

## Openstatsware: who we are and what we build together

Liming Li, Roche (China) Holding Ltd.

### ABSTRACT

Openstatsware is an official working group of the American Statistical Association (ASA) Biopharmaceutical Section. Formed in August 2022, we are now a world-wide cross-industry collaborative organization of 56 members from 35 organizations. The aim of openstatsware is to engineer packages that implement important statistical methods to fill the gap in the open-source statistical software landscape and develop and disseminate best practices for engineering high-quality open-source statistical software. We now have three workstreams, including Mixed Models for Repeated Measures (MMRM), Bayesian MMRM and Health Technology Assessment (HTA). We have also held the workshop "Good Software Engineering Practice for R Packages" multiple times across the world, and the video available at YouTube.

### INTRODUCTION

Open-source software has become very popular in biostatistics, and to fill the gaps in the open-source statistical software landscape, and to bring good statistical engineering work practices to the industry, we formed openstatsware, a working group under American Statistical Association (ASA) Biopharmaceutical Section. Under openstatsware we formed multiple sub-teams to work on multiple statistical software and we collaborated to share our software engineering experience to the industry.

### INTRODUCTION TO OPENSTATSWARE

Open-source software has gained increasingly popularity in biostatistics. A lot of novel statistical methods are implemented in R, and there are unprecedented opportunities to have collaboration with methodologists and software engineers across the industry. In addition, those implementations are transparent: we can review the source code from public code repositories. However, there is large variability in the quality of these open-source software. For example, there is no statistical quality assurance on open-source software and there is no guarantee that the result you obtain from these implementations are correct.

To fill the gap between the current open-source software landscape in biopharmaceutical industry, we formed openstatsware, an official working group of the American Statistical Association (ASA) Biopharmaceutical Section in 2022. We now have 56 members from 35 organizations.

Our goal is to engineer R packages that implement important statistical methods, focusing on the methods that are highly needed for biopharmaceutical applications. In addition, we also want to develop and disseminate best practices for engineering high-quality open-source statistical software, by actively doing statistical engineering together.

### CURRENT WORKSTREAMS

Currently, there are three workstreams in openstatsware, including Mixed Models for Repeated Measures (MMRM), Bayesian MMRM and Health Technology Assessment.

MMRM workstream was formed in 2022 to develop a high-quality R package for mixed models for repeated measures (MMRM). MMRM is a very popular choice for analyzing longitudinal continuous outcomes in randomized clinical trials. Although there are already many packages available, they do not fully satisfy the users' need, for example, there are often convergence problems, the supported covariance structures are limited, the adjusted degree of freedom methods is not well supported, and the

computational efficiency is not satisfactory. To overcome these issues, we developed mmmr package with aligned best practices of software engineering, with support from the pharmaceutical industry. This package is now publicly available on CRAN.

While mmmr package is treating these repeated measures in frequentist way, Bayesian have a different view, and we lack such a software to deal with this issue. In 2023, we formed Bayesian MMRM workstream to work on this. With brms, a powerful and versatile package for fitting Bayesian regression models, brms.mmmr is created. Extensive unit tests and integration tests are added to ensure the robustness of the package. This package is also available on CRAN.

Health Technology Assessment (HTA) workstream is formed in 2024. Sponsors are often required to submit evidence of relative effectiveness of their treatment compared to relevant comparators that may not be included in their clinical trial, for HTA in different countries. Matching-adjusted indirect comparison (MAIC) method is a prevalent and well-accepted method to drive population-adjusted treatment effect in such case for two trials, one of which has Individual patient data, and the other has only aggregate data. We are now developing maicplus package to solve this problem.

## GOOD PRACTICE OF SOFTWARE ENGINEERING

During our package development work, we gain a lot of experience on good practice of software engineering. We found that there are four key components: development process, coding standards, robust test suite and documentation. For the development process, we believe that as an open-source project, we should go public as soon as possible, and never wait until the product is finished. Other good-hands might be interested and join your work. Public code repositories like github is a good place to open-source your project, and we can also benefit from the version control system, which is the corner stone of effective collaboration. For coding standards, we should use consistent and readable code style to simplify joint work, and a written contribution guideline will help a lot. Robust test suite is crucial to quality of the package. We should use both unit tests and integration tests to prevent possible regressions. Continuous integration and continuous development will simplify this work. Documentation, while it requires lots of work, are extremely important. It has to be kept up-to-date, but it helps to build trust with the users, and also help provide developers a good reference over time.

To disseminate these good practices, we held multiple workshops across the world, and we also have a YouTube channel about tips and tricks in good statistical software engineering practices.

## CONCLUSION

Our working group, openstatsware, is working on statistical methods that is important for biopharmaceutical industry to fill in gaps in the open-source statistical software landscape, and we are also sharing our experience with the industry to improve the way software engineering is done, to help improve the efficiency, reliability and innovation within Biostatistics.

## REFERENCES

Daniel Sabanes Bove, et al. "openstatsware." <openstatsware home page>. <2024-07-05>. Available at <<https://www.openstatsware.org>>

Ya Wang. "Introducing openstatsware." <openstatsware home page>. <2024-02-23>. Available at <<https://www.openstatsware.org/slides/asa-biop-webinar-feb2024-quarto.html>>.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Liming Li  
Roche (China) Holding Ltd.  
liming.li@roche.com