SCDM Live

{/5a ھوي
9a9! /hbC9w9b/9

9-11 April, 2025 | Brussels, Belgium

Open-source analytics tools for Clinical Data Management

Why open-source?

- Cost-efficiency
- Customization
- Community support
  - Faster problem solving
- Complete transparency
  - Improved trust
  - Improved security

SCDM Live

{♫śĂ╫śƞ

Japie
van Tonder

/9h†
{ľ╫ś■╬ t Ǵ[Ǵ

Björn
Koneswarakantha

Principle
Quality Data
Scientist,
Roche

Isabel
Glauß

Clinical Data
Engineer,
Boehringer
Ingelheim

t ś╫Ă
Ç╫╫Ă■ś■

Principal
Clinical Data
Scientist,
Bayer

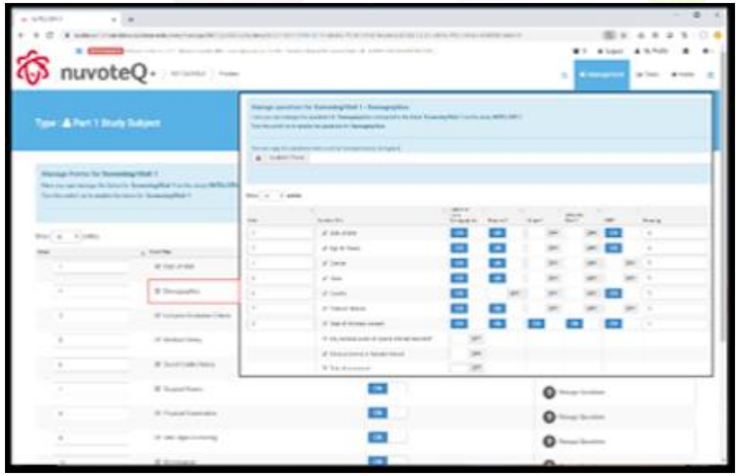
SCDM Live

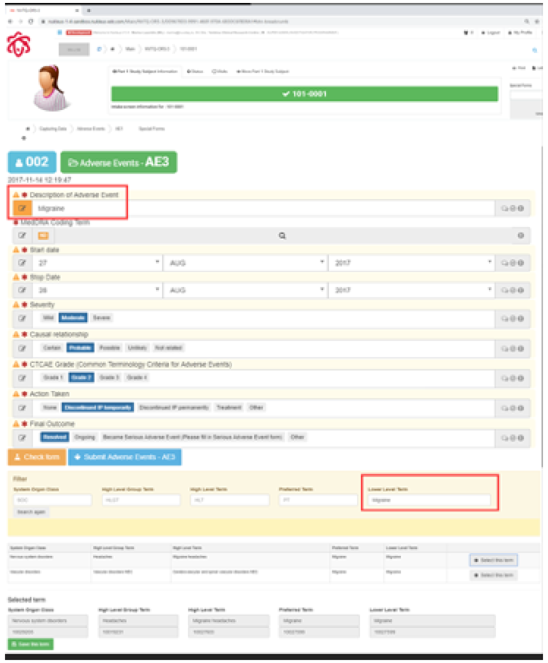
Automation of study-specific clinical database build and validation by leveraging common open-source software development tools

Japie van Tonder, Scigenix Pty Ltd

Introduction

- [illegible]
- [illegible]
- [illegible]





YśŦ / ╚Ă▄ś▬┼śℓ╚▬ ëĂ▄ŕĂĠ╚◘▬

- **Manual validation** → Lack of automation leads to inconsistencies
- **Slow process** → Delayed study timelines
- **Expensive & resource-heavy** → Labor costs
- **Limited adaptability** → A lack of automation makes it difficult to scale processes efficiently

