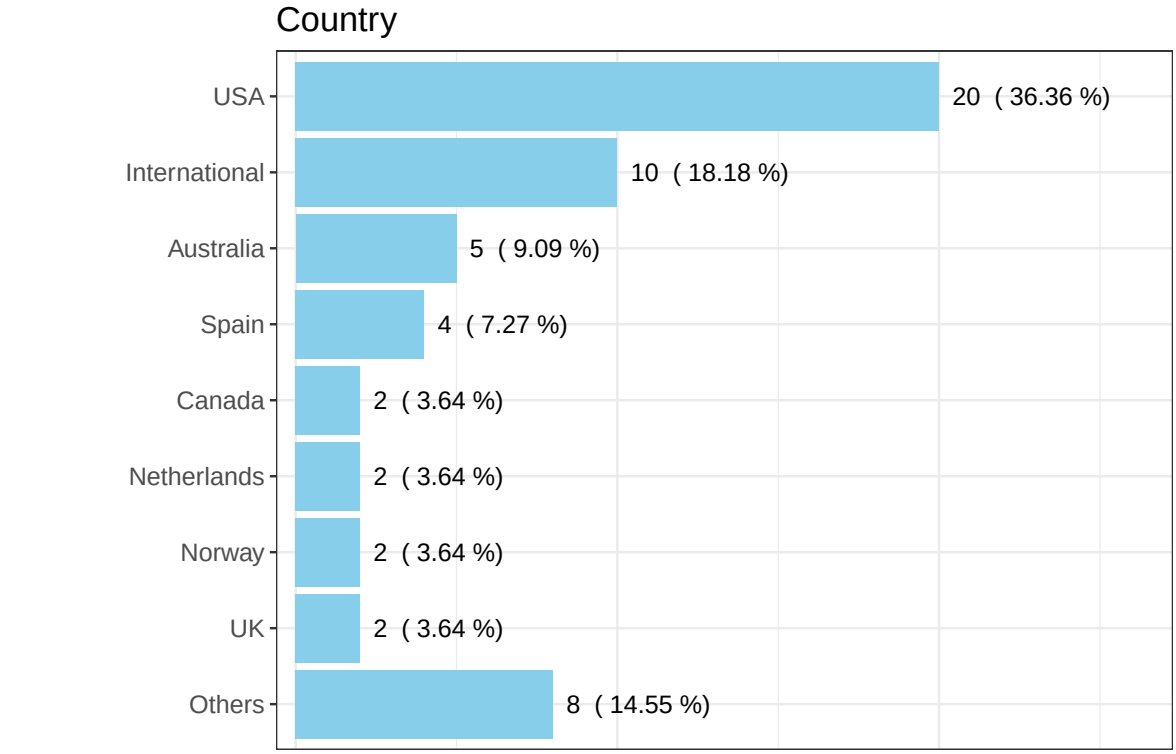
Exploratory analysis of the Methodological Quality and Reporting of GLMM in Sports: a Scoping Review

General characteristics of the selected articles

(A)



(B)

