Reporting in Sports Analytics: Current State and Future Directions

Jordi Cortés · · Daniel Fernández · · Martí Casals

Advances in Sports Analytics. Statistical Methods, Data Science, and Research Practices

8 July 2026, University of Brescia (Italy)

Outline

1. Our team
2. Reproducibility crisis
3. Metaresearch: Reporting guidelines
4. Metaresearch in the **sports** field
   - Clustering
   - Ordinal response
   - Survival analysis
   - Generalized linear models
   - Synthetic data
   - R packages
5. Conclusions

GRBIO: Sports analytics group. Who are we?

Klaus Langohr (UPC) | David Blanco (UIC) | Martí Casals (UB) | Daniel Fernández (UPC) | Jordi Cortés (UPC)

Main current topics

- Current State of Sports Analytics
  - **Systematic/Scoping reviews** of reporting standards and methodological practices.
  - First **census** of sports analytics researchers to assess field adoption.

GRBIO: Sports analytics group. Who are we?

Klaus Langohr (UPC) | David Blanco (UIC) | Martí Casals (UB) | Daniel Fernández (UPC) | Jordi Cortés (UPC)

Main current topics

- Current State of Sports Analytics
  - **Systematic/Scoping reviews** of reporting standards and methodological practices.
  - First **census** of sports analytics researchers to assess field adoption.
- Modeling Approaches
  - **Survival analysis** for analyzing mortality of NBA players.

GRBIO: Sports analytics group. Who are we?

Klaus Langohr (UPC) | David Blanco (UIC) | Martí Casals (UB) | Daniel Fernández (UPC) | Jordi Cortés (UPC)

Main current topics

- Current State of Sports Analytics
  - **Systematic/Scoping reviews** of reporting standards and methodological practices.
  - First **census** of sports analytics researchers to assess field adoption.
- Modeling Approaches
  - **Survival analysis** for analyzing mortality of NBA players.
- Exploratory Analytics
  - Player typology identification (profiling) using unsupervised learning (clustering/**archetypoid analysis**).

GRBIO: Sports analytics group. Who are we?

Klaus Langohr (UPC) | David Blanco (UIC) | Martí Casals (UB) | Daniel Fernández (UPC) | Jordi Cortés (UPC)

Main current topics

- Current State of Sports Analytics
  - **Systematic/Scoping reviews** of reporting standards and methodological practices.
  - First census of sports analytics researchers to assess field adoption.
- Modeling Approaches
  - Survival analysis for analyzing mortality of NBA players.
- Exploratory Analytics
  - Player typology identification (profiling) using unsupervised learning (clustering/archetypoid analysis).

Why does reporting matter?

The problem in biomedical research

Biomedical research has faced a **reproducibility crisis** for more than two decades. Incomplete reporting, poor study design and lack of transparency reduce the value of scientific evidence.

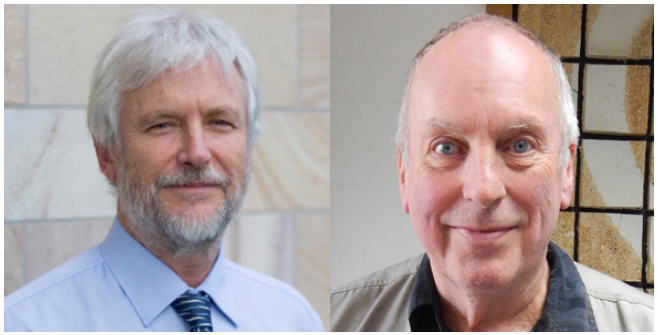
Why does reporting matter?

The problem in biomedical research

Biomedical research has faced a **reproducibility crisis** for more than two decades. Incomplete reporting, poor study design and lack of transparency reduce the value of scientific evidence.

Key evidence

- Ioannidis (2005): **Why Most Published Research Findings Are False.**
- Glasziou & Chalmers (2009): Up to **85% of research investment** may be wasted because of avoidable deficiencies in research.



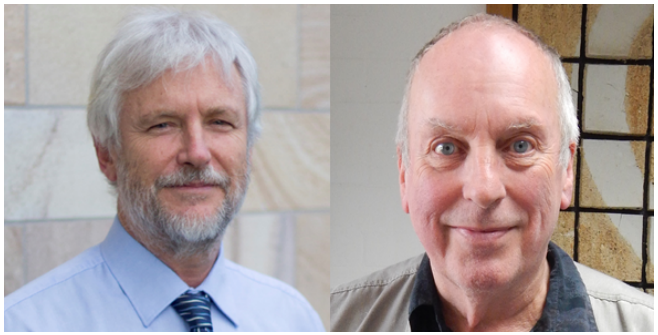
Why does reporting matter?

The problem in biomedical research

Biomedical research has faced a **reproducibility crisis** for more than two decades. Incomplete reporting, poor study design and lack of transparency reduce the value of scientific evidence.

Key evidence

- Ioannidis (2005): **Why Most Published Research Findings Are False.**
- Glasziou & Chalmers (2009): Up to **85% of research investment** may be wasted because of avoidable deficiencies in research.



Sports analytics is not immune to these problems.

Reporting guidelines

Reporting guidelines

Reporting guidelines are **evidence-based checklists** designed to improve the completeness, transparency, and reproducibility of research reporting.

equator network

CONSORT
Randomized Trials
Framework for reporting randomized controlled trials.

STROBE
Observational Studies
Guidelines for cohort, case-control, and cross-sectional studies.

PRISMA
Systematic Reviews
Minimum set of items for reporting systematic reviews.

SAMPL
Statistical Analyses
Guidelines for statistical methods and analyses in biomedical journals.

TRIPOD
Prediction Models
Standards for reporting clinical prediction models.

From Reporting Guidelines to Meta-Research

What is meta-research?

Meta-research is the scientific study of **how research is designed, conducted, analysed, reported and disseminated**.
In other words,

"Research on research"

From Reporting Guidelines to Meta-Research

What is meta-research?

Meta-research is the scientific study of **how research is designed, conducted, analysed, reported and disseminated**.
In other words,

"Research on research"

Reporting guidelines

- The **EQUATOR Network** includes many reporting guidelines to do metaresearch.
- Other initiatives also exist, such as **CHAMP**.
- Although most were developed for biomedical research, their principles are equally applicable to

"Sports Sciences"

From Reporting Guidelines to Meta-Research

What is meta-research?

Meta-research is the scientific study of **how research is designed, conducted, analysed, reported and disseminated**.
In other words,

"Research on research"

Next

We will now review several meta-research studies conducted in Sports Analytics and identify their main reporting limitations.

Reporting guidelines

- The **EQUATOR Network** includes many reporting guidelines to do metaresearch.
- Other initiatives also exist, such as **CHAMP**.
- Although most were developed for biomedical research, their principles are equally applicable to

"Sports Sciences"

From Reporting Guidelines to Meta-Research

What is meta-research?

Meta-research is the scientific study of **how research is designed, conducted, analysed, reported and disseminated**.
In other words,

"Research on research"

Next

We will now review several meta-research studies conducted in Sports Analytics and identify their main reporting limitations.

Reporting guidelines

- The **EQUATOR Network** includes many reporting guidelines to do metaresearch.
- Other initiatives also exist, such as **CHAMP**.
- Although most were developed for biomedical research, their principles are equally applicable to

"Sports Sciences"

Our approach

All reviews presented in this talk were conducted following the

PRISMA guideline

PRISMA Reporting Guideline and Good Reporting Practices

Most important items

✓ **Item 5**. **Eligibility criteria**. PICO framework and inclusion/exclusion criteria.

