MRC Clinical Trials Unit

Smarter Studies
Global Impact
Better Health

UCL

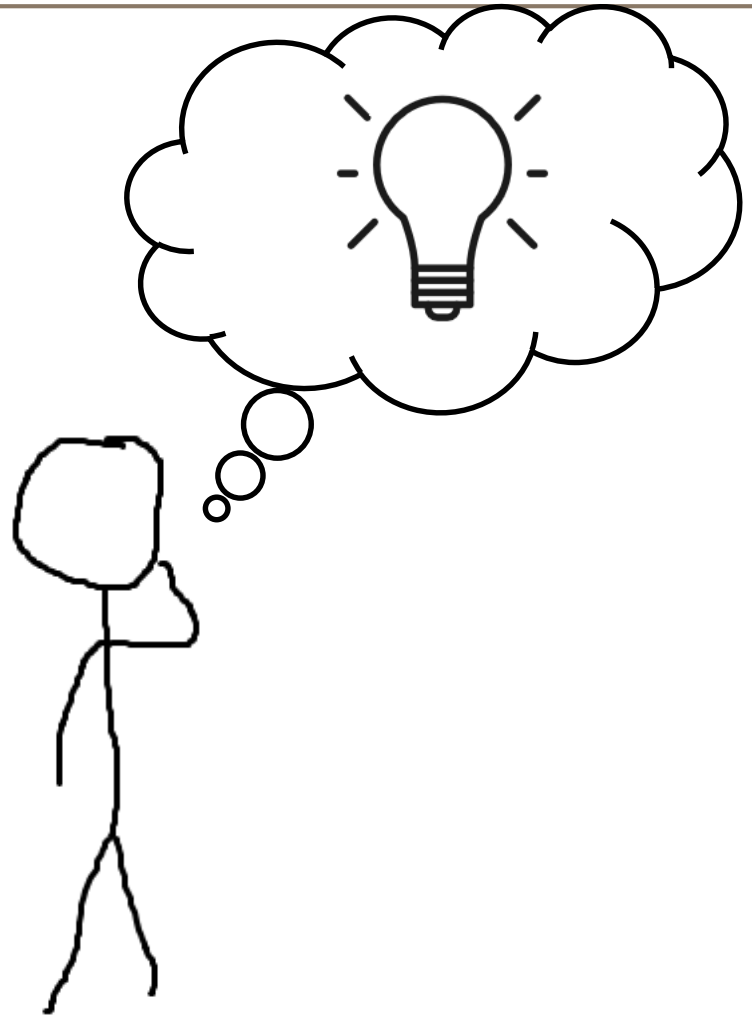
Towards stronger simulation studies in statistical research