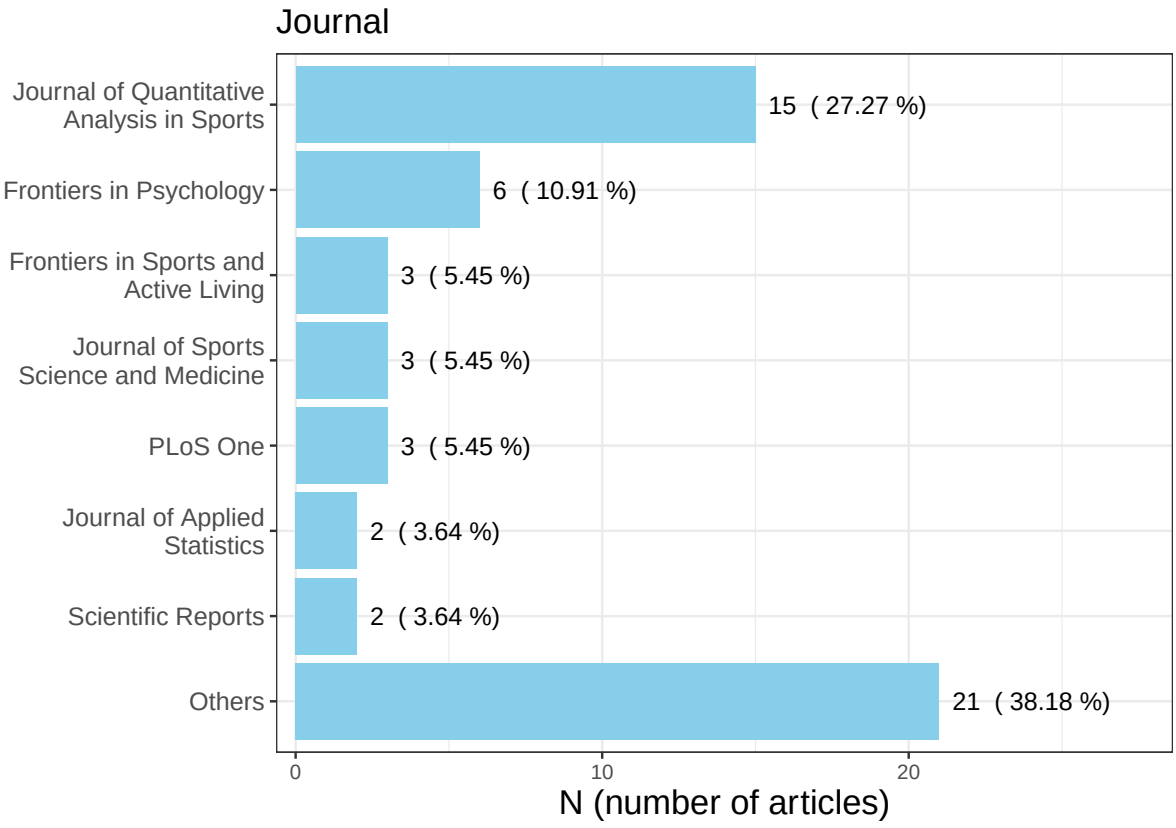


Table 1: Characteristics of the N = 55 selected articles in the scoping review, frequencies and percentages.

	N (%)
Type of Journal:	
Multidisciplinary	19 (34.5%)
Sports	32 (58.2%)
Statistics	4 (7.27%)
Type of design:	
Cross-sectional	5 (9.09%)
Longitudinal/repeated measures	35 (63.6%)
Not Reported	15 (27.3%)
Multilevel (nested design):	
No	12 (21.8%)
Yes	43 (78.2%)
Gender:	
Both	15 (27.3%)
Female	3 (5.45%)
Male	27 (49.1%)
Not Reported	10 (18.2%)
Participants category:	
Amateur	18 (32.7%)
Both	1 (1.82%)
Not Reported	1 (1.82%)
Professional	35 (63.6%)
Category classification:	
Academic achievement	3 (5.45%)
Health	13 (23.6%)
Sports Performance Analysis	39 (70.9%)
Test for fixed effects:	
Chi-square test	1 (1.82%)
NF	1 (1.82%)
Not Reported	46 (83.6%)
Wald test	5 (9.09%)
Z test	2 (3.64%)
Test for random effects: Not Reported	55 (100%)
Variance estimates of random effects:	
Not Reported	40 (72.7%)
Yes	15 (27.3%)
Statistical software function or macro:	
glmer	2 (3.64%)
glmm Lasso	1 (1.82%)
Not Reported	48 (87.3%)
PROC GLIMMIX	3 (5.45%)
PROC MIXED	1 (1.82%)
Data shared:	
No	46 (83.6%)
Yes	9 (16.4%)
Code shared:	
No	47 (85.5%)
Yes	8 (14.5%)
Repository of Data or Code shared:	
No	44 (80.0%)

