

Automation of ADaM Dataset & TFL Visualization Through an R Package – GenAdam2

Mahesh Divakaran, Genpro Research, Trivandrum, India
Mohin K Mohanan, Genpro Research, Trivandrum, India
Limna Salim, Genpro Research, Trivandrum, India

ABSTRACT

In recent years, research studies involving R software has become a prevalent topic in the pharmaceutical industry. Its extensive library of user-defined packages including methods that effectively handle complicated data sets and generate tables, figures, and listings makes it perfect for Biostatisticians and programmers in the pharmaceutical and biotechnology industries. Within the scope of this research, we present an R package capable of generating ADaM domains with a single R function. GenADaM2 also allows the user to analyze ADaM or visualize TFLs using the R shiny, and from within the program, allows us to apply filters and save the data in several formats, such as .docx, .rtf, .html, .pdf, etc.

INTRODUCTION

The pharmaceutical industry relies heavily on data-driven decision-making processes to develop and launch new drugs. Clinical trials generate large amounts of data, and the analysis of this data is critical to determine the safety and efficacy of drugs. Clinical Data Interchange Standards Consortium (CDISC) has developed a set of standards for organizing and presenting clinical trial data to ensure that data is consistent, complete, and accurate. The Analysis Data Model (ADaM) is one such standard that provides a framework for organizing and analyzing clinical trial data.

Over the past few years, R has become a popular choice for data analysis and visualization in the pharmaceutical and biotechnology industries. Its powerful library of user-defined packages provides efficient and effective ways to handle complex data sets and generate tables, figures, and listings. The GenADaM2 R package was developed to address the challenges faced by biostatisticians and programmers in the pharmaceutical industry in generating ADaM domains, analyzing ADaM data, and visualizing Tables, Figures, and Listings (TFLs).

In this paper, we present an overview of the GenADaM2 R package and demonstrate its functionality in generating ADaM domains with a single R function, analyzing ADaM data, and visualizing TFLs using R Shiny. We also describe the different file formats in which data can be exported and how the package adheres to CDISC standards. We compare our package to other similar packages available in R, and provide examples of how GenADaM2 can be used to facilitate the analysis of clinical trial data.

The remainder of the paper is organized as follows: Section 1 provides an overview of the ADaM standard and the need for automation. Section 2 describes the GenADaM2 R package and its features. Section 3 provides examples of using the package to generate ADaM domains, analyze ADaM data, and visualize TFLs. Finally, Section 4 provides a summary of the paper and future directions for the GenADaM2 package.

BACKGROUND

Throughout clinical trials, the pharmaceutical sector produces enormous amounts of data. To get regulatory agencies to approve new pharmaceuticals, this data must be standardized and presented. ADaM is a standard format created by the Clinical Data Interchange Standards Consortium for the submission of clinical trial data (CDISC). In order to construct analytical datasets, ADaM defines a number of variables and data formats. Subsets of the clinical data utilized for statistical analysis are called analysis datasets.

A vital task for programmers and biostatisticians is the creation of ADaM databases. It entails converting unprocessed data into analytical datasets that meet the ADaM standard. When dealing with enormous amounts of data, creating ADaM datasets can be time-consuming and difficult. In order to generate the datasets, the procedure also needs program writing skills and a solid comprehension of the ADaM standard.

For programmers and biostatisticians working in the pharmaceutical sector, R software has become a vital tool. It is ideal for handling challenging data sets and producing tables, figures, and listings due to its substantial library of user-defined packages. The analysis of clinical trial data makes considerable use of R software.

GENADAM

SAS offers a potent programming tool called Macro that lets us avoid writing repetitive code and reuse it repeatedly when required. Additionally, it aids in the creation of dynamic variables within the code that can accept various values depending on the run instances of the same code. Similarly, R has dynamic functions, which makes the programmer's life easier. An R package collects R libraries, datasets, and documentation. R is an open-source programming language allowing users to create and publish their packages. Many packages are available today to derive parts of clinical data

reporting, which helps reduce programming time and effort. So we introduced the GenAdam package last year which allows us to generate the standard ADaM variables by just a function call. The GenAdam R package can be a helpful substitute for constructing clinical trial ADaM datasets by calling a single function along with visualizing the resulting dataset in a shiny app interface.