Biometrisches Kolloquium 2021

Tim Morris on behalf of the simulation panel

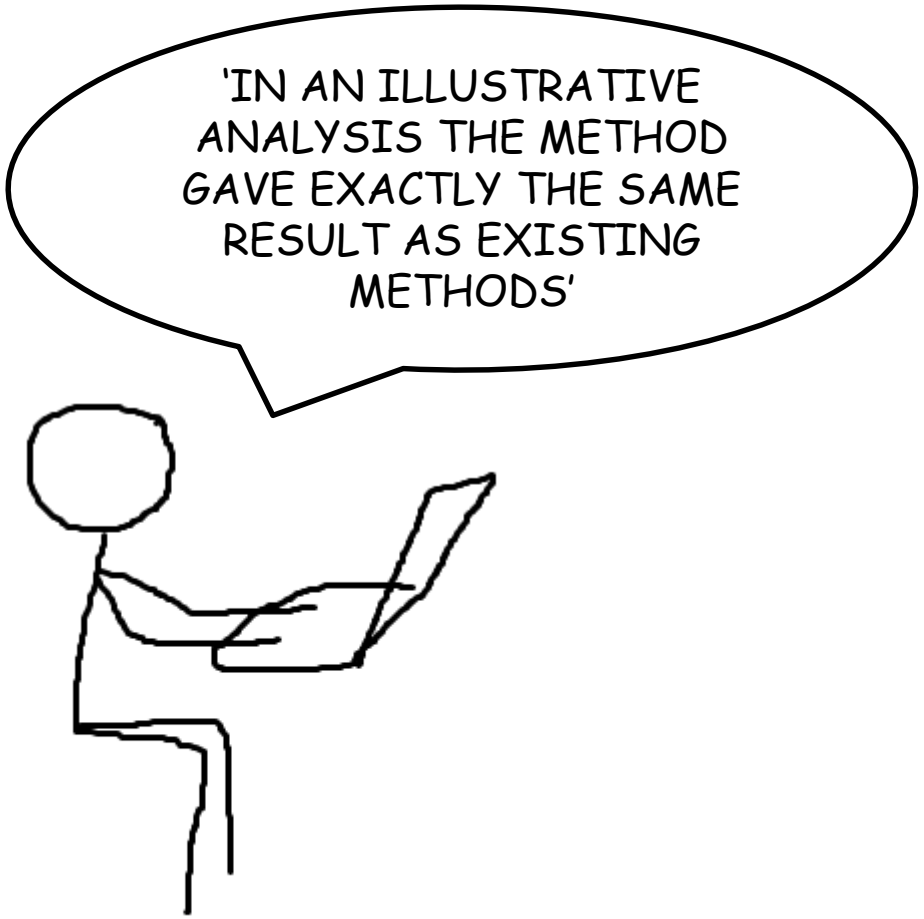
A statistician

A statistician



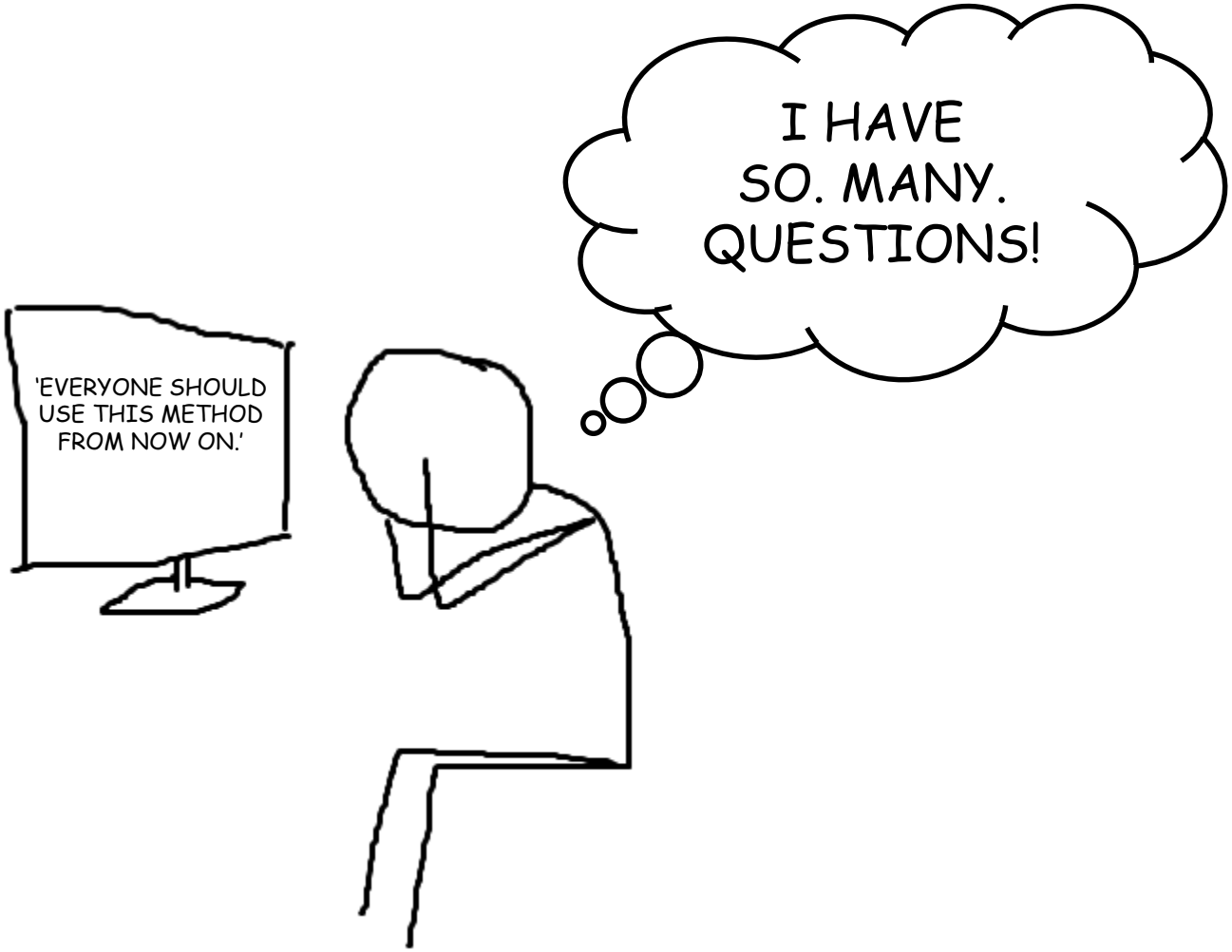
A statistician

A statistician



A statistician

What's the problem?



Why does this matter?

Always useful to remember (imagine?) that a simulation study could be used to justify what people do!

We want to get things right (whether or not we are incentivised).

Received: 15 August 2017 | Revised: 20 October 2017 | Accepted: 22 October 2017

DOI: 10.1002/bimj.201700129

LETTER TO THE EDITOR

Biometrical Journal

On the necessity and design of studies comparing statistical methods

Anne-Laure Boulesteix[1]
Harald Binder[2]
Michal Abrahamowicz[3]
Willi Sauerbrei[2]
for the Simulation Panel of the STRATOS Initiative

Regression Models for Prognostic Prediction: Advantages, Problems, and Suggested Solutions[1]

Frank E. Harrell, Jr,[2,*] Kerry L. Lee,[2] David B. Matchar,[2] and Thomas A. Reichert[3,4]

Multiple regression models have wide applicability in predicting the outcome of patients with a variety of diseases. However, many researchers are using such models without validating the necessary assumptions. All too frequently, researchers also "overfit" the data by developing models using too many predictor variables and insufficient sample sizes. Models developed in this way are