Paper DV02

Automating TLG Generation for Clinical Trials Leveraging R/R Shiny Technology

Margaret Wishart, Bristol Myers Squibb, Lawrenceville, United States
Vijay Sharma, Bristol Myers Squibb, Lawrenceville, United States

ABSTRACT

Scalability, consistency, and efficiency in creating Tables, Listings, and Graphs (TLGs) have proven impactful within Full Development studies across Therapeutic Areas (TA). This paper presents a use case for clinical programming to rapidly build an application utilizing R Shiny to automate TLG generation. The foundation of the increased efficiency in quickly generating outputs depends on pre-defined global and TA-level standards. To achieve standardization, a Governance Team was instituted with TA-level experts to define standards and maintain a library of outputs. Leveraging an R-based interactive web tooling to utilize standard output templates on a study-by-study basis with a point-and-click selection mechanism. Introducing R within the industry has created a need for commonly used statistical methods to be defined and preserved for uniformities across studies, requiring experts with cross Therapeutic Area knowledge. Such open-source technologies aid in building systems that supplement current tools and packages while maintaining the original processes and methodologies.

INTRODUCTION

Tables, Listings, and Graphs (TLGs) are essential components throughout the life of clinical trials. TLGs provide study teams with pivotal information regarding the safety and efficacy of a respective drug. The study team relies on TLGs to guide decision-making and provide evidence in support of a decision. Numerous rationales exist for TLGs due to many contributing factors within the life of a clinical trial.

Traditional methods of producing TLGs can be time-consuming, often requiring complex coding, manual data manipulation, and output generation processes. Extended turnover times can be critical, especially when quick data analysis and decision-making are vital.

The pharmaceutical industry is continuously changing and experiencing significant transformations, propelled by advancements in emerging technologies aimed at creating efficiencies that aid in the development of innovative therapies that enhance patients' lives. These changes have resulted in the creation of new products and services that automate manual tasks. Consequently, Bristol Myers Squibb (BMS) has started integrating open-source technologies, such as R. The R programming language, coupled with the Shiny and other add-on R packages, can create interfaces that are easy to use. These interfaces can be used to quickly create TLGs with dynamic grouping, statistics, and filters for exploratory analysis. This enables data to be viewed and analyzed quickly. Interfaces can also be used for minor customization of standard output templates, which can be locked down with validation to meet exploratory or regulatory needs. With dedicated resources, time, and energy, a tool has been generated to rapidly build a customized application per individual analyses utilizing R Shiny to review and generate TLGs interactively.

OBJECTIVES

The current objectives of the TLG Automation tool created with the foundation of R/R Shiny, include exploratory analysis, supporting event-based requests (EBRs), and validating outputs.

Typically, TLGs are created in a way that doesn't allow further data exploration through different scenarios. However, R Shiny can help end-users interactively explore their data by allowing them to toggle through different populations, analyze subgroups, select desired percent cutoffs, and even customize the colors of graphs, all within a single output.

The TLG Automation tool at BMS is crucial in analyzing data and in addressing EBRs that arise during a clinical trial. EBRs, triggered by the need for internal decision-making or requests from health authorities, are characterized by their urgency and time sensitivity. The tool’s agility enables programmers to develop output requests quickly and provide them to the necessary stakeholders.

The TLG Automation tool is a robust solution for exploratory analysis and an interactive platform that can significantly enhance safety monitoring and early signal detection in clinical trials.

DEVELOPMENT

Establishing an appropriate business structure was critical in generating a stable and scalable product, balancing technical expertise, standards, and processes.

GOVERNANCE TEAM

BMS has taken a step forward in enhancing its analytical capabilities by establishing a specialized governance team for the development and implementation of TLG automation. Sponsored by management from relevant functions, this team exemplifies a multidisciplinary approach, combining expertise from various areas critical to the success of this initiative.

