66

Session: Statistical Software Engineering

Wilmar Igl[1], Pravin Madhavan[2], Isaac Gravestock[3], Brian Lang[4]
[1]ICON PLC, UPPSALA, Sweden. [2]Berry Consultants, Oxford, United Kingdom. [3]Roche, Basel, Switzerland. [4]MSD, Zurich, Switzerland

Wilmar Igl
Please provide a brief biography for the Presenting author(s)
Wilmar Igl, PhD, is a statistician, psychologist and former researcher at the interface between statistics and data science with over 22 years of experience in biomedicine.
He has been working as a Biostatistics Consulting Director at ICON, Sweden, since November 2022. Here, he provides biostatistical consulting services to biotech and pharma companies. Before he started this role, he served as a statistical assessor at the Swedish Medical Products Agency between 2018 and 2022. He collected experience in the pharma industry between 2010 and 2012 at Bayer Pharma (Berlin, Germany) and at AstraZeneca (Cambridge, UK) between 2015 and 2017.
His main interests include clinical trial designs (including adaptive designs), Bayesian statistics, patient-reported outcomes, open-source software, and statistical programming.

Pravin Madhavan
Please provide a brief biography for the Presenting author(s)
Pravin Madhavan is a Senior Software Engineer at Berry Consultants where he is responsible for the development of various clinical trial simulation software; including FACTS (the Fixed and Adaptive Clinical Trial Simulator), ADDPLAN (Adaptive Design Planner) and QUOTES (Quantification and Optimization of Trial Expectations Simulator). His interests lie in software engineering and DevOps best practices in scientific computing.

Isaac Gravestock
Please provide a brief biography for the Presenting author(s)
Isaac Gravestock PhD is a statistician and software developer who leads the Statistical Engineering team at Roche. His interests include Bayesian methodology for external and hybrid controls, causal inference for observation data using Target Trial Emulation and statistical computing. Previously, he was part of the Real World Data Science group at Roche for 5 years designing and analysing studies using real-world and observational data.

Brian Lang
Please provide a brief biography for the Presenting author(s)
Brian Lang, Ph.D., is a Statistician at MSD in the Health Technology Assessment (HTA) space. He also leads teams developing research software and process automation using R and R Shiny. As a workstream lead on standard toolchain and CI/CD, Brian focuses on improving development processes and training his team members, including statisticians and statistical programmers, in software development practices.

Single topic, multi-speaker session, Workshop or Single presentation submission

A single topic, mutli-speaker session/workshop

Single topic session or workshop abstracts

The session is organized by the Openstatsware SIG (see https://www.openstatsware.org).

“The Mythical Man Month (1975-2025) – Planning, Implementing, and Managing Statistical Software Projects”
Wilmar Igl, PhD, ICON PLC, Uppsala, Sweden

Fifty years have passed since Brooks (1975) wrote his classic book entitled “The Mythical Man-month: Essays on Software Engineering”. Brooks laid out fundamental principles of and his personal insights into software engineering and management, of which Brooks’ Law (“Adding manpower to a late software project makes it later.") has become most popular. Here, an overview of the current state-of-the-art of planning, implementing, and managing software engineering projects will be presented with a focus on developing open-source statistical software in the pharmaceutical industry.

The central work by Brooks will be used as a starting point and complemented with modern concepts of software engineering and management including DevOps and Agile software development (e.g., Scrum) and, thereby, highlighting established truths and novel insights. Special attention will be given to the estimation of resources, timelines and quality metrics of statistical software within the larger context of ultimately managing human resources (“peopleware”). Selected examples of projects developing open-source statistical software in the pharmaceutical industry which illustrate the described concepts from a practical perspective will be presented including problems and solutions.

Following the adage by Watt S. Humphrey, the father of software quality, who said “Every business is a software business”, the pharmaceutical industry is on the verge of experiencing a fundamental change of generating, analyzing and presenting the required evidence for regulatory approval of their products from commercial, closed-source or within-company, custom-specific software to cross-industry, open-source software. This talk will help to provide statisticians, statistical programmers and their managers with concepts to navigate this transformation.