unlikely to stand the test of validation on a separate patient sample. Without attempting such a validation, the researcher remains unaware that overfitting has occurred. When the ratio of the number of patients suffering endpoints to the number of potential predictors is small (say < 10), data reduction methods are available that can greatly improve the performance of regression models. Regression models can make more accurate predictions than other methods such as stratification and recursive partitioning, when (*a*) model assumptions are thoroughly examined; (*b*) steps are taken (ie, choosing another model or transforming the data) when assumptions are violated; and (*c*) the method of model formulation does not result in overfitting the data. [Cancer Treat Rep 69:1071–1077, 1985]

J Clin Epidemiol Vol. 49, No. 12, pp. 1373–1379, 1996

S0895-4356/96/$15.00
PII S0895-4356(96)00236-3

A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis

Peter Peduzzi,[1,4,*] *John Concato*,[2,3] *Elizabeth Kemper*,[1,4] *Theodore R. Holford*,[4] *and Alvan R. Feinstein*[2,3,4]

[1]COOPERATIVE STUDIES PROGRAM COORDINATING CENTER AND THE [2]MEDICAL SERVICE, VETERANS AFFAIRS MEDICAL CENTER, WEST HAVEN CONNECTICUT 06516; AND THE DEPARTMENTS OF [3]MEDICINE (CLINICAL EPIDEMIOLOGY UNIT) AND [4]EPIDEMIOLOGY AND PUBLIC HEALTH, YALE UNIVERSITY SCHOOL OF MEDICINE, NEW HAVEN, CONNECTICUT 06510

