

Abstract

The recently revised *Study Data Technical Conformance Guide* (Section 4.1.2.10) states that sponsors should provide software programs used:

- to create ADaM datasets,
- to generate tables and figures for primary and secondary efficacy analyses, and
- to generate prescribing information presented in product labels.

Submission of sponsor software programs contributes to traceability and reviewer understanding of the derivation of analysis data set variables, and enhances transparency of the results obtained from algorithms and tests employed by the sponsor. We shall explore the guidelines in section 4.1.2.10 in greater depth. To further enhance readability and reproducibility, we recommend that sponsor submissions conform to good programming practices as developed by a PhUSE working group, and incorporate best practices from the reproducible research community.

Study Data Technical Conformance Guide

4.1.2.10 Software Programs Sponsors should provide the software programs used to create all ADaM datasets and generate tables and figures associated with primary and secondary efficacy analyses. Furthermore, sponsors should submit software programs used to generate additional information included in Section 14 CLINICAL STUDIES of the Prescribing Information (PI) if applicable. The specific software utilized should be specified in the ADRG. The main purpose of requesting the submission of these programs is to understand the process by which the variables for the respective analyses were created and to confirm the analysis algorithms. Sponsors should not submit software programs with executable file extensions, and these programs should be submitted in ASCII text format.

Need for Sponsor Programs**Why do FDA reviewers need sponsor programs?**

- 1. Traceability.** Can the ADaM data sets be traced back to SDTM data, and CRFs?
- 2. Transparency.** Study protocols and Statistical Analysis Plans (SAPs) may not completely specify implementation, while the code does.
- 3. Reproducible and Repeatable Results.** Can FDA reviewers generate the same results as the sponsor? If not, why not? FDA reviewers need to be confident that labelling information is accurate and an appropriate representation of the available evidence.

Recommendations**Observe Good Programming Practices:**

1. Write all programs and documentation in grammatically correct English.
2. Include detailed headers at the beginning of each program. (see below).
3. Comment liberally (endeavor to make code self-documenting).
4. Practice version control.
5. Document all dependencies such as macros and packages. Use readme files. Flowcharts can be helpful.
6. Specify language, version and build.
7. Specify Operating System and hardware configuration.
8. Document any configuration file changes, document changes to the BLAS (Basic Linear Algebra System) if appropriate.
9. Use standard naming conventions for variables (CDISC) and follow a style guide (e.g. The tidyverse style guide, <http://style.tidyverse.org/>).
10. Do not submit code that contains errors. Avoid spaghetti code.

Documentation in Program Headers

The PhUSE Good Programming Practice Guidance suggests that program headers can be an important component of submission documentation.

The following should be included in all program headers:

- Identification of the project of which the program is a part.
- Program name.
- Author identification which should be human readable and unique.
- Short description of program purpose.
- List of macros used in the program.
- Date program was first put into production, was finalized, or first passed validation.
- This date will be chosen based on the operational procedures used within the company /organization creating the program. The date should indicate the first date when the program was released for final use."

Source: http://phusewiki.org/wiki/index.php?title=Good_Programming_Practice_Guidance

Reproducible Research Considerations**Make Results Repeatable and Reproducible.**

1. Be a seed saver, save the seed for random simulations, including bootstrap estimates, resampling, etc.
2. Don't claim more accuracy for your simulations than they can reasonably provide; if necessary increase the number of simulations.
3. No algorithm or method is perfect, and all have limitations. Can results of one algorithm be cross checked with another?
4. Avoid hard coding data and/or variables. Beware of cutting and pasting.
5. Workflow planning. Plan to make the end results both transparent and reproducible. Take the time to do it right, rather than just doing it quickly.
6. Modular programs that do not call macros or other programs are easier to use when verifying results.

Conclusions

As trial designs become more complicated, and simulations play a greater role in analyses for submissions functional code and programs assume greater importance. Providing quality, well documented code to support sponsor analyses and results is mutually beneficial to both FDA and sponsors. Observing good programming practices and incorporating reproducible research methods can improve the quality of submitted programs.

Poorly documented, poorly written, erroneous, or incomplete code can be indicators of poor quality control which may result in multiple information requests, additional meetings, and delays in the review clock.