Marburg University

PMλ.

Simulation studies in psychology: Status quo, problems, and potential solutions

MSMR · August 2026

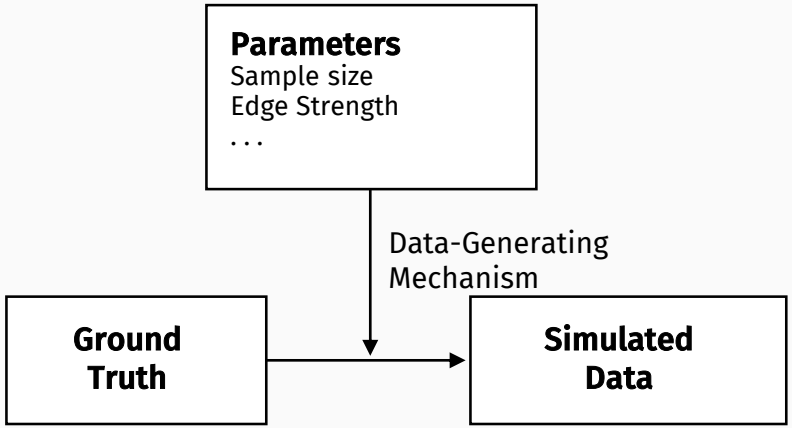
Björn Siepe

Psychological Methods Lab, Department of Psychology, Marburg University

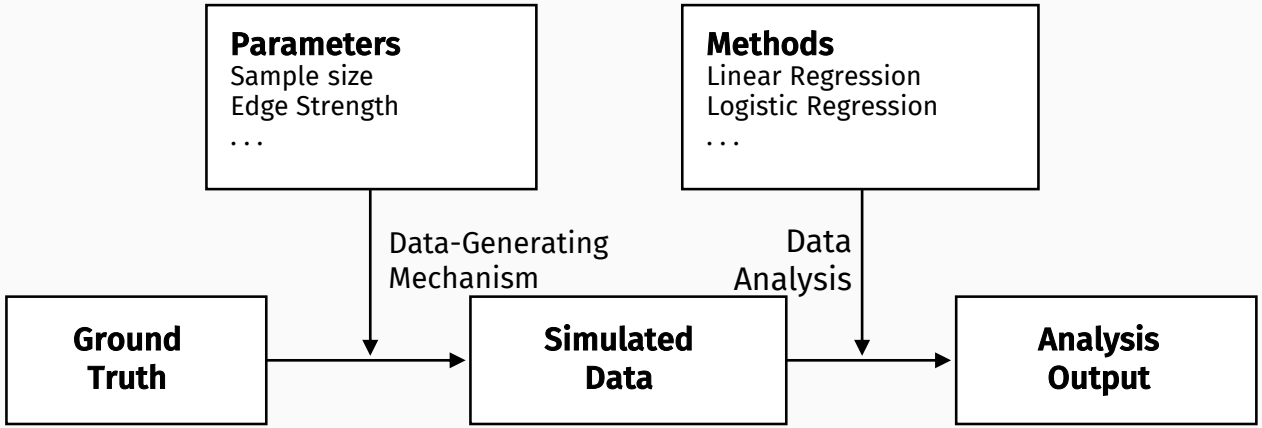
Part I: Background and Issues

Simulation studies

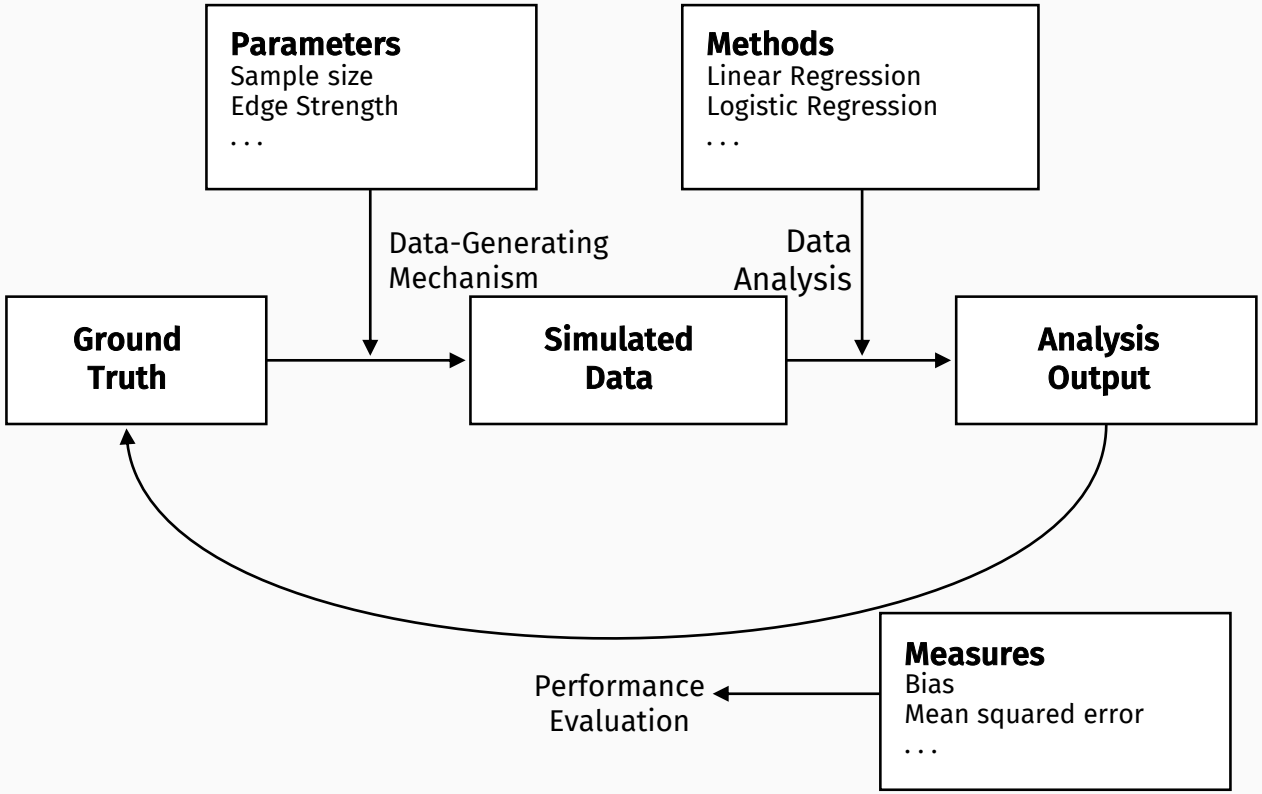
Simulation studies



Simulation studies



Simulation studies



Issues in simulation studies

Issues in simulation studies

Marcin Wichary - Wikimedia

Issues in simulation studies

Marcin Wichary - Wikimedia

> **Statisticians ... often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

Meta-Science on Simulation Studies

Meta-Science on Simulation Studies

Studied across **(bio)statistics, ecology, and psychometrics**

e.g., Koehler et al., 2009; Morris et al., 2019; Williams et al., 2024; Feinberg & Rubright, 2016; Bouma et al., 2026

Meta-Science on Simulation Studies

Studied across **(bio)statistics, ecology, and psychometrics**

e.g., Koehler et al., 2009; Morris et al., 2019; Williams et al., 2024; Feinberg & Rubright, 2016; Bouma et al., 2026

Flexibility

Many degrees of freedom

Pawel et al., 2024

Over-optimism towards new methods

Ullmann et al., 2023

Meta-Science on Simulation Studies

Studied across **(bio)statistics, ecology, and psychometrics**

e.g., Koehler et al., 2009; Morris et al., 2019; Williams et al., 2024; Feinberg & Rubright, 2016; Bouma et al., 2026

Flexibility

Many degrees of freedom

Pawel et al., 2024

Over-optimism towards new methods

Ullmann et al., 2023

Opacity

Insufficient reporting

Hoaglin & Andrews, 1975

Monte Carlo error often ignored

Koehler et al., 2009

Meta-Science on Simulation Studies

Studied across **(bio)statistics, ecology, and psychometrics**

e.g., Koehler et al., 2009; Morris et al., 2019; Williams et al., 2024; Feinberg & Rubright, 2016; Bouma et al., 2026

Flexibility

Many degrees of freedom

Pawel et al., 2024

Over-optimism towards new methods

Ullmann et al., 2023

Opacity

Insufficient reporting

Hoaglin & Andrews, 1975

Monte Carlo error often ignored

Koehler et al., 2009

Complexity

Reproducibility failures

Luijken et al., 2024

Unclear non-convergence handling

Pawel et al., 2026; Wünsch et al., 2025

Meta-Science on Simulation Studies

Studied across **(bio)statistics, ecology, and psychometrics**

e.g., Koehler et al., 2009; Morris et al., 2019; Williams et al., 2024; Feinberg & Rubright, 2016; Bouma et al., 2026

Flexibility

Many degrees of freedom

Pawel et al., 2024

Over-optimism towards new methods

Ullmann et al., 2023

Opacity

Insufficient reporting

Hoaglin & Andrews, 1975

Monte Carlo error often ignored

Koehler et al., 2009

Complexity

Reproducibility failures

Luijken et al., 2024

Unclear non-convergence handling

Pawel et al., 2026; Wünsch et al., 2025

Overall: Issues similar to other empirical research

Boulesteix et al., 2020

Simulation studies in psychology

Simulation studies in psychology

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

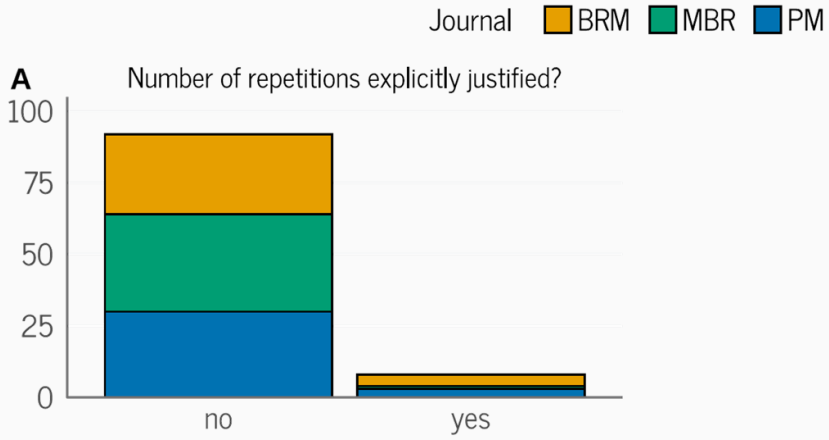
https://doi.org/10.1037/met0000695

Simulation Studies for Methodological Research in Psychology: A Standardized Template for Planning, Preregistration, and Reporting

