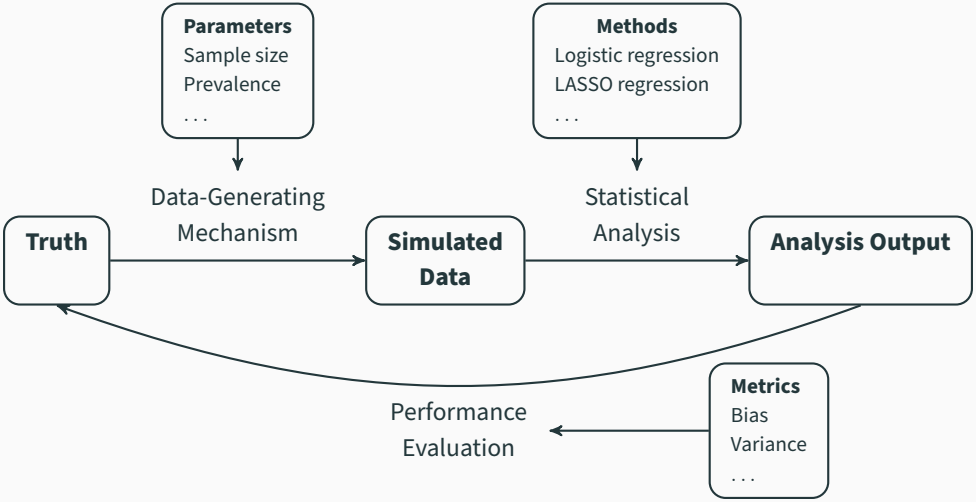
Improving the quality of simulation studies with simulation protocols

Björn S. Siepe, Frantisek Bartos, Tim P. Morris, Anne-Laure Boulesteix, Daniel Heck, **Samuel Pawel**⋆

⋆ Department of Biostatistics, Center for Reproducible Science, University of Zurich

June 10, 2024, Swiss Reproducibility Conference 2024

Simulation studies



Simulation studies can have huge impact

A **simulation study** of the number of events per variable in logistic regression analysis

P **Peduzzi**, J Concato, E Kemper, TR Holford… - Journal of clinical …, 1996 - Elsevier

… In a **simulation study** of forward stepwise multiple linear regression, Freedman and Pee [3] demonstrated that the … In **simulation studies** of the effect of EPV on proportional … Peter **Peduzzi**. …

☆ Save 99 Cite Cited by 8827 Related articles All 9 versions

Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives

L **Hu**, PM **Bentler** - Structural equation modeling: a …, 1999 - Taylor & Francis

This article examines the adequacy of the "rules of thumb" conventional cutoff criteria and several new alternatives for various fit indexes used to evaluate model fit in practice. Using a 2-…

☆ Save 99 Cite Cited by 116305 Related articles All 9 versions

Collinearity: a review of methods to deal with it and a **simulation study** evaluating their performance

CF Dormann, J Elith, S Bacher, C Buchmann… - …, 2013 - Wiley Online Library

… In the fourth part we carry out a large **simulation study** to compare all reviewed methods. We provide complementary case studies on real data in Supplementary material Appendix 1.2. …

☆ Save 99 Cite Cited by 8455 Related articles All 30 versions

Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo **simulation study**

KL Nylund, T Asparouhov… - … equation modeling: A …, 2007 - Taylor & Francis

… This article presents the results of a **simulation study** that examines the performance of likelihood-based tests and the traditionally used Information Criterion (ICs) used for determining …

☆ Save 99 Cite Cited by 10707 Related articles All 14 versions

Evidence from simulation studies

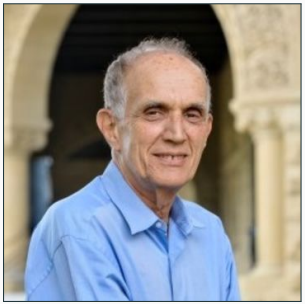
> *"…extensive simulation studies show that the proposed method performs on par or* ***better than existing methods*** *…"*

- Trustworthy?
- Replicable?
- Reproducible?
- Robust?