GENADAM2

GenADaM2 is an R package that automates the generation of ADaM datasets. It entails converting unprocessed data into analysis datasets that adhere to the ADaM standard. When working with huge amounts of data, creating ADaM datasets can be time-consuming and difficult. Additionally, the procedure needs program writing skills and a solid comprehension of the ADaM standard in order to generate the datasets.

For programmers and biostatisticians in the pharmaceutical sector, R software has evolved into a crucial tool. Its vast library of user-defined packages makes it ideal for managing challenging data sets and producing tables, figures, and listings. The analysis of clinical trial data makes substantial use of the R software.

VISUALIZATION OF TFLS

GenADaM2 supports the creation of ADaM datasets as well as TFL (Tables, Figures, and Listings) display using the R Shiny program. Filtering, sorting, and visualization are all possible with the app while analyzing ADaM datasets interactively. Histograms, box plots, and scatter plots are just a few of the visualization types that the program supports. In the pharmaceutical industry, being able to visualize data is crucial. It makes it possible for programmers and biostatisticians to comprehend the data and spot patterns and trends more easily. Outliers can be found and many scenarios can be explored thanks to the data's interaction capabilities.

SESSION 1: OVERVIEW OF THE ADAM STANDARD AND THE NEED FOR AUTOMATION

The Analysis Data Model (ADaM) standard was developed by the Clinical Data Interchange Standards Consortium (CDISC) to provide a consistent structure and format for clinical trial data analysis. The ADaM standard includes a set of standard variables that define the structure of the data and facilitate the analysis of clinical trial data. The standard variables include subject-level variables, visit-level variables, and treatment-level variables.

The pharmaceutical and biotechnology industries have been using R software for the analysis of clinical trial data. R's extensive library of user-defined packages includes methods that effectively handle complicated data sets and generate tables, figures, and listings, making it perfect for Biostatisticians and programmers in the pharmaceutical and biotechnology industries. However, the process of generating ADaM datasets and visualizing TFLs can be time-consuming and error-prone, especially when dealing with large and complex datasets.

This is where the GenADaM2 R package comes into play. It is an R package specifically designed to automate the generation of ADaM datasets and visualization of TFLs. The package provides an easy-to-use interface that allows users to generate ADaM datasets and analyze them efficiently, saving time and reducing the likelihood of errors.

SESSION 2: OVERVIEW OF THE GENADAM2 R PACKAGE AND ITS FEATURES

The GenADaM2 package is an R package that provides an easy-to-use interface to generate ADaM datasets and visualize TFLs. With a single function, `gen_adam()`, the package generates an ADaM dataset with standard variables and derived variables. The package also provides an easy-to-use interface to visualize TFLs and apply filters to the data. The data can be saved in several formats, such as .docx, .rtf, .html, .pdf, etc.

The package comes with two main functions: `gen_adam()` and `tfl_viz()`. The `gen_adam()` function generates the ADaM dataset with standard variables and derived variables, given the specification path and the name of the ADaM dataset. The `tfl_viz()` function generates TFLs using R shiny. It opens a shiny dashboard where users can create TFLs and apply filters to the data.

The GenADaM2 package is compatible with R 4.2 and depends on the following packages: data.table, shiny, shinythemes, shinyjs, tidyverse, ggplot2, Plot.ly, haven, shinyWidgets, DT, reshape2, openxlsx, tidyr. The package is available under the Apache License (≥ 2).

SESSION 3: EXAMPLES OF USING THE GENADAM2 PACKAGE

The GenADaM2 package provides an easy-to-use interface to generate ADaM datasets and visualize TFLs. In this section, we will provide some examples of how to use the package to generate ADaM domains, analyze ADaM data, and visualize TFLs.