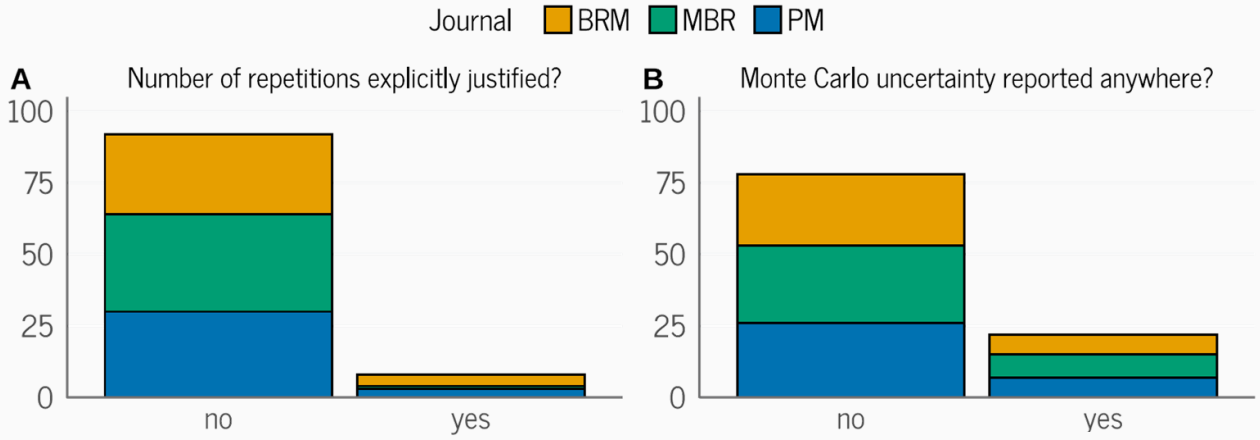
Björn S. Siepe[1], František Bartoš[2], Tim P. Morris[3], Anne-Laure Boulesteix[4, 5], Daniel W. Heck[1], and Samuel Pawel[6, 7]

Simulation studies in psychology

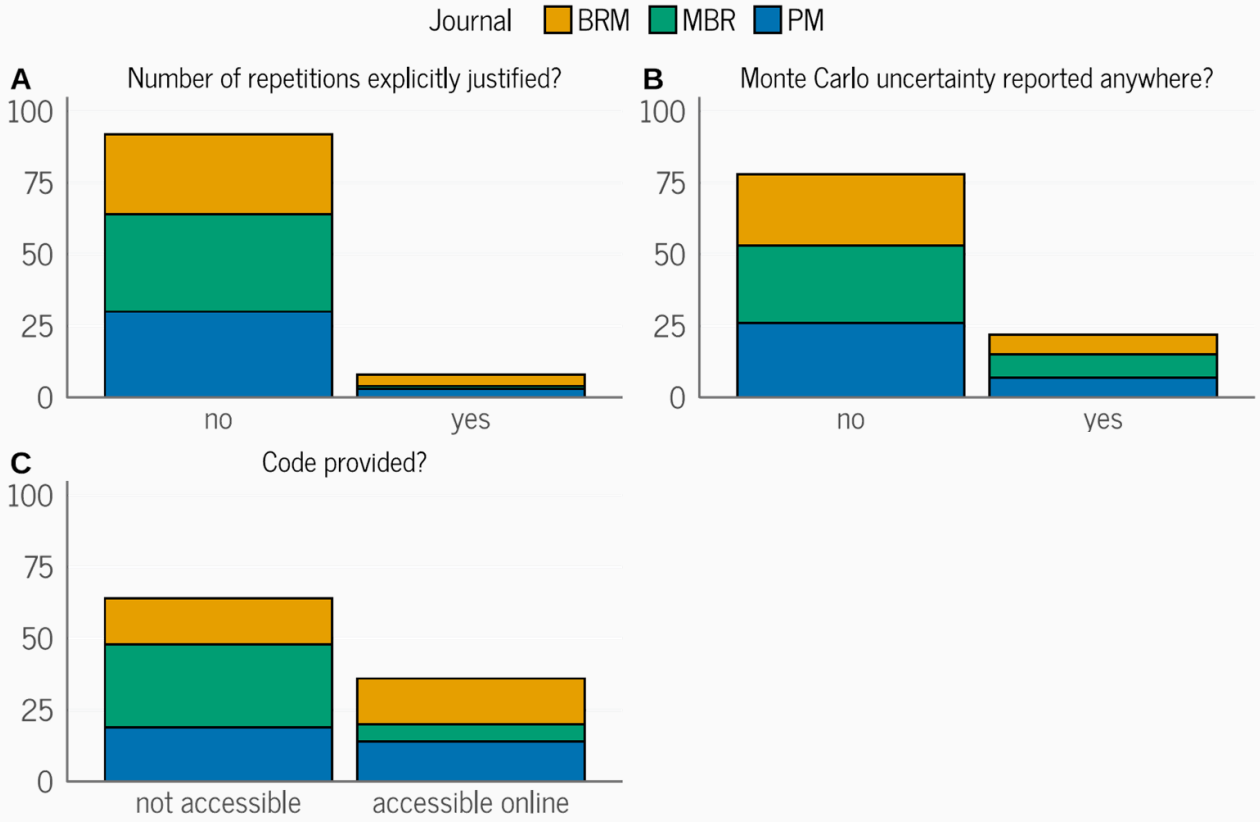
Simulation studies in psychology



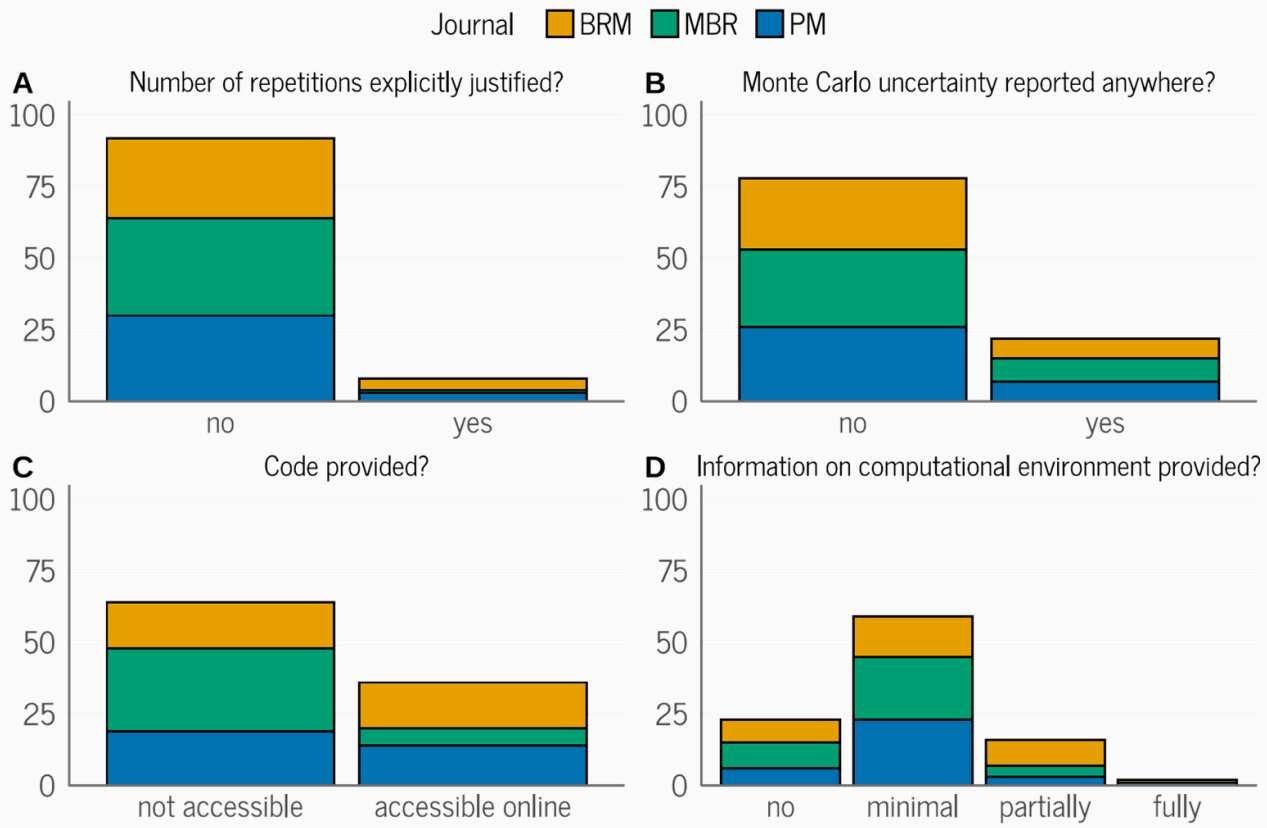
Simulation studies in psychology



Simulation studies in psychology



Simulation studies in psychology



Improving reporting (based on Morris et al., 2019)

Improving reporting (based on Morris et al., 2019)

Table 3

Definitions of Common PerformanceMeasures, Their Estimates, MCSEs, and Number of Simulation Repetitions n_{sim} to Achieve a Desired MCSE$_*$ (Morris et al., 2019, Table A1)