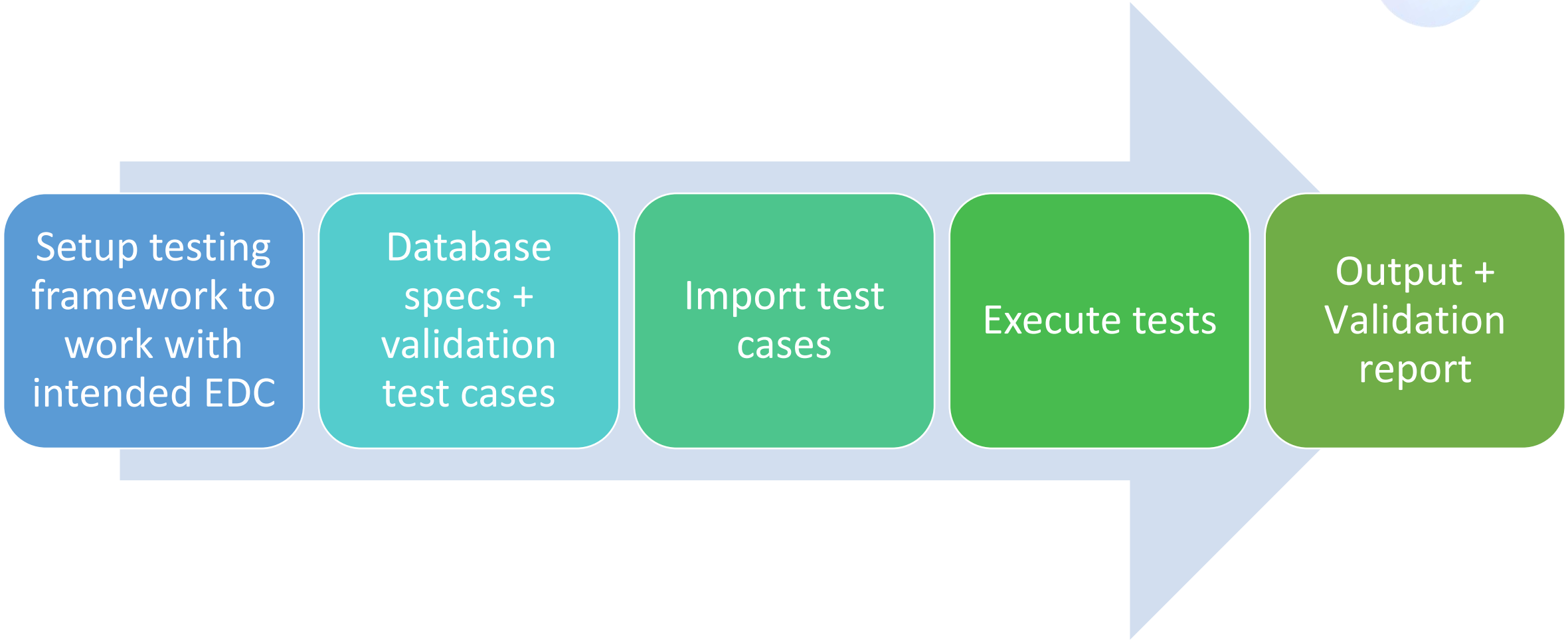
The Automation Framework – Overview

- Goal
  - Automate database validation (“UAT”)
- How?
  - Programmatic approach using open-source tools
- Outcome
  - Faster, more reliable, and scalable database workflows

Modern Front-End Testing Frameworks

- [illegible]
  - Web development teams minimize manual testing by using automated front-end testing frameworks
  - Reduces human error, increases efficiency, and ensures consistency
- [illegible]
  - **Cypress –** Fast, reliable end-to-end testing for modern web apps
  - **Playwright –** Multi-browser automation, ideal for UI consistency checks
  - **Jest & Mocha –** JavaScript-based testing frameworks for unit and integration tests
  - Many other frameworks

I ▪Ŏ LĊí ◪n╫ℓ □ t'n▪l'śℓℓ .'nśĂ╫ŕ▪Ŏ■



/Ăℓś {Ċʃ ŕ₣ wśĂ▄í ժn▄ŕ L▒♫ĂIĊ

- Initial setup
  - Programming link up specifications / testing framework / EDC
  - 2 months

- After automation
  - Execution of all tests reduced from roughly 2 days to 2 hours
  - **92% increase in efficiency**
  - Test ⟶ Update ⟶ Re-test cycle significantly faster
  - Easily perform a re-test of the entire database with each post-production change
  - Ultimately increased build quality (no human error in testing)

Benefits of Automation

- Faster EDC deployment → Saves days/weeks of effort
- Improved accuracy & consistency → Reduces human errors
- Lower licensing costs → Open-source
- Scalable & adaptable → Works with various EDC platforms

Conclusion

- / Ă▬Ġ! l’Ġ■
  - Consider adopting open-source tools in your workflows

- YśŦÇĂ╢śĂŎĂŦ

  "By leveraging automation and established software development practices, we can reduce human effort, improve accuracy, and accelerate clinical database deployment."