The governance team comprises of key representatives from distinct but interconnected functions within BMS. This includes the Product Owner/Technical Lead and statistical function SMEs from different Therapeutic Areas. Additionally, the team integrates Data Visualization Developers and the BMS R statistical package developer. The primary objective of this team is twofold: to develop the tools necessary for TLG automation and to establish the processes that will govern its implementation.

The Governance Team is also responsible for the ongoing maintenance. This includes regular updates and revisions to the templates and configurations, ensuring that the tool remains relevant and effective in the face of evolving data analysis needs and regulatory requirements.

STANDARDS

Statistical Function SMEs from Therapeutic Areas, which are a part of the governance team, play a pivotal role in identifying, specifying, and configuring the outputs for TLG Automation.

- **Identification of Commonly Analyzed Outputs**: The initial step involves the statistical function SMEs identifying the range of commonly analyzed outputs that the TLGs tool needs to create. This identification process is crucial as it ensures that the tool is tailored to meet the most common and critical needs of clinical trial data analysis.
- **Specification Provision to R Developers**: Once these outputs are identified, the specifications are provided to the R developers. This collaboration is key to ensure that the technical development of the tool aligns with the practical requirements of data analysis in clinical trials. The Data Visualization Developers then use these specifications to create a unique template that will serve as the foundation for generating the different required outputs.
- **Configuration of Outputs:** After the Data Visualization Developers have created the initial template, the statistical function SMEs configure these outputs based on the analysis requirements. This step is essential in customizing the tool to handle the specificities and nuances of different analyses.
- **Publication and Maintenance in the Standard Library**: The final configured outputs, based on de-identified datasets, are then published in a standard library. This library serves as a central repository, accessible for future projects and trials, thereby promoting efficiency and consistency in data analysis across different teams and studies.

TECHNICAL

The Technical Development team plays a distinct yet interconnected role, ensuring the creation of a tool that is efficient, compliant, methodologically sound, and customizable for various therapeutic areas.

- **Technical Lead and Data Visualization Developers**: Pivotal in ensuring that the development environments are Good Practice (GxP) compliant, adhering to the regulatory standards, and following prescribed SDLC methodologies. Their primary responsibility is to develop templates for the TLGs based on the requirements provided by the Statistical Function SMEs. These templates are the backbone of the TLGs automation process, ensuring that the outputs are standardized, reliable, and meet the specific requirements.
- **BMS R Statistical Package Developer**: Charged with the crucial task of maintaining and developing the BMS R package, which contains statistical methods. The package is a central resource for the various statistical methods used in TLGs, ensuring standardization of statistical methods in R across the organization. Additionally, provides guidance on the implementation of complex statistical methods, ensuring that the TLG templates produced are not only technically sound but also methodologically robust.

Collaborative development of TLG templates is crucial. R programmers work closely with the R Statistical Package Developer, as well as the Statistical Function SMEs, to ensure that the templates are not only compliant and efficient but also tailored to the specific needs of each therapeutic area. This collaboration is crucial as it allows for the integration of technical programming skills with in-depth statistical knowledge and therapeutic area expertise. (**Figure 1**)