Example 1: Generating an ADaM Dataset using the GenADaM2 Package

The following code demonstrates how to generate an ADaM dataset using the `gen_adam()` function in the GenADaM2 package:

```
# Load the GenADaM2 package
library(GenADaM2)

# Set the path to the ADaM specification file
spec_path <- "/path/to/specification/file.xml"
```

```
# Generate the ADaM dataset
adam_name <- "adam_dataset"
gen_adam(adam_name, spec_path)
```

Example 2: Analyzing ADaM Data with GenADaM2

In this example, we demonstrate the use of GenADaM2 for analyzing ADaM data. For this purpose, we will use a sample ADaM dataset generated using the `gen_adam()` function of the package. The dataset contains data on the efficacy of a drug in a clinical trial, with endpoints such as mean change in a particular measure from baseline to week 4 and week 8.

We start by loading the dataset into R using the `read_xpt()` function:

```
adam_data <- read_xpt("sample_adam.xpt")
```

Once the data is loaded, we can use various functions from the `dplyr` package to perform data manipulations and aggregations. For example, to calculate the mean change in the measure from baseline to week 4 and week 8, we can use the following code:

```
library(dplyr)
adam_data %>%
  group_by(TRTP) %>%
  summarise(mean_change_wk4 = mean(MEASURE_WK4 - MEASURE_BASELINE),
            mean_change_wk8 = mean(MEASURE_WK8 - MEASURE_BASELINE))
```

This code groups the data by treatment (TRTP) and calculates the mean change in the measure from baseline to week 4 and week 8 for each treatment group.

In addition to data manipulation and aggregation, GenADaM2 provides several functions for data visualization. For example, we can use the `ggplot2` package to create box plots showing the distribution of the measure at each visit by treatment group:

```
library(ggplot2)
ggplot(adam_data, aes(x = VISIT, y = MEASURE, fill = TRTP)) +
  geom_boxplot() +
  facet_wrap(~TRTP)
```

This code creates a box plot for the measure at each visit, with treatment group (TRTP) determining the fill color. The `facet_wrap` function is used to create separate plots for each treatment group.

In addition to static visualizations, GenADaM2 also provides an easy-to-use interface for creating interactive TFLs using the `tfl_viz()` function. This function opens a Shiny dashboard where users can select variables and specify filter criteria to generate custom TFLs. We can use this function to create a TFL without any code, just drag and drop the variables.