CITED >6,000 TIMES!

van Smeden *et al. BMC Medical Research Methodology* (2016) 16:163
DOI 10.1186/s12874-016-0267-3

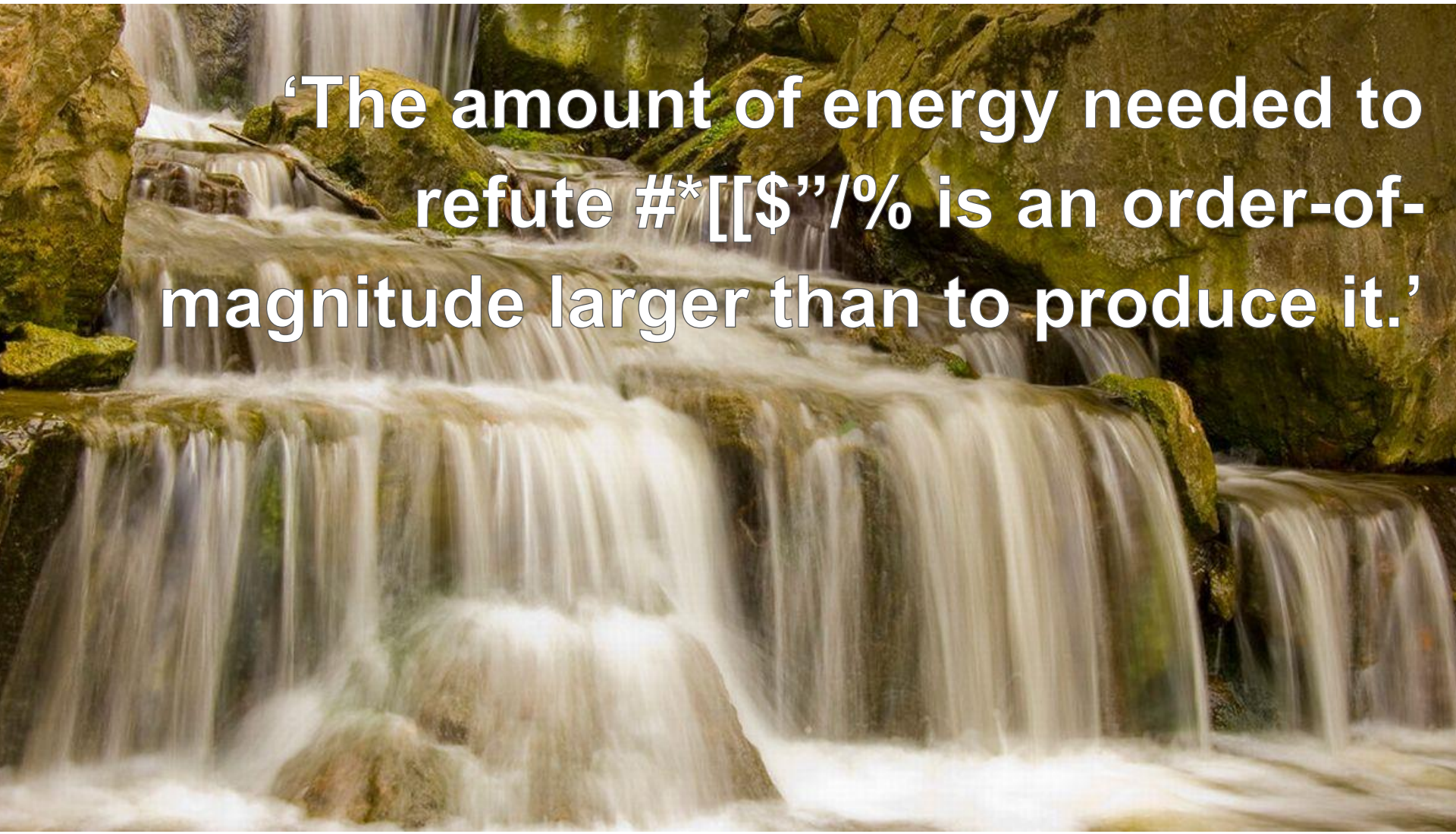
BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