✓ **Item 7**. **Search strategy**. Complete search strings and information sources.

✓ **Item 8**. **Selection process**. Independent screening of records and full texts.

✓ **Item 9**. **Data collection process**. Standardized extraction of study characteristics.

✓ **Item 16**. **PRISMA flow diagram**. Transparent reporting of the study selection process.

PRISMA Reporting Guideline and Good Reporting Practices

Most important items

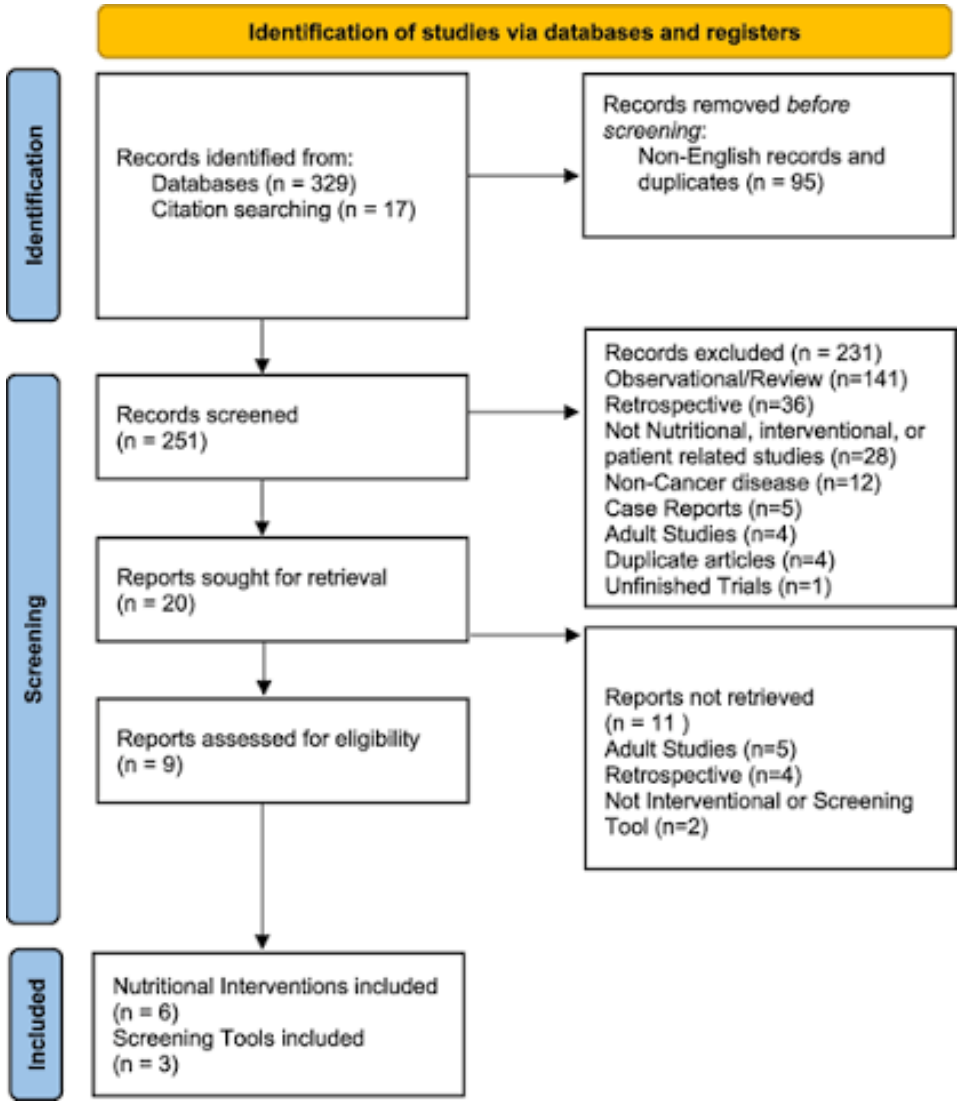
✓ **Item 5**. **Eligibility criteria**. PICO framework and inclusion/exclusion criteria.

✓ **Item 7**. **Search strategy**. Complete search strings and information sources.

✓ **Item 8**. **Selection process**. Independent screening of records and full texts.

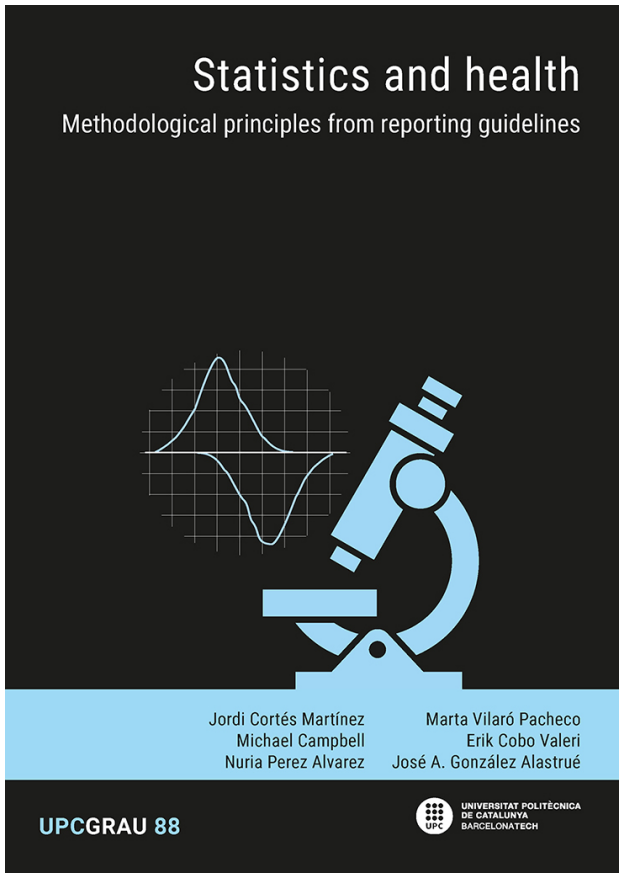
✓ **Item 9**. **Data collection process**. Standardized extraction of study characteristics.

✓ **Item 16**. **PRISMA flow diagram**. Transparent reporting of the study selection process.



PRISMA 2020 flow chart.

Recommended reading



Open-access book

Statistics in Sports Analytics

Available free at:

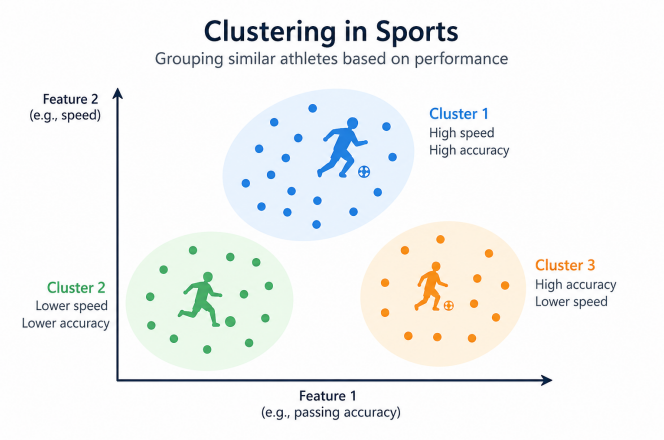
https://upcommons.upc.edu/entities/publication/5dfeccb6-3756-4065-b36c-c79393622c08

Reporting of Clustering Methods (I): Why clustering?