Evidence from simulation studies

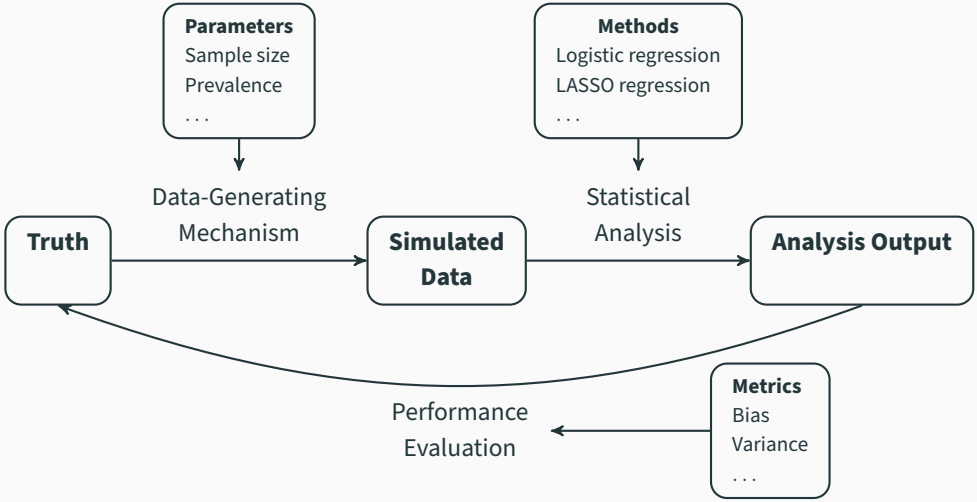
"*In fact it is* ***very difficult to run an honest simulation*** *comparison, and* ***easy to inadvertently cheat*** *by choosing favorable examples, or by not putting as much effort into optimizing the dull old standard as the exciting new challenger.*"

Brad Efron (2001)

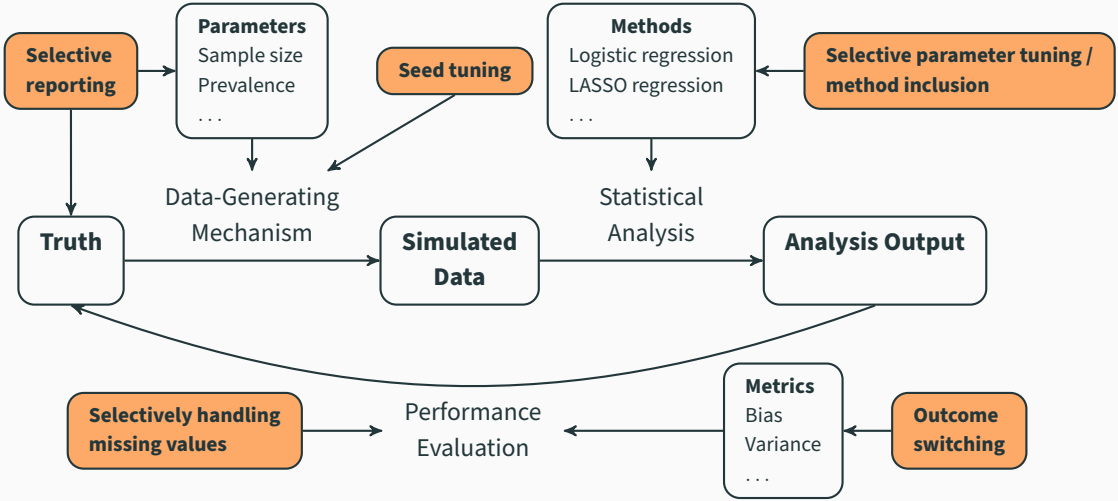


https://statistics.stanford.edu/people/bradley-efron

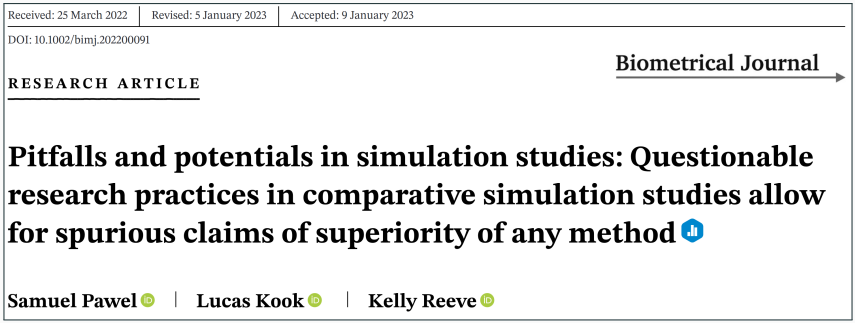
Questionable research practices in simulation studies



