

# Clinical Statistical Reporting in a Multilingual World



## Project Scope

Several discrepancies have been discovered in statistical analysis results between different programming languages, even in fully qualified statistical computing environments. Subtle differences exist between the fundamental approaches implemented by each language, yielding differences in results which are each correct in their own right. The fact that these differences exist causes unease on the behalf of sponsor companies when submitting to a regulatory agency, as it is uncertain if the agency will view these differences as problematic. Understanding the agency's expectations will contribute significantly to enabling the broader adoption of multiple programming languages in the production of data submission packages for regulatory review

The Clinical Statistical Reporting in a Multilingual World project seeks to clearly define this problem and provide a framework for assessing the fundamental differences for a particular statistical analysis across languages. In this context, the risk of interpreting numerical differences in analysis results due solely to differences in programming language can be mitigated, instilling confidence in both the sponsor company and the agency during the review period. This will be accomplished by:

1. Identifying common statistical analyses performed during submissions to narrow the scope of where discrepancies must be identified (e.g., continuous summaries, frequency counts, hazard models, bioequivalence testing, steady-state assessments, bioavailability testing, ANOVA)
2. Providing necessary documentation to produce equivalence in results between separate statistical analysis software packages/languages (where possible)
3. Evaluating and documenting differences in results between popular statistical analysis implementations as use cases
4. Provision of sample code for use cases through a publicly accessible code repository for both review and consumption
5. Promoting the notion that the 'right' implementation of a particular statistical analysis should be based sound statistical reasoning and not limited by the capabilities of a specific programming language or statistical analysis software package, nor its default settings

Project Leads	Email
Michael Rimler (GSK)	<a href="mailto:michael.s.rimler@gsk.com">michael.s.rimler@gsk.com</a>
Mike Stackhouse (Atorus Research)	<a href="mailto:mike.stackhouse@atorusresearch.com">mike.stackhouse@atorusresearch.com</a>
Paula Rowley (PHUSE Project Assistant)	<a href="mailto:paula@phuse.global">paula@phuse.global</a>

Objectives & Deliverables	Timelines
GitHub Repository documenting identified differences between statistical analysis implementations (based on R and SAS use cases as a starting point)	Q22022
Expand repository to provide comparable syntax across languages (based on R and SAS use cases as a starting point)	Q22022
Expand GitHub repository to incorporate Python and /or Julia	Q22022
White Paper providing framework for addressing language discrepancies in statistical analysis implementations, including specific use cases as examples	Q32022

Project Members	Organisation
Aiming Yang	Merck, Sharp & Dohme
Amrit Singh	Bayer
Andy Miskell	Eli Lilly
Andy Nicholls	GSK
Brian Varney	Experis
Chung-kai Sun	Janssen Research & Development
Clara Beck	Chrestos
Doug Thompson	GSK
Harshal Khanolkar	Novo Nordisk
Joseph Rickert	RStudio
Karnika Dalal	Regeneron
Kevin Putschko	Experis

**CURRENT STATUS** Q12022

Started draft writing of White Paper with new sub-team. Group are now combining White Paper draft pieces together for wider team review.

Kyle Lee	FDA
Ke Wang	Novartis
Michael Kane	Yale University
Matthew Kumar	Bayer
Mia Qi	Janssen Research & Development
Min-Hua Jen	Eli Lilly
Soren Lophaven	Omicron
Suhas Ravindra	GSK
Steve Walker	Experis