Electronic Journal of Applied Statistical Analysis
Vol. 17, Issue 03, December 2024, 653-675
DOI: 10.1285/i20705948v17n3p653

Reporting of clustering techniques in sports sciences: a scoping review

Daniel Fernández[a], Martí Casals*[b,c], Martí Oliver[a], Montse Plensa[a], and Marica Manisera[d]



Objective	Example
Player profiling	Identify groups of players with similar characteristics
Talent identification	Detect emerging player profiles
Performance analysis	Identify patterns of technical or tactical behaviour
Injury prevention	Group athletes with similar injury risk profiles

Reporting of Clustering Methods (II): Current Challenges

Current situation

No reporting guideline
currently exists for
unsupervised clustering
analyses.

A dedicated guideline,
TRoCA
(*Transparent Reporting of Cluster Analyses*),
is currently under development.

Protocol published in 2025.

Reporting of Clustering Methods (II): Current Challenges

Current situation

No reporting guideline currently exists for unsupervised clustering analyses.

A dedicated guideline, **TRoCA** (*Transparent Reporting of Cluster Analyses*), is currently under development.

Protocol published in 2025.

Minimum reporting checklist

- ✓ Clustering algorithm
- ✓ Variables included (cf. STROBE)
- ✓ Number of clusters and justification
- ✓ Internal/external validation procedure
- ✓ Software and package (cf. SAMPL)
- ✓ Availability of code and data (FAIR / Open Science)

Reporting of Clustering Methods (II): Current Challenges

Current situation

No reporting guideline currently exists for unsupervised clustering analyses.

A dedicated guideline, **TRoCA** (*Transparent Reporting of Cluster Analyses*), is currently under development.

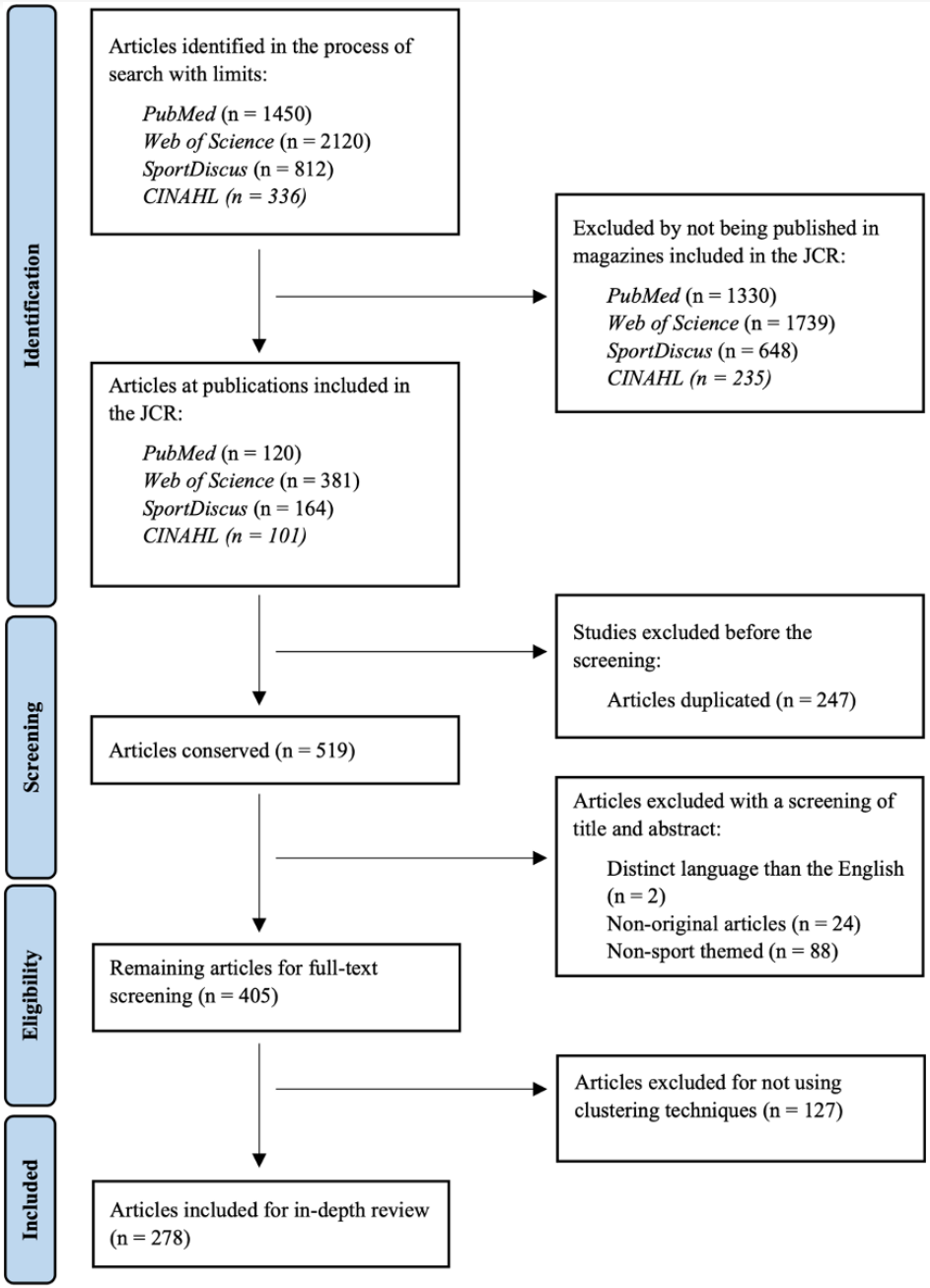
Protocol published in 2025.

Minimum reporting checklist

- ✓ Clustering algorithm
- ✓ Variables included (cf. STROBE)
- ✓ Number of clusters and justification
- ✓ Internal/external validation procedure
- ✓ Software and package (cf. SAMPL)
- ✓ Availability of code and data (FAIR / Open Science)

Until a guideline becomes available, authors should report sufficient methodological details to ensure **transparency** and **reproducibility** of clustering analyses.

Reporting of Clustering Methods (III): Study selection



Review characteristics

- 278 included studies
- Four databases
- PRISMA methodology
- Sports journals indexed in JCR

Reporting of Clustering Methods (IV): Main characteristics

Review

$n = 278$ articles from 1996 to 2023

Study context

Country	
USA	**12.6%**
UK	5.8%
Australia	5.8%
Sex	
Male	37.7%
Female	**7.2%**
Both	43.9%
Participants	
Professional	**47.8%**
Amateur	19.4%