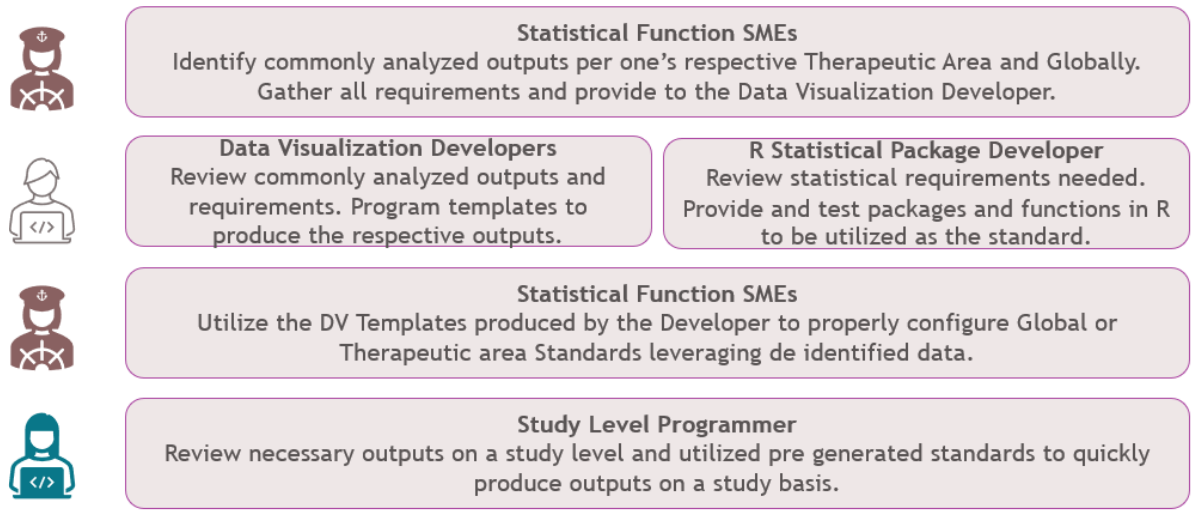

Figure 1: Roles and responsibilities in establishing TLG Automation.

IMPLEMENTATION

The implementation process of TLG Automation involves a series of steps, each designed to ensure that the automated system is efficient, accurate, and seamlessly integrated into the existing workflow. (***Figure 2***)

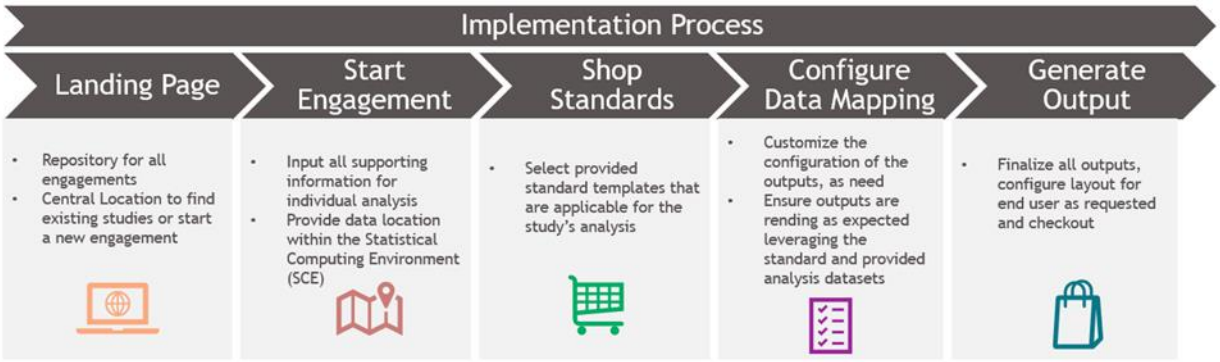

Figure 2: Study level implementation of TLG Automation.

Implementing the TLG Automation tool at the study level involves a detailed and systematic process that should be completed by the protocol statistical programmer. This process ensures that the automated system functions optimally within the specific context of each study.

- **Start Engagement**: The first step in deploying the TLG Automation tool at the study level is for the user to create an engagement for the analysis. This step establishes the application environment for the subsequent stages of implementation. The user is responsible for providing all relevant metadata related to the analysis. Alongside the metadata, the user also specifies the location of the analysis data within the Statistical Computing Environment (SCE).
- **Shop Standards**: The tool provides users with the ability to browse through a collection of standard templates, which are grouped by Global or Therapeutic Area (TA) Standards. This categorization aids users in quickly finding the most relevant templates for their specific needs. Users can select their desired templates and add them to a 'shopping cart.' This feature also offers the convenience of replicating template layouts from other studies that have already been configured, making the process more efficient and consistent across multiple but similar projects. If a desired output is not available in the standard library, the change request process would be initiated. The Study level Programmer would alert their TA representative within the Governance Team of this request.

