative Binomial</i>	1 (2,9%)
	<i>NR*</i>	34 (97,1%)
<i>Method of variable selection</i>	<i>Stepwise</i>	3 (8,6%)
	<i>Lasso</i>	1 (2,9%)
	<i>NR*</i>	31 (88,5%)
<i>Method of model selection</i>	<i>Akaike Information Criteria (AIC)</i>	5 (14,3%)
	<i>2 methods: AIC &amp; BIC</i>	2 (5,7%)
	<i>Bayesian Information Criteria (BIC)</i>	1 (2,9%)
	<i>Deviance Information Criteria (DIC)</i>	1 (2,9%)
	<i>NR*</i>	26 (74,3%)

GLMM validation	Yes	2 (5,7%)
	NR*	33 (94,3%)

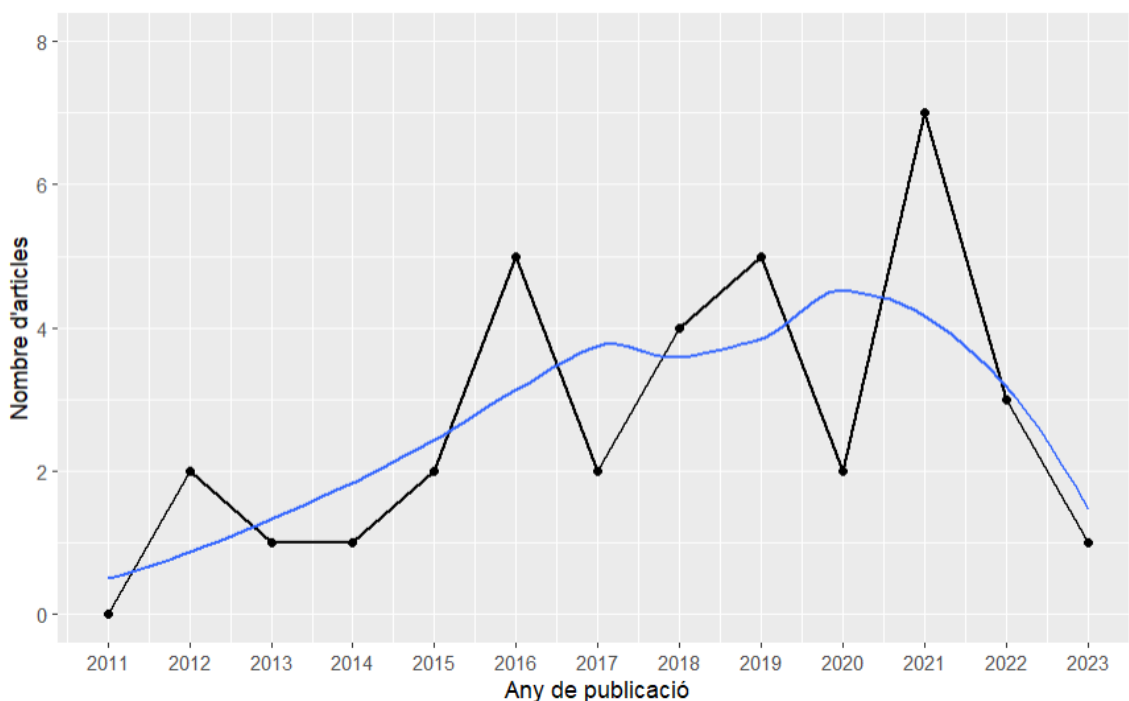
\*NR: Not reported.

#### 4.1.1. Característiques generals dels articles seleccionats

La Taula 6 mostra els resultats que fan referència a les característiques generals dels 35 articles seleccionats. D'aquests, 10 (28,6%) han utilitzat dades provinents d'Estats Units, 9 (25,7%) han utilitzat dades de diversos països, 4 (11,4%) d'Espanya, 3 (8,6%) d'Austràlia, 2 (5,7%) de Regne Unit i sis països en els quals s'han recollit dades en una ocasió. Hi ha 1 article que no informa d'on provenen les dades.

Pel que fa a l'any de publicació podem observar en el Gràfic 1 que tots els articles han estat publicats a partir de 2012 i cada cop amb més freqüència, ja que 29 dels 35 articles (82,6%) s'han publicat a partir de 2016. Així doncs sembla que el nombre de publicacions segueix una tendència creixent al llarg dels darrers anys almenys fins l'any 2020, tal i com també s'havia vist en les revisions sistemàtiques dutes a terme en els àmbits de medicina (14) i psicologia (15).

