

Reproducible Research (RR) with Biometrical Journal (BIMJ)

Dankmar Böhning

Editor of Biometrical Journal &
Director of Southampton Statistical Sciences Research Institute (S3RI)
University of Southampton

December 6, 2017

BIMJ

- ▶ Biometrical Journal publishes articles on statistical methods and their applications in life sciences including clinical and experimental medicine, public health, environmental sciences, biology, agriculture and forestry.
- ▶ Ideally the manuscript should include a description of the problem and a section detailing the application of the new methodology to the problem.
- ▶ Case studies, review articles and letters to the editors are also welcome.
- ▶ Papers containing only extensive mathematical theory are not suitable for publication in Biometrical Journal.

co-editor: Marco Alfó (Rome)

40 Associate Editors

Reproducible Research

a quote from the BIMJ website:

"The journal strongly supports Reproducible Research. Authors are therefore vigorously encouraged to submit computer code and data sets used to illustrate new methods and to reproduce the results of the paper. Special attention should be devoted to simulation work and its preparation for reproducibility. These will be checked for reproducibility and are, after a successful check, published as Supporting Information on the journal's webpage."

what is RR?

- ▶ A result can be **reproduced** if the provided code can be used on the provided data to recreate tables, figures, etc., which were published in an article.
- ▶ **Replication** of an experiment is much stronger action, as it requires that the same findings can be derived in an independent experiment.
- ▶ We consider reproducibility to be a **first and important step** towards replicability as it is the basis to enable independent researchers to replicate the experiment described in a publication.

what is RR?

- ▶ In computer experiments such as simulation studies, replicability could, for example, be checked by modifying the code or by using the code as basis for new experiments.
- ▶ Data analysis can be replicated by using the provided code on a new data set, possibly after modifications that are required by the new data. Thus, **reproducibility** is or at least should be the minimum standard in science.

why do we support RR?

- ▶ Very helpful in controversies how certain results (simulation) were achieved.
- ▶ Code (and data) makes it easier (or even possible) to reproduce the results and thus facilitates understanding of the methods and results.
- ▶ Publicly available code helps to promote the results and methods of the article and increases scientific impact.
- ▶ Good code increases the precision of the description of methods and results and thus improves the article.
- ▶ Code facilitates communication with other researches who intend to apply the proposed methods or develop alternatives to them.

issues with RR: reproducibility

- ▶ Missing data or code 44.6%
- ▶ Code produced errors 37.5%
- ▶ Code ran but . . .
 - . . .did not reproduce results 16.1%
 - . . .did not reproduce all tables/figures 51.8%
- ▶ No seed used for RNG 25.0%
- ▶ Code and paper were difficult to match 41.1%

issues with RR: documentation/programming style

- ▶ Missing README 80.4%
- ▶ Bad documentation of code and functions 41.1%
- ▶ Unnecessary comments and code 30.4%
- ▶ Code supplied as PDF/Word/. . . 14.3%

- ▶ Usage of absolute paths 17.9%
- ▶ Bad coding style 21.4%
- ▶ Too many code files/difficult file names 10.7%

RR in practice

- ▶ papers are reviewed initially
- ▶ if major/minor revision, authors are invited to re-submit and **and informed about reproducibility policy**, and to prepare for it
- ▶ after acceptance authors are invited for the RR audit by submitting data and code according to published guidelines
- ▶ if successful then paper gets published with appropriate RR documentation

problems RR delivery

- ▶ codes do come with diversity of languages: R, SAS, STATA, MATLAB, ...
- ▶ RR editor (Fabian Scheipl) works with team of experts on these submitted codes
- ▶ not for all submitted codes expertise exists
- ▶ authors are unwilling to submit data and code
- ▶ sometimes problems are quite complex and do not allow full RR audit

in summary

RR audit helps

- ▶ to improve quality
- ▶ to improve communication
- ▶ to improve citation of works
- ▶ to improve impact
- ▶ to improve visibility