No rationale for 1 variable per 10 events criterion for binary logistic regression analysis

Maarten van Smeden[1*], Joris A. H. de Groot[1], Karel G. M. Moons[1], Gary S. Collins[2], Douglas G. Altman[2], Marinus J. C. Eijkemans[1] and Johannes B. Reitsma[1]

CITED 145 TIMES


‘The amount of energy needed to refute #*[[$”/% is an order-of-magnitude larger than to produce it.’

Simulation studies need to be done well first time!

Received: 29 November 2017 | Revised: 23 August 2018 | Accepted: 2 November 2018

DOI: 10.1002/sim.8086

Check for updates

TUTORIAL IN BIOSTATISTICS

WILEY Statistics in Medicine

Using simulation studies to evaluate statistical methods

Tim P. Morris[1] | Ian R. White[1] | Michael J. Crowther[2]

What steps can we take?

- A collection of idle thoughts on some steps that would make a huge difference
- All opinions!
- You will be able to think of more; write them down and use them to give your simulation studies a competitive edge!

1. Structure for your readers

Do you know about reporting guidelines?

These are guidelines guiding researchers on how to report various types of study. The things the investigators think are obvious will not be to outsiders.

www.equator-network.org/

equator network

Enhancing the QUAlity and Transparency Of health Research

EQUATOR resources in German | Portuguese | Spanish

Home | About us | Library | Toolkits | Courses & events | News | Blog | Librarian Network | Contact

Your one-stop-shop for writing and publishing high-impact health research

find reporting guidelines | improve your writing | join our courses | run your own training course | enhance your peer review | implement guidelines

Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.

- Search for reporting guidelines
- Not sure which reporting guideline to use?
- Reporting guidelines under development
- Visit the library for more resources

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	Extensions
Economic evaluations	CHEERS	

See all 455 reporting guidelines

1. Structure for your readers

It makes it much easier to read any study if you have a framework.

When randomised trials are reported using Consort, I know roughly what to expect to read. (The fact that no trial is ‘vanilla’ does not detract from the usefulness.)

1. Structure for your readers

A – Aims

D – Data-generating mechanisms

E – Estimands

M – Methods of analysis

P – Performance measures

1. Structure for your readers

A – Aims

D – Data-ge

E – Estimar..s

M – Methods of analysis

P – Performance measures

To make life even easier for your readers, try putting these in a table.

See e.g. ‘Remark profiles’

1. Structure for your readers

Table 1. Simulation profile

a) Design	
Question	Comparing the prediction ability of strategies that combine clinical and molecular variables (C and M variables)
Combinations	Seven strategies to combine C and M variables, five methods to construct a prediction model, preliminary screening (yes/no), giving 70 strategy/method/screening combinations
Strategies	Naive, Clinical offset, Favoring, Dimension reduction. All with/without clinical variable selection, apart from Naive
Methods	Boosting, Lasso, Ridge, Elastic net, Smoothly clipped absolute deviation penalty (SCAD)
Screening	Sure independent screening (SIS). We tried with iterative SIS (ISIS), but it never converged. Will be ignored
Variables	15 clinical variables (5 with and 10 without effect) 10000 molecular variables in 50 independent blocks, 28 variables with effect (see Table 2)
Correlation	Structured within blocks of C and M variables and between the blocks (no [0], moderate [0.5], strong [0.8] correlation) Nine settings (see Table 3), 3 settings presented in detail, others in the Supplementary Material.
Sample Size	500 (100 and 1000 in the Supplementary Material)
Outcome	Mean Square Prediction Error (MSPE), Sensitivity (true positive rate) and Specificity (true negative rate).