Gràfic 1: Nombre d'articles publicats per any



Dels 35 articles seleccionats, 20 d'ells (57,1%) es van publicar en revistes catalogades com a esportives, 13 (37,1%) en revistes multidisciplinàries i els 2 (5,7%) articles restants es van publicar en revistes d'estadística.

Els estudis longitudinals o de mesures repetides s'han dut a terme en 29 (82,9%) articles, mentre que 4 (11,4%) eren transversals (*cross-sectional*) i en 2 (5,7%) articles no s'ha pogut identificar el tipus de disseny.

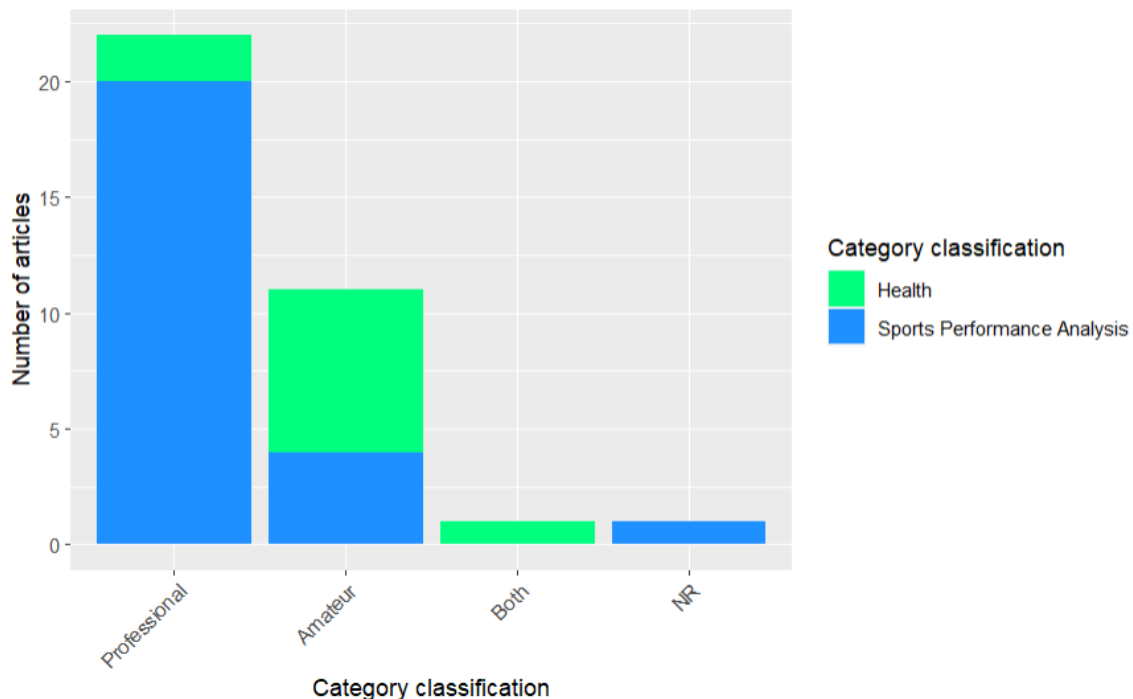
I en 24 (68,6%) articles s'ha detectat una estructura multinivell en les dades.

#### 4.1.2. Característiques generals de l'esport

En la Taula 7 s'observa que l'esport més analitzat és el futbol, present en 9 (25,7%) articles. El segueixen el bàsquet amb 6 (17,1%), el rugbi amb 3 (8,6%) i el futbol australià, la natació i el beisbol amb 2 (5,7%) cadascun. En 4 (11,4%) articles s'han analitzat atletes de diverses disciplines i finalment uns altres 7 (20,0%) articles han treballat amb dades d'esports que només s'han observat un cop. Un total de 17 (48,6%) articles han recollit informació només d'homes, mentre que només 1 (2,9%) ho ha fet amb dones. Hi ha 8 (22,8%) que han recollit dades d'ambdós sexes i 9 (25,7%) que no ho especifiquen.

