Paper AR02

OpenRBQM Extended: Endpoint Monitoring, Root-cause Analysis, and Workflows

Spencer Childress, Gilead Sciences, Durham, NC, United States
Jeremy Wildfire, Gilead Sciences, Durham, NC, United States

ABSTRACT

OpenRBQM is a PHUSE-sponsored collaborative effort to create a toolbox of open-source R packages focused on risk-based quality management (RBQM). In the past year the OpenRBQM framework has been refactored and extended to support many new use cases, including endpoint monitoring and root-cause analysis. This presentation describes the process of creating new tools that are compatible with OpenRBQM and introduces several new R packages using this framework.

INTRODUCTION

RBQM is a regulatory-recommended approach to detect and manage risk to a clinical trial that requires continuous oversight and execution throughout the life of a study, from protocol development to database lock. A comprehensive approach to risk management includes predefined, study-specific risk identification and mitigation. OpenRBQM is a suite of analytics tools that specializes in the types of risks that can be evaluated with clinical data, like the site-level protocol deviation rate, which a centralized monitoring team can then triage for follow up with clinical sites. The approach described below can be adapted to any study and deployed on a regular cadence to enable organizations to identify and mitigate risks in their own trials.

WORKFLOW FRAMEWORK

The core {gsm} package, an R package in the OpenRBQM ecosystem which provides a framework for risk assessments (Wu et al., 2024), has been refactored to support a structured, extensible analysis pipeline from source data to final outputs. Each step of the analysis pipeline falls into one of four layers – mapping, analysis, reporting, and module – and each layer depends on the layer before it. This approach provides the scaffolding to easily add new metrics and outputs by adhering to a sequential, metadata-driven process.

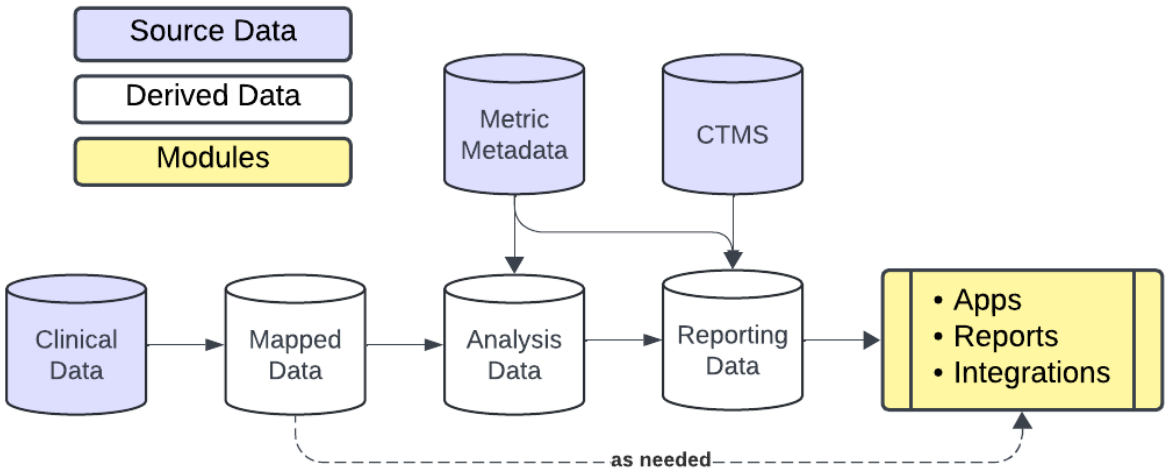

Figure 1. Analysis Pipeline

The mapping layer is a data processing step that generates all data dependencies required in the analysis layer. A mapping workflow corresponds to a single output data domain, such as demographics or adverse events, and includes a data specification for one or more input data domains. It also lists a series of steps that restructure the input data domains in some way, including applying a subset, merging data domains, or deriving new columns. Essentially the mapping workflow yields analysis-ready data.

The analysis layer is a data analysis step that evaluates a metric of interest, like the adverse event reporting rate across sites or the subject discontinuation rate at the country or study level. Like a mapping workflow, an analysis workflow includes a data specification of required data domains, although these data domains correspond to the output of a mapping workflow. Each workflow applies a standard series of steps to evaluate the metric:

1. Calculate the metric at the participant level.
2. Aggregate the metric by the group or stratum of interest: site, country, or study.
3. Apply a statistical method to assess the magnitude of the metric, captured as a score.
4. Flag each group by its score using predefined thresholds, e.g. -3, -2, 2, and 3.