Transparency

× **88.8%**
report the **type** of clustering

× **73.4%**
report the clustering **method**

× **45.0%**
report the method to determine the **number** of clusters

× **0%**
shared the analysis **code**

× **0%**
shared the analysed **dataset**

Reporting of Ordinal Regression Models in Sports Sciences (I): utility

Wiley Interdisciplinary Reviews: Computational Statistics

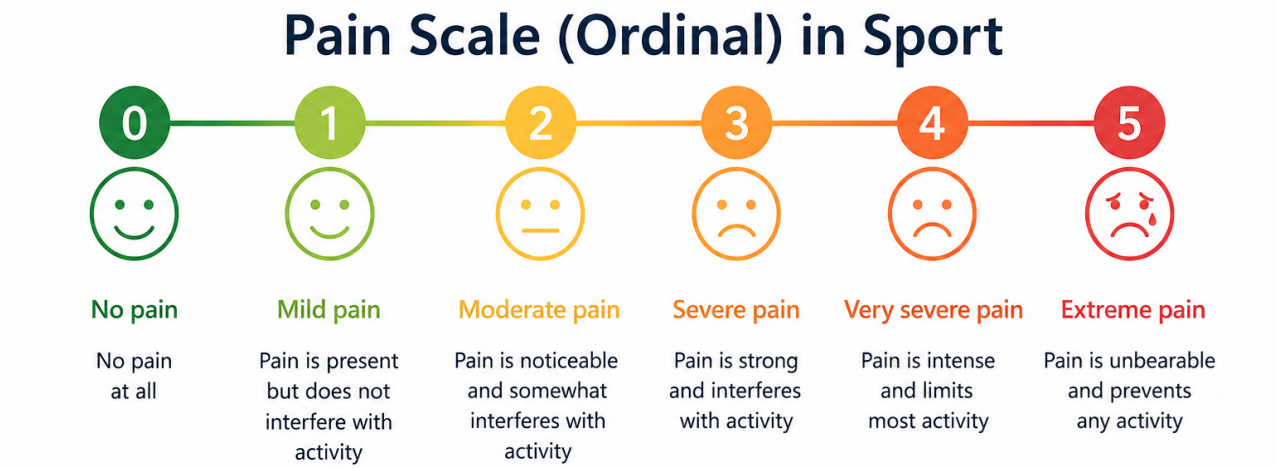
WILEY

SYSTEMATIC REVIEW

WIREs COMPUTATIONAL STATISTICS

Reporting of Regression Models for Ordinal Responses in Sports Sciences Field: A Systematic Review

Daniel Fernández[1] | Mireia Estopañan[1] | Benjamin Baumer[2] | Martí Casals[3,4,5]



Objective	Example
Injury severity	Mild – Moderate – Severe
Perceived exertion	Likert scale (1–10)
Performance	Poor – Fair – Good – Excellent
Competition ranking	1st – 2nd – 3rd – ...

Reporting of Ordinal Regression Models in Sports Sciences (II): Good reporting practices

Reporting guidelines

- **STROBE** is the most appropriate guideline.
- For prediction studies, more insights are provided by **TRIPOD**.
- Beyond, the RGs, **additional statistical information is essential** for ordinal regression analyses.

Reporting of Ordinal Regression Models in Sports Sciences (II): Good reporting practices

Reporting guidelines

- **STROBE** is the most appropriate guideline.
- For prediction studies, more insights are provided by **TRIPOD**.
- Beyond, the RGs, **additional statistical information is essential** for ordinal regression analyses.

STROBE recommendations

- ✓ Clearly define the **outcome**
- ✓ Describe **variables** and measurements
- ✓ Report statistical **methods**
- ✓ Present effect estimates with **precision**
- ✓ Discuss study **limitations**

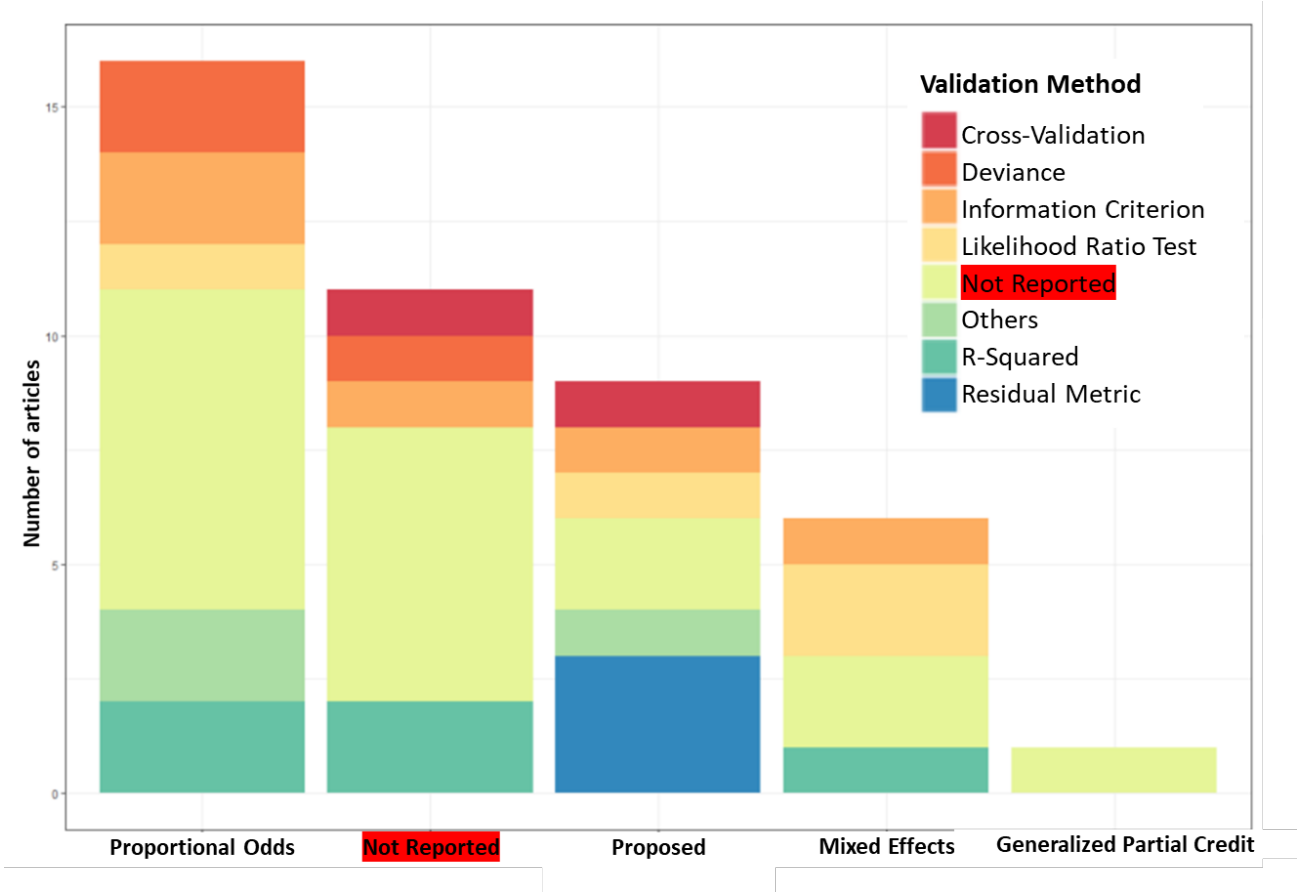
Additional information for ordinal models

- ✓ **Choice of ordinal** regression model
- ✓ Proportional odds **assumption**
- ✓ Model **validation**
- ✓ **Software** and package

Reporting of Ordinal Regression Models (III): Main findings

Review

$n = 34$ articles up to 2023



Transparency

× **71%**
report the ordinal **model** used.

× **47%**
assess the **assumptions**.

× **18%**
share the **data**.

× **75%**
report the **software** used.