Per altra banda, la categoria dels participants dels 35 articles fa referència a l'esport professional (62,8%), *amateur* (31,4%), ambdós (2,9%) o el cas en que no s'especifica la categoria dels seus participants (2,9%). En darrer lloc s'han observat dues categories de classificació: 25 (71,4%) orientats a l'anàlisi del rendiment i 10 (28,6%) orientats a la salut. En el Gràfic 2 es presenta un anàlisi bivariable entre aquestes últimes variables comentades. Dels 22 articles amb esportistes professionals, 20 tenien un objectiu d'anàlisi del rendiment. En canvi, dels 11 articles amb esportistes *amateur*, 7 estaven més enfocats en analitzar temes de salut.

Gràfic 2: Nombre de participants de cada categoria segons la classificació de l'estudi



#### 4.1.3. Característiques generals del GLMM ajustat

Dels 35 articles inclosos, tan sols 8 (22,9%) han mencionat el test dut a terme per calcular els efectes fixes, 5 (14,3%) amb el *Wald test*, 2 (5,7%) amb el *Z test* i 1 (2,9%) amb el *chi-square test*. Pel que fa al component dels efectes aleatoris, no hi ha cap article que expressi el test dut a terme per aquest component (Taula 8).

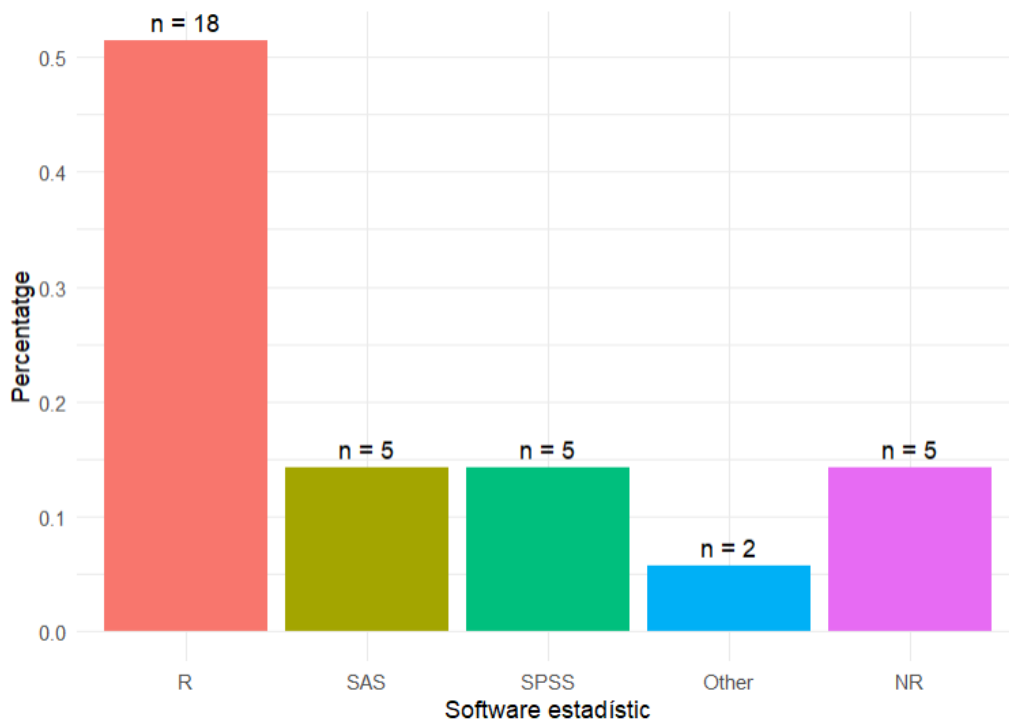
Les estimacions de les variàncies d'aquests efectes aleatoris s'han reportat en 12 (34,3%) dels 35 articles.

El mètode d'estimació s'ha reportat en 5 articles: *Gauss-Hermite Quadrature* (n=3), *Monte Carlo Markov Chain* (n=1) i un altre article on s'ha emprat tant PQL com l'aproximació de *Laplace*.

En el Gràfic 3 s'observa que més de la meitat dels articles (n=18) han realitzat el seu estudi mitjançant el *software* R, s'han fet 5 (14,3%) tant amb SAS com amb SPSS, 2 corresponen a *softwares* només observats un cop i 5 (14,3%) no fan menció del *software* utilitzat. Dins d'aquests *softwares*, el paquet que més s'ha fet servir és *lme4*, en 8 (22,8%) articles. També s'han utilitzat els paquets *brms*, *ordinal*, *glmmLasso* en 1 article cadascun, a part d'un altre on s'han ajustat diversos models amb els paquets *lme4*, *hglm* i INLA. Les funcions d'aquests *softwares* estadístics que s'han utilitzat i mencionat han estat: *PROC GLIMMIX* (n=3), *glmer* (n=2) i *glmmLasso* (n=1).