Performance measure	Definition	Estimate	MCSE	n_{sim}
Bias	$\mathrm{E}(\hat{\theta}) - \theta$	$(\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i / n_{\text{sim}}) - \theta$	$\sqrt{S_{\hat{\theta}}^2 / n_{\text{sim}}}$	$S_{\hat{\theta}}^2 / \text{MCSE}_*^2$
Relative bias	$\{\mathrm{E}(\hat{\theta}) - \theta\}/\theta$	$\{(\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i / n_{\text{sim}}) - \theta\}/\theta$	$\sqrt{S_{\hat{\theta}}^2 / (\theta^2 n_{\text{sim}})}$	$S_{\hat{\theta}}^2 / (\text{MCSE}_*^2 \theta^2)$
MSE	$\mathrm{E}\{(\hat{\theta} - \theta)^2\}$	$\sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)^2 / n_{\text{sim}}$	$\sqrt{S_{(\hat{\theta}-\theta)^2}^2 / n_{\text{sim}}}$	$S_{(\hat{\theta}-\theta)^2}^2 / \text{MCSE}_*^2$
RMSE	$\sqrt{\mathrm{E}\{(\hat{\theta} - \theta)^2\}}$	$\sqrt{\sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)^2 / n_{\text{sim}}}$	$\sqrt{S_{(\hat{\theta}-\theta)^2}^2 / (4 n_{\text{sim}} \widehat{\text{MSE}})}$	$S_{(\hat{\theta}-\theta)^2}^2 / (4 \widehat{\text{MSE}}\, \text{MCSE}_*^2)$
Empirical variance	$\mathrm{Var}(\hat{\theta})$	$S_{\hat{\theta}}^2$	$S_{\hat{\theta}}^2 \sqrt{2/(n_{\text{sim}} - 1)}$	$1 + 2(S_{\hat{\theta}}^2)^2 / \text{MCSE}_*^2$
Empirical standard error	$\sqrt{\mathrm{Var}(\hat{\theta})}$	$\sqrt{S_{\hat{\theta}}^2}$	$\sqrt{S_{\hat{\theta}}^2 / \{2(n_{\text{sim}} - 1)\}}$	$1 + S_{\hat{\theta}}^2 / (2\text{MCSE}_*^2)$
Coverage	Pr (CI includes θ)	$\sum_{i=1}^{n_{\text{sim}}} \mathbb{1}(\text{CI}_i \text{ includes } \theta)/n_{\text{sim}}$	$\sqrt{\widehat{\text{Cov}}(1 - \widehat{\text{Cov}})/n_{\text{sim}}}$	$\widehat{\text{Cov}}(1 - \widehat{\text{Cov}})/\text{MCSE}_*^2$
Power (or Type I error rate)	Pr (Test rejects H_0)	$\sum_{i=1}^{n_{\text{sim}}} \mathbb{1}(\text{Test}_i \text{ rejects } H_0)/n_{\text{sim}}$	$\sqrt{\widehat{\text{Pow}}(1 - \widehat{\text{Pow}})/n_{\text{sim}}}$	$\widehat{\text{Pow}}(1 - \widehat{\text{Pow}})/\text{MCSE}_*^2$
Mean CI width	$\mathrm{E}(\text{CI}_{\text{upper}} - \text{CI}_{\text{lower}})$	$\sum_{i=1}^{n_{\text{sim}}} (\text{CI}_{i,\text{upper}} - \text{CI}_{i,\text{lower}})/n_{\text{sim}}$	$\sqrt{S_W^2 / n_{\text{sim}}}$	S_W^2 / MCSE_*^2
Mean of generic statistic G	$\mathrm{E}(G)$	$\sum_{i=1}^{n_{\text{sim}}} G_i / n_{\text{sim}}$	$\sqrt{S_G^2 / n_{\text{sim}}}$	S_G^2 / MCSE_*^2

Improving reporting (based on Morris et al., 2019)

Table 3

Definitions of Common PerformanceMeasures, Their Estimates, MCSEs, and Number of Simulation Repetitions n_{sim} to Achieve a Desired MCSE$_*$ (Morris et al., 2019, Table A1)

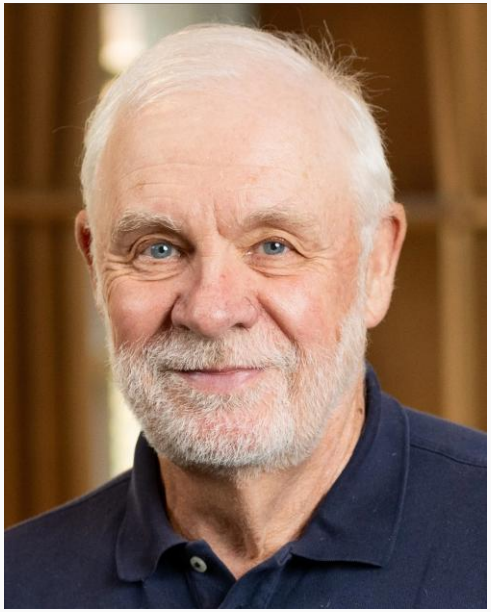
Performance measure	Definition	Estimate	MCSE	n_{sim}
Bias	$\mathrm{E}(\hat{\theta}) - \theta$	$(\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i / n_{\text{sim}}) - \theta$	$\sqrt{S_{\hat{\theta}}^2 / n_{\text{sim}}}$	$S_{\hat{\theta}}^2 / \text{MCSE}_*^2$
Relative bias	$\{\mathrm{E}(\hat{\theta}) - \theta\}/\theta$	$\{(\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i / n_{\text{sim}}) - \theta\}/\theta$	$\sqrt{S_{\hat{\theta}}^2 / (\theta^2 n_{\text{sim}})}$	$S_{\hat{\theta}}^2 / (\text{MCSE}_*^2 \theta^2)$
MSE	$\mathrm{E}\{(\hat{\theta} - \theta)^2\}$	$\sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)^2 / n_{\text{sim}}$	$\sqrt{S_{(\hat{\theta}-\theta)^2}^2 / n_{\text{sim}}}$	$S_{(\hat{\theta}-\theta)^2}^2 / \text{MCSE}_*^2$
RMSE	$\sqrt{\mathrm{E}\{(\hat{\theta} - \theta)^2\}}$	$\sqrt{\sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)^2 / n_{\text{sim}}}$	$\sqrt{S_{(\hat{\theta}-\theta)^2}^2 / (4 n_{\text{sim}} \widehat{\text{MSE}})}$	$S_{(\hat{\theta}-\theta)^2}^2 / (4 \widehat{\text{MSE}}\, \text{MCSE}_*^2)$
Empirical variance	$\mathrm{Var}(\hat{\theta})$	$S_{\hat{\theta}}^2$	$S_{\hat{\theta}}^2 \sqrt{2/(n_{\text{sim}} - 1)}$	$1 + 2(S_{\hat{\theta}}^2)^2 / \text{MCSE}_*^2$
Empirical standard error	$\sqrt{\mathrm{Var}(\hat{\theta})}$	$\sqrt{S_{\hat{\theta}}^2}$	$\sqrt{S_{\hat{\theta}}^2 / \{2(n_{\text{sim}} - 1)\}}$	$1 + S_{\hat{\theta}}^2 / (2\text{MCSE}_*^2)$
Coverage	Pr (CI includes θ)	$\sum_{i=1}^{n_{\text{sim}}} \mathbb{1}(\text{CI}_i \text{ includes } \theta)/n_{\text{sim}}$	$\sqrt{\widehat{\text{Cov}}(1 - \widehat{\text{Cov}})/n_{\text{sim}}}$	$\widehat{\text{Cov}}(1 - \widehat{\text{Cov}})/\text{MCSE}_*^2$
Power (or Type I error rate)	Pr (Test rejects H_0)	$\sum_{i=1}^{n_{\text{sim}}} \mathbb{1}(\text{Test}_i \text{ rejects } H_0)/n_{\text{sim}}$	$\sqrt{\widehat{\text{Pow}}(1 - \widehat{\text{Pow}})/n_{\text{sim}}}$	$\widehat{\text{Pow}}(1 - \widehat{\text{Pow}})/\text{MCSE}_*^2$
Mean CI width	$\mathrm{E}(\text{CI}_{\text{upper}} - \text{CI}_{\text{lower}})$	$\sum_{i=1}^{n_{\text{sim}}} (\text{CI}_{i,\text{upper}} - \text{CI}_{i,\text{lower}})/n_{\text{sim}}$	$\sqrt{S_W^2 / n_{\text{sim}}}$	S_W^2 / MCSE_*^2
Mean of generic statistic G	$\mathrm{E}(G)$	$\sum_{i=1}^{n_{\text{sim}}} G_i / n_{\text{sim}}$	$\sqrt{S_G^2 / n_{\text{sim}}}$	S_G^2 / MCSE_*^2

ADEMP-PreReg
Template for Simulation Studies

November 27, 2024

Deeper issues in simulation studies

Deeper issues in simulation studies



https://src.isr.umich.edu/research/faculty-profiles/profiles/littleroderick-j/

> Good simulation studies are not given the respect they deserve. Often the **design is perfunctory and simplistic** [...] and results are over-generalized.
>
> Roderick Little, 2013

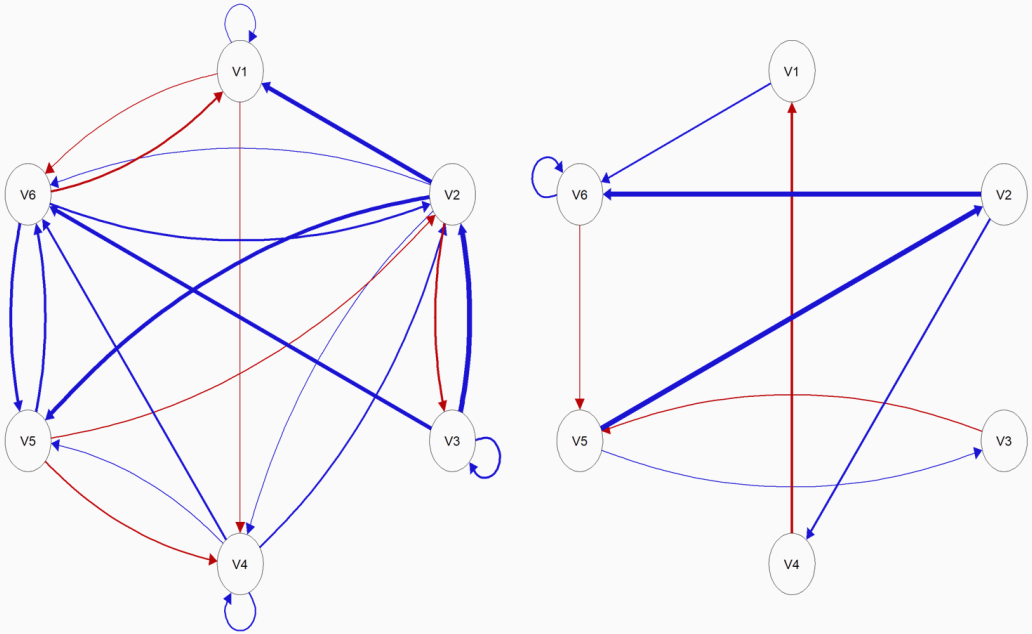
Few constraints on plausible DGMs

Few constraints on plausible DGMs

What is a useful/realistic data-generating mechanism?