Table 1: Characteristics of the N = 55 selected articles in the scoping review, frequencies and percentages. *(continued)*

	N (%)
Yes	11 (20.0%)
Variable response distribution:	
Binomial	18 (32.7%)
Multinomial	3 (5.45%)
Negative-Binomial	1 (1.82%)
Normal	8 (14.5%)
Not Reported	7 (12.7%)
Ordinal	6 (10.9%)
Poisson	10 (18.2%)
Poisson, Negative-Binomial	2 (3.64%)
Overdispersion evaluation:	
Not Reported	50 (90.9%)
Yes	5 (9.09%)
Overdispersion measurement:	
Not Reported	52 (94.5%)
Pearson residuals	1 (1.82%)
Pearson's dispersion parameter	2 (3.64%)
Proposed alternative for overdispersion:	
Negative Binomial	1 (1.82%)
Not Reported	54 (98.2%)
Method of variable selection:	
Holdout sample	1 (1.82%)
Lasso	1 (1.82%)
Not Reported	50 (90.9%)
Stepwise	3 (5.45%)
Method of model selection:	
AIC	5 (9.09%)
AIC, BIC	2 (3.64%)
BIC	1 (1.82%)
Cross Validation	1 (1.82%)
Deviance Information Criteria (DIC)	1 (1.82%)
Holdout sample	1 (1.82%)
LRT	1 (1.82%)
Not Reported	43 (78.2%)
GLMM Validation:	
Not Reported	50 (90.9%)
Yes	5 (9.09%)

In general, the quality and reporting of Generalized Linear Mixed Models in sports is **poor**:

- 27.3% of the selected articles did not report the type of study design (whether longitudinal or cross-sectional).
- 18.2% of the articles did not report the sex of the use case participants.
- **83.6%** of the articles did not report any test for fixed effects and **NO** articles reported test for random effects. **72.7%** of the articles did not report the variance estimates of random effects.
- 12.7% of the articles did not report distribution of the response variable. 91% of the articles did not

assess overdispersion in their variables, 78.2% did not report the method for the model selection, and 91% did not report that any validation had been performed on their GLMMs.

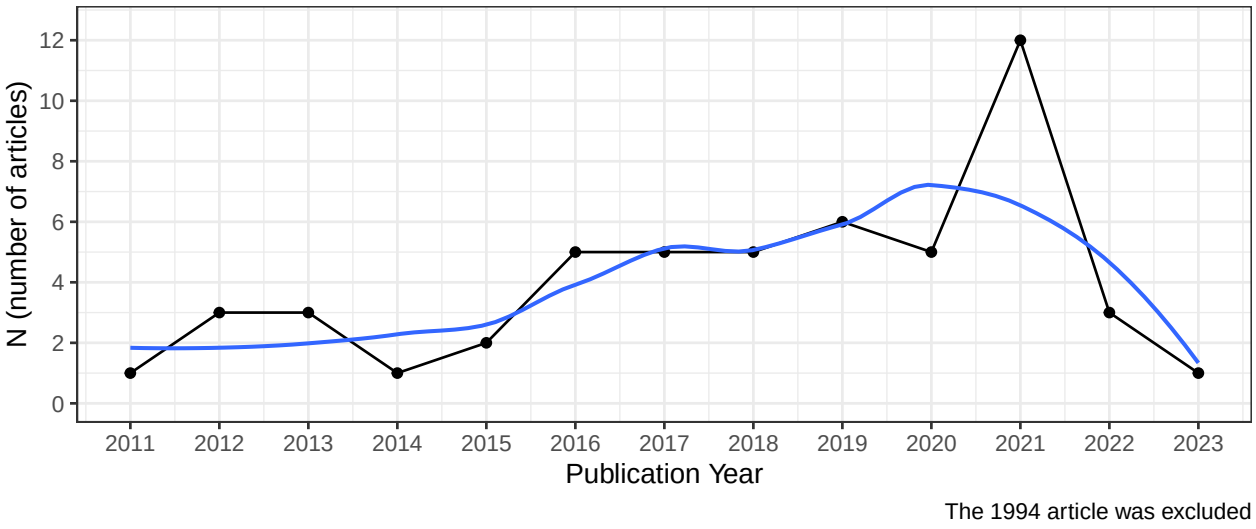


Figure 1: Number of papers published among the 55 selected over the years. The blue line represents the temporal trend.