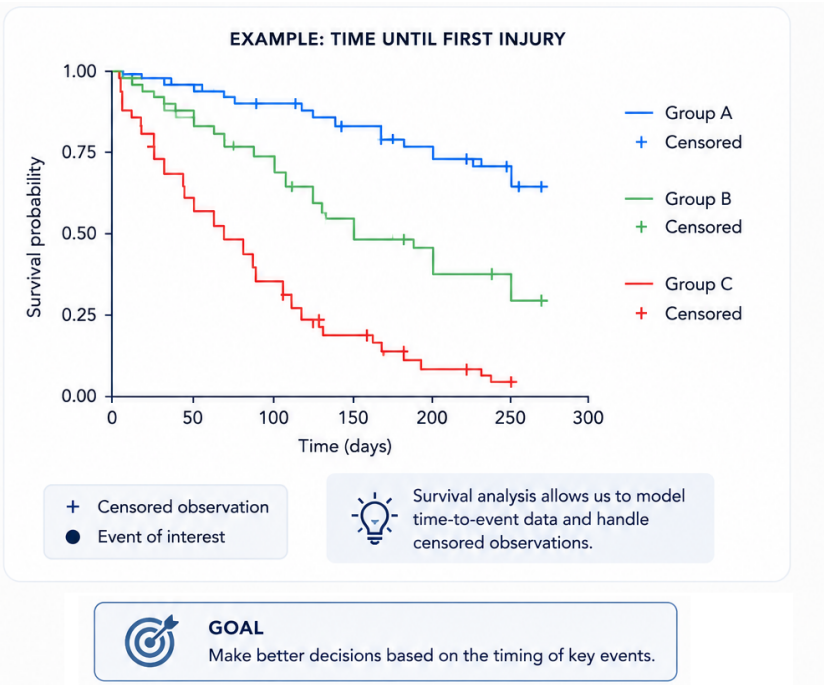
Reporting of Time-to-Event Models in Sports Sciences (I): utility

Open access Review

BMJ Open Sport & Exercise Medicine

Methodological quality and reporting using time-to-event data in sports science and medicine: a scoping review

Jordi Cortés [1], Rasmus Oestergaard Nielsen [2,3], Martí Casals [4,5]



Objective	Example
Injury	Time to first injury or re-injury
Mortality	Survival of former elite athletes
Performance	Time until reaching a performance milestone
Adherence	Time until dropout from an intervention

Reporting of Time-to-Event Models in Sports Sciences (II): SAMPL

Reporting survival (time-to-event) analyses

- Describe the purpose of the analysis.
- Identify the dates or events that mark the beginning and the end of the time period analyzed.
- Specify the circumstances under which data were censored.
- Specify the statistical methods used to estimate the survival rate.
- Confirm that the assumptions of survival analysis were met.
- For each group, give the estimated survival probability at appropriate follow-up times, with confidence intervals, and the number of participants at risk for death at each time. It is often more helpful to plot the cumulative probability of not surviving, especially when events are not common.
- Reporting median survival times, with confidence intervals, is often useful to allow the results to be compared with those of other studies.
- Consider presenting the full results in a graph (e.g., a Kaplan-Meier plot) or table.
- Specify the statistical methods used to compare two or more survival curves.
- When comparing two or more survival curves with hypothesis tests, report the *P* value of the comparison
- Report the regression model used to assess the associations between the explanatory variables and survival or time-to-event.
- Report a measure of risk (e.g., a hazard ratio) for each explanatory variable, with a confidence interval.

Reporting of Time-to-Event Models in Sports Sciences (II): SAMPL

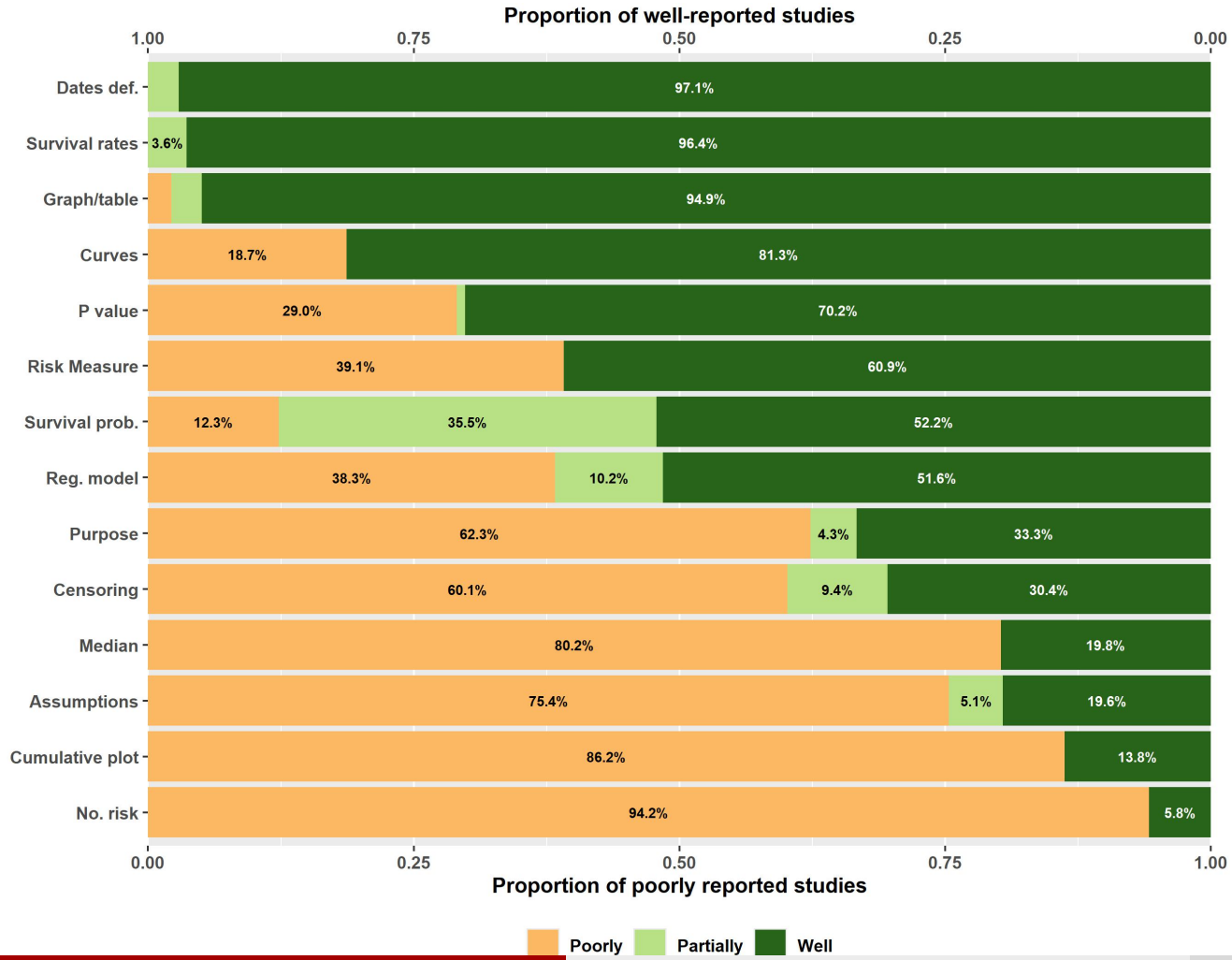
SAMPL recommendations for Time-to-Event analyses

- ✓ Purpose
- ✓ Event dates
- ✓ **Censoring reasons**
- ✓ Survival methods
- ✓ **Assumptions**
- ✓ Survival rates + 95% CI
- ✓ **Participants at risk**
- ✓ Cumulative plots
- ✓ **Median survival + 95% CI**
- ✓ Survival plot
- ✓ Comparison methods
- ✓ P-value comparison
- ✓ Regression model coefficients
- ✓ Risk measures + 95% CI

Reporting of Time-to-Event Models in Sports Sciences (III): Compliance with SAMPL

Scoping review

$n = 138$ articles from 2013 to 2023



Main deficiencies

× **94%**
did not report the **number of participants at risk**

× **80%**
did not report the **median survival time** with 95% CI

× **75%**
did not assess the **proportional hazards assumption**

× **60%**
did not report the **censoring type**

Reporting of GLMMs in Sports Sciences (I): Why GLMMs?