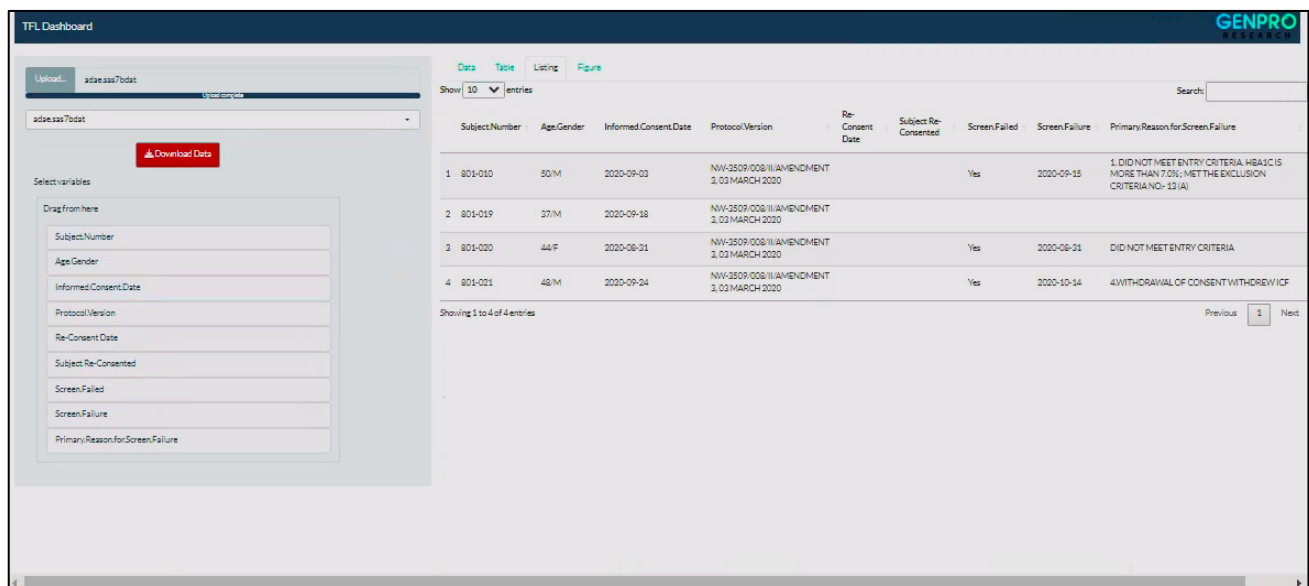


Figure 1: Sample output of Listing using `tfl_viz()` in R Shiny

Listing 16.2.1.1
Subject Enrollment
All Subjects Screened

Subject Number	Age/Gender	Informed Consent Date	Protocol Version	Re-Consent Date	Protocol Version Subject Re-Consented	Screen Failed	Date of Screen Failure	Primary Reason for Screen Failure
801-010	50/M	2020-09-03	NW-3509/008/I I/AMENDMENT 3, 03 MARCH 2020			Yes	2020-09-15	1. DID NOT MEET ENTRY CRITERIA. HBA1C IS MORE THAN 7.0% ; MET THE EXCLUSION CRITERIA NO.- 13 (A)
801-019	37/M	2020-09-18	NW-3509/008/I I/AMENDMENT 3, 03 MARCH 2020					
801-020	44/F	2020-08-31	NW-3509/008/I I/AMENDMENT 3, 03 MARCH 2020			Yes	2020-08-31	DID NOT MEET ENTRY CRITERIA
801-021	46/M	2020-09-24	NW-3509/008/I I/AMENDMENT 3, 03 MARCH 2020			Yes	2020-10-14	4.WITHDRAWAL OF CONSENT WITHDREW ICF

Figure 2: Sample Listing Output generated by Genadam2

Overall, these examples demonstrate the flexibility and usefulness of GenADaM2 for analyzing ADaM data and generating custom TFLs.

SESSION 4: SUMMARY AND FUTURE DIRECTIONS

In this paper, we introduced the GenADaM2 R package, which provides an efficient and automated solution for generating ADaM datasets and analyzing clinical trial data. The package has several features, including the ability to generate standard variables and derived variables, visualize TFLs, and save data in various formats.

We demonstrated how easy it is to use the GenADaM2 package to generate ADaM domains and analyze clinical trial data. With just a few lines of code, the package can handle complex data sets and provide users with results in multiple formats. This automation can greatly reduce the time required for data cleaning and analysis, which can ultimately speed up the drug development process.

As for future directions, we plan to continue developing the GenADaM2 package to further enhance its capabilities. We are working on incorporating additional features such as generating CDISC-compliant outputs and integrating other clinical data standards. We also plan to expand the package's compatibility with other R packages and data sources.

In conclusion, the GenADaM2 package provides a user-friendly solution for generating ADaM datasets and analyzing clinical trial data. Its ease of use and automation can greatly reduce the time and effort required for data cleaning and analysis, ultimately leading to faster drug development and improved patient outcomes.

EXPORTING DATA

GenADaM2 allows for the export of TFLs in several formats, such as .docx, .rtf, .html, and .pdf. The package includes support for the creation of tables and listings in these formats, making it easy to share results with colleagues and regulatory authorities.

The ability to export data in different formats is essential in the pharmaceutical industry. It allows for the sharing of results with different stakeholders who may have different requirements. For example, regulatory authorities may require results to be submitted in a specific format, while colleagues may prefer to view results in a different format.