After all metrics have been evaluated the reporting layer compiles and enriches the analysis results with metric and group metadata. Reporting data is ready-made for visualization and presentation in a chart, report, or application, which leads directly to the module layer. A module workflow produces an output that surfaces the analysis results in some manner, like a summary of metric results captured in a report, an interactive data application, or the propagation of information to an external system for assessment.

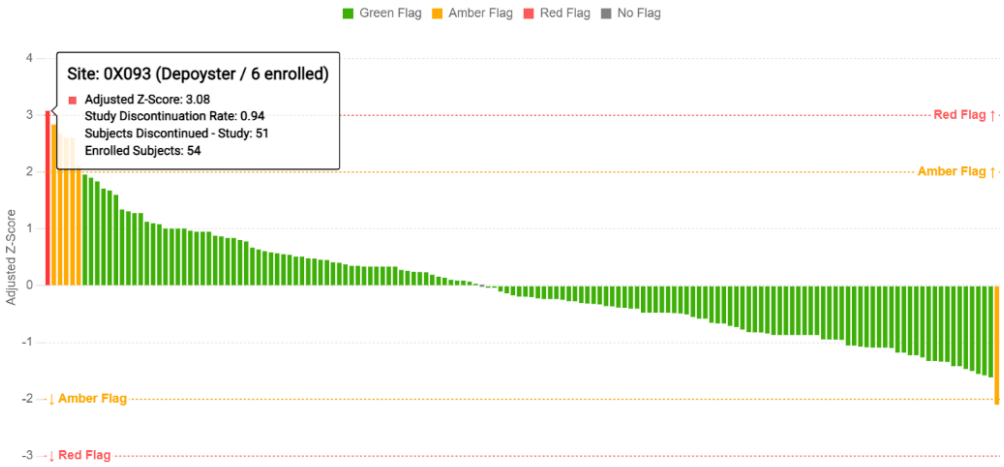

Figure 2. Bar Chart of Analysis Results with Metric and Site Metadata

This extensible metadata-driven analysis pipeline allows for study-specific customizations for all aspects of RBQM analytics from data ingestion and mapping to reporting, drill-down, and central statistical monitoring. Taken together, this framework represents the start of a “qualityverse” that provides a robust open-source analytics framework for RBQM. A full list of metrics, reports, applications, and packages in the OpenRBQM is available on the working group’s website.

EXTENSIONS

{gsm} itself has been split out into component packages to improve maintainability. References to {gsm} in this paper refer to the suite of packages formerly encompassed in the singular package named {gsm}. Component packages include {gsm.core}, the statistical analysis and workflow engine; {gsm.mapping}, the data ingestion and mapping piece; {gsm.kri}, which houses the metric workflows and reporting code; and {gsm.qc}, a qualification framework.

{gsm.app} is a package tailored to the {gsm} analysis data model that connects risk signals with the underlying data. It allows subject matter experts such as central monitors and statisticians to drill down into the data by metric, site, participant, and data domain. Built with Shiny, the app is highly interactive, with two-way links between inputs and widgets, and surfaces every data visualization present in {gsm}.

Much like {gsm} the app has been developed with extensibility in mind. By default, an instance of {gsm.app} surfaces data available in each layer of the analysis pipeline:

- A summary of the analysis results, including a standard set of visualizations by metric
- Drilldown into the underlying mapped data by study, site, or participant
- Study metadata including key protocol information, site activation, and participant enrollment
- A listing of site-level attributes – principal investigator, status, location, number of active participants, etc.
- Participant-level statistics