Finalment, les dades i el codi amb els que s'han dut a terme els estudis només s'han compartit en 5 (14,3%) articles cadascun, tot i que en 6 (17,1%) articles s'ha compartit un repositori de dades i/o codi.

Gràfic 3: *Softwares estadístics utilitzats en els 35 articles seleccionats.*



A la Taula 8 també es presenten els resultats amb informació referent a característiques



d'especificació, validació i construcció dels models ajustats. Les distribucions de variable resposta més vistes entre els 35 articles seleccionats han estat la *Poisson* i la binomial, observades en 9 (25,7%) articles cadascuna. Les segueixen l'ordinal amb 5 articles (14,3%), la *multinomial* amb 3 (8,6%), la binomial negativa amb 2 (5,7%), i 1 article en el que s'ha provat tant la distribució *Poisson* com la binomial negativa. Hi ha 6 (17,1%) articles en els quals no s'ha fet menció de la distribució que segueix la variable resposta de l'estudi.

Pel que fa a la sobredispersió, tan sols s'ha avaluat en 5 (14,3%) articles. D'aquests, en 2 articles s'ha mesurat mitjançant el paràmetre de dispersió de *Pearson* i en 1 mitjançant residus de *Pearson*, mentre que en els altres no s'ha mencionat. A més a més, només en 1 article s'ha proposat una alternativa per fer front a la sobredispersió. El mètode de selecció de variables en els models ajustats tampoc s'ha vist gaire entre tots els articles: 3 articles expliquen que van seleccionar les variables del model o models mitjançant el mètode *stepwise* i un mitjançant *Lasso*.

Hi ha 26 (74,3%) articles que no expliquen en quin mètode s'han basat per comparar models. Dels 9 articles que sí que ho han fet, el que més s'ha repetit ha estat l'*Akaike Information Criteria* (n=5). També s'ha utilitzat *Bayesian Information Criteria* (BIC) i *Deviance Information Criteria* (DIC) en 1 article cadascun i 2 articles on s'ha utilitzat tant AIC com BIC.

Per últim, s'ha observat si els articles han realitzat una validació del GLMM, cosa que només ha esdevingut en 2 (5,7%) articles, mentre que en els altres 33 (94,3%) no s'ha especificat.

## 5. Discussió

L'objectiu d'aquest estudi de recerca era revisar la freqüència i la qualitat d'ús dels models GLMM en l'esport. Per fer-ho possible, s'ha realitzat una revisió d'abast dels estudis publicats en aquest àmbit que ajustaven models GLMM, podent d'aquesta manera proporcionar informació més específica sobre com s'estan duent a terme i reportant aquests models en ciències de l'esport. Això ha permès identificar llacunes en com es reporta informació d'aquests models i, per tant, avaluar la qualitat dels GLMM en l'esport. A més a més, també s'han pogut comparar els resultats amb els que es van obtenir en les revisions realitzades en les àrees de medicina (14) i psicologia (15).

Els 35 articles seleccionats en aquesta revisió, després d'aplicar una selecció de publicació de R, mostren que el nombre de referències bibliogràfiques que utilitzen els GLMM en el camp de l'esport ha anat augmentant amb el pas del temps, com també es va observar en les revisions de medicina (14) i psicologia (15). Les primeres publicacions daten de 2012, des de llavors cada any s'ha publicat almenys un article. Un 82,9% dels articles s'han publicat a partir de 2016, la qual cosa fa pensar que hi ha una tendència a l'alça, donada pel creixent interès de l'estadística esportiva del qual s'ha parlat, tot i que només hi ha un article d'aquest any 2023, probablement degut a que en la cerca s'han seleccionat articles publicats com a molt tard el març d'aquest any. Entre els articles seleccionats predominen els longitudinals o de mesures repetides (82,9%), la qual cosa és coherent amb l'ús dels GLMM per modelar dades correlacionades al llarg del temps, així com també la proporció d'estudis (68,6%) que han detectat una estructura multinivell en les dades. Són resultats semblants als que es van veure en la revisió de medicina (14), en canvi en la de psicologia (15) es van seleccionar més estudis transversals.