Journal of Sports Analytics
Volume 11, 2025

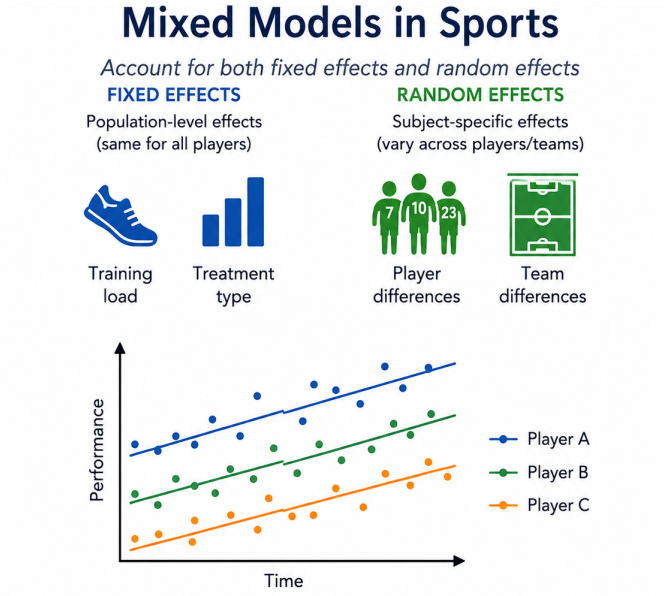
https://doi.org/10.1177/22150218251384557

Sage Journals

Review

Reporting of generalized linear mixed models (GLMM) in sports sciences: A scoping review

Martí Casals [1,2,3], Daniel Fernández [4], Lore Zumeta-Olaskoaga [5], Arnau Sánchez [6], and Paola Zuccolotto [7,8]



Application	Example
Repeated measures	Training load measured throughout a season
Clustered data	Players nested within teams
Match analysis	Goals, passes or tackles per match
Injury studies	Repeated injury events over time

Reporting of GLMMs in Sports Sciences (II): SAMPL

Reporting regression analyses

- Describe the purpose of the analysis.
- Identify the variables used in the analysis and summarize each with descriptive statistics.
- Confirm that the assumptions of the analysis were met. For example, in linear regression indicate whether an analysis of residuals confirmed the assumptions of linearity.
- If relevant, report how any outlying values were treated in the analysis.
- Report how any missing data were treated in the analyses.
- For either simple or multiple (multivariable) regression analyses, report the regression equation.
- For multiple regression analyses: 1) report the alpha level used in the univariate analysis; 2) report whether the variables were assessed for a) colinearity and b) interaction; and 3) describe the variable selection process by which the final model was developed (e.g., forward-stepwise; best subset).
- Report the regression coefficients (beta weights) of each explanatory variable and the associated confidence intervals and *P* values, preferably in a table.
- Provide a measure of the model's "goodness-of-fit" to the data (the coefficient of determination, r^2, for simple regression and the coefficient of multiple determination, R^2, for multiple regression).
- Specify whether and how the model was validated.
- For primary comparisons analyzed with simple linear regression analysis, consider reporting the results graphically, in a scatter plot showing the regression line and its confidence bounds. Do not extend the regression line (or the interpretation of the analysis) beyond the minimum and maximum values of the data.
- Name the statistical package or program used in the analysis.

Reporting of GLMMs in Sports Sciences (II): SAMPL

SAMPL recommendations for GLMM

- ✓ Purpose
- ✓ Variables + EDA
- ✓ Outliers treatment
- ✓ Missing data
- ✓ Model expression
- ✓ Alpha level
- ✓ Collinearity treatment
- ✓ **Variable selection**
- ✓ Regression coefficients + 95
- ✓ Goodness of fit (R^2)
- ✓ **Validation**
- ✓ Plot of the regression effects
- ✓ **Statistical package**

Reporting of GLMMs in Sports Sciences (III): Main findings

Review

55 studies published before March 2023

A. Reporting of methodological aspects of model implementation	
Estimation method	
Gauss-Hermite quadrature (GHQ)	3 (5.5%)
Monte Carlo Markov Chain (MCMC)	1 (1.8%)
Penalized quasi-likelihood (PQL) Laplace approximation	1 (1.8%)
Not reported	50 (90.9%)
Overdispersion evaluation	
Yes	4 (25%)
Not Reported	12 (75%)
Not applicable	39
GLMM Validation	
Yes	5 (9.09%)
Not Reported	50 (90.9%)
B. Transparency in reporting and reproducibility	
Code shared	
No	47 (85.5%)
Yes	8 (14.5%)
Data shared	
No	46 (83.6%)
Yes	9 (16.4%)

Main deficiencies

× **91%**
did not report the **estimation** method.

× **75%**
did not assess **overdispersion**.

× **78.2%**
did not report the **model selection** strategy.

× Only **34.3%**
reported the **statistical package**.

× Only **16%**
shared **data** and/or **code**.

R Packages for Sports Analytics (I): Why Software Matters

Review

International Journal of Sports Science & Coaching

A systematic review of sport-related packages within the R CRAN repository

Martí Casals[1,2,3], José Fernández[4], Victor Martínez[5], Michael Lopez[6], Klaus Langohr[7], and Jordi Cortés[7]

International Journal of Sports Science & Coaching
2023, Vol. 18(2) 621–629

DOI: 10.1177/17479541221136238
journals.sagepub.com/home/spo
SAGE

Why R packages matter

R packages make statistical methods **transparent**, **reproducible**, **reusable**, and easily disseminated across the scientific community.

This review

- Systematic review following the **PRISMA** guideline.
- Review of all sport-related packages available in CRAN.
- Focus on software characteristics and reproducibility.

R Packages for Sports Analytics (II): Reporting Beyond the Manuscript

Reproducibility

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

Transparency

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

R Packages for Sports Analytics (II): Reporting Beyond the Manuscript

Reproducibility

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

Transparency

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

Main findings

2006
First active sports R package **published**

× **17.3%** × **14.8%**
Basketball and **soccer** dominate the ecosystem.

× **56.8%**
do not provide **vignettes**.

× **51.6%**
do not implement **statistical methods**.

R Packages for Sports Analytics (II): Reporting Beyond the Manuscript

Reproducibility

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

Transparency

- ✓ Open-source code
- ✓ CRAN / GitHub repository
- ✓ Version control
- ✓ Example datasets
- ✓ Reproducible examples

Main findings

2006
First active sports R package **published**

× **17.3%** × **14.8%**
Basketball and **soccer** dominate the ecosystem.

× **56.8%**
do not provide **vignettes**.

× **51.6%**
do not implement **statistical methods**.