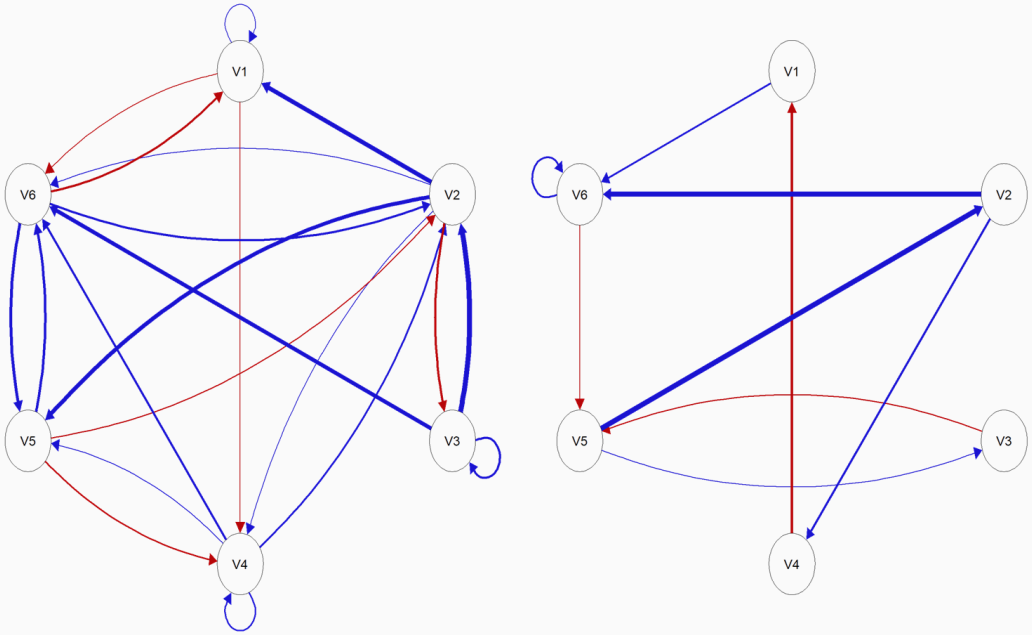
Few constraints on plausible DGMs

What is a useful/realistic data-generating mechanism?



Few constraints on plausible DGMs

What is a useful/realistic data-generating mechanism?

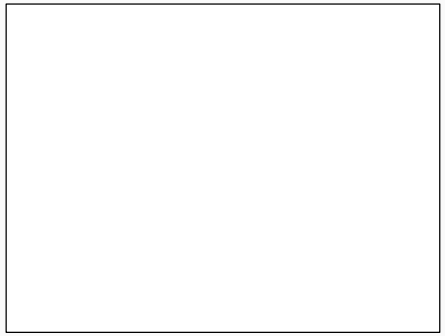


→ Hard to understand when performance differences actually matter

Part II: Solutions?

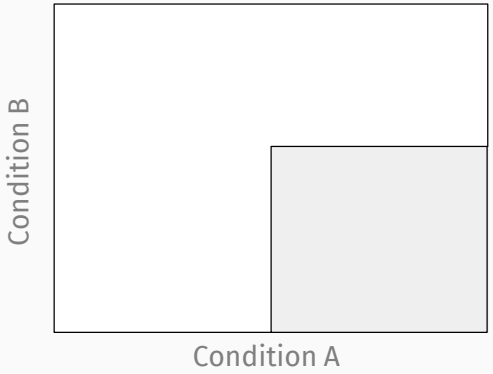
Exploring the data-generating space

Exploring the data-generating space

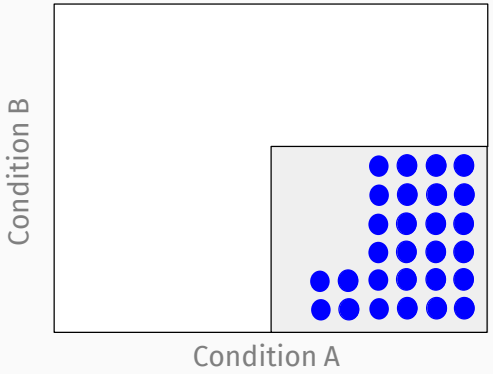


Exploring the data-generating space

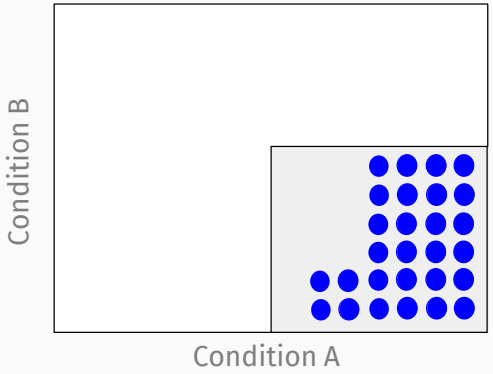
Exploring the data-generating space



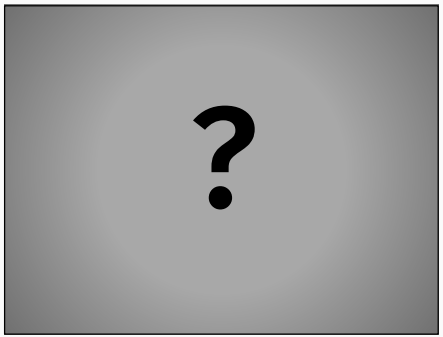
Exploring the data-generating space



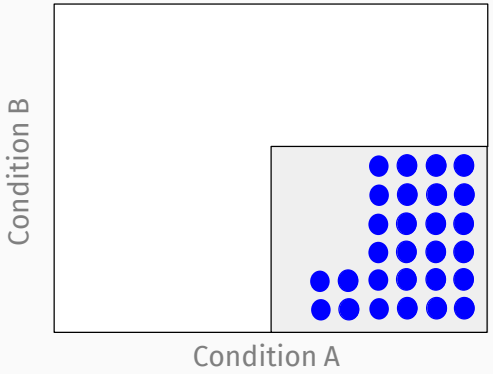
Exploring the data-generating space



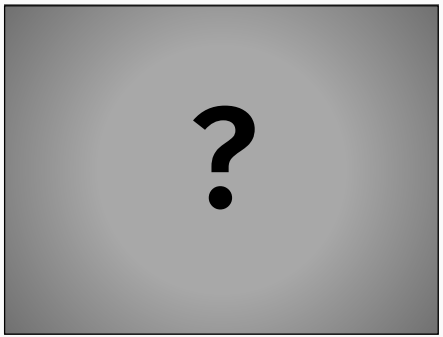
Transparency



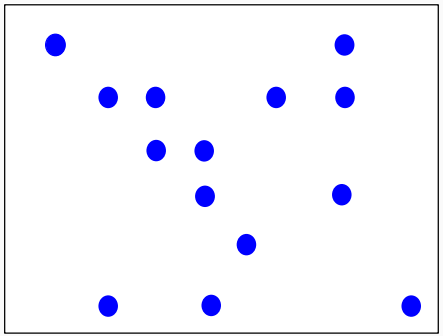
Exploring the data-generating space



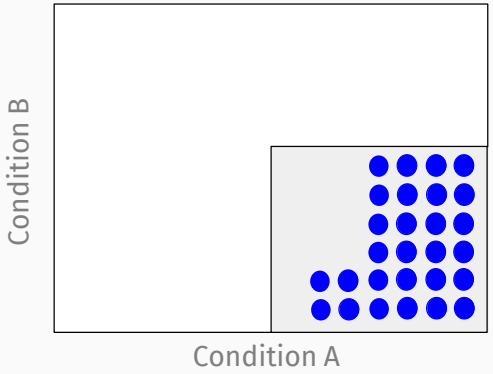
Transparency



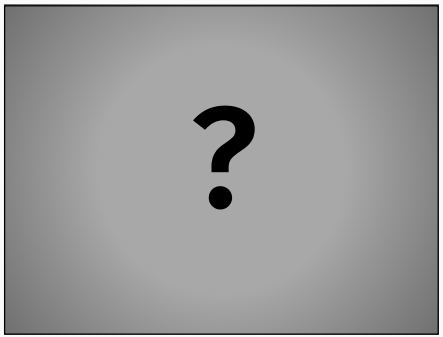
Efficient exploration



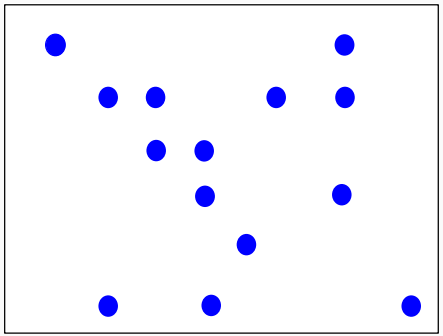
Exploring the data-generating space



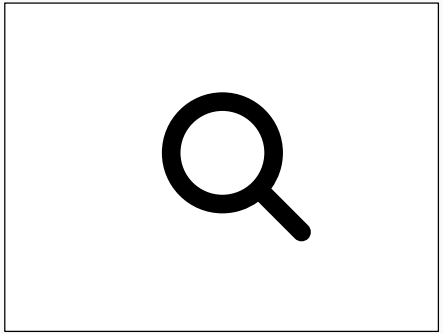
Transparency



Efficient exploration

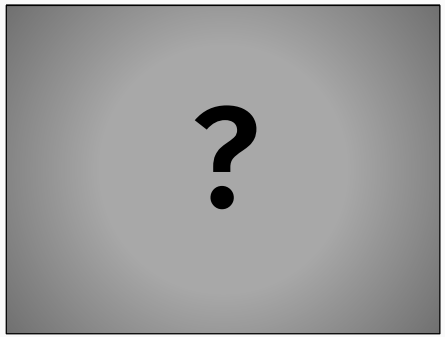


Benchmarking

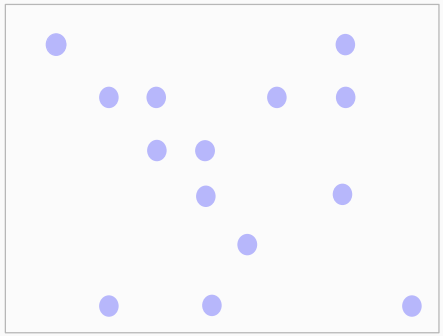


The role of preregistration for transparency

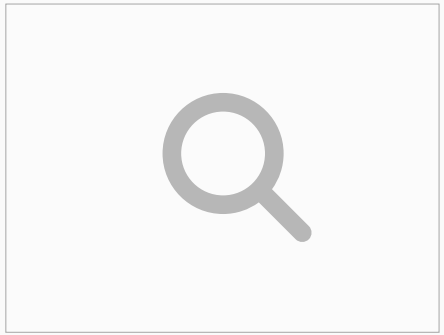
Transparency



Efficient exploration

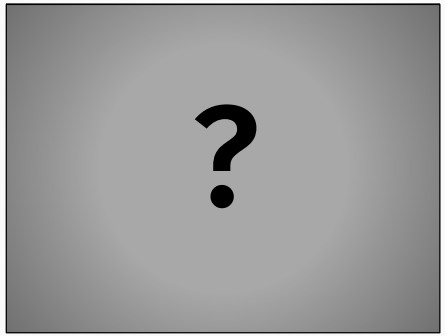


Benchmarking

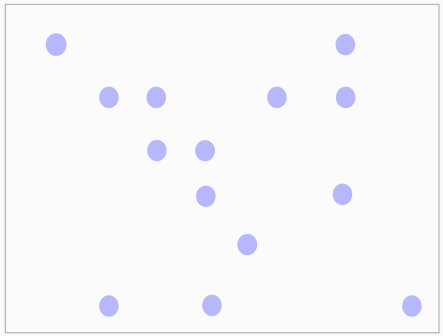


The role of preregistration for transparency

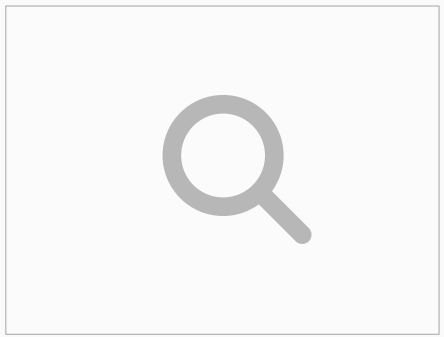
Transparency



Efficient exploration



Benchmarking



Why, when, and how to (or not to) preregister a simulation study

Björn S. Siepe[†] [1], František Bartoš [2*], Anne-Laure Boulesteix [3,4*], Daniel W. Heck [1*], Aaron Peikert [5,6*], Alexandra Sarafoglou [2*], Samuel Pawel [7]