“Continuous Integration (CI) practices for statistical software development”
Pravin Madhavan, PhD, Berry Consultants, Oxford, UK

When creating a large software package, one of the main issues we face is maintaining the quality of the software over time; in particular, its correctness and its release frequency. This is of particular importance as the development team inevitably grows and/or changes over time. In this talk I will describe my experience with Continuous Integration (CI), which is the practice of automatically integrating code changes into a shared repository multiple times a day. CI is a crucial practice in modern (statistical) software development, promoting early defect detection, faster feedback loops, and improved collaboration within teams, and thus ensuring that development processes remain predictable, streamlined and scalable.

This talk explores some of the key aspects of CI and shares findings from years of experience of developing large commercial statistical software solutions; including automated testing, build automation, and version control. It will cover best practices for successful CI adoption, such as integrating frequent commits, maintaining a robust test suite, and ensuring clear communication within teams. In particular, it will talk about a simple automated testing strategy used in large commercial statistical software solutions that can be used in open-source software packages.

“Scaling Statistical Innovation and Open Source Collaborations”
Isaac Gravestock, PhD, Roche, Basel, Switzerland

“It is a truth universally acknowledged, that a pharmaceutical industry statistician in possession of a good statistical method, must be in want of a user friendly software implementation.”

Statistical innovations do not gain broad adoption in industry without robust and user friendly software implementations. Rufibach *et al* (2024) give examples demonstrating how high-quality user friendly software and external collaboration have been integral to the adoption of statistical methods in pharmaceutical companies.

More broadly, Heinze *et al* (2024) argue that novel statistical methods need to have a good evidence base built around them before they can be adopted and that robust and user-friendly software implementations are essential to conduct the simulation studies and comparisons with existing methods to build this evidence.

In our experience through the openstatsware working group (and more broadly the openpharma collaborative space), industry and academic development collaborations have been especially fruitful in building software packages which have led to the adoption of novel statistical methods. Open source collaborative development has allowed us to avoid bureaucratic contracting hurdles and build software that works in different companies' business and technical settings, thus avoiding any one company's idiosyncrasies.

We share some examples of how software has enabled or boosted adoption of statistical methods:

- Adverse event rates in ongoing blinded trials
- Matching adjusted indirect comparisons
- Dynamic borrowing with hybrid controls
- Extrapolation based paediatric sample size calculation
- Reference based multiple imputation

"Analysis Specification to Execution in R/Shiny"

Brian M. Lang[1], David Maher-McWilliams[2], Gianluca Mortari[1]
[1]MSD, Zurich, Switzerland
[2]MSD (UK) Limited, London, UK

In the pharmaceutical industry, a significant amount of statistical analysis involves repeated analyses, with each organization adopting its own approach to standard tables, listings, and figures (TLFs). Accurate and efficient TLF generation is critical for regulatory submissions as well as the health technology assessment decision-making processes. The journey from specifying the required analysis, including population and endpoints, to the execution of the code which creates these TLFs, is a collaborative effort between statisticians and statistical programmers.

To streamline this collaboration and the overall analysis and reporting process, we adopted a high-level, database-first approach to study analysis planning and execution. Despite deep knowledge of clinical data and analysis and reporting knowledge, carrying out this effort internally is a large challenge. However, by following rigorous software development standards, including modular programming, version control, and best practices in DevOps, we developed a module-based R/Shiny application. This application, paired with a transformation of our standard analysis TLFs and their generating code into a robust database structure, creates an explicit link between the analysis specifications and the multilingual code execution needed to fulfill them. This method not only enhances traceability and efficiency but also enables easy updates and scalability for future analyses.

This presentation will include an illustration of the database we developed and the app we've created to interact with it, highlighting key features such as analysis specification and code execution. We will also discuss the challenges of such a large undertaking and how we strive to follow best practices for software development.