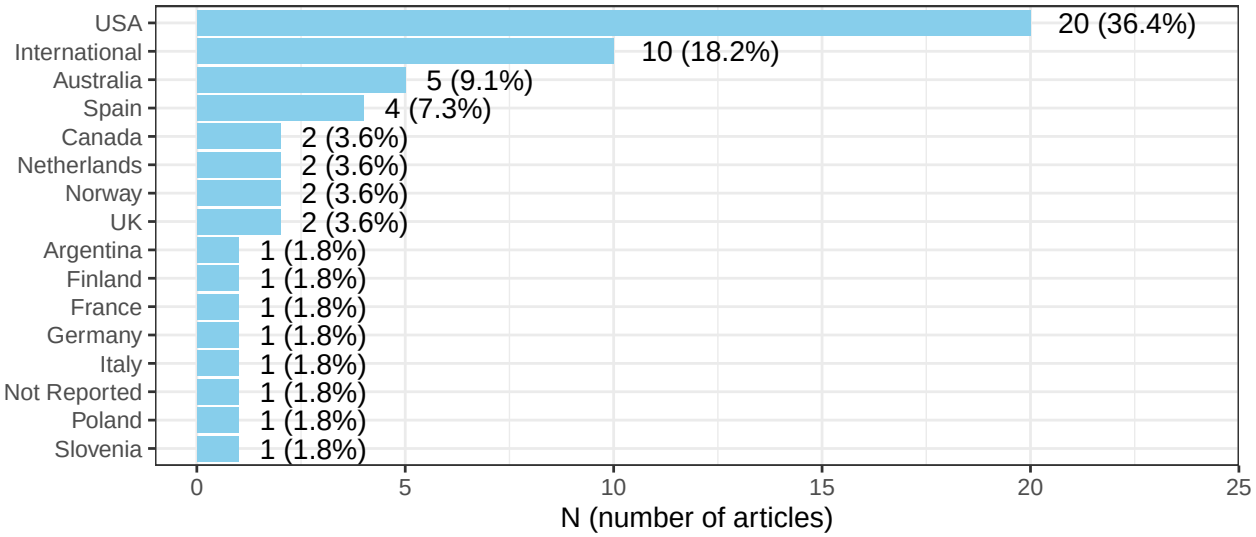


Figure 2: Distribution of the countries from which the data of the 55 selected articles come from.