- **Configure Data Mapping:** Once a template is selected, the tool displays predefined DataSpec mappings for the chosen panels. These specifications detail the parameters passed to the selected R template to generate outputs that are standard for certain types of analyses. The tool offers flexibility in updating these mappings to accommodate any changes in the study's data requirements. This adaptability helps maintain the relevance and accuracy of the analysis for the specific study. (***Figure 3***)

Figure 3: Data Mapping for TLG Automation.

- **Generate Output:** After setting and updating the mappings, the study programmer can preview the expected outputs**.** This feature allows the programmer to confirm that the mappings are correct, and that the data is processed as intended. Subsequently, the programmer can generate the final outputs tailored to the study's specific requirements, ensuring that the automated system functions optimally within the specific context of each study.

OVERALL SUMMARY

Implementing innovative technology can be tricky, but with proper structure in place, achieving a smooth and successful transition is possible.

Analysis Standards ensure reproducibility, providing a reliable and consistent framework for analyzing the results. Identifying a Governance Team consisting of members with deep knowledge in each therapeutic area was crucial to successfully achieving this implementation.

The R developers gathered all the requirements necessary to generate flexible templates to produce interactive outputs. With many outputs having a similar layout but requiring unique variables, it made it the perfect use case for implementing parameter-driven code. The underlying R templates or shiny modules were deliberately developed to allow for a one-to-many relationship, where one template can be used for many outputs reflecting the information prompted by the parameters.

Combining the templates and identified outputs enabled the Governance Team members to configure all commonly used outputs based on the deidentified data and publish them to unique standard libraries. These published libraries are the foundation of the scalability of outputs, driving efficiency in an automated way.

In addition to establishing standards, the Governance Team created a process to be followed when generating TLGs using this new capability.

KEY FEATURES

Generating TLGs interactively in clinical trials requires tools that offer flexibility, robust data handling, and interactive features. Here are some key features of the tool that are particularly useful for interactive TLG generation:

- **Interactivity**: Allows for the creation of TLGs interactively. Users can interact with the data and visualizations in real time.
- **Customization**: Since the tool is parameter-driven, it offers extensive customization options for the layout and appearance of tables and graphs.

- **Dynamic Data Processing**: Capable of processing data in response to user inputs, enabling dynamic and near real-time data analysis.
- **Traceability**: The tool enables reproducibility of outputs and maintains an audit trail with locked code versions and metadata.
- **Export**: The user has the capability to generate submission-compliant RTF outputs from a given output. (***Figure 4***)

