PHUSE India Summer SDE 2021

Interactive Data Visualizations for Data Exploration, Decision Making and Publication

Abhijit Roy
Sr. Manager
GlaxoSmithKline

Disclaimer

“The views and opinions expressed in this presentation and on the following slides are solely those of the presenter and do not necessarily reflect the views of GlaxoSmithKline.”

Scope of Presentation

- Interactive Data Visualizations (IDV) in Biopharma
- Its benefits & Usage
- GSK's own IDV tool called RAPIDO Data Viewer
  - Vision
  - Design
  - Process Flow
  - Features
  - Quality Control
  - Considerations & Challenges
  - Computer System Validation
  - R Validation
  - Why ADaM?
  - Why R/Shiny?

Interactive Data Visualizations in Biopharma

> Ther Innov Regul Sci. 2018 Nov;52(6):696-700. doi: 10.1177/2168479018754846. Epub 2018 Feb 5.

The Safety Explorer Suite: Interactive Safety Monitoring for Clinical Trials

Jeremy Wildfire [1], Ryan Bailey [1], Rebecca Z Krouse [1], Spencer Childress [1], Britt Sikora [1], Nathan Bryant [1], Shane Rosanbalm [1], Emily Wilson [1], Jack G Modell [1]

Interactive Analysis for Decision Making & Submissions

- An update from PHUSE Data Visualization Project

tidyCDISC

visR

Interactive Data Visualizations in Biopharma

Benefits

- Quick insights – don't have to wait for programmer to produce and QC a new display
- Data Democratization
- Reduce number of unused outputs for CSRs – eliminate "just-in-case analyses"
- Can reduce cycle time and improve efficiency

Main Uses In Biopharma Industry

- Data Exploration
- Instream data review
- Risk-based monitoring and Medical review
- Safety Review Boards
- Data Monitoring Committees
- Online data review tools to replace listings

Common Usage: Data Exploration

Common Warning: All outputs from this application are unvalidated. For validated outputs, please contact the study statistician and programmer

In general: If results from Interactive Data Visualizations are included in publications or submissions, they are reprogrammed & validated in Statistical Computing Environment (SCE)

- Results for publication and submission should be produced within a validated environment
- Follows Quality Control/validation SOP
- SCE creates audit trail and logs of all work
- Link between SCE and Publishing system

Data Visualizations for Data Exploration, Decision-Making and Publishing

Vision: Incorporate interactive data visualization directly into Good Clinical Practice (GCP) governed processes such as:

- Writing Clinical Study Reports & publications
- Decision-making
- Inclusion in submission packages to regulatory agencies

RAPIDO Data Viewer

Vision: Faster, Smaller, Interactive

Project RAPIDO
Transforming our analyses

gsk

Modernize and transform data exploration

- Non-Biostatisticians can explore data on their own
- Improve collaboration - Workflows for sharing of data insights
- Better understanding of data by all study team members
- Rich & compelling user experience with interactive features

Increase Efficiency

- Reduce the number of EOS Analysis TLFs to those truly needed for CSRs and submissions - Eliminate “just in case” analyses
- Replace static listings with an Interactive Patient Profile
- Automate as much as possible

Deliver Faster

- Deliver EOS Analysis and Submissions faster
- Speed-up time to insight from data ➔ Top Quartile Performance