[illegible]

Björn Koneswarakantha, Hoffmann-La Roche

Cross Industry Organisation

IMPALA

Intercompany quality analytics consortium

- Founded 2022
- 19 industry members
- https://impala-consortium.org/
- Statistical Open Source Packages
- Scientific Publications
- Quality Frameworks (RAPID, Quality Briefs)
- Joint Health Authority Engagement

- Founded 2024
- > 3 industry members
- RBQM focus (Critical to Quality, Centralized Monitoring, Quality Tolerance Limits)
- 6 Projects
- Open Source Packages

Open Source Landscape – Quality Monitoring Clinical Trials

ctas

[illegible]

simaerep

[illegible]

rbqmR

Quality Tolerance Limits.

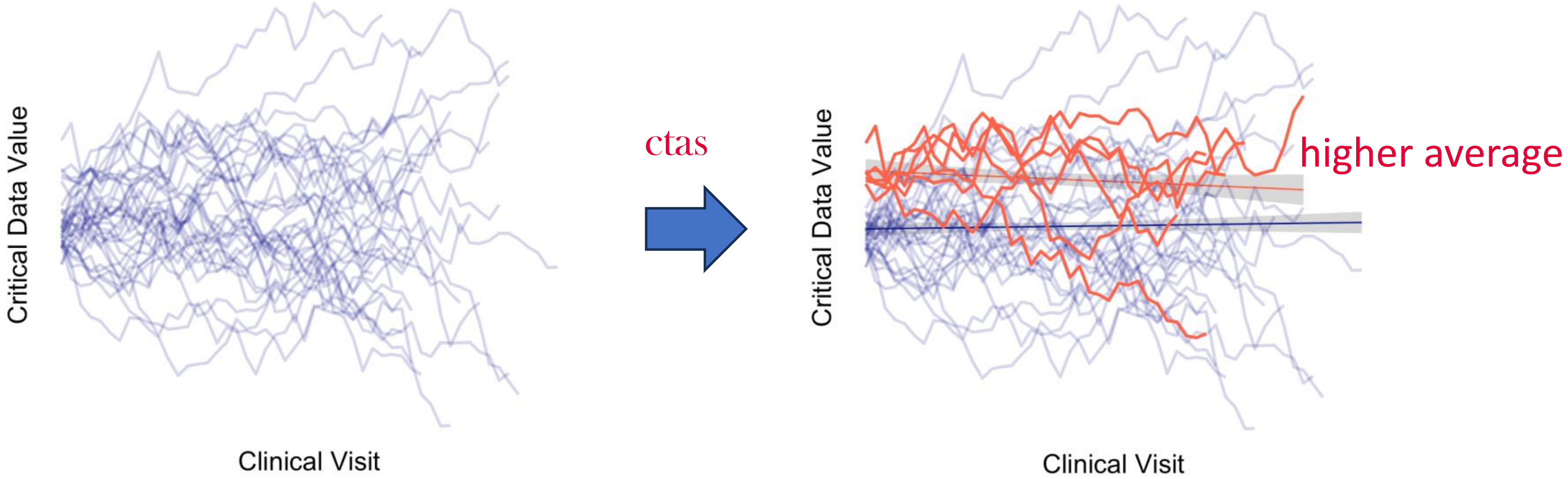
https://github.com/openpharma/rbqmR

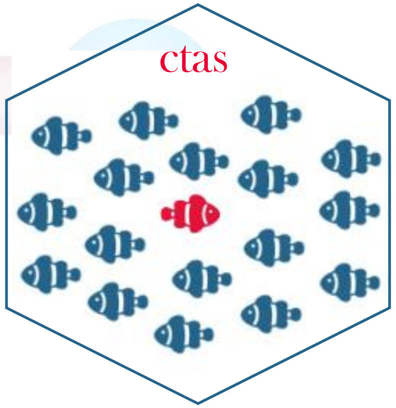
gsm

End to End RBQM Framework

- Mapping
- Processing
- Reporting

https://github.com/Gilead-BioStats/gsm

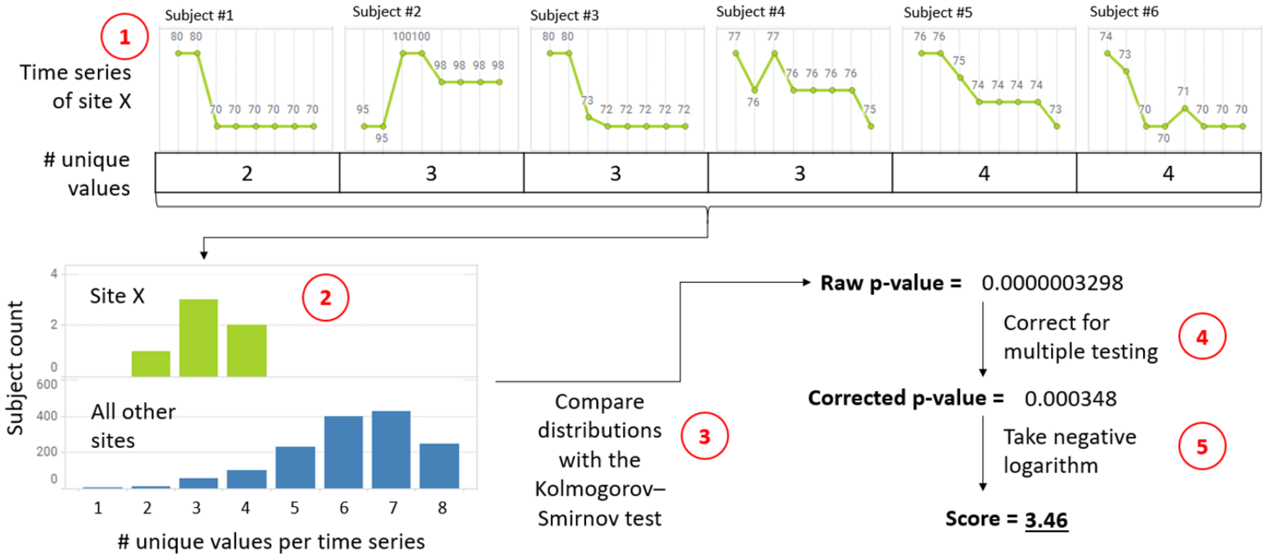




Collaborative Scientific Evaluation Within IMPALA
Webinar recording on CTAS

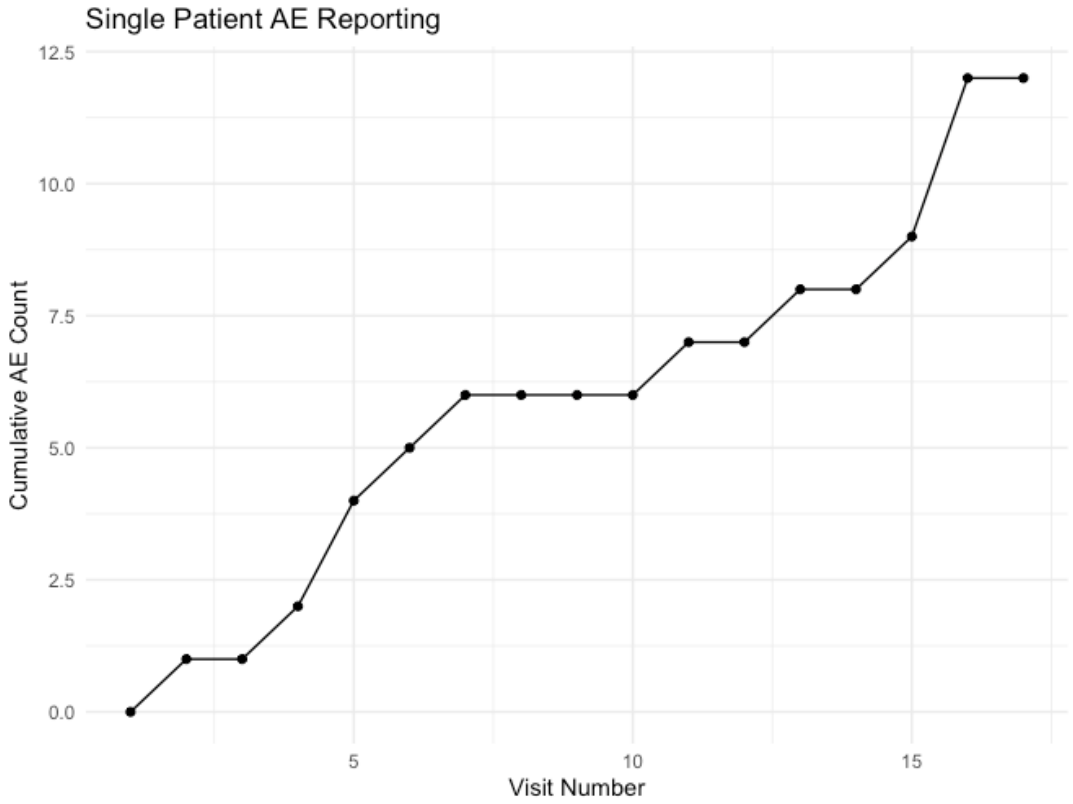
https://github.com/IMPALA-Consortium/ctas