Good Reporting Example

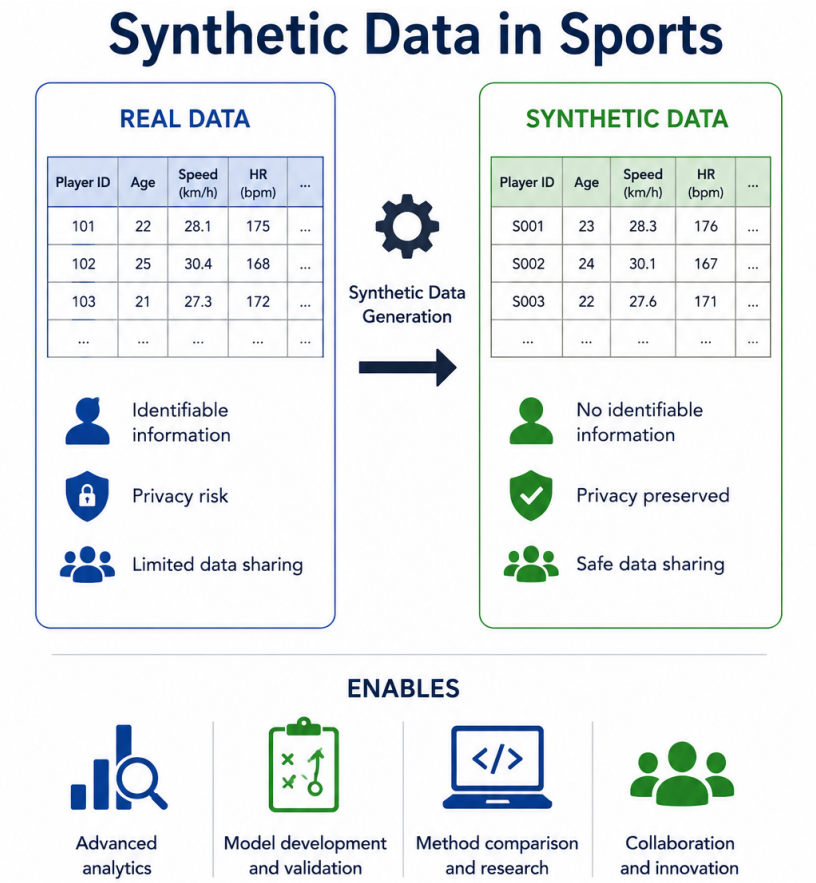
BasketballAnalyzeR (Manisera & Zuccolotto)

Synthetic Data in Sports (I): A New Research Area

Synthetic Data Application on Competitive Sports Context: A Scoping Review

Lander Rodriguez-Idiazabal[a*], Liukuan Yu[b,c], Daniel Fernández[b,c,d], Jordi Cortés Martínez[b,c], Martí Casals[e,f]

Not yet published.



Why synthetic data?

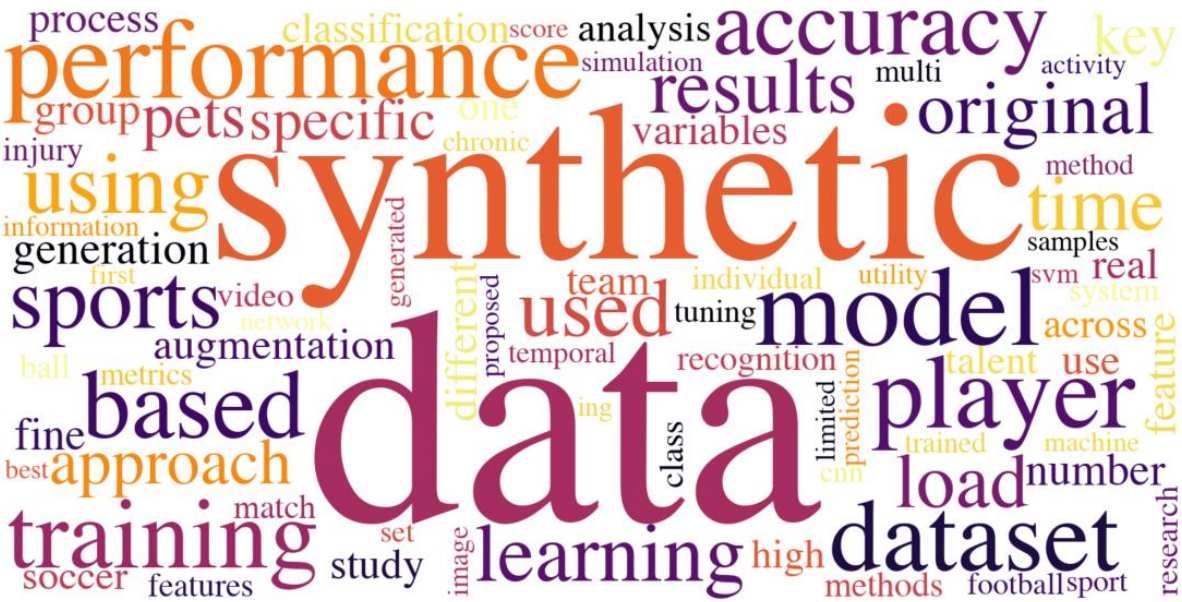
- Privacy-preserving data sharing
- Open science and reproducibility
- Data augmentation
- Benchmarking statistical methods

Synthetic Data in Sports (II): Reporting Synthetic Data Studies

Main focus

- Synthetic data
- Data augmentation
- Machine learning
- Prediction
- Team and player performance

Synthetic Data in Sports (II): Reporting Synthetic Data Studies



Main focus

- Synthetic data
- Data augmentation
- Machine learning
- Prediction
- Team and player performance

Current situation

There are currently

NO reporting guideline

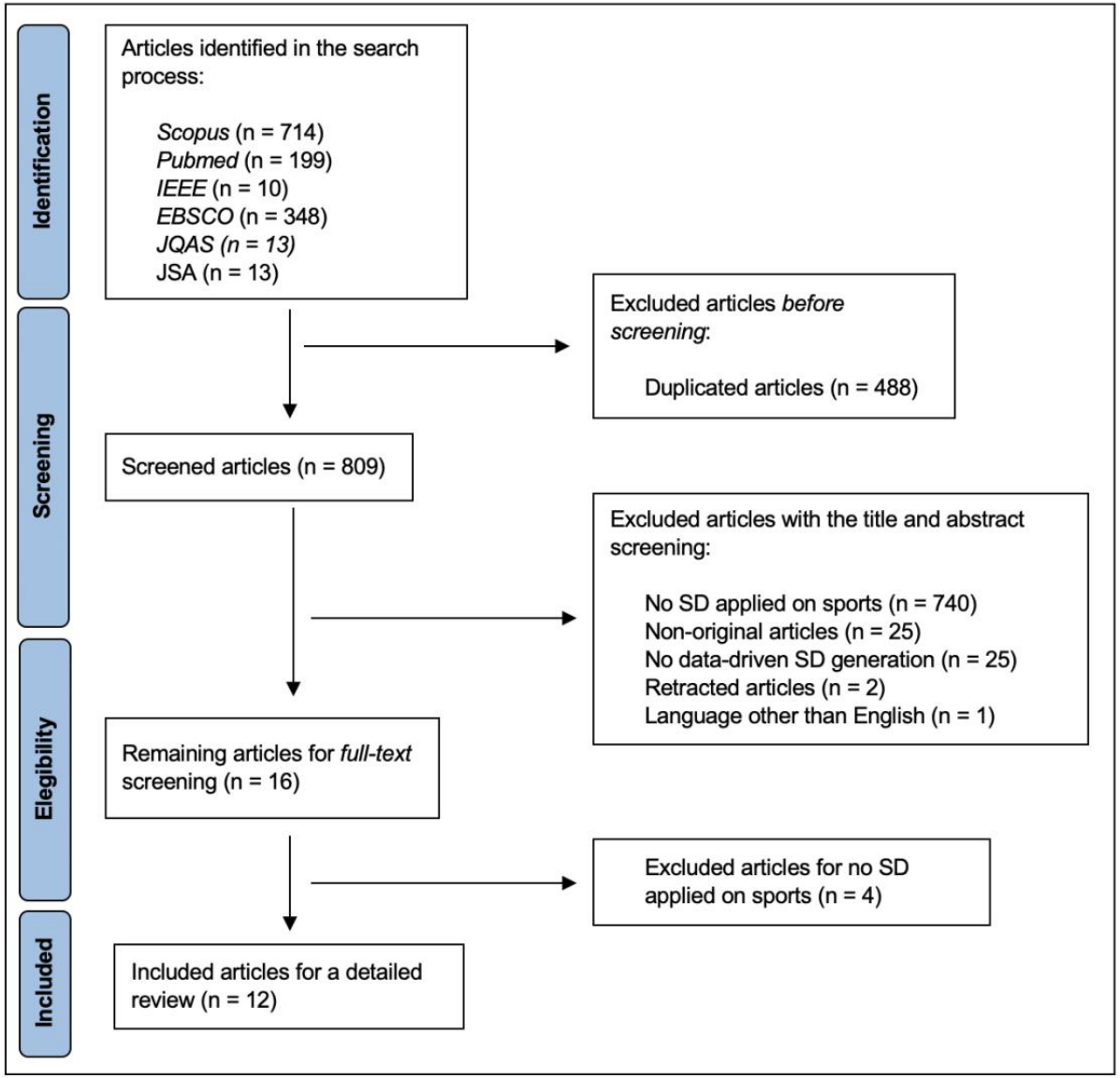
specifically designed for synthetic data studies.