Pel que fa a les característiques de l'esport, el futbol (25,7%) és l'esport on s'ha ajustat més el model GLMM, seguit del bàsquet (17,1%). Això pot ser degut a la popularitat, disponibilitat de les dades i evolució tecnològica creixent en aquests esports els darrers anys (35). També és interessant veure que gran part dels estudis estan centrats en l'anàlisi del rendiment (71,4%) i que en aquests han participat majoritàriament esportistes professionals, en canvi en els estudis més enfocats a temes de salut sembla que més aviat es recorre a esportistes *amateurs*. A més a més, s'ha observat que en un 48,6% dels articles els participants són homes, mentre que en un 22,8% dels articles s'han recollit dades d'ambdós sexes i en un altre 25,7% no s'ha especificat el sexe, fet destacable ja que tant les característiques físiques com l'estil de joc dels dos sexes són diferents (36).

Tal i com ja es va observar en els camps de medicina (14) i psicologia (15), les distribucions de variable resposta més comunes entre els articles analitzats han estat la *Poisson* i la Binomial, en 9 articles cadascuna (25,7%). En aquest cas es pot justificar per la naturalesa de les dades que es poden observar en el camp de l'esport, on el més habitual és treballar amb variables resposta que, o bé són un comptatge d'events (*Poisson*) com el nombre de gols, faltes rebudes/comeses o targetes rebudes, o sinó corresponen a si s'ha produït una ocurrència (Binomial), com victòria/derrota o si s'ha patit o no una lesió.

És important reportar el mètode d'estimació utilitzat perquè pot influenciar en la validació de les

estimacions dels models GLMM. En aquesta revisió tan sols un 14,3% dels articles l'han mencionat, tot i que en les revisions en medicina i psicologia tampoc es va arribar al 20%. L'únic que s'ha mencionat en més d'un article és el *Gauss-Hermite Quadrature* (n=3), el qual produeix estimacions més precises dels efectes fixes i dels components de la variància (37), encara que no és molt aplicable per anàlisis amb més de 2 o 3 efectes aleatoris (16).

El mètode de selecció de variables s'ha reportat en un 25,7% dels articles, semblant al 15,7% i 30,5% vistos en els camps de medicina (14) i psicologia (15) respectivament. De la mateixa manera que en aquestes dues revisions, el mètode més observat ha estat l'AIC (6 dels 9 articles que han reportat la variable).

També s'ha reportat insuficientment la inferència d'efectes fixes i aleatoris, ja que, tan sols s'ha reportat el test per efectes fixes en 8 articles (22,9%) i en alguns casos sí que s'ha reportat les estimacions de les variàncies d'aquests efectes aleatoris (34,3%). Els procediments inferencials han de ser coherents amb el mètode d'estimació utilitzat, però 6 dels 8 articles que han mencionat el test per efectes fixes no han reportat el mètode d'estimació. A més a més, tot i que en les revisions de medicina (14) i psicologia (15) tampoc es va reportar pràcticament el test per efectes aleatoris (per sota del 5,0%), cal que es mencioni aquesta informació perquè per exemple no tots els tests son aplicables a qualsevol estructura jeràrquica.

El mètode de selecció variables només s'ha esmentat en 4 articles (11,4%). És important mencionar si s'ha fet ús d'aquest recurs perquè tot i que els models sencers no donen estimacions esbiaixades, sí que contenen variables predictores no significatives.

Pel que fa a la sobredispersió, tot i que el fet d'ignorar-la pot causar problemes, només s'ha avaluat en 5 articles (14,3%), dels quals 3 mencionen com s'ha mesurat i 1 únic article que proposa una alternativa per fer front a aquesta sobredispersió (17). Són resultats semblants als vistos en les revisions de medicina (14) i psicologia (15), però segueix sent preocupant que no es tingui en compte, i més quan s'han vist diverses variables resposta que segueixen una *Poisson*, un tipus de distribució que habitualment presenta sobredispersió i que conseqüentment escau d'una distribució alternativa per ajustar les dades.