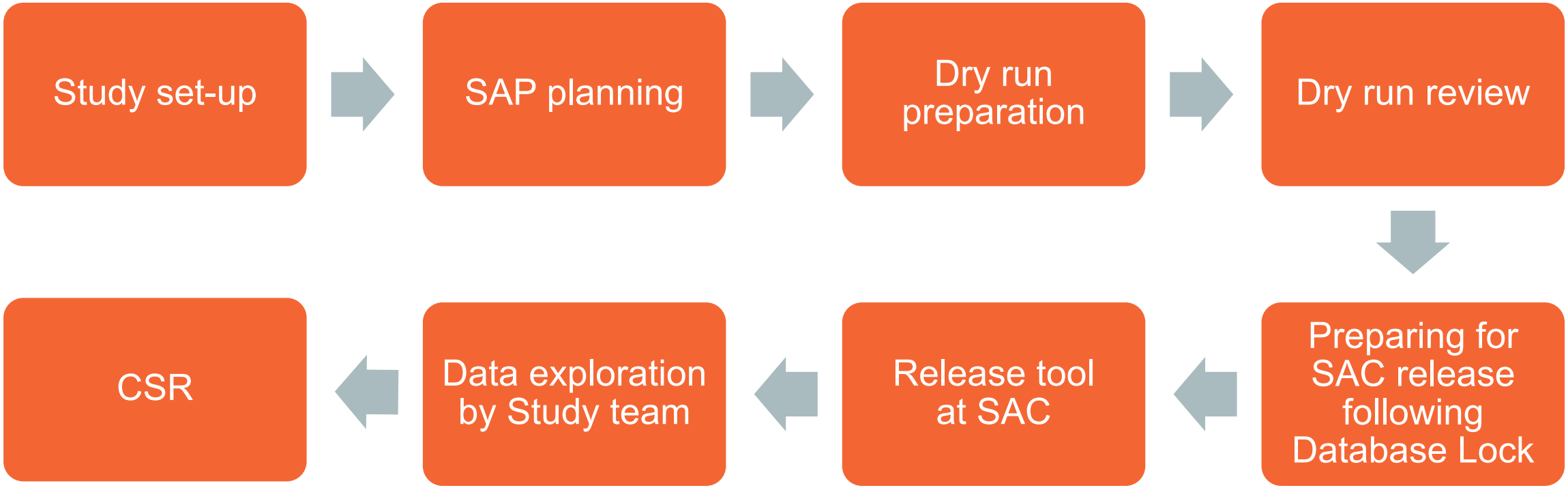
RAPIDO Data Viewer

Design

- Validated Enterprise Software with audit trail and access control
- Bespoke R / Shiny app hosted on Azure cloud
- Integrated into SCE with direct link to data pipeline & regular data refreshes
- Compliant with Study Results Dissemination, Data Reuse, Disclosure, Informed Consent & Privacy requirements
- User-Friendly Data Exploration and Collaboration Features
- Library of validated R packages that can be reused in SCE or other Shiny apps

RAPIDO Data Viewer

Process Flow



SAC: Statistical Analysis Complete

RAPIDO Data Viewer

Features

Initial Features

- Reduce SAC subgroup displays & listings
- Access to study pop and safety data
- Data exploration and subject-level data review
  - Filters, Search, Sorting
  - Dynamically create subsets of patients to follow across different data domains
- Bookmarking of displays to promote collaboration
- Drill-down for subject details

Future Features

- Access to all data
- Dynamic exploration of data
  - Custom table generation (drag and drop)
  - Add new sub-group variables
  - Select visualization from library and then apply visualization to selected data
- Automations
  - e.g., Direct link to LSC, archiving
- Direct link to publishing
- Displays with stat inference

Possible Future Use

- Instream use such as data review or dose escalation decisions
- Use by external partners of GSK (e.g., Global Health)
- Advisory Committees
- Narrative writing
- Highlight outliers
- Pooled data for integrated summaries
- Create validated reports for submissions

RAPIDO Data Viewer

Quality Control

- RAPIDO Data Viewer is a validated IT system
- QC is needed for setup of RAPIDO DV at reporting effort level and for new non-validated displays
- If data exploration in RAPIDO DV results in an adhoc analysis request, the adhoc analysis is produced and QCed in SCE following Biostatistics processes for analysis and QC

Quality Control for RAPIDO DV	How to Perform QC
Programmer sets up RAPIDO DV for reporting effort (configuration and data mapping). **Independent QCer** will QC the setup.	• Visual check of Configuration and Data Mapping screens in RAPIDO DV • Document QC
Independent QCer will QC any new non-validated RAPIDO DV displays (if approved to be added to RAPIDO DV)	• QC in SCE following Biostatistics QC SOP/WI • Document QC in QC trackers

Consideration and Challenges

- Enterprise software with Access Control and Security
- Computer system validation, including R validation
- Compliance with disclosure and Human Subject research requirements
- Statistical inference in interactive displays
- Standardization Challenges
  - ADaM variability
  - Standardizing analyses and derivations
- New skills and responsibilities

Computer System Validation

From FDA Glossary of Computer System Software Development Terminology:

> **validation.** (1) (FDA) Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes. Contrast with data validation.

Requirements

- Regulatory agencies require that computer systems used in clinical trials (including statistical analysis) be validated with appropriate software testing procedures to demonstrate they are reliable
- Validation must be documented
- Additional requirements: security safeguards, audit trails, date/time stamps, Standard Operation Procedures (SOPs), controls for system changes, and training of personnel

Computer System Validation

Key Steps in Computer System Validation

- User requirement specifications
- Functional requirement specifications
- Validation Plan and Risk assessment
- Unit Testing
- Systems Integration Testing
- User Acceptance Testing
- Validation Traceability Matrix
- Network and Infrastructure Qualification

Take Home Messages

- Documentation is key
- Essential to have IT and Computer System Quality Assurance expertise
- Allocate plenty of time and resource

R Validation

R Consortium Working Groups for R Validation and Use of R for Regulatory Submissions

- R Validation Hub (Pharmar): a cross-industry initiative whose mission is to enable the use of R by the Bio-Pharmaceutical Industry in a regulatory setting, where the output may be used in submissions to regulatory agencies.
- R Tables for Regulatory Submission (RTRS): Develop standards for creating tables that meet the requirements of FDA submission documents
- Submissions: Focus on IT and platform challenges that must be addressed in order to make “all R” regulatory submissions.