Questionable research practices in simulation studies



Questionable research practices in simulation studies



Received: 25 March 2022 | Revised: 5 January 2023 | Accepted: 9 January 2023

DOI: 10.1002/bimj.202200091

Biometrical Journal

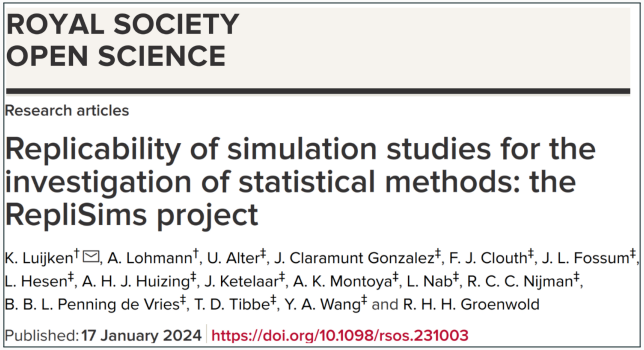
RESEARCH ARTICLE

Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method

Samuel Pawel | Lucas Kook | Kelly Reeve

"*By **deliberately using several QRPs**, we were able to **present a method with no expected benefits** […] **as an improvement** over […] well-established competitors.*"

Replicability of simulation studies



ROYAL SOCIETY OPEN SCIENCE

Research articles

Replicability of simulation studies for the investigation of statistical methods: the RepliSims project

K. Luijken[†], A. Lohmann[†], U. Alter[‡], J. Claramunt Gonzalez[‡], F. J. Clouth[‡], J. L. Fossum[‡], L. Hesen[‡], A. H. J. Huizing[‡], J. Ketelaar[‡], A. K. Montoya[‡], L. Nab[‡], R. C. C. Nijman[‡], B. B. L. Penning de Vries[‡], T. D. Tibbe[‡], Y. A. Wang[‡] and R. H. H. Groenwold

Published: 17 January 2024 | https://doi.org/10.1098/rsos.231003

"*the information provided in the original publication of highly cited and influential simulation studies was* ***often insufficient for complete replication***"

Simulation study protocols

STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:4279–4292
Published online 31 August 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2673

The design of simulation studies in medical statistics

Andrea Burton[1, 2, *, †], Douglas G. Altman[1], Patrick Royston[1,3] and Roger L. Holder[4]

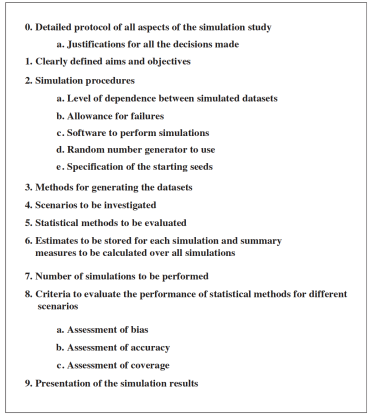
"***When planning*** *a simulation study, it is* ***recommended that a detailed protocol be produced****, giving full details of how the study will be performed, analysed and reported.*"

Simulation study protocols

Advantages

+ Planning and reporting
+ Transparency and replicability
+ Can be preregistered
? Less/more work

→ **How to structure protocol?**



0. Detailed protocol of all aspects of the simulation study
    a. Justifications for all the decisions made
1. Clearly defined aims and objectives
2. Simulation procedures
    a. Level of dependence between simulated datasets
    b. Allowance for failures
    c. Software to perform simulations
    d. Random number generator to use
    e. Specification of the starting seeds
3. Methods for generating the datasets
4. Scenarios to be investigated
5. Statistical methods to be evaluated
6. Estimates to be stored for each simulation and summary measures to be calculated over all simulations
7. Number of simulations to be performed
8. Criteria to evaluate the performance of statistical methods for different scenarios
    a. Assessment of bias
    b. Assessment of accuracy
    c. Assessment of coverage
9. Presentation of the simulation results

Proposal from Burton et al. (2006)

The ADEMP-PreReg template

ADEMP-PreReg
Template for Simulation Studies

Version: 0.1.0
Last updated: 2023-10-31
Preregistration template designed by
Björn S. Siepe, František Bartoš, Tim P. Morris, Anne-Laure Boulesteix, Daniel W. Heck, and Samuel Pawel

- Protocol template based on **ADEMP structure** (Morris et al., 2019) + **open science** + **reproducibility** aspects
- **Different versions**: LaTeX, Overleaf, MS/Libre office, Google docs
- **Living document**: https://github.com/bsiepe/ADEMP-PreReg