Our proposal

Future studies should always report

✓ Utility
✓ Fidelity
✓ Privacy
✓ Data/Code availability
✓ Ethical considerations

Synthetic Data in Sports (III): Current State



Review

Only **12 studies** published between 2021–2025.

Main findings

× **83%**
data augmentation.

× **17%**
privacy-preserving data sharing adressed.

× **0%**
privacy risk evaluated when applicable.

× **25%**
shared **data**

× **33%**
shared **code**.

Main Conclusions

Take-home messages

✓ Reporting quality in Sports Analytics is still **highly heterogeneous**.

✓ Similar deficiencies were identified across different statistical methods.

✓ Reporting guidelines should be used as **research tools**, not only publication checklists.

✓ Existing guidelines (PRISMA, STROBE, TRIPOD, SAMPL...) are largely applicable to Sports Science.

✓ New reporting guidelines are still needed for several methods (e.g., clustering).

Better reporting $\Longrightarrow$ Better reproducible Sports Analytics

Final thought

"Less research,

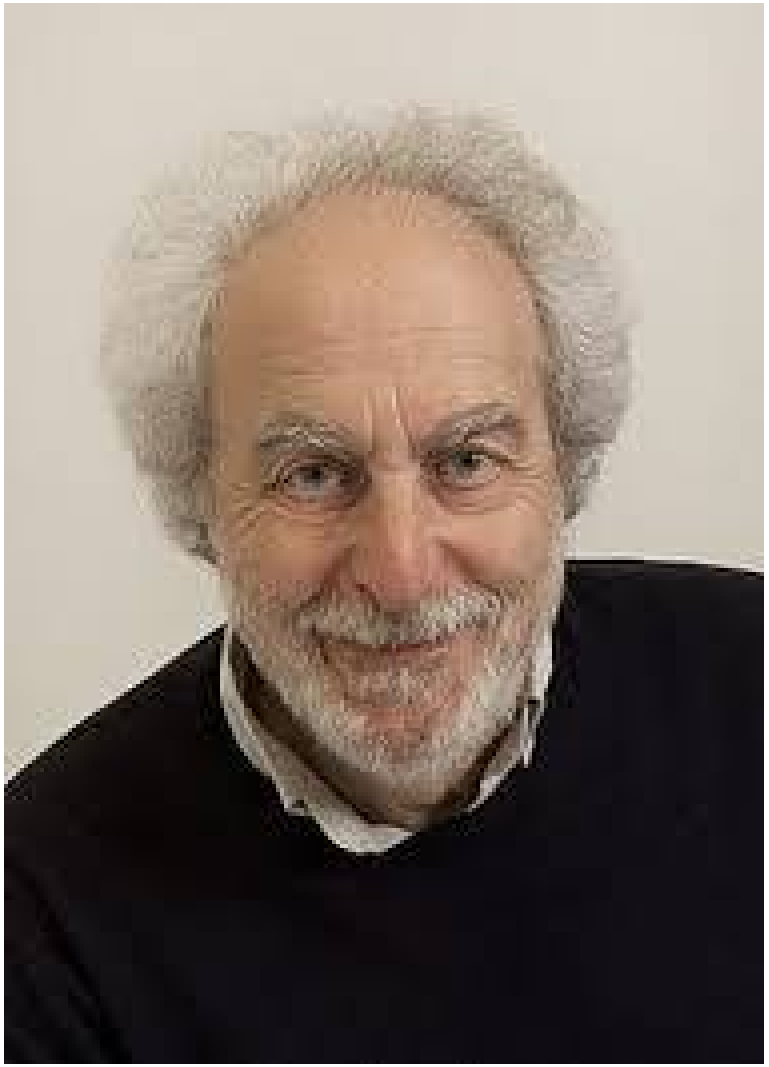
but

***better research**,*
and research done
for the right reasons."

Douglas G. Altman

Our challenge

Better reporting
$\Longrightarrow$ **Better science**
$\Longrightarrow$ **Better Sports Analytics**



Douglas G. Altman (1948–2018)

References

1. Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

2. Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. The Lancet, 374(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

3. Fernández, D., Casals, M., Oliver, M., Plensa, M., & Manisera, M. (2024). Reporting of clustering techniques in sports sciences: A scoping review. Electronic Journal of Applied Statistical Analysis, 17(3), 653–675. https://doi.org/10.1285/i20705948v17n3p653

4. Fernández, D., Estopañan, M., Baumer, B., & Casals, M. (2025). Reporting of regression models for ordinal responses in sports sciences field: A systematic review. WIREs Computational Statistics, 17(1), e70012. https://doi.org/10.1002/wics.70012

5. Cortés, J., Nielsen, R. O., & Casals, M. (2025). Methodological quality and reporting using time-to-event data in sports science and medicine: a scoping review. BMJ Open Sport & Exercise Medicine, 11, e002800. https://doi.org/10.1136/bmjsem-2025-002800

6. Casals, M., Fernández, D., Zumeta-Olaskoaga, L., Sánchez, A., & Zuccolotto, P. (2025). Reporting of generalized linear mixed models (GLMM) in sports sciences: A scoping review. Journal of Sports Analytics. Advance online publication. https://doi.org/10.1177/22150218251384557

7. Casals, M., Fernández, J., Martínez, V., Lopez, M., Langohr, K., & Cortés, J. (2023). A systematic review of sport-related packages within the R CRAN repository. International Journal of Sports Science & Coaching, 18(2), 621–629. https://doi.org/10.1177/17479541221136238

jordi.cortes@upc.edu https://jordicortes.netlify.app/