Characteristics of the sport and data

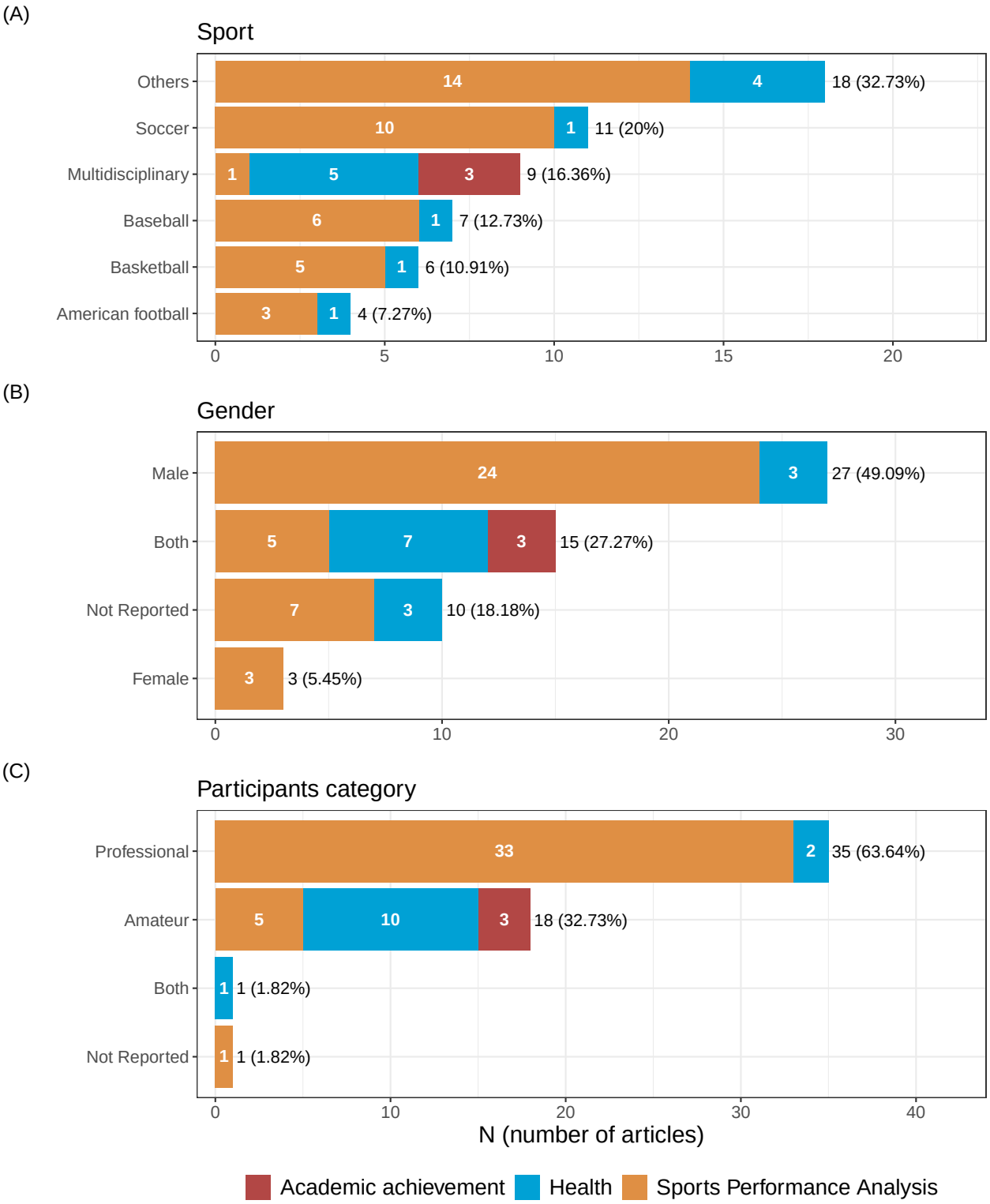


Figure 3: Distribution of the sports studied (panel (A)), the gender of the participants (panel (B)) and the professional category of the participants (panel (C)) in the 55 selected articles.