- CTAS (Clinical Trial Anomaly Spotter) is a CSM tool focused on identifying anomalous time series of continuous variables.
- “Unsupervised” tool which does not use pre-specified KRIs.
- Originally developed by Bayer where it is available for all company studies.
- Code base shared with IMPALA for co-development in July 2023.

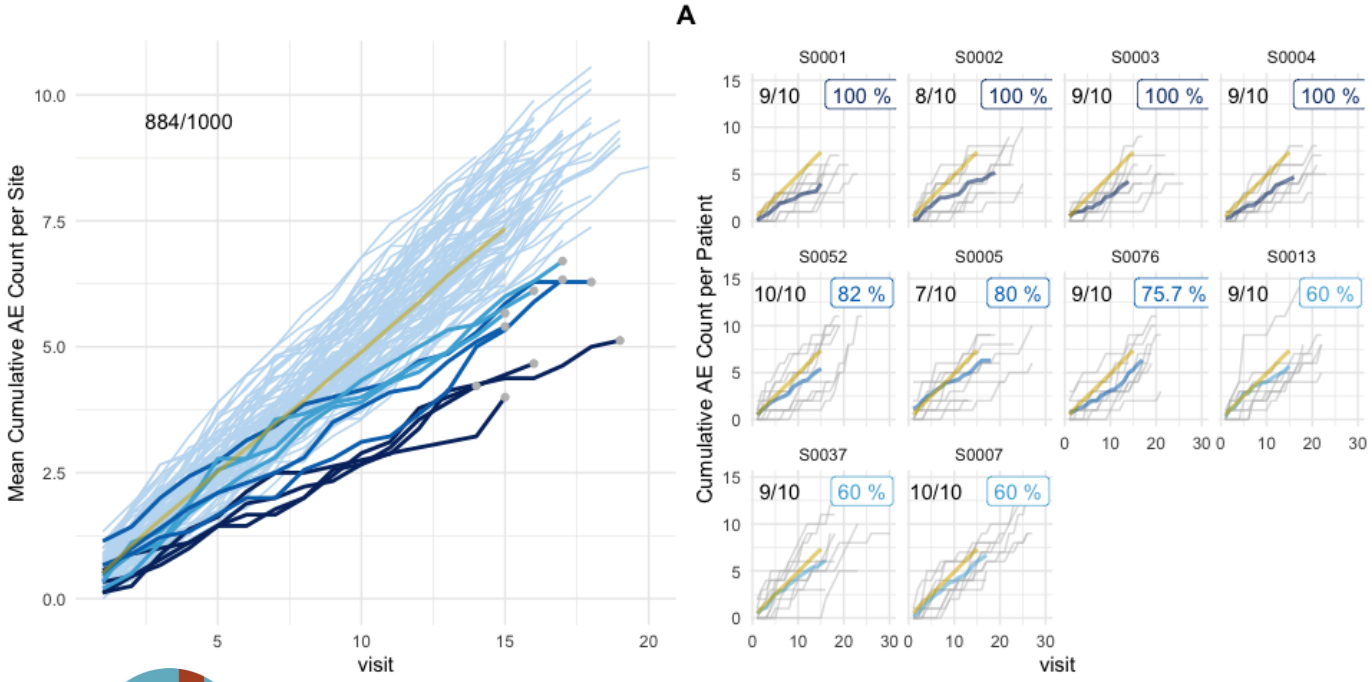


Adverse Event Reporting in Clinical Trials - Patient

Clinical Trial - Patient Visit



{simaerep} Calculate Event Reporting Probabilities
Bootstrap - Simulations



Bootstrap Simulations
Redistribute Patients between sites

Collaborative Scientific Evaluation Within IMPALA
https://doi.org/10.1007/s43441-024-00631-8

https://github.com/IMPALA-Consortium/ctas
https://impala-consortium.github.io/gsm.simaerep/