Per acabar de completar la informació relacionada amb l'ús dels GLMM, també és imprescindible assenyalar el *software* estadístic amb el qual s'ha ajustat el model, ja que presenten diferències en els mètodes d'estimació i temps de computació i permet la reproduïbilitat. El *software* més utilitzat per ajustar aquests models ha estat R (51,4%) seguit de SAS (14,3%) i SPSS (14,3%). El paquet més utilitzat ha estat lme4 (25,7%, implementat a R) i la funció més vista PROC GLIMMIX (8,6%, implementada a SAS). Tot i la disponibilitat oferta per aquests *softwares*, els GLMM no son fàcils d'ajustar, sovint passa que aquests models no convergeixen, especialment quan hi ha molts efectes fixes i aleatoris (15). És per això la importància de reportar el *software*, paquet i/o funció del *software*, així com també proporcionar les dades, codi o un repositori que contingui aquests, per tal que un expert interessat en l'estudi tingui la possibilitat de replicar-lo o assegurar-se que s'han seguit les pautes necessàries. En els articles seleccionats per a aquesta revisió no s'ha tingut molt en compte aquest darrer punt, ja que

només s'han compartit les dades, codi i/o repositori d'aquests en 9 articles diferents.

El principal objectiu d'aquest estudi s'ha centrat en avaluar la qualitat dels estudis publicats en ciències de l'esport que involucren models GLMM. Així doncs, es pot afirmar que, com ja es va observar en els camps de la medicina (14) i la psicologia (15), la qualitat reportada en l'esport també es pobla.

### Limitacions de l'estudi

Una limitació d'aquest estudi és el nombre relativament baix d'articles identificats, malgrat la haver tingut en consideració més de 4000 articles en una cerca inicial. Tot i que es va procurar abastar el major nombre de publicacions possible amb la incorporació de grans bases de dades i revistes àmpliament conegudes com WoS, PubMed, JQSA i JAS, pot ser que s'hagi prescindit d'articles procedents d'altres fonts com les mencionades per Swartz (34). Per tant, és important reconèixer que els resultats d'aquesta revisió poden no representar completament tot el panorama de la investigació que utilitzen models lineals generalitzats mixtes en ciències de l'esport. Aquesta limitació posa de manifest la necessitat d'estudis futurs per explorar bases de dades i fonts addicionals per garantir una cobertura més completa de la literatura.

Una altra limitació potencial és el possible biaix de cobertura provocat pels termes i criteris de cerca. La selecció de paraules clau i els criteris d'inclusió/exclusió utilitzats per identificar articles rellevants en aquest estudi poden introduir un cert nivell de subjectivitat i, potencialment, excloure alguns estudis rellevants. Tot i que es va intentar tenir present tota la literatura contemplada en el món de l'esport i els GLMM per garantir una estratègia de cerca integral, no es pot assegurar que s'hagi descartat o obviat algun estudi rellevant.

### Direccions futures de recerca

Cal contextualitzar que aquest estudi encara té la capacitat de ser més enriquidor a llarg termini. Per una banda, es podria afegir un cas pràctic per complementar tot el contingut proporcionat, ja sigui amb una base de dades pròpia o replicant un dels casos vistos en la revisió que comparteixen les dades. Amb la realització d'aquesta tasca s'aconseguiria una visió més clara per als investigadors interessats en redactar un estudi com els que s'han seleccionat en aquesta revisió, facilitant d'aquesta manera els passos que s'han de seguir i la informació que cal reportar per poder publicar un article amb bona reputació tant en l'àmbit esportiu com en l'estadístic. Més endavant també es podria tornar a dur a terme una revisió del mateix estil, analitzant els nous articles que compleixin els requisits per ser inclosos, primer de tot pel fet que explorar bases de dades i fonts addicionals garantiria una cobertura més completa de la literatura, i també per la possibilitat de poder observar si hi hagut una millora de la qualitat dels GLMMs ajustats i comparar-los amb els resultats obtinguts en aquesta revisió, podent així analitzar la incidència que ha tingut des de llavors en els estudis d'aquestes característiques.

## Conclusions

En els darrers anys, l'estadística ha anat agafant cada cop més importància en ciències de l'esport, fet que s'ha vist reflectit en el nombre de publicacions que tracten aquest tòpic. En aquest sentit, una de les eines estadístiques més útils que s'ha aplicat ha estat el model lineal mixte generalitzat (GLMM). En aquesta revisió d'abast s'inclouen articles en anglès, que es poden trobar en grans bases de dades com són *PubMed* i *Web of Science* (WoS) o en una revista reconeguda com és el *Journal of Quantitative Analysis in Sport* (JAQS) i que han ajustat algun GLMM en el camp de l'esport.