ADEMP: doi.org/10.1214/ss/1009213726, ADEMP-PreReg: doi.org/10.31234/osf.io/ufgy6

The ADEMP-PreReg template

1. Instructions
2. General information
3. **A**ims
4. **D**ata-generating mechanism
5. **E**stimands and targets
6. **M**ethods
7. **P**erformance Measures
8. Computational details

7 Performance Measures

7.1 Which performance measures will be used?

Explanation: Please provide details on why they were chosen and on how these measures will be calculated. Ideally, provide formulas for the performance measures to avoid ambiguity. Some models in psychology, such as item response theory or time series models, often contain multiple parameters of interest, and their number may vary across conditions. With a large number of estimated parameters, their performance measures are often combined. If multiple estimates are aggregated, specify how this aggregation will be performed. For example, if there are multiple parameters in a particular condition, the mean of the individual biases of these parameters or the bias of each individual parameter may be reported.

Example

Our primary performance measures are the type I error rate (in conditions where the true effect is zero) and the power (in conditions where the true effect is non-zero) to reject the null hypothesis of no difference between the control and treatment condition. The null hypothesis is rejected if the p-value for the null hypothesis of no effect is less than or equal to the conventional threshold of 0.05. The rejection rate (the type I error rate or the power, depending on the data generating mechanism) is estimated by

$$\widehat{\text{RRate}} = \frac{\sum_{i=1}^{n_{\text{sim}}} 1(p_i \leq 0.05)}{n_{\text{sim}}}$$

where $1(p_i \leq 0.05)$ is the indicator of whether the p-value in simulation i is equal to or less than 0.05. We use the following formula to compute the MCSE of the rejection rate

$$\text{MCSE}_{\widehat{\text{RRate}}} = \sqrt{\frac{\widehat{\text{RRate}}(1 - \widehat{\text{RRate}})}{n_{\text{sim}}}}.$$

The ADEMP-PreReg template

Purposes

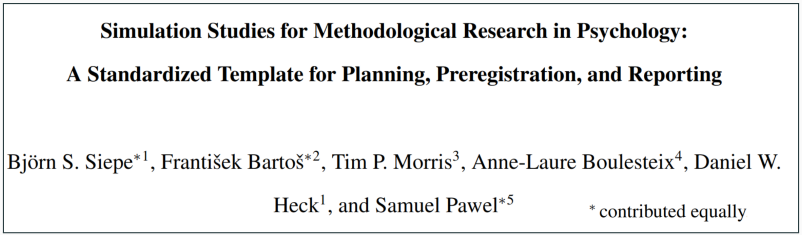
- **Planning** of simulation studies
- **Preregistration**
- **Blueprint** for reporting
- **Reviewing** of simulation studies

Limitations

- Preregistration could be **faked**
- May **slow down** exploratory research

doi:10.5281/zenodo.7994221

Conclusions



Simulation Studies for Methodological Research in Psychology:

A Standardized Template for Planning, Preregistration, and Reporting

Björn S. Siepe*[1], František Bartoš*[2], Tim P. Morris[3], Anne-Laure Boulesteix[4], Daniel W. Heck[1], and Samuel Pawel*[5]

* contributed equally

- Simulation studies can have **big impact**, should be **conducted carefully**
- **Protocols** can make simulation studies **more reliable**
- **ADEMP-PreReg template** helps in preregistration, planning, reporting, reviewing of simulation studies

doi.org/10.31234/osf.io/ufgy6

References i

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231. doi:10.1214/ss/1009213726.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292. doi:10.1002/sim.2673.

Luijken, K., Lohmann, A., Alter, U., Gonzalez, J. C., Clouth, F. J., Fossum, J. L., Hesen, L., Huizing, A. H. J., Ketelaar, J., Montoya, A. K., Nab, L., Nijman, R. C. C., de Vries, B. B. L. P., Tibbe, T. D., Wang, Y. A., and Groenwold, R. H. H. (2023). Replicability of simulation studies for the investigation of statistical methods: The replisims project. doi:10.48550/ARXIV.2307.02052. arXiv preprint.

Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102. doi:10.1002/sim.8086.

Pawel, S., Kook, L., and Reeve, K. (2023). Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biometrical Journal*, (e2200091):1–19. doi:10.1002/bimj.202200091.

Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., and Pawel, S. (2023). Simulation studies for methodological research in psychology: A standardized template for planning, preregistration, and reporting. doi:10.31234/osf.io/ufgy6.