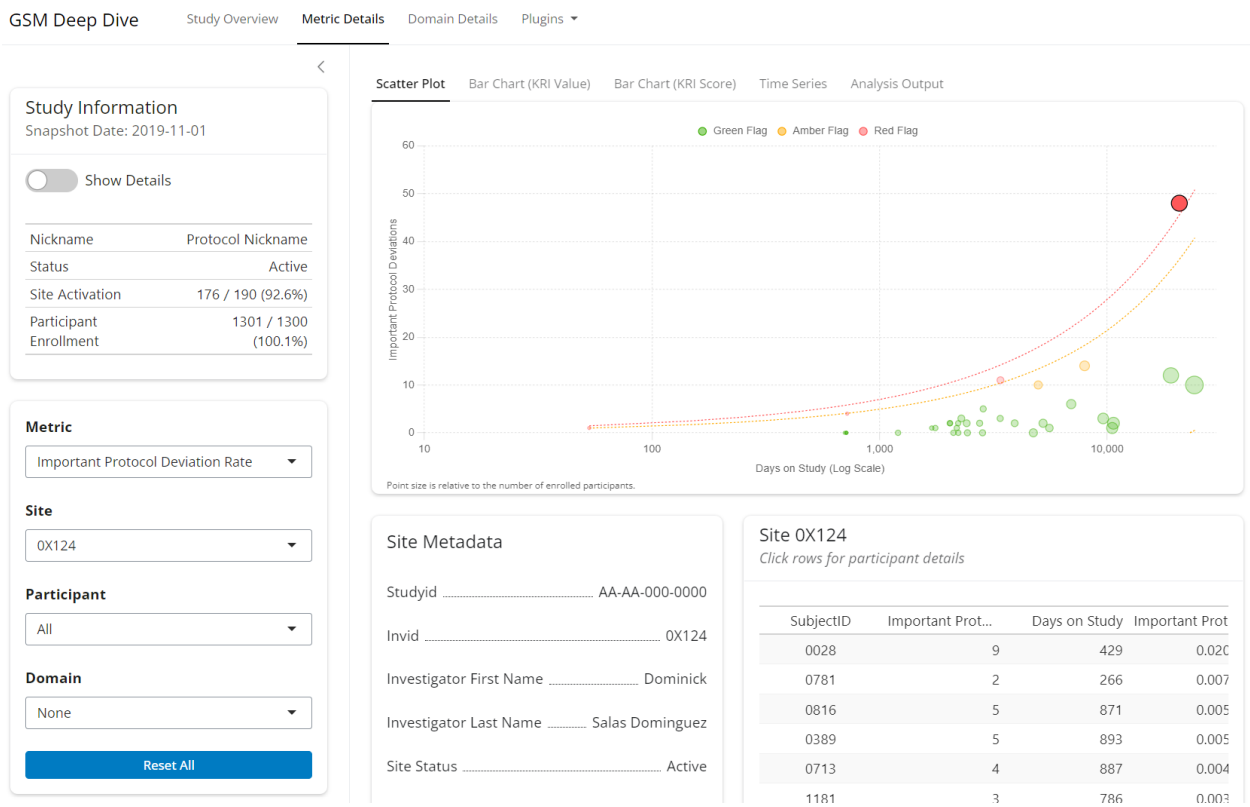

Figure 3. Deep Dive Application

The extensibility comes via “plugins”, self-contained Shiny modules with dedicated server and user interface (UI) functions that present the data underlying the metrics in a meaningful manner. By working with subject matter experts to understand how they interpret and investigate risk signals each plugin is developed in a fit-for-purpose manner to facilitate risk-based monitoring. The first plugin the team prototyped focuses on adverse events and includes a high-level summary of AE frequency by seriousness, severity, and relationship; a table of AE prevalence by system organ class and preferred term; and a participant-level timeline that visualizes when adverse event onset and resolution. Future plugins will include a patient profile, a site health assessment, and a treatment disposition explorer.

{gsm.endpoints} provides a framework for clinical trial endpoint monitoring. It evaluates both general risks to study endpoints such as patient retention and therapeutic area-specific risks such as overall- and progression-free survival in oncology studies, enabling early detection of censoring. Endpoint monitoring typically involves complex data derivation in the mapping layer and is more study-specific, but it provides a huge benefit by identifying critical risks to the trial.