Un cop analitzats i revisats els articles, s'ha detectat que les dades més freqüents per ajustar GLMMs són les pertanyents a futbol i bàsquet, a més a més de participants masculins i professionals. Aquests estudis tenen com a propòsit analitzar el rendiment esportiu o la salut dels atletes. Els GLMMs ajustats presenten una manca d'informació important, tant de característiques d'inferència i mètodes d'estimació com d'especificació validació i construcció del model. Així doncs, s'ha de conscienciar dels aspectes clau del GLMM que s'han de reportar en un informe d'aquestes característiques, així com també de la necessitat de compartir les dades i el codi utilitzat. Per tant, si s'aconsegueix que tots els investigadors segueixin aquestes pautes, es podran garantir estudis i resultats de qualitat, precisos i transparents, que es podran reproduir i validar sempre que un ho trobi convenient.

Cal destacar que, segons el que se sap, a dia d'avui aquesta és la primera revisió d'abast d'articles que apliquen models lineals generalitzats mixtes l'àmbit de l'esport. Per tant, aquesta revisió pot esdevenir un recurs per promoure la publicació d'articles d'aquestes característiques i també serà útil a l'hora de desenvolupar-lo i comprendre'l.

## 6. Referències

1. ASA Community [Internet]. [cited 2023 May 8]. Available from: <https://community.amstat.org>
2. Moneyball. Columbia Pictures, Scott Rudin Productions, Michael De Luca Productions; 2012.
3. Concussion. LStar Capital, Scott Free Productions, The Cantillon Company; 2016.
4. Alamar B. Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers. Columbia University Press; 2013. 150 p.
5. Mathletics: How Gamblers, Managers, and Sports Enthusia... [Internet]. [cited 2023 May 8]. Available from: <https://www.goodreads.com/book/show/6732000-mathletics>
6. Lopez MJ. Bigger data, better questions, and a return to fourth down behavior: an introduction to a special issue on tracking data in the National football League. *J Quant Anal Sports* [Internet]. 2020 Jun 1 [cited 2023 May 8];16(2):73–9. Available from: <https://www.degruyter.com/document/doi/10.1515/jqas-2020-0057/html>
7. Casals M, Daunis-i-Estadella P. Violinboxplot and enhanced radar plot as components of effective graphical dashboards: An educational example of sports analytics. *Int J Sports Sci Coach* [Internet]. 2023 Apr 1 [cited 2023 May 8];18(2):572–83. Available from: <https://doi.org/10.1177/17479541221099638>
8. Nielsen RO, Simonsen NS, Casals M, Stamatakis E, Mansournia MA. Methods matter and the ‘too much, too soon’ theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference? *Br J Sports Med* [Internet]. 2020 Nov 1 [cited 2023 May 8];54(22):1307–9. Available from: <https://bjsm.bmj.com/content/54/22/1307>
9. Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *J Am Stat Assoc* [Internet]. 1993 Mar 1 [cited 2023 May 8];88(421):9–25. Available from: <https://doi.org/10.1080/01621459.1993.10594284>
10. McGilchrist CA. Estimation in Generalized Mixed Models. *J R Stat Soc Ser B Methodol* [Internet]. 1994 [cited 2023 May 15];56(1):61–9. Available from: <https://www.jstor.org/stable/2346027>
11. Casals M, Martinez J. Modelling player performance in basketball through mixed models. *Int J Perform Anal Sports*. 2013 Apr 1;13:64–82.
12. Newans T, Bellinger P, Drovandi C, Buxton S, Minahan C. The Utility of Mixed Models in Sport Science: A Call for Further Adoption in Longitudinal Data Sets. *Int J Sports Physiol Perform*. 2022 Aug 1;17(8):1289–95.
13. Broatch J, Karl A. Multivariate Generalized Linear Mixed Models for Joint Estimation of Sporting Outcomes. *Ital J Appl Stat*. 2018 Jan 1;30:189–211.
14. Casals M, Girabent-Farrés M, Carrasco JL. Methodological Quality and Reporting of Generalized Linear Mixed Models in Clinical Medicine (2000–2012): A Systematic Review. *PLoS ONE* [Internet]. 2014 Nov 18 [cited 2023 May 8];9(11):e112653. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4236119/>