CONCLUSION

The use of R software has revolutionized the way pharmaceutical companies conduct their clinical trials. The ability to write code in R allows biostatisticians and programmers to automate the analysis of data and generate results quickly. The GenADaM2 package is an example of how R software can be used to automate the generation of ADaM datasets and the analysis of clinical trial data.

The GenADaM2 package significantly reduces the time required to generate ADaM datasets and ensures that the datasets conform to the ADaM standard. The package also allows for the visualization of TFLs using the R Shiny app, making it easy to interact with data and identify patterns and trends.

The ability to export data in different formats is essential in the pharmaceutical industry, allowing for the sharing of results with different stakeholders who may have different requirements. The GenADaM2 package includes support for the export of ADaM datasets and TFLs in several formats, making it easy to share results with colleagues and regulatory authorities.

Validation is an essential step in the generation of ADaM datasets. The GenADaM2 package includes support for the validation of ADaM datasets, saving time and reducing the risk of errors.

In conclusion, the GenADaM2 package is a valuable tool for biostatisticians and programmers in the pharmaceutical industry. It significantly reduces the time required to generate ADaM datasets and ensures that the datasets conform to the ADaM standard. The package also allows for the visualization of TFLs, the export of data in different formats, and the validation of datasets, making it an essential tool for the analysis of clinical trial data.

REFERENCES

1. SAS Institute. (2017). ADaM Implementation Guide. Cary, NC: SAS Institute.
2. R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
3. Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Sebastopol, CA: O'Reilly Media.
4. Kuhn, M. (2021). caret: Classification and Regression Training. R package version 6.0-88. URL: <https://CRAN.R-project.org/package=caret>.
5. Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2021). shiny: Web Application Framework for R. R package version 1.7.1. URL: <https://CRAN.R-project.org/package=shiny>.
6. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer.
7. ICH. (1996). Guideline for Good Clinical Practice E6(R1). Geneva, Switzerland: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use.
8. FDA. (2020). Clinical Data Interchange Standards Consortium (CDISC) Standards. Retrieved from <https://www.fda.gov/industry/fda-resources-data-standards/clinical-data-interchange-standards-consortium-cdisc-standards>
9. Mahesh, D., & Mohin, K.M. (2022). SM02: Automation of Standard ADaM Variables Through an R Package – GenAdam. Paper presented at the PHUSE EU Connect, Belfast. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/PAP_SM02.pdf.

ACKNOWLEDGMENTS

We would like to thank the management teams from Genpro Research, who supported this research. This paper and the research behind it would not have been possible without the exceptional support of our managers Roshan Stanly (Manager - Clinical Data Analytics), Ajith B Nair (Senior Manager - Clinical Data Standards), Limna Salim (India Head - Biometrics & Operations), Anoop P Ambika (CTO), and Sachin Marulkar (President & CEO). We would also like to acknowledge our colleagues Pranav M Sreejith, who helped us write and review this paper.

RECOMMENDED READING

Mahesh, D., & Mohin, K.M. (2022). SM02: Automation of Standard ADaM Variables Through an R Package – GenAdam. Paper presented at the PHUSE EU Connect, Belfast. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/PAP_SM02.pdf.

CONTACT INFORMATION

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

Author Name : Mahesh Divakaran
Company : Genpro Research
Address : Second Floor, Nila Building
Technopark Campuss
City / Postcode : Thiruvananthapuram , Kerala – 695581
Work Phone : +91.4714.026.700
Email : mahesh.divakaran@genproresearch.com
Web : www.genproresearch.com

Co-Author Name : Mohin K Mohanan
Company : Genpro Research
Address : Second Floor, Nila Building
Technopark Campus
City / Postcode : Thiruvananthapuram , Kerala – 695581
Work Phone : +91.4714.026.700
Email : mohin.mohanan@genproresearch.com
Web : www.genproresearch.com

Brand and product names are trademarks of their respective companies.