{gsm} – Good Statistical Monitoring

- Results
  - AE Reporting Rate
  - SAE Reporting Rate
  - Non-important PD Rate
  - Important PD Rate
  - G3+ Lab Abnormality Rate
  - Study Discontinuation Rate
  - Treatment Discontinuation Rate
  - Query Rate
  - Outstanding Query Rate
  - Outstanding Data Entry Rate
  - Data Change Rate
  - Screen Failure Rate
- KRI Glossary
- Error Log

AA-AA-000-0000 Assessment Overview

Generated with the Good Statistical Monitoring {gsm} package

Study Overview

24 Red KRIs | 65 Amber KRIs

Showing Red Sites

16 of 176 sites with at least one **red** KRI (9.1% of total).
60 of 176 sites with at least one **red or amber** KRI (34.1% of total).

Site	Red KRIs	Amber KRIs	AE	SAE	PD	IPD	LB	SDSC	TDSC	QRY	OQRY	ODAT	CDAT	SF
28	4	0	✓	⌃⌃	✓	✓	⌃⌃	⌃⌃	⌃⌃	✓	✓	✓	✓	✓
173	2	1	✓	✓	⌃⌃	⌃⌃	✓	✓	✓	✓	✓	✓	✓	⌄
43	2	1	⌃⌃	✓	✓	✓	✓	✓	✓	⌄	✓	⌃⌃	✓	✓
75	2	1	⌃⌃	✓	⌃⌃	⌃	✓	✓	✓	✓	✓	✓	✓	✓
83	2	1	⌃⌃	✓	✓	✓	⌃⌃	✓	✓	✓	✓	✓	⌃	✓
161	2	0	✓	✓	✓	✓	✓	⌃⌃	⌃⌃	✓	✓	✓	✓	✓
114	1	1	✓	✓	⌃⌃	⌃	✓	✓	✓	✓	✓	✓	✓	✓
186	1	1	✓	✓	⌃	⌃⌃	✓	✓	✓	✓	✓	✓	✓	✓
20	1	1	✓	✓	✓	⌃⌃	✓	–	–	✓	–	✓	✓	⌃
60	1	1	✓	✓	✓	✓	✓	⌃⌃	⌃	✓	✓	✓	✓	✓
68	1	1	✓	✓	⌃⌃	✓	✓	✓	✓	✓	✓	✓	⌃	✓
78	1	1	✓	✓	✓	✓	✓	–	–	✓	⌃⌃	✓	⌄	✓
145	1	0	✓	✓	✓	⌃⌃	✓	–	–	✓	–	–	✓	✓
171	1	0	✓	⌃⌃	✓	✓	✓	–	–	✓	✓	✓	✓	–
32	1	0	✓	✓	✓	✓	⌃⌃	✓	✓	✓	✓	✓	✓	✓
62	1	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⌄⌄