15. Bono R, Alarcón R, Blanca MJ. Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review. *Front Psychol* [Internet]. 2021 [cited 2023 May 8];12. Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.666182>
16. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol*. 2009 Mar;24(3):127–35.
17. Casals M, Langohr K, Carrasco JL, Ronnega L. Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a simulation study.
18. Aitkin M. A general maximum likelihood analysis of overdispersion in generalized linear models. *Stat Comput* [Internet]. 1996 Sep 1 [cited 2023 May 8];6(3):251–62. Available from: <https://doi.org/10.1007/BF00140869>
19. Aitkin M. A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. *Biometrics* [Internet]. 1999 [cited 2023 May 8];55(1):117–28. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.1999.00117.x>
20. Booth JG. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood by Y. LEE, J. A. NELDER, AND Y. PAWITAN. *Biometrics* [Internet]. 2007 [cited 2023 May 8];63(4):1296–7. Available from: [https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2007.00905\\_1.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2007.00905_1.x)
21. Booth JG, Hobert JP. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J R Stat Soc Ser B Stat Methodol* [Internet]. 1999 [cited 2023 May 8];61(1):265–85. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00176>
22. Jang W, Lim J. A Numerical Study of PQL Estimation Biases in Generalized Linear Mixed Models Under Heterogeneity of Random Effects. *Commun Stat - Simul Comput* [Internet]. 2009 Feb 24 [cited 2023 May 16];38(4):692–702. Available from: <https://doi.org/10.1080/03610910802627055>
23. Bates DM, DebRoy S. Converting a large R package to S4 classes and methods. 2003;
24. Thiele J, Markussen B. Potential of GLMM in modelling invasive spread. *CABI Rev* [Internet]. 2012 Jan [cited 2023 May 9];2012:1–10. Available from: <http://www.cabidigitallibrary.org/doi/10.1079/PAVSNNR20127016>
25. Gilmour A, Gogel B, Cullis B, Welham S, Thompson R. ASREML user guide release 1.0. In 2002 [cited 2023 May 16]. Available from: <https://www.semanticscholar.org/paper/ASREML-user-guide-release-1.0-Gilmour-Gogel/bd1ef9825029655d534cff4000dd8f90ac902600>
26. 2021: Introduction to Multilevel Modelling Using MLwiN | School of Education | University of Bristol [Internet]. [cited 2023 May 16]. Available from: <https://www.bristol.ac.uk/education/events/2021/introduction-to-multilevel-modelling-using-mlwin.html>
27. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* [Internet]. 2008 Nov 1 [cited 2023 May 16];59(4):390–412. Available from: <https://www.sciencedirect.com/science/article/pii/S0749596X07001398>

28. MCiT\_TESI.pdf [Internet]. [cited 2023 May 8]. Available from: [https://www.tdx.cat/bitstream/handle/10803/384612/MCiT\\_TESI.pdf?sequence=1&isAllowed=y](https://www.tdx.cat/bitstream/handle/10803/384612/MCiT_TESI.pdf?sequence=1&isAllowed=y)
29. Dean CB, Nielsen JD. Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal.* 2007 Dec;13(4):497–512.
30. Schwarz CJ. Mixed Effects Models and Extensions in Ecology with R by ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A., and SMITH, G. M. *Biometrics* [Internet]. 2009 [cited 2023 May 16];65(3):992–3. Available from: [https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01315\\_4.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01315_4.x)
31. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf Libr J* [Internet]. 2009 [cited 2023 May 10];26(2):91–108. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-1842.2009.00848.x>
32. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* [Internet]. 2018 Nov 19 [cited 2023 May 8];18(1):143. Available from: <https://doi.org/10.1186/s12874-018-0611-x>
33. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med* [Internet]. 2009 Aug 18 [cited 2023 May 8];151(4):264–9. Available from: <https://www.acpjournals.org/doi/10.7326/0003-4819-151-4-200908180-00135>
34. Swartz T. Where Should I Publish My Sports Paper? *Am Stat.* 2018 Apr 9;74:1–18.
35. Sports Analytics: How Different Sports Use Data Analytics [Internet]. [cited 2023 May 16]. Available from: <https://www.datacamp.com/blog/sports-analytics-how-different-sports-use-data-analysis>
36. Thibault V, Guillaume M, Berthelot G, Helou NE, Schaal K, Quinquis L, et al. Women and Men in Sport Performance: The Gender Gap has not Evolved since 1983. *J Sports Sci Med.* 2010;9(2):214–23.
37. Pan J, Thompson R. Gauss-Hermite Quadrature Approximation for Estimation in Generalised Linear Mixed Models. *Comput Stat* [Internet]. 2003 Mar 1 [cited 2023 Jun 2];18(1):57–78. Available from: <https://doi.org/10.1007/s001800300132>



## 7. Annex

Repositori de dades i codi: <https://github.com/arnausanaix/TFG-Scoping-Review-GLMM>