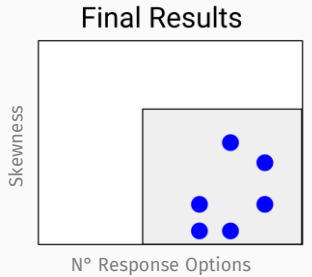
Simulation studies make generalizable claims

Simulation studies make generalizable claims

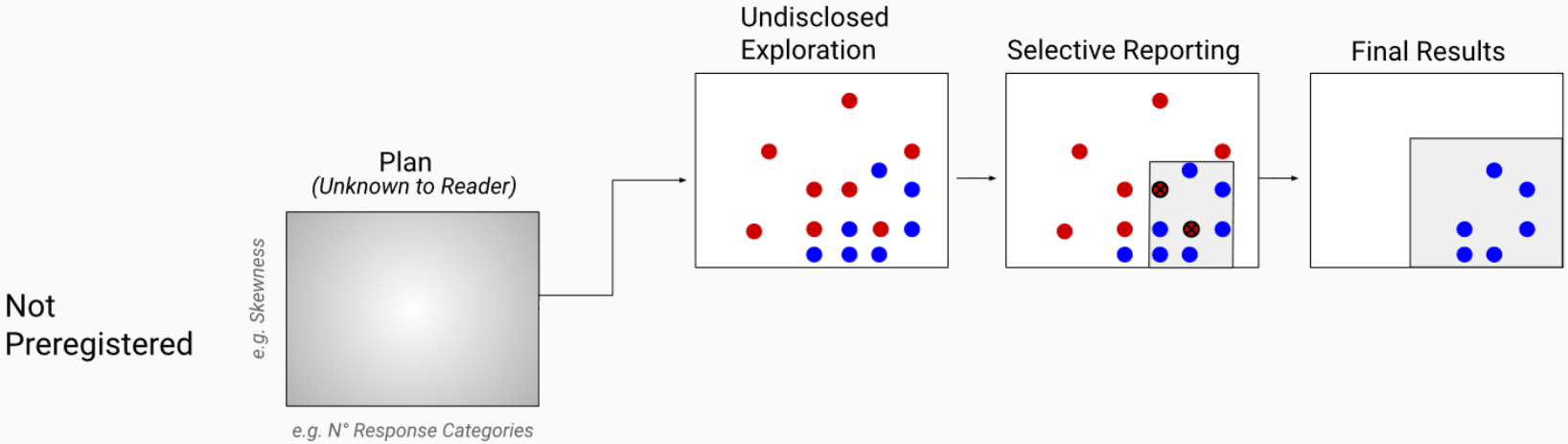
> **We recommend method XYZ for ordinal items**
>
> Hypothetical authors

Simulation studies make generalizable claims

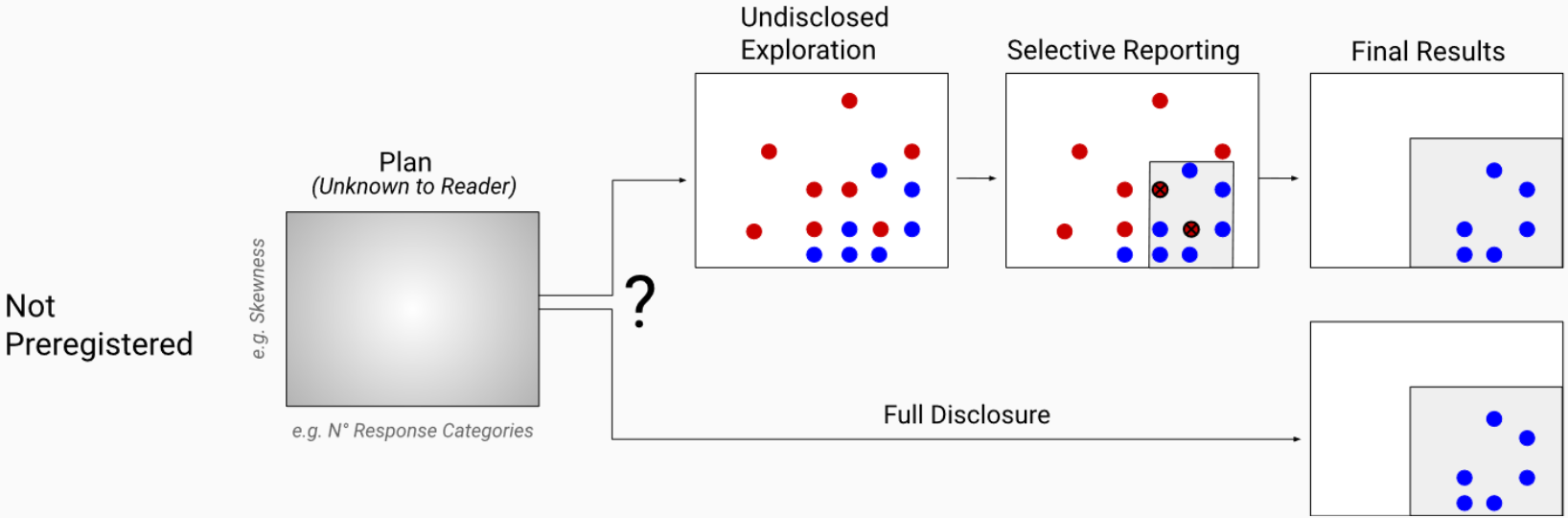
> **We recommend method XYZ for ordinal items**
>
> Hypothetical authors



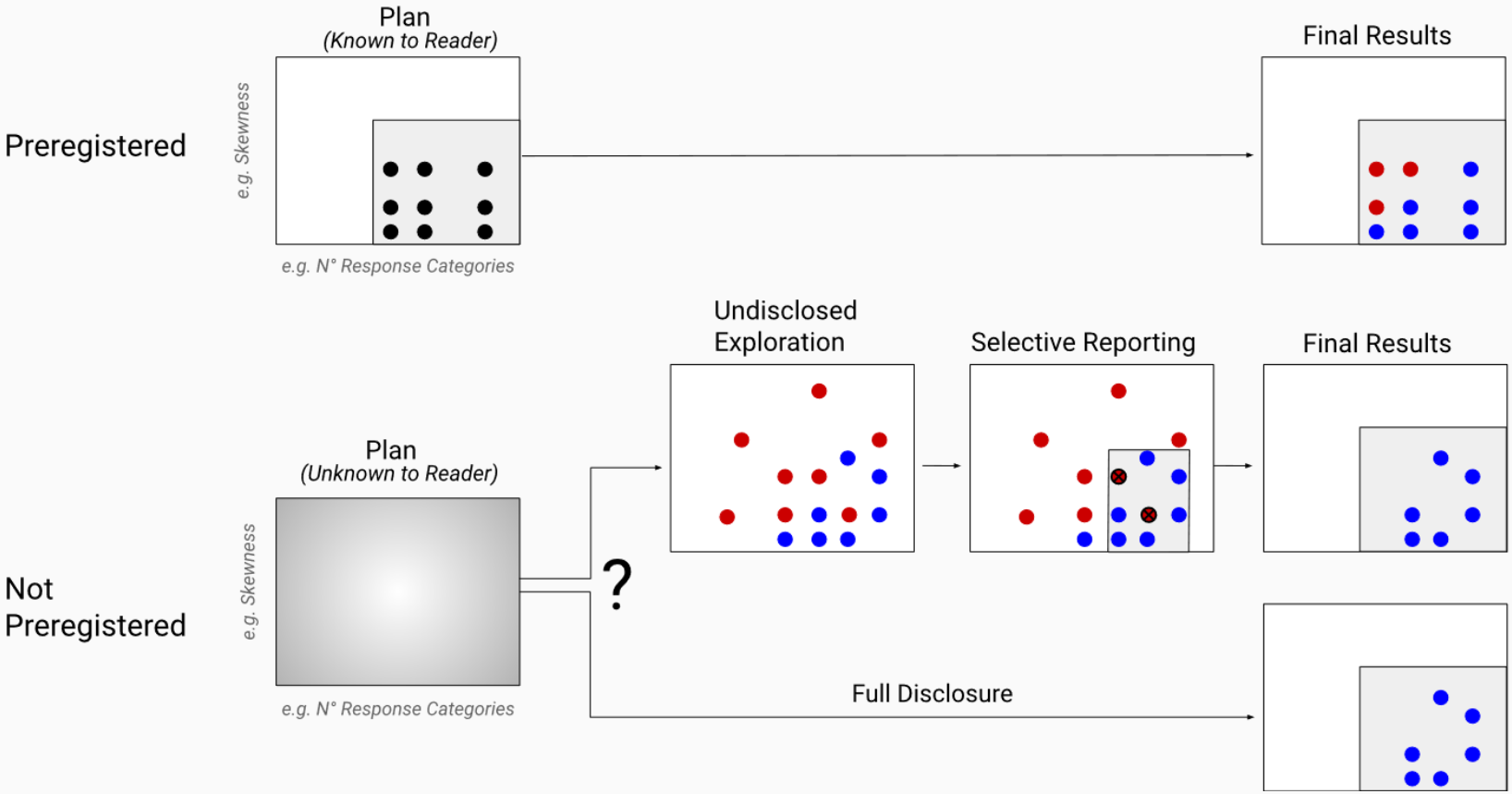
Preregistration reduces epistemic uncertainty