GitHub

https://github.com/Gilead-BioStats/gsm

SCIENCE

https://doi.org/10.1007/s43441-024-00651-4

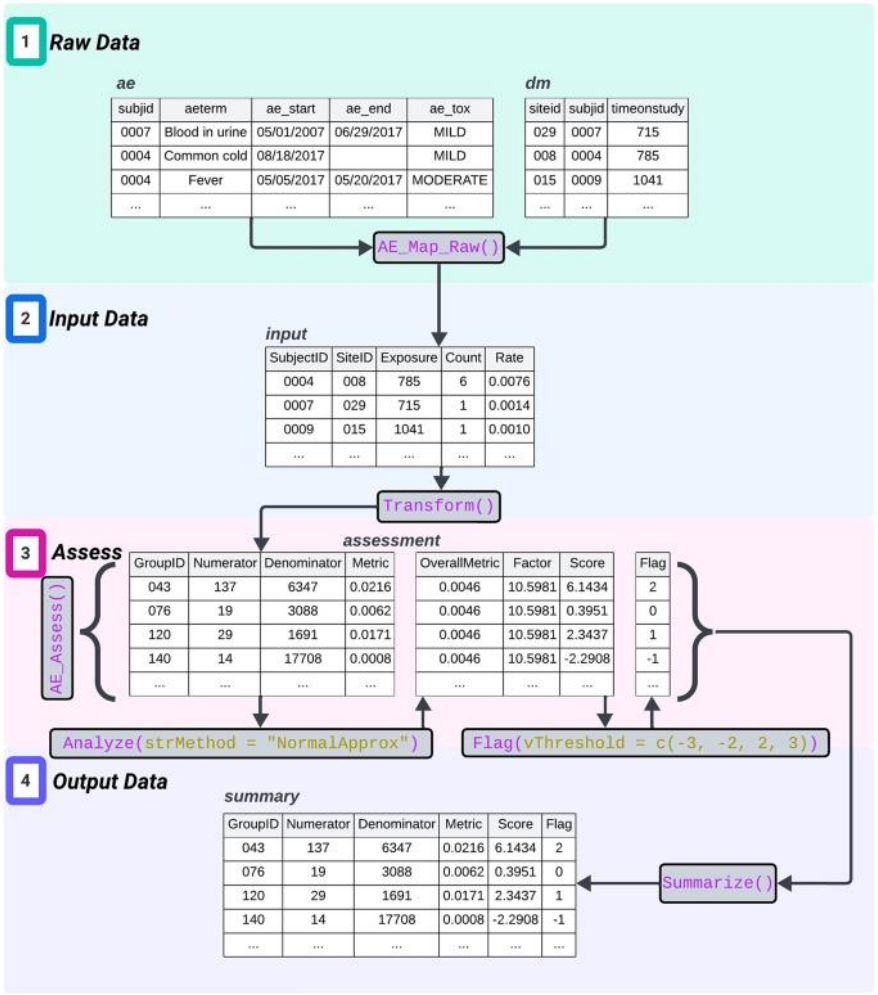
{gsm} – Good Statistical Monitoring

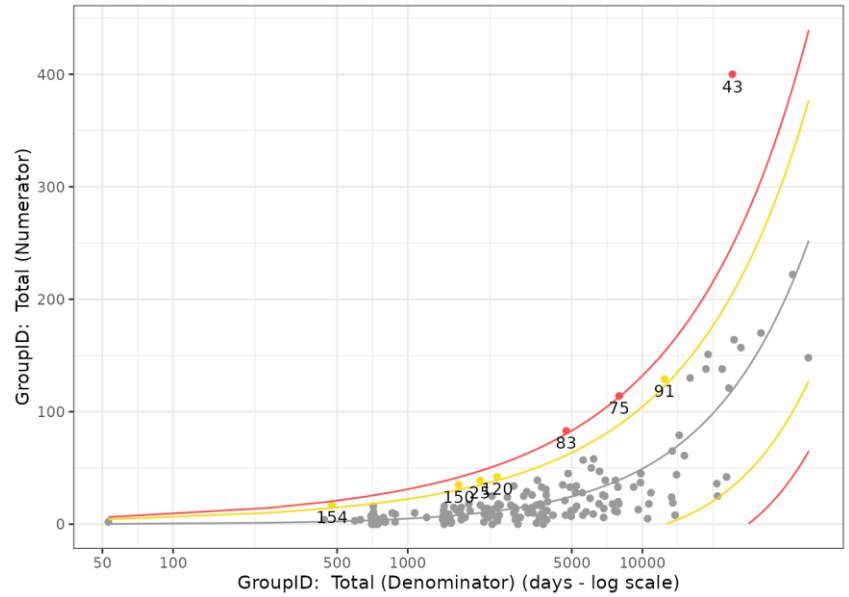
gsm

End to End RBQM Framework

- Mapping
- Processing
- Reporting

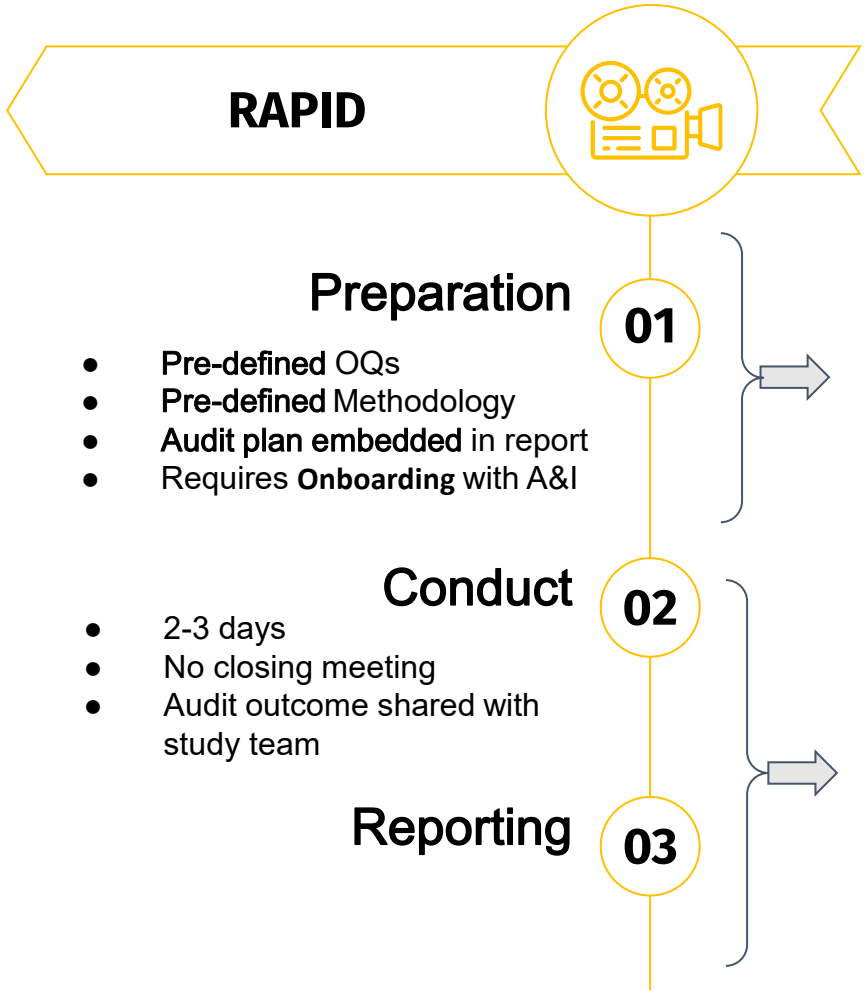
https://github.com/Gilead-BioStats/gsm





RAPID Audits