Characteristics of the methodology

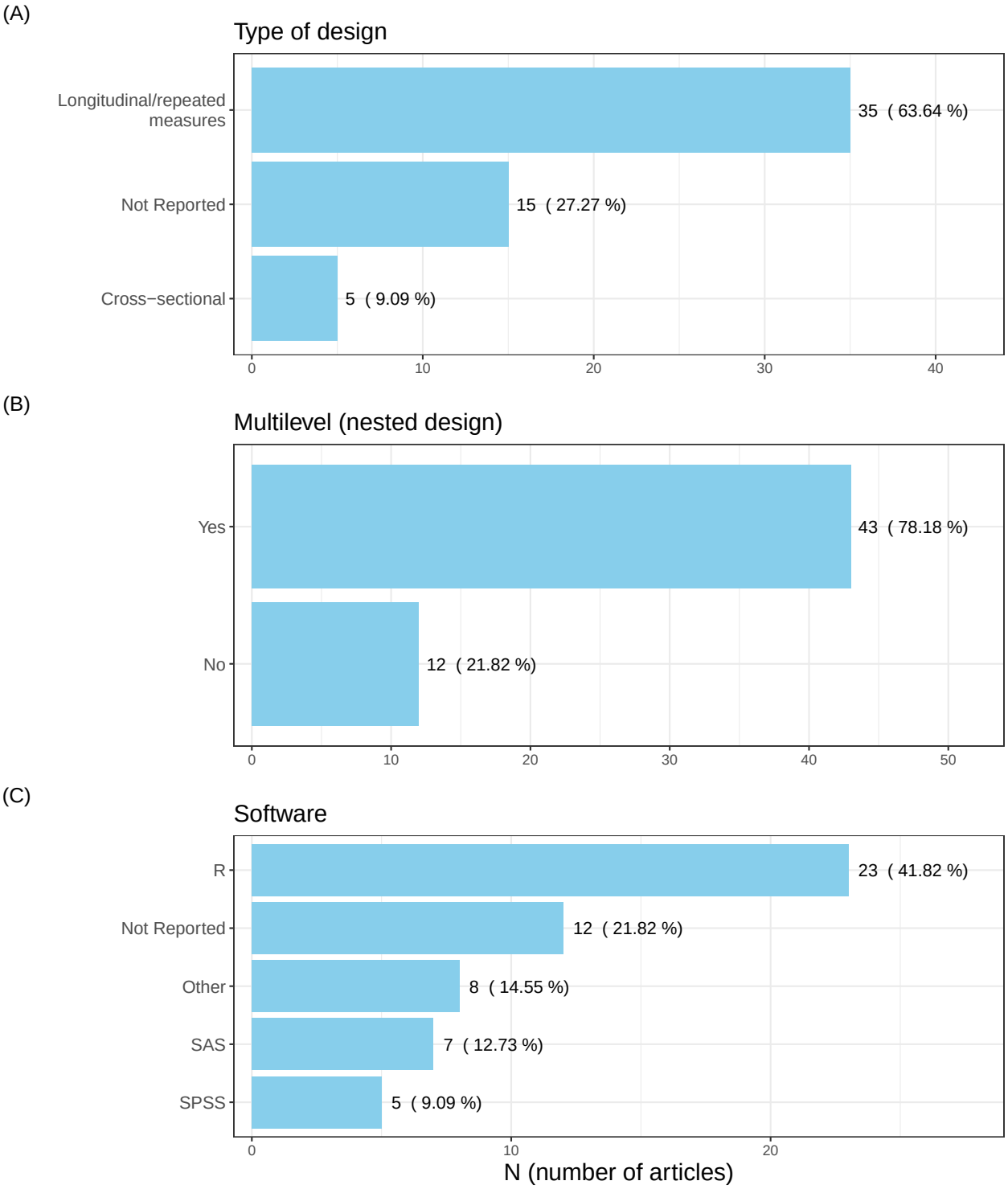


Figure 4: Distribution of the type of study design (panel (A)), whether multilevel models are used (panel (B)) and of the statistical software used (panel (C)) in the 55 selected articles.

Main conclusions

- Out of 55 selected articles 32 (58.2%) were published in sports journals, 19 (34.5%) in multidisciplinary journals and 4 (7.28%) in statistics journals. Regarding the multilevel design, 43 (78%) of the selected articles reported to have a multilevel design. Moreover, over sixty percent were longitudinal/repeated measure studies, whereas 9% (5 articles) were cross-sectional studies, and 27% (15 articles) did not report their study design.
- On the use case sports data, the three most predominant sports were soccer (20%), multidisciplinary sports (16.4%) and baseball (12.7%). Over seventy percent of the articles were devoted to the study of sports performance analysis, 23.6% to health and 5.5% to academic performance. In addition, the use case participants were professionals in the 64% of the selected articles, while in 33% they were amateurs. Only three articles (5.4%) included female participants, 27 articles (49%) included male participants and 10 articles (18.2%) did not report the sex of the participants.
- The majority of the selected articles did not share their data nor their code, only 9 and 8 articles did so, respectively. 42% of the articles used R open-source software in their analyses.