Preregistration reduces epistemic uncertainty

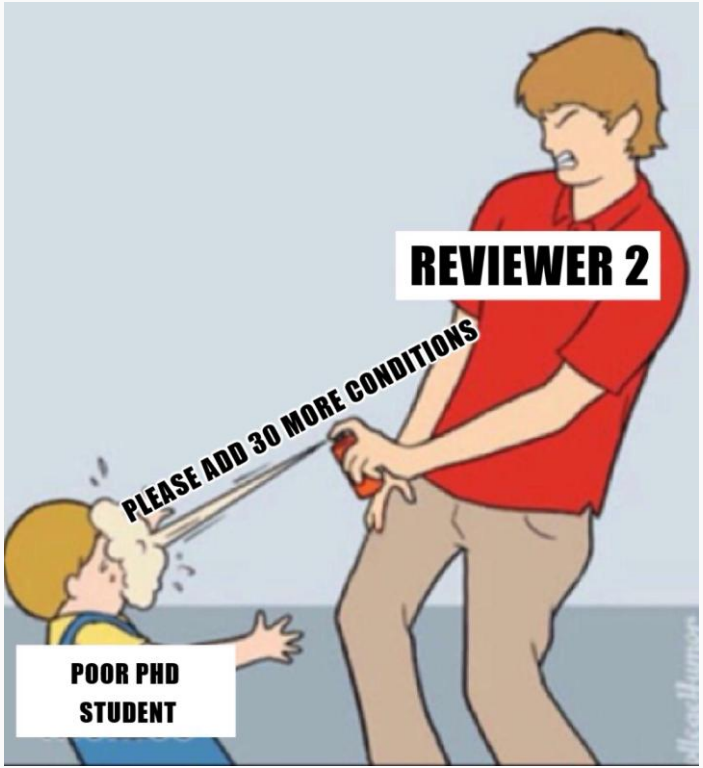


Preregistration reduces epistemic uncertainty



Registered Reports to the rescue

Registered Reports to the rescue



Created with ImageResizer

Registered Reports to the rescue

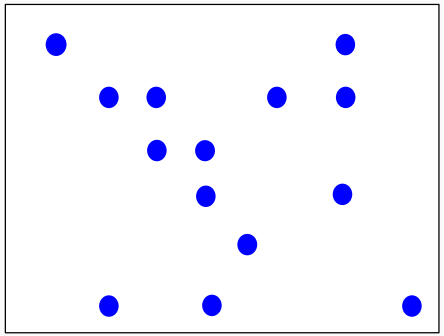
Created with ImageResizer

Centre for Open Science

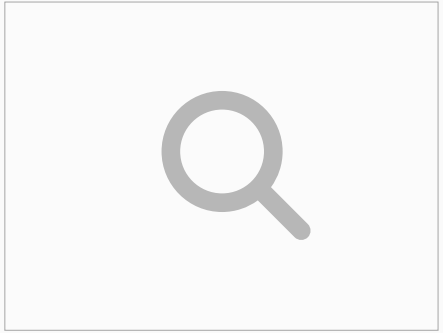
Exploring the data-generating space

Transparency

Efficient exploration



Benchmarking



Exploring more efficiently

Exploring more efficiently

Multivariate Behavioral Research, *35* (2), 137-167

Design and Analysis of Monte Carlo Experiments:
Attacking the Conventional Wisdom

Anders Skrondal
Department of Epidemiology
National Institute of Public Health, Oslo

Exploring more efficiently

Multivariate Behavioral Research, *35* (2), 137-167

Design and Analysis of Monte Carlo Experiments:
Attacking the Conventional Wisdom

Anders Skrondal
Department of Epidemiology
National Institute of Public Health, Oslo

Less but the same? Methods for Reducing the Size of Simulation Study Designs

Angelike & Papenberg (in prep.)

Exploring more efficiently

Multivariate Behavioral Research, *35* (2), 137-167

Design and Analysis of Monte Carlo Experiments:
Attacking the Conventional Wisdom

Anders Skrondal
Department of Epidemiology
National Institute of Public Health, Oslo

Less but the same? Methods for Reducing the Size of Simulation Study Designs

Angelike & Papenberg (in prep.)

Design of experiments for simulation studies: Comparing strategies using surrogate models

Hafiz, ..., & Peikert (in prep.)

Incentivizing cumulative simulation research

Incentivizing cumulative simulation research

Living Synthetic Benchmarks:
A Neutral and Cumulative Framework for Simulation Studies

František Bartoš [1], Samuel Pawel [2], Björn S. Siepe [3]

Incentivizing cumulative simulation research

Separate Studies (Status Quo)

Paper 1
(new method)

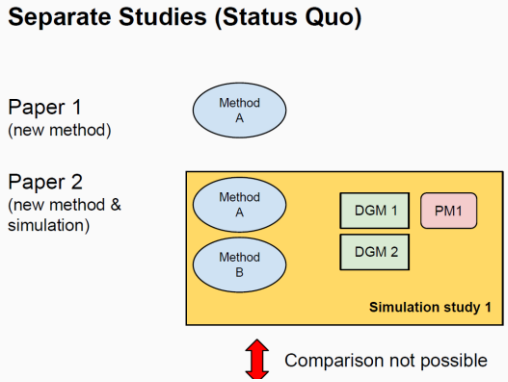
Living Synthetic Benchmarks: A Neutral and Cumulative Framework for Simulation Studies

František Bartoš [1], Samuel Pawel [2], Björn S. Siepe [3]

Incentivizing cumulative simulation research

Living Synthetic Benchmarks:
A Neutral and Cumulative Framework for Simulation Studies

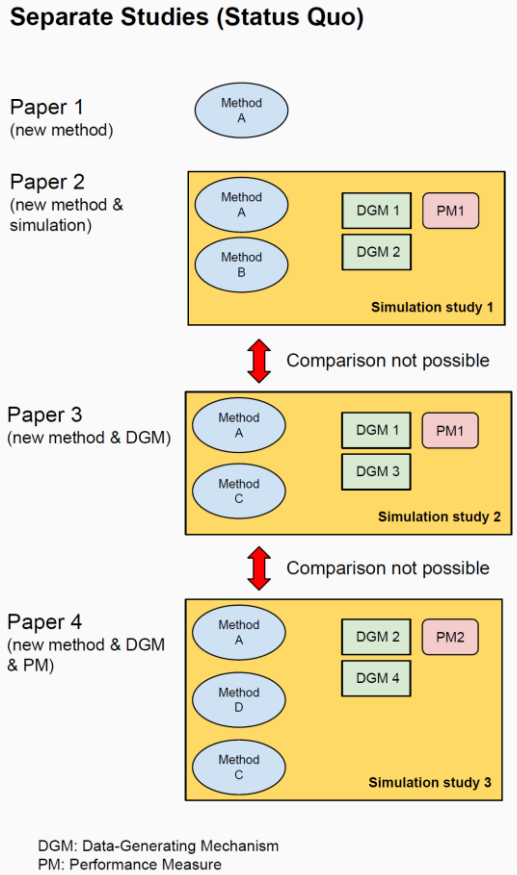
František Bartoš [1], Samuel Pawel [2], Björn S. Siepe [3]



Incentivizing cumulative simulation research

Living Synthetic Benchmarks:
A Neutral and Cumulative Framework for Simulation Studies

František Bartoš [1], Samuel Pawel [2], Björn S. Siepe [3]

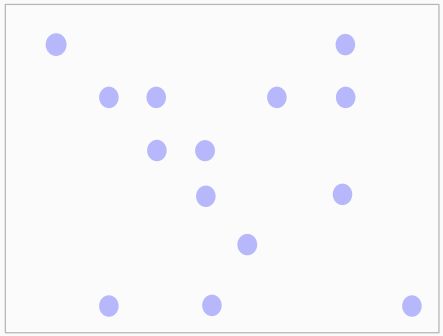


Exploring the data-generating space

Exploring the data-generating space

Transparency

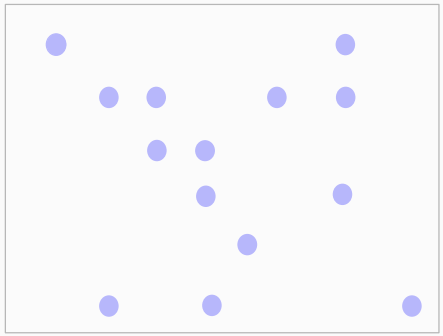
Efficient exploration



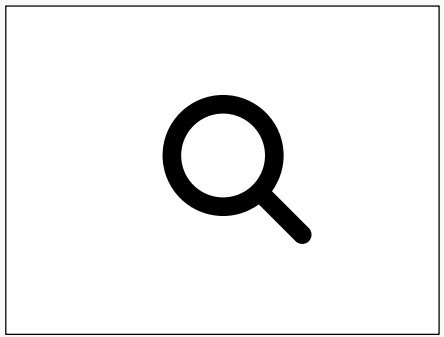
Exploring the data-generating space

Transparency

Efficient exploration

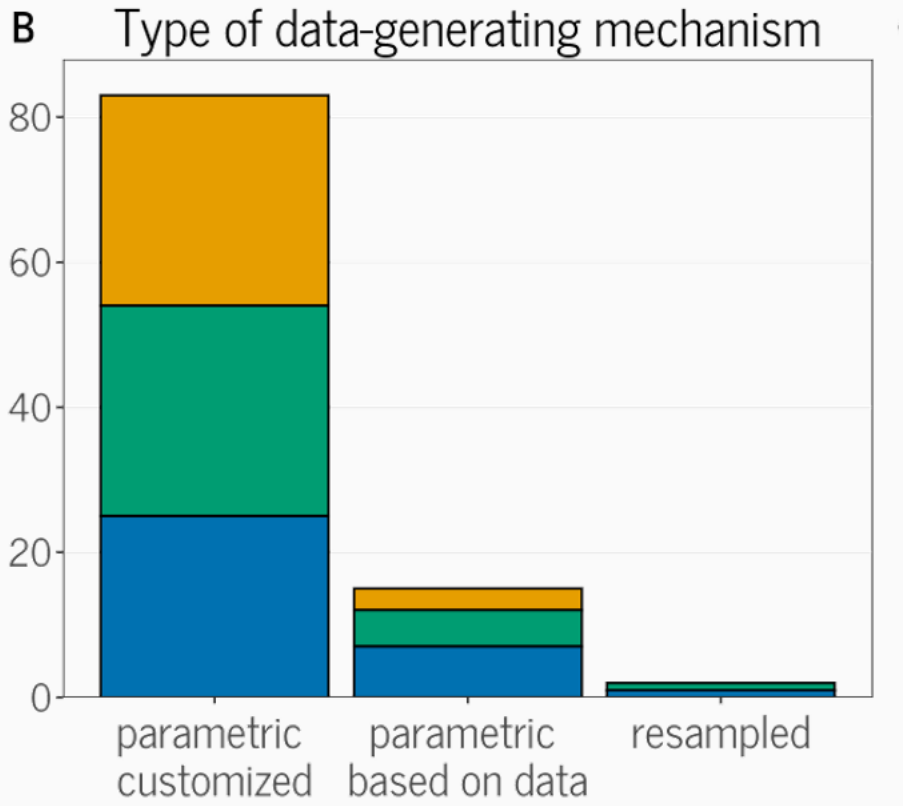


Benchmarking



What about *real* benchmarking?