White Papers

- R: Regulatory Compliance and Validation Issues A Guidance Document for the Use of R in Regulated Clinical Trial Environments
- A Risk-Based Approach for Assessing R Package Accuracy within a Validated Environment:

RAPIDO Data Viewer

Why Use ADaM Datasets as Input to RAPIDO?

- Allows us to include efficacy results in RAPIDO
  - These are in the ADaM datasets but not in the SDTM datasets
- Contains derived data that is critical for analyses. Some examples are:
  - Population flags, analysis visits, numeric visit values for sorting, imputed values, censoring values, treatment failure flags, etc.
  - PRO: change from baseline, proportion of subjects with change from baseline above a threshold
  - Biomarkers: change from baseline, % change from baseline, normalization of some biomarkers
- Ensures that the results in RAPIDO match the results in the SAC
- Efficient to use existing derivations in ADaM datasets
- Using ADaM datasets as input to data visualization tools aligns with other pharma companies, e.g. Biogen (tidyCDISC), Roche (TEAL), Novartis (TTE Module), Novo Nordisk (Link)
- SDTM datasets can also be used as input, if needed

RAPIDO Data Viewer

Why R/Shiny?

Powerful and Flexible – Allows an Ambitious Vision

- Integrate with our future SCE
  - Use SCE data domains, access control, and publishing features
  - Automate production of core SAC displays
  - Output-agnostic code
- Advanced Functionality Beyond Data Visualization, e.g.
  - Dynamically determine subgroups from ADSL
  - Dynamically create subsets of patients and follow across analyses/visuals
  - Highlight outliers or unusual values for user to focus on
  - Bookmark displays & add notes to displays
  - Add draft watermark to displays when printed
- More Control of User Interface
  - Search function on different sections of a page
  - Alternatives to scroll bar for long listings
  - Dynamic scaling of axes

- Popular with statisticians, programmers, data scientists
  - R has powerful data analytics capabilities
  - More interest in learning R and Python than Spotfire
  - New hires more likely to know R and Python than Spotfire

- Industry trending to R/Shiny
  - Possible collaboration with other pharma companies
  - Open Source – so not reliant on a single vendor
  - R Studio Enterprise is powerful environment for code development

Trailblazers

- **PHUSE Best Practices for Interactive Analyses for Decision Making and Submission**
  - Pilot 1: eSub with embedded interactive data displays (Eli Lilly - Mary Nilsson, Zachary Skrivanek)
  - Pilot 2: Docker POC with R/Shiny-based interactive tool (Genentech - Nilesh Narayan, Doug Kelkhoff, Xiangyun Wang)
  - 2020 PHUSE US Connect Presentation: https://www.lexjansen.com/phuse-us/2020/dv/DV13_ppt.pdf

- **bioWARP from Roche Diagnostics**
  - "bioWARP (biostatistical Web-Applications and R Procedures) is a Shiny application enabling employees at Roche Diagnostics to create validated reports for regulatory authorities' submissions."
  - R/Pharma 2018: R/Pharma 2018 (rinpharma.github.io), PSI Web 2019: PowerPoint Presentation

- **Accumulus Synergy (https://www.accumulus.org/)**
  - Goal: Global Cloud-based Data Sharing Platform between Biopharmaceutical industry & health authorities
  - Formed in 2020 with 10 leading Biopharma companies

References

- TidyCDISC https://rmed-tidycdisc.netlify.app/#18
- VisR https://visrsoftware.github.io/
- Safety Explorer suite https://rhoinc.github.io/safety-explorer-suite/docs/
- Interactive Analysis for Decision making and Submissions https://advance.phuse.global/display/WEL/Best+Practices+for+Interactive+Analysis+for+Decision+Making+Submissions
- Shiny – R Studio https://shiny.rstudio.com/
- Computer System Validation FDA https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-principles-software-validation

Acknowledgement

- Maria Dalton (RAPIDO Business Lead)

  Director of Programming,

  Tech Excellence and Innovations

  GSK

Thank You

[Questions, Discussions, Feedback are Welcome]