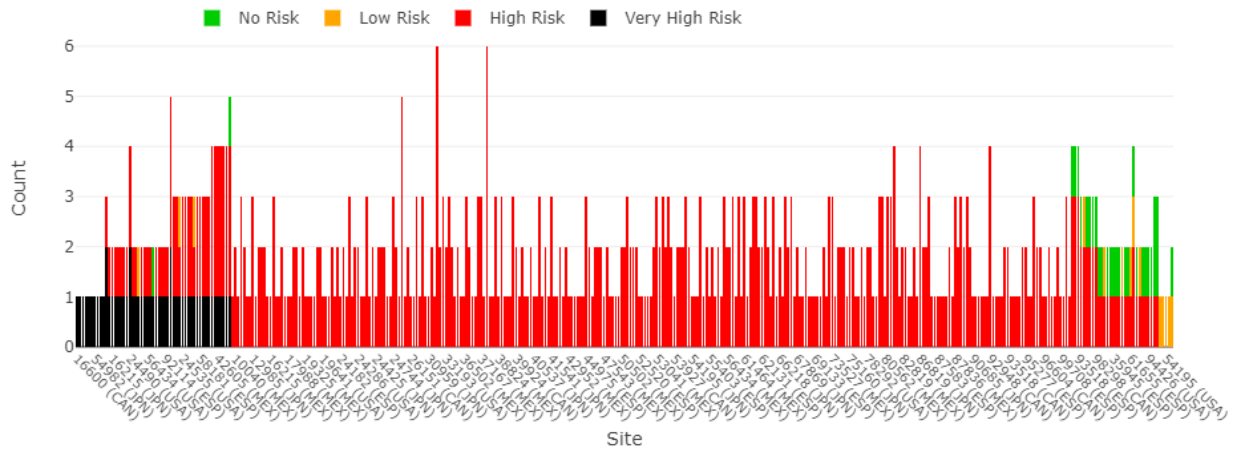

Figure 4. Sites at with Potential Premature Discontinuations (Anonymized Clinical Trial Data)

{gsm.simaerep}, recently developed by an OpenRBQM collaborator from Roche, aims to detect sites that may be under-reporting adverse events. The package applies a bootstrap-based simulation method to assign a probability of under-reporting to each site. Especially relevant early in a study, this novel approach facilitates rapid, comprehensive, and near-real-time detection of AE under-reporting at each clinical trial site (Koneswarakantha et al., 2024).

{gsm.digitpref} helps identify data entry bias by detecting patterns in the actual numbers present in a numerical attribute. By comparing the data entry tendencies of sites with the overall study-level distribution for a specific data field, the tool reveals which digits the site favors. By focusing on critical fields, the study management team can follow up with the site on best practices and reduce potential bias in the data.

CONCLUSION

These extensions are made possible by an extensible, workflow-driven framework that standardizes the inputs and outputs of each metric and module. With the contributions of OpenRBQM collaborators the initiative has developed several new tools, including domain-specific modules: {gsm.simaerep}, {gsm.ae}, and {gsm.query}; dedicated therapeutic area-specific metrics with {gsm.endpoints}; and generalized evaluation of data quality with {gsm.digitpref}. Moving forward the team plans to continue expanding the reach of the *qualityverse* to new clinical data domains and additional therapeutic areas for a more comprehensive understanding and awareness of risks to clinical trials. Those interested in implementing and/or contributing are encouraged to reach out to the working group leads of OpenRBQM.

REFERENCES

Wu G., Wang Z., Wildfire J. *et al.* Good Statistical Monitoring: A Flexible Open-Source Tool to Detect Risks in Clinical Trials. Ther Innov Regul Sci. 2024 Sep;58(5):838-844. doi: 10.1007/s43441-024-00651-4. Epub 2024 May 9. PMID: 38722529; PMCID: PMC11335794.

Koneswarakantha, B., Adyanthaya, R., Emerson, J. *et al.* An Open-Source R Package for Detection of Adverse Events Under-Reporting in Clinical Trials: Implementation and Validation by the IMPALA (Inter coMPany quALity Analytics) Consortium. *Ther Innov Regul Sci* **58**, 591–599 (2024). https://doi.org/10.1007/s43441-024-00631-8

ACKNOWLEDGMENTS

Huge thanks to the OpenRBQM working group, including but not limited to:

- Jon Harmon, Atorus Research
- Jeremy Howell, Roche
- Björn Koneswarakantha, Roche
- Jenn Krohn, Gilead Sciences
- Laura Maxwell, Atorus Research
- Devin Pastoor, A2-Ai
- Zhongkai Wang, Gilead Sciences

RECOMMENDED READING

- The Basics of Clinical Trial Centralized Monitoring
- Comprehensive Assessment of Risk-Based Quality Management Adoption in Clinical Trials

CONTACT INFORMATION

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

Author Name: Spencer Childress
Company: Gilead Sciences
Address: 333 Lakeside Dr. Foster City, CA 94404
Email: spencer.childress@gilead.com