Real-time Audit Package Informed by Data



RAPID Audits

- Maintain Industry Standard
- Pre-Defined Methodology
- Based on Data Analytics
- High Coverage
- Fast
- Minimal Preparation Time

w! tL5 !9 □! ŕōśnℓś 9ōś■Ċℓ

Audit Process

Objective Questions:
- AE Late-Reporting
- AE Under-Reporting

Statistical Signal

simaerep + z-scores

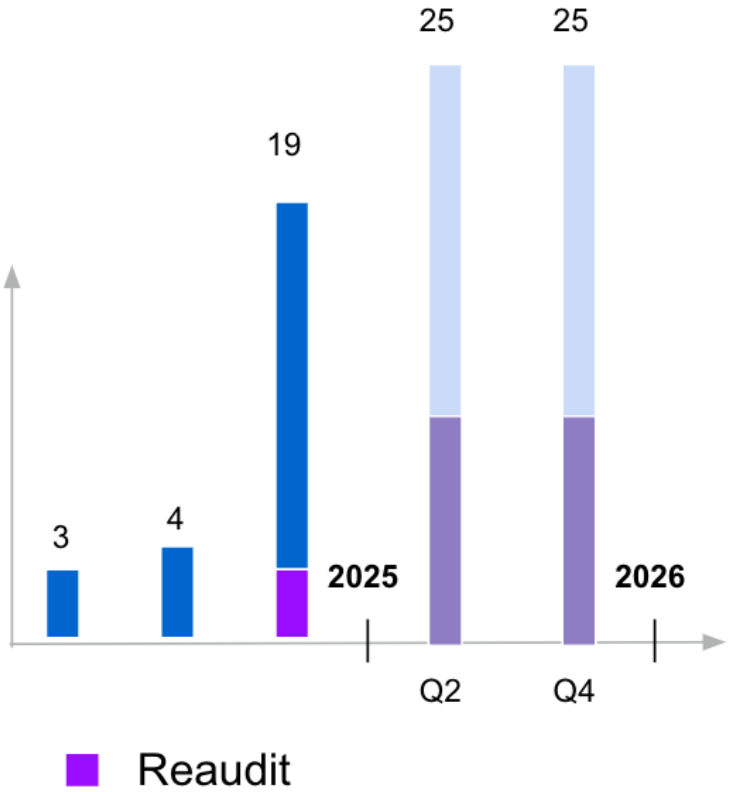
Document Review

Issues
Site Monitoring
Reports