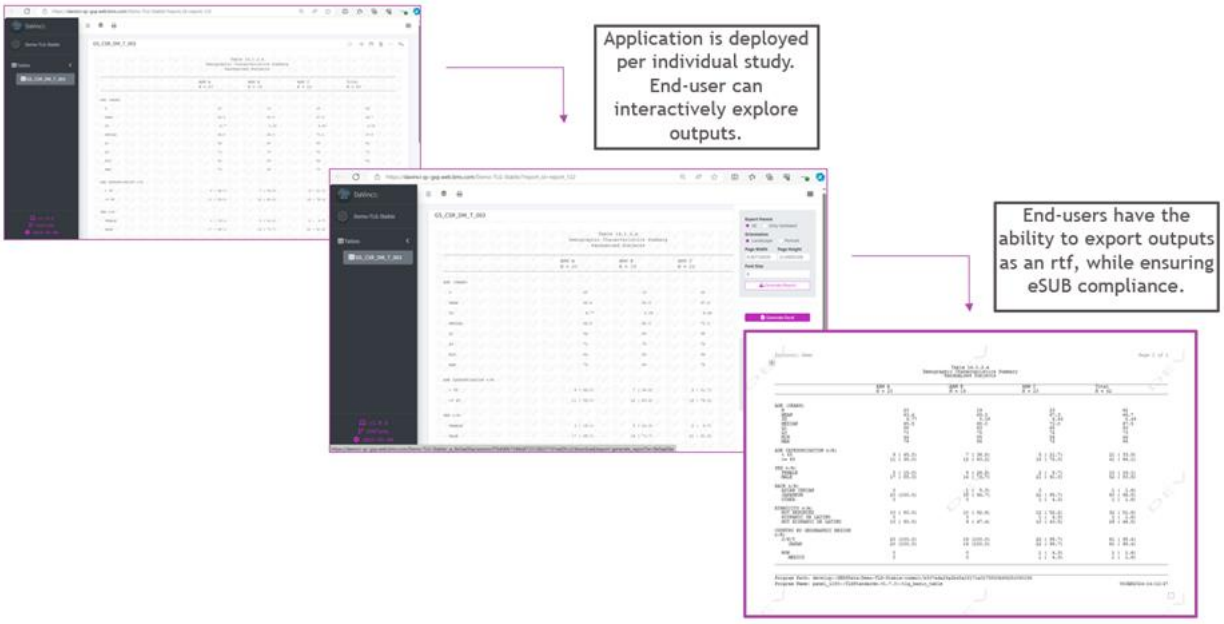

Figure 4: End-user capability to export TLGs as an RTF file

CHALLENGES

Creating TLGs in clinical trials using R, the programming language and environment, may present some challenges. Clinical trials often require sophisticated statistical methods for analysis. Implementing these methods correctly in R, especially for less common or newly developed statistical techniques, can be challenging and requires a deep understanding of statistics and R programming. Challenges can be introduced while ensuring the correct package is being selected, more specifically, identifying the underlying functions and parameters to be used.

Shifting towards R from a traditionally used programming language in the context of producing TLGs for clinical studies indeed presents some challenges and considerations; it may require niche resources, updating current processes, and training current resources.

CONCLUSION

Numerous advantages can be attributed to implementing the TLG Automation tool, including but not limited to efficiency, consistency, flexibility, and cost-effectiveness. Using an interactive tool to explore data improves productivity, reduces waste, and optimizes resources in a significant way. The TLG Automation tool streamlines the process of generating the outputs, significantly reducing the time and effort required compared to traditional methods.

The tool provides predefined templates and suggested configurations that generate the standards across the organization. The structure allows for enforcing organizational standards to ensure consistency across reports and trials. Abiding by the standards results in consistency and reduces the time, energy, and resources required. While the standards are provided, a consideration that each Clinical Trial will differ depending on the disease, indication, and analysis is accounted for. The tool allows for flexibility in configuring individual outputs and generating the user interface for the stakeholders to review the TLGs.

The decision to leverage R as a programming language for generating Outputs has brought several advantages, both in terms of process efficiency and cost-effectiveness.

ACKNOWLEDGEMENTS

We extend our heartfelt gratitude to our colleagues at Bristol Myers Squibb (BMS) and the dedicated Subject Matter Experts (SMEs) who have been instrumental in developing and implementing this tool. It is an honor to recognize the key contributors, including Samar Noor (Project Sponsor), Rajan Vohra (Product Owner/Technical Lead) , Nicole Thorne (Change Management Lead), Joe Zientek (R Development Lead), Lina Spokauskaite (R Development Lead), Abraham Zhao (R Developer / Statistical Function SME), Sangeetha Mani (R Developer), Xiaohan Zou (Statistical Function SME), Rudra Sheth (Statistical Function SME), Tamara Martin (Statistical Functional SME), and Taniya Muliyil (Statistical Functional SME) who played an integral role in making this possible. Their commitment and expertise have been pivotal in promoting and advancing the TLG Automation tool within BMS.

CONTACT INFORMATION

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

Author Name: Margaret Wishart (Product Manager)
Company: Bristol Myers Squibb
Email: margaret.wishart@bms.com

Author Name: Vijay Sharma (Statistical Function SME)
Company: Bristol Myers Squibb
Email: vijay.sharma@bms.com