

Comparing Analysis Method Implementations in Software (CAMIS)



Project Scope

Comparing Analysis Method Implementations in Software (CAMIS) has evolved from the Clinical Statistical Reporting in a Multilingual World project.

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 CAMIS 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

The [CAMIS repository](#) to document known differences is now live and open for community contributions.

Project Leads		Objectives & Deliverables	Timelines
Lyn Taylor, <i>Parexel</i>	lyn.taylor@parexel.com	Expand repository to provide comparable syntax across languages (based on R and SAS use cases as a starting point)	Q2 2023
Christina Fillmore, <i>GSK</i>	christina.e.fillmore@gsk.com	Expand GitHub repository to incorporate Python and /or Julia	Q3 2023
Harshal Khanolkar, <i>Novo Nordisk</i>	hlsk@novonordisk.com	White Paper providing framework for addressing language discrepancies in statistical analysis implementations, including specific use cases as examples	Ongoing
Kerry Robson, <i>PHUSE Project Assistant</i>	kerry@phuse.global		
CURRENT STATUS Q1 2024			
Launched CAMIS-ONCO. Expanded membership and industry awareness of the project. Increased content loaded into the repository CAMIS (psiaims.github.io)			