b) Results			
Setting	MSPE	Sens/spec	Remarks
B1: set 1, no correlation, no pre-screening	Table 5 for SCAD (Figure 1A) for favor.2 (Figure 1B) (ridge	For SCAD clin. dat. (Figure 3) mol. dat. (Figure 4) for favor.2	SCAD/favor.2 best performance MSPE
B2: set 2, high correlation, no pre-screening			
B3: set 3, mod. correlation, no pre-screening	excluded)	(Figure 5)	
B4: effect of pre-screening	Figure 6		Only beneficial for ridge regression

De Bin et al. doi:10.1093/bib/bbz136

2. Justify your choices

A D E M P

In explaining what you did, you should also include justifications for your choices: **why did I do this?**

‘We did 1,000 repetitions [that’s what everyone does]’

2. Justify your choices

I like to think of an analogy of simulation studies to drug development.

Think of different phases: proof-of-concept (like pre-clinical work), trying to hone a method (like dose-finding), comparison of competing methods in non-ideal situations (phase III), understanding when a method breaks (phase IV)…

3. The data-generating mechanisms you can imagine do not represent all possible DGMs

4. Know your performance measures

- Do you ever read a paper and think something is missing?
- For me, a classic is ‘the new method is unbiased and demonstrates good coverage’… which makes the reader infer that the authors are hiding its inefficiency
- Always remember that MSE and coverage depend heavily on sample size

4. Know your performance measures

Some commonly-used performance measures:

Bias Empirical SE MSE	Properties of estimator $\hat{\theta}$
Average model-based SE	Property of SE $\widehat{se}(\hat{\theta})$
Coverage Power	Properties of confidence interval
Convergence Computational speed	Computational/planning

5. If you used code, make it actually available!

The **second worst** thing authors can write is ‘code not available’

The **worst** is ‘code available on request’

Have you ever approached authors asking for their code?

5. If you used code, make it actually available!

Simulation studies are one of the few contexts where replication is a concrete thing.

If you are interested, see the **replisims.org/** initiative of Anna Lohmann, Rolf Groenwold and Kim Luijken.

6. Quantify Monte Carlo error

Simulation studies involve drawing [pseudo-]random numbers. Results will depend (to some extent) on the particular numbers that were drawn.

We need to quantify uncertainty due to using a finite number of repetitions (Monte Carlo error).

7. Neutral schmeutral

- Have you come across the ‘methodological attribution problem’?
- From Gelman’s ‘Bayesian statistics then and now’ *Statistical Science* 2010 (doi:10.1214/09-STS308)

My second meta-principle of statistics is *the methodological attribution problem,* which is that the many useful contributions of a good statistical consultant, or collaborator, will often be attributed to the statistician's methods or philosophy rather than to the artful efforts of the statistician himself or herself. Don Rubin has told me that scientists are fundamentally Bayesian (even if they do not realize it), in that they interpret uncertainty intervals Bayesianly. Brad Efron has talked vividly about how his scientific collaborators find permutation tests and p-values to be the most convincing form of evidence. Judea Pearl assures me that graphical models describe how people really think about causality. And so on. I am sure that all these accomplished researchers, and many more, are describing their experiences accurately. Rubin wielding a posterior distribution is a powerful thing, as is Efron with a permutation test or Pearl with a graphical model, and I believe that

8. Be ready to explain to people who don't know about simulation studies!

Two types of people who don't understand simulation studies:

1. Those who don't understand
2. Those who think they do understand and think it's all very intuitive

8. Be ready to explain to people who don't know about simulation studies!

Open access Communication

BMJ Open

Introduction to statistical simulations in health research

Anne-Laure Boulesteix ,[1] Rolf HH Groenwold,[2,3] Michal Abrahamowicz,[4] Harald Binder,[5] Matthias Briel,[6,7] Roman Hornung,[1] Tim P Morris ,[8] Jörg Rahnenführer,[9] Willi Sauerbrei,[5] for the STRATOS Simulation Panel

A final thought…

Some of us find it helpful to think about the analogy of simulation studies to drug development

I described one reason. The second is that it forces us to consider how we would do them if they were regulated and the burden was on us to verify to a cautious neutral party that a method does in fact work.