What about *real* benchmarking?



Psychology has been slow to adopt benchmarking

Psychology has been slow to adopt benchmarking

Computer Vision/ML

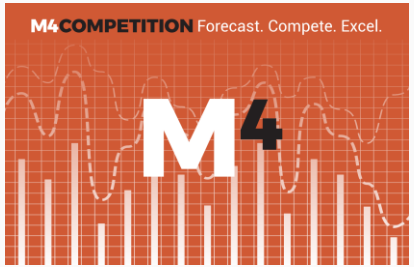
Source: Steni Sebastian, Medium

Psychology has been slow to adopt benchmarking

Computer Vision/ML

Source: Steni Sebastian, Medium

Forecasting



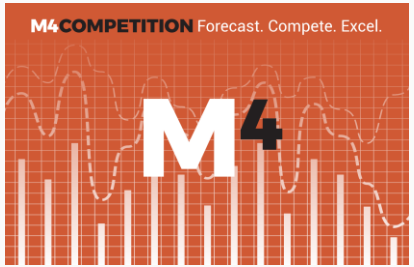
Source: Rob Hyndman

Psychology has been slow to adopt benchmarking

Computer Vision/ML

Source: Steni Sebastian, Medium

Forecasting



Source: Rob Hyndman

Biostatistics

Robinson and Vitek *Genome Biology* (2019) 20:205
https://doi.org/10.1186/s13059-019-1846-5

EDITORIAL

Benchmarking comes of age

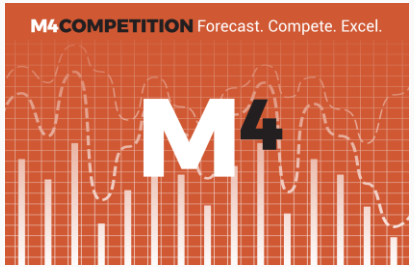
Mark D. Robinson[1*] and Olga Vitek[2*]

Psychology has been slow to adopt benchmarking

Computer Vision/ML

Source: Steni Sebastian, Medium

Forecasting



Source: Rob Hyndman

Biostatistics

Robinson and Vitek *Genome Biology* (2019) 20:205
https://doi.org/10.1186/s13059-019-1846-5

EDITORIAL

Benchmarking comes of age

Mark D. Robinson[1*] and Olga Vitek[2*]

Why not in psychology?

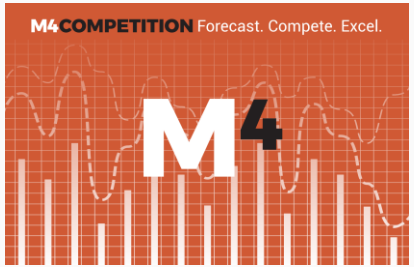
(Rocca & Yarkoni, 2021)

Psychology has been slow to adopt benchmarking

Computer Vision/ML

Source: Steni Sebastian, Medium

Forecasting



Source: Rob Hyndman

Biostatistics

Robinson and Vitek *Genome Biology* (2019) 20:205
https://doi.org/10.1186/s13059-019-1846-5

EDITORIAL

Benchmarking comes of age

Mark D. Robinson[1*] and Olga Vitek[2*]

Why not in psychology?

(Rocca & Yarkoni, 2021)

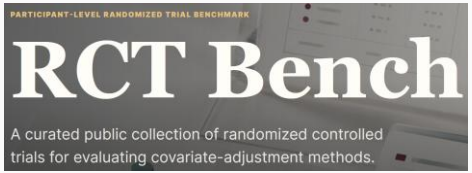
1. Lack of openly available data
2. Focus on in-sample performance

The time is ripe for benchmarking

The time is ripe for benchmarking

We now have lots of open data

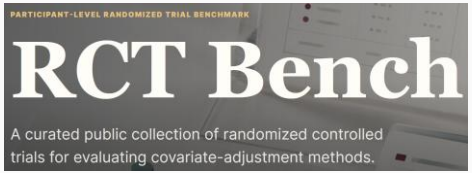
Item Response Warehouse



The time is ripe for benchmarking

We now have lots of open data + **Some of us care about prediction**

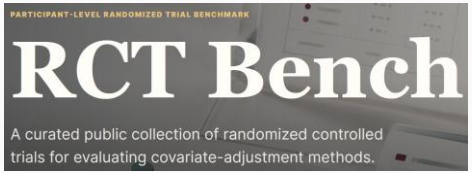
Item Response Warehouse



The time is ripe for benchmarking

We now have lots of open data + **Some of us care about prediction**

Item Response Warehouse



Perspectives on Psychological Science
Volume 12, Issue 6, November 2017, Pages 1100-1122

https://doi.org/10.1177/1745691617693393

SAGE journals

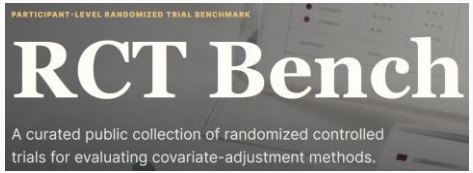
Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Tal Yarkoni and Jacob Westfall

The time is ripe for benchmarking

We now have lots of open data + **Some of us care about prediction**

Item Response Warehouse



Perspectives on Psychological Science
Volume 12, Issue 6, November 2017, Pages 1100-1122
© The Author(s) 2017, Article Reuse Guidelines
https://doi.org/10.1177/1745691617693393

SAGE journals

Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

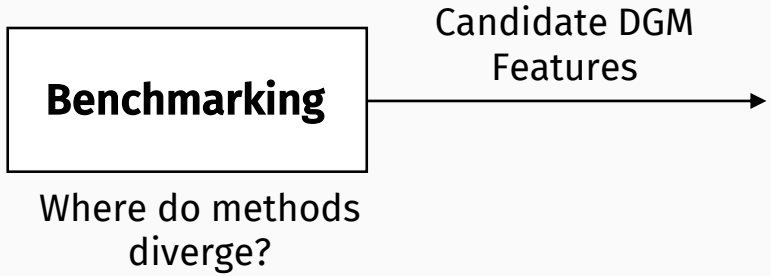
Tal Yarkoni and Jacob Westfall

=

Benchmarking and simulations

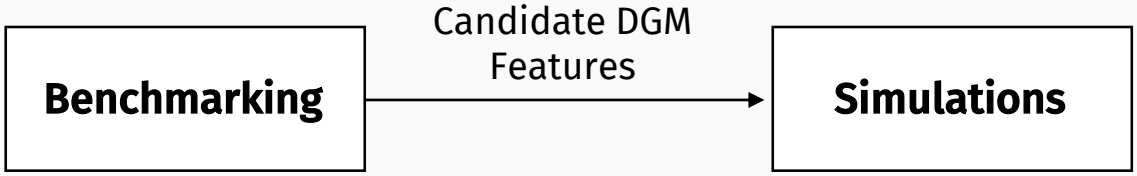
Benchmarking

Benchmarking and simulations



Benchmarking and simulations

Benchmarking and simulations




Biometrical Journal

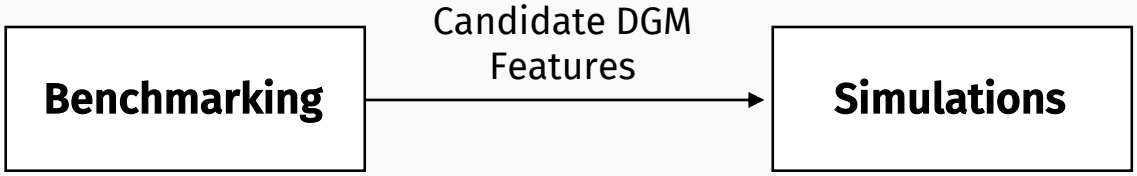
RESEARCH ARTICLE | Open Access

On the role of benchmarking data sets and simulations in method comparison studies

Sarah Friedrich, Tim Friede

First published: 21 February 2023 | https://doi.org/10.1002/bimj.202200212 | VIEW METRICS

Benchmarking and simulations



Where do methods diverge?

Which aspects drive differences?

Biometrical Journal

RESEARCH ARTICLE | Open Access

On the role of benchmarking data sets and simulations in method comparison studies

Sarah Friedrich, Tim Friede

First published: 21 February 2023 | https://doi.org/10.1002/bimj.202200212 | VIEW METRICS

WIREs COMPUTATIONAL STATISTICS

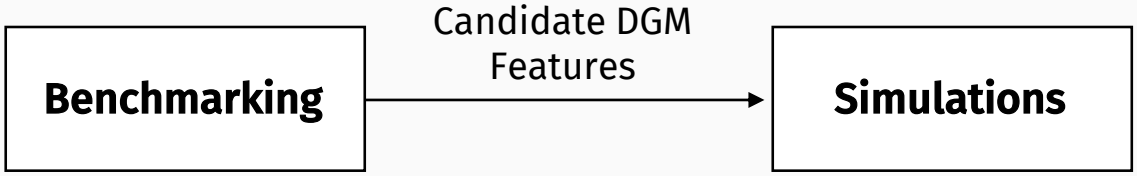
ADVANCED REVIEW | Open Access

Statistical Parametric Simulation Studies Based on Real Data

Christina Sauer, F. Julian D. Lange, Maria Thurow, Ina Dormuth, Anne-Laure Boulesteix

First published: 19 August 2026 | https://doi.org/10.1002/wics.70074 | VIEW METRICS

Benchmarking and simulations



Where do methods diverge?

Which aspects drive differences?

Biometrical Journal

RESEARCH ARTICLE | Open Access

On the role of benchmarking data sets and simulations in method comparison studies

Sarah Friedrich, Tim Friede

First published: 21 February 2023 | https://doi.org/10.1002/bimj.202200212 | VIEW METRICS

WIREs COMPUTATIONAL STATISTICS

ADVANCED REVIEW | Open Access

Statistical Parametric Simulation Studies Based on Real Data

Christina Sauer, F. Julian D. Lange, Maria Thurow, Ina Dormuth, Anne-Laure Boulesteix

First published: 19 August 2026 | https://doi.org/10.1002/wics.70074 | VIEW METRICS

Weber *et al. Genome Biology* (2019) 20:125
https://doi.org/10.1186/s13059-019-1738-8

Genome Biology

REVIEW Open Access

Essential guidelines for computational method benchmarking

Check for updates

Lukas M. Weber[1,2], Wouter Saelens[3,4], Robrecht Cannoodt[3,4], Charlotte Soneson[1,2,8], Alexander Hapfelmeier[5], Paul P. Gardner[6], Anne-Laure Boulesteix[7], Yvan Saeys[3,4*] and Mark D. Robinson[1,2*]

Simulations, 51 years later

Simulations, 51 years later

> **Statisticians ... often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

Simulations, 51 years later

> **Statisticians ... often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

The field is moving fast:

Simulations, 51 years later

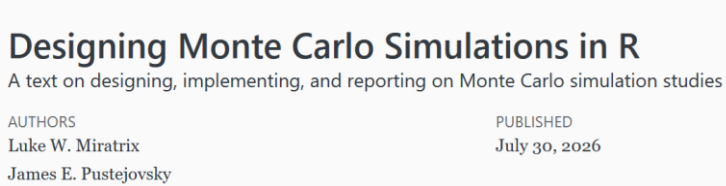
> **Statisticians … often pay too little attention to their own principles of design**
> Hoaglin & Andrews, 1975

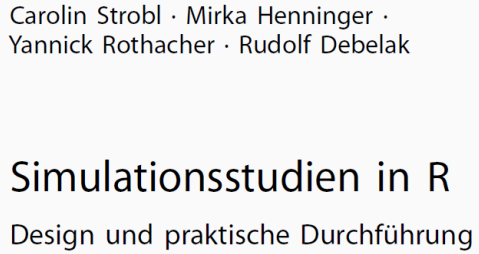
The field is moving fast:

Simulations, 51 years later

> **Statisticians … often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

The field is moving fast:

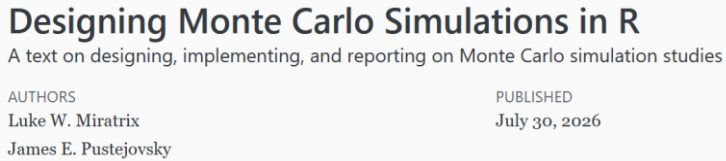


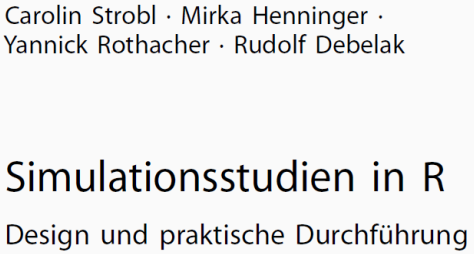


Simulations, 51 years later

> **Statisticians … often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

The field is moving fast:





Adopt Registered Reports at Psychological Methods

Simulations, 51 years later

> **Statisticians … often pay too little attention to their own principles of design**
>
> Hoaglin & Andrews, 1975

The field is moving fast:

EDITORIAL | Free Access

Editorial for the special collection "Towards neutral comparison studies in methodological research"

Anne-Laure Boulesteix, Mark Baillie, Dominic Edelmann, Leonhard Held, Tim P. Morris, Willi Sauerbrei

First published: 17 February 2024 | https://doi.org/10.1002/bimj.202400031 | VIEW METRICS

Designing Monte Carlo Simulations in R

A text on designing, implementing, and reporting on Monte Carlo simulation studies

AUTHORS
Luke W. Miratrix
James E. Pustejovsky

PUBLISHED
July 30, 2026

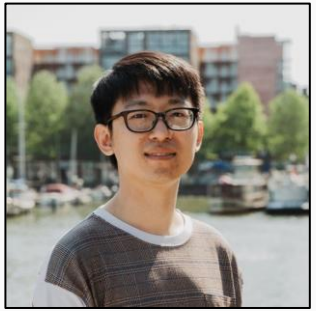
Carolin Strobl · Mirka Henninger · Yannick Rothacher · Rudolf Debelak

Simulationsstudien in R

Design und praktische Durchführung

Adopt Registered Reports at Psychological Methods

Co-Authors








References

Bartoš, F., Pawel, S., & Siepe, B. S. (2025). Living Synthetic Benchmarks: A Neutral and Cumulative Framework for Simulation Studies. *arXiv preprint. arXiv:2510.19489*. https://doi.org/10.48550/arXiv.2510.19489
Boulesteix, A. L., Hoffmann, S., Charlton, A., & Seibold, H. (2020). A replication crisis in methodological research?. Significance, 17(5), 18-21. https://doi.org/10.1111/1740-9713.01444
Bouma, A., van Assen, M. A. L. M., van Aert, R. C. M., & Voncken, L. (2026, February 6). Reporting Practices, Open Science Practices, and Trustworthiness of Simulation Studies in Psychology: A Questionnaire Study. Retrieved from osf.io/preprints/psyarxiv/jn9sy_v2
Efron, B. (2001). Statistical modeling: The two cultures: Comment. Statistical Science, 16(3), 218-219.
Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. Educational Measurement: Issues and Practice, 35(2), 36-49. https://doi.org/10.1111/emip.12111
Heinze, G., Boulesteix, A. L., Kammer, M., Morris, T. P., White, I. R., & Simulation Panel of the STRATOS initiative (2024). Phases of methodological research in biostatistics-Building the evidence base for new methods. Biometrical journal. Biometrische Zeitschrift, 66(1), e2200222. https://doi.org/10.1002/bimj.202200222
Hoaglin, D. C., & Andrews, D. F. (1975). The reporting of computation-based results in statistics. The American Statistician, 29(3), 122-126. https://doi.org/10.1080/00031305.1975.10477393
Koehler, E., Brown, E., & Haneuse, S. J. P. A. (2009). On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses. The American Statistician, 63(2), 155–162. https://doi.org/10.1198/tast.2009.0030
Little, R. J. (2013). In Praise of Simplicity not Mathematistry! Ten Simple Powerful Ideas for the Statistical Scientist. Journal of the American Statistical Association, 108(502), 359–369. https://doi.org/10.1080/01621459.2013.787932
Luijken, K., Lohmann, A., Alter, U., Gonzalez, J. C., Clouth, F. J., Fossum, J. L., ... & Groenwold, R. H. H. (2024). Replicability of simulation studies for the investigation of statistical methods: the RepliSims project. Royal Society Open Science, 11(1), 231003. https://doi.org/10.1098/rsos.231003
Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in medicine, 38(11), 2074-2102. https://doi.org/10.1002/sim.8086
Pawel, S., Kook, L., & Reeve, K. (2024). Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. Biometrical journal. Biometrische Zeitschrift, 66(1), e2200091. https://doi.org/10.1002/bimj.202200091
Pawel, S., Bartoš, F., Siepe, B. S., & Lohmann, A. (2026). Handling Missingness, Failures, and Non-Convergence in Simulation Studies: A Review of Current Practices and Recommendations. The American Statistician, 80(1), 31–48. https://doi.org/10.1080/00031305.2025.2540002
Robinson, M.D., Vitek, O. (2019). Benchmarking comes of age. Genome *Biol* 20, 205 https://doi.org/10.1186/s13059-019-1846-5
Rocca, R., & Yarkoni, T. (2021). Putting Psychology to the Test: Rethinking Model Evaluation Through Benchmarking and Prediction. Advances in Methods and Practices in Psychological Science, 4(3), 10.1177/25152459211026864. https://doi.org/10.1177/25152459211026864
Sauer, C., F. J. D. Lange, M. Thurow, I. Dormuth, & A.-L. Boulesteix. (2026). Statistical Parametric Simulation Studies Based on Real Data. *Wiley Interdisciplinary Reviews: Computational Statistics* 18, no. 3: e70074. https://doi.org/10.1002/wics.70074.
Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A. L., Heck, D. W., & Pawel, S. (2024). Simulation studies for methodological research in psychology: A standardized template for planning, preregistration, and reporting. Psychological methods, 10.1037/met0000695. Advance online publication. https://doi.org/10.1037/met0000695
Siepe, B.S., Haslbeck, J.M., Kloft, M. et al. (2026). Introducing openESM: A database of openly available experience sampling datasets. Behav Res 58, 240. https://doi.org/10.3758/s13428-026-03112-y
Siepe, B. S., Bartoš, F., Boulesteix, A., Heck, D. W., Peikert, A., Sarafoglou, A., & Pawel, S. (2026). Why, when, and how to (or not to) preregister a simulation study. PsyArXiv Preprint. https://doi.org/10.31234/osf.io/cxjrb_v1
Ullmann, T., Beer, A., Hünemörder, M., Seidl, T., & Boulesteix, A. L. (2023). Over-optimistic evaluation and reporting of novel cluster algorithms: an illustrative study. Advances in Data Analysis and Classification, 17(1), 211-238. https://doi.org/10.1007/s11634-022-00496-5
Williams, C., Yang, Y., Lagisz, M., Morrison, K., Ricolfi, L., Warton, D. I., & Nakagawa, S. (2024). Transparent reporting items for simulation studies evaluating statistical methods: Foundations for reproducibility and reliability. Methods in Ecology and Evolution, 15(11), 1926-1939. https://doi.org/10.1111/2041-210X.14415
Wünsch, M., Herrmann, M., Noltenius, E., Mohr, M., Morris, T. P., & Boulesteix, A. L. (2025). Rethinking the handling of method failure in comparison studies. Statistics in Medicine, 44(23-24), e70257. https://doi.org/10.1002/sim.70257.

Thank You!

https://openesmdata.org

bjoernsiepe@gmail.com

https://bsiepe.github.io

bsiepe.bsky.social

Slides & Papers

Resources

- All icons either from PowerPoint or fontawesome
  - https://fontawesome.com/icons/bluesky
  - https://fontawesome.com/icons/glasses
  - https://fontawesome.com/icons/code
  - https://fontawesome.com/icons/dna
  - https://fontawesome.com/icons/classic/solid/magnifying-glass
- Python Software Foundation. (n.d.). *The Python logo*. https://www.python.org/community/logos/ (PSF Trademark Policy)
- Rob Hyndman - https://robjhyndman.com/hyndsight/m4comp/
- Steni Sebastian - https://medium.com/@ssteni/imagenet-what-why-and-how-e5ebed04abb5
- R Foundation for Statistical Computing. (2016). *R logo*. https://www.r-project.org/logo/ (CC-BY SA 4.0)
- James Grellier – Own Work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=10253357
- Marcin Wichary from San Francisco, U.S.A. - [1]Uploaded by Partyzan_XXI, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=8198235
- Carbon (https://carbon.now.sh/) for pretty code screenshots
- The Splintered Mind - https://schwitzsplinters.blogspot.com/2021/02/adversarial-collaboration.html
- Sandbox Studio Artwork - https://www.symmetrymagazine.org/article/the-facts-and-nothing-but-the-facts?language_content_entity=und
- Registered Reports graphic: Centre for Open Science, https://cdn.cos.io/media/images/registered_reports.width-800.png

Resources