

Software Corner

Quarto: A New Tool for Reproducible Reporting

The landscape of computational statistics is dynamic, with new tools constantly reshaping how we conduct and report research. Posit (formerly RStudio) recently introduced a new open-source scientific and technical publishing system, [Quarto](#), which presents biometricians with an attractive alternative to the tried-and-tested R Markdown for creating reproducible reports or an epic leap forward if you're still using Sweave!

Quarto helps generate reproducible, production quality articles, presentations, dashboards, websites, blogs, and books in HTML, PDF, MS Word and ePub. Under the hood, [Pandoc](#) markdown is used to generate beautiful and detailed documents that include equations, citations, crossrefs, figure panels, callout blocks and much more.

Language Compatibility

One of Quarto's most notable features is its support for multiple languages like Python, Julia, and ObservableJS, in addition to R. This inclusivity means that a researcher can weave together different threads of analysis into a single narrative, regardless of the programming language used. Importantly, Quarto is not tied directly to R in any way. Quarto is its own program.

Flexible Output Formats

While the sprawling R Markdown ecosystem allows for a significant degree of document customization, Quarto offers much of this in a more coherent package. For example, the [revealjs](#) document type is a capable replacement for [xaringan](#), and Quarto natively supports a [book](#) document type which effectively replaces the capabilities of the [bookdown](#) package. There are also various [journal](#) formats already available and there is a growing range of [extensions](#) available.

How do I get started?

If you're an RStudio user Quarto comes bundled with recent versions and you can create a new Quarto document or presentation in the same way you'd create a new R Markdown document. I highly recommend the [Quartostamp](#) addin for RStudio which helps with the syntax. You can also install [Quarto](#) as a stand-alone application and use it with VS code, Jupyter Rstudio, Neovim or directly from the command line.

If you'd like to know more, see the [Quarto homepage](#) to get started and for inspiration see the [Awesome Quarto GitHub](#) repository which call itself "The most up to date curated list of Quarto docs, talks, tools, examples & articles the internet has to offer."

In Summary

Quarto offers features that could enhance the reproducibility and presentation of statistical analysis, particularly for those engaged in multi-language projects or seeking advanced customization. It represents another step forward in our collective effort to communicate statistical findings with clarity and precision in a reproducible way.

Garth Tarr
Sydney, Australia

BIOMETRIC BULLETIN
International Biometric Society Internationale Biometrische Gesellschaft Societate Internationala de Biometrie

Advertise With Us!

 GREAT OPEN RATE! Large, loyal readership. Over 40% Open Rate	 GLOBAL AUDIENCE Reach IBS Members in Over 60 Countries	 CONTACT US! Interested? Email Us! ibs@biometricsociety.org
---	---	--

STrengthening Analytical Thinking for Observational Studies (STRATOS): Six foci for the next three years

James Carpenter (1), Michal Abrahamowicz (2), Nan van Geloven (3), Paul Gustafson (4), Marianne Huebner (5), Ruth Keogh (1), Willi Sauerbrei (6), Pamela Shaw (7), Els Goetghebeur (8)

(1) Department of Medical Statistics, London School of Hygiene & Tropical Medicine, and MRC Clinical Trials Unit at UCL, London, UK.

(2) Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine and Health Sciences, McGill University, Montreal, Canada

(3) Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

(4) Department of Statistics, University of British Columbia, Vancouver, Canada

(5) Department of Statistics and Probability, Michigan State University, East Lansing, USA

(6) Institute of Medical Biometry and Statistics, Medical Center, University of Freiburg, Freiburg, Germany

(7) Biostatistics Unit of Kaiser Permanente Washington Health Research Institute, Seattle, USA

(8) Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium In August and September 2023, the STRATOS initiative held symposia at the ISCB (Milan) and the IBS Central European Network (Basel), where plans were made for the direction and foci for the next 3 years of STRATOS research. Here, we bring this thinking to the broader Biometric community, with the hope that further colleagues may support the initiative.

Background

Notwithstanding some recent improvements, major concerns remain about the methodological quality and standards of empirical research. Especially in the analysis of observational data, problems arise from poor application and reporting of statistical methodology. In the recent COVID crisis, van Calster et al (2021) discussed research waste resulting from the current organization of scientific research, which tends to prioritize the 'need to publish' over publication quality. Their proposals for change refer to initiatives aiming to improve the methodology and reproducibility of research. Among these, our Strengthening Analytical Thinking for Observational Studies (STRATOS, www.stratos-initiative.org) initiative is the only group with a primary focus on developing and disseminating evidence-supported guidance concerning statistical analyses.

Since its inception in 2013, the STRATOS initiative has grown to more than 100 members from 20 countries. Its activities are organised into 9 topic groups (TG) and 12 cross-cutting panels. Activities of all TGs and some panels were presented in articles in the Biometric Bulletin (BB) over the last 5+ years. In BB March 2022 we published a brief update on achievements of the STRATOS initiative in the last 5 years.

Future foci

In discussions during the last 6 months, including our symposia in Milan and Basel, we identified the following six foci:

I. Simulation studies

Simulation studies are key tools for validating and comparing statistical methods, and hence critical to the development of evidence-based statistical guidance. STRATOS will maintain a focus on simulation studies and prioritize improving their methodology over the coming years. We will build on (i) the ADEMP framework for simulation (Morris et al, 2019), and (ii) recently described 'phases' of methodological research (Heinze et al, 2023), to delineate a framework for (a) appropriate evaluation of new methodology at each phase of its development and (b) performing and reporting

neutral simulation studies for comparing statistical methods in relevant settings and under plausible assumptions (Boulesteix et al 2018).

2. Open science

The importance of open science is evident, but it is an extremely broad topic, and still in its infancy. A STRATOS Open Science panel, started in 2022, is working on accessible guidance for making research more transparent, reproducible and hence credible. This goes beyond reporting guidelines, e.g. including advocating routine publication of peer-reviewed analysis and simulation code as supplementary material to each research article. For more details, see the short introduction of the panel in BB February 2023.

3. Initial Data analysis (IDA)

The 'Initial data analysis' TG3 aims to improve awareness of IDA as a critical component of the research process, and develop guidance on conducting IDA in a systematic, reproducible manner. TG3 proposed a framework for this (Huebner et al 2018), and showed in a review that many IDA analyses are not reported or not done systematically, if at all (Huebner et al 2020). Current work with TG2 (Selection of variables and functional forms in multivariable analysis), highlights the impact of poor IDA for model building. Looking forward, a comprehensive range of experiences from using our IDA framework will stimulate and inform its development and hence widen its utility and applicability.

4. Machine learning (ML) enhanced statistical methods

While ML methodologies promise quick automated data driven answers to many questions, it is obvious that both ML and established statistical methodologies have their specific strengths and weaknesses. Each could benefit from the insights offered by the other. How to do that best and when is not obvious. We plan to identify the ML enhanced statistical methods that are most important for different TG's, and systematically assess their properties in realistic settings.

Causal ML is such a highly active research topic. Sophisticated integration of statistical and ML methods has enhanced robustness and efficiency through using carefully crafted estimating equations (Chernozhukov et al., 2018). We seek to refine and apply analogous integrated methodology for survival analysis and high dimensional data, and to elucidate general properties of integrated methods in prediction and diagnostics.

5. Estimands in observational data analysis

The term 'estimand' essentially refers to what is being estimated and for whom. In the trials context, the ICH E9 addendum (ICH, 2019) formally defines it in terms of five components which make for clear targets and more transparent reporting.

The insights and benefits which the estimands framework is bringing to trials research are equally needed in observational studies, where (slightly ironically) much of the relevant methodological expertise was originally developed. New estimands and adapted estimation approaches all too rarely find their way into applied health research (Liu et al (2023)). STRATOS will continue to work on bridging the gap in this area, not least in partnership with the EU IMI-SISAQOL project (<https://www.sisaqol-imi.org/>). This is developing methods and guidance for the evaluation of patient

reported outcomes in randomized trials and single arm oncology studies, where truncation due to death is a frequent complication.

6. More guidance for researchers with limited statistical knowledge and experience

At the start of the STRATOS initiative, Sauerbrei et al. (2014) highlighted that many methodological developments are not implemented in practice. Lack of guidance on practical issues is presumed to be an important hurdle. Researchers with only basic statistical knowledge and limited experience in using statistical methodology need much more help. To this end, STRATOS has provided an overview of papers published in the medical literature for such 'level 1' researchers (Wallisch et al 2022), and, importantly, has started to develop guidance in the form of Shiny apps, short videos and more. However, these projects are still in their infancy; an overview is given BB (January 2023). Discussing how to improve this work and initiate further educational contributions is an important aim for the near future. Development of a shared, curated, data set alongside these foci, STRATOS is working to develop a series of shared, curated real-world datasets, that present several analytical challenges, such as the one used in a recent pharmacoepidemiology study (Danieli et al, 2023). A procedure will be worked out to transform the resource into a fully anonymized synthetic dataset, which may be widely shared for open sciences purposes. The aim is that future STRATOS papers will use these data both to keep the methods anchored in the context where important problems co-emerge, and promote transparency and reproducibility.

Cross TG collaboration will enhance guidance development on co-occurring 'thorny issues' such as discovering data properties that result in updating planned statistical analyses, e.g. missing data, competing risks, measurement error, time-varying effects, confounding in causal questions, high dimensionality and more.

Finally, STRATOS is working on a glossary (Biometric Bulletin April 2020) including a flexible database interface. Input from other experienced methodologists and groups working on related projects are most welcome and appreciated.

Conclusion

The above foci have emerged from discussions across what is now an established international network. They have been broadly embraced by the scientific community and we believe they represent priority areas for moving observational research methodology forward. The challenge remains to effectively operationalize solutions and get these integrated into routine data science. This requires distinct guidance for researchers with different levels of statistical training. Ultimately, the provision of such guidance will also help enhancing the review process, so that fewer poor and erroneous results will be published, which should accelerate scientific progress and help restore damaged public confidence in science. We believe the six foci described above provide a stimulating and exciting prospect for the STRATOS initiative in the next 3+ years. As we seek to drive forward both development and dissemination of guidance for researchers, we look forward to partnering with a range of colleagues to progress this work: for details of how to get involved, visit the STRATOS website: we will be pleased to hear from you.

References

- Boulesteix AL, Binder H, Abrahamowicz M, Sauerbrei W for the Simulation Panel of the STRATOS Initiative (2018): On the necessity and design of studies comparing statistical methods. *Biometrical Journal*, 60, 216-218. DOI: 0.1002/bimj.201700129
- Van Calster, B., Wynants, L., Riley, R. D., van Smeden, M. & Collins, G. S. (2021). Methodology over metrics: current scientific standards are a disservice to patients and society. *Journal of Clinical Epidemiology*, 138, 219-226.
- Chernozhukov, V; Chetverikov, D, Demirer, M, Duflo, E., Hansen, C., Newey, W. and Robins, J (2017) Double/debiased machine learning for treatment and structural parameters. (2018) *The Econometrics Journal*, 21, C1-C68.
- Danieli C, Moura CS, Pilote L, Bernatsky S, and Abrahamowicz M.(2023) Importance of accounting for timing of time-varying exposures in association studies: Hydrochlorothiazide and non-melanoma skin cancer. *Pharmacoepidemiology and Drug Safety*, in press.
- Heinze G, Boulesteix AL, Kammer M, Morris TP, White IR for the Simulation Panel of the STRATOS initiative (2023): Phases of methodological research in biostatistics - Building the evidence base for new methods. *Biometrical Journal*, to appear. DOI: <https://doi.org/10.1002/bimj.202200222>
- Huebner M, Vach W, le Cessie S, Schmidt C, Lusa L. (2020) Hidden Analyses: a review of reporting practice and recommendations for more transparent reporting of initial data analyses. *BMC Med Res Meth*, 20:61. <https://doi.org/10.1186/s12874-020-00942-y>
- Huebner M, le Cessie S, Schmidt CO and Vach W (2018). A contemporary conceptual framework for initial data analysis. *Observational Studies*, 4, 171-192. <https://doi.org/10.1353/obs.2018.0014>
- International Council For Harmonisation Of Technical Requirements For Pharmaceuticals For Human Use (2019) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). Amsterdam: European Medicines Agency.
- Liu L., Choi J., Musoro J. Z., Sauerbrei W., Amdal C. D., Alanya A., Barbachano Y., Cappelleri J. C., Falk R. S., Fiero M. H., Regnault A., Reijneveld J. C., Sandin R., Thomassen D., Roychoudhury S., Goetghebeur E., and le Cessie S., on behalf of the SISAQOL-IMI Work Package. Single-arm studies involving patient-reported outcome data in oncology: a literature review on current practice. *The Lancet Oncology*, 24(5), e197-e206.
- Morris, TP, White, IR, Crowther, MJ. (2019) Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 38: 2074–2102. <https://doi.org/10.1002/sim.8086>
- Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S and Carpenter J on behalf of STRATOS initiative (2014): STRengthening Analytical Thinking for Observational Studies: The STRATOS initiative. *Statistics in Medicine*, 33: 5413-5432.
- Wallisch C, Bach P, Hafermann L, Klein N, Sauerbrei W, Steyerberg EW, Heinze G, Rauch G, topic group 2 of the STRATOS initiative (2022). Review of guidance papers on regression modeling in statistical series of medical journals. *PloS one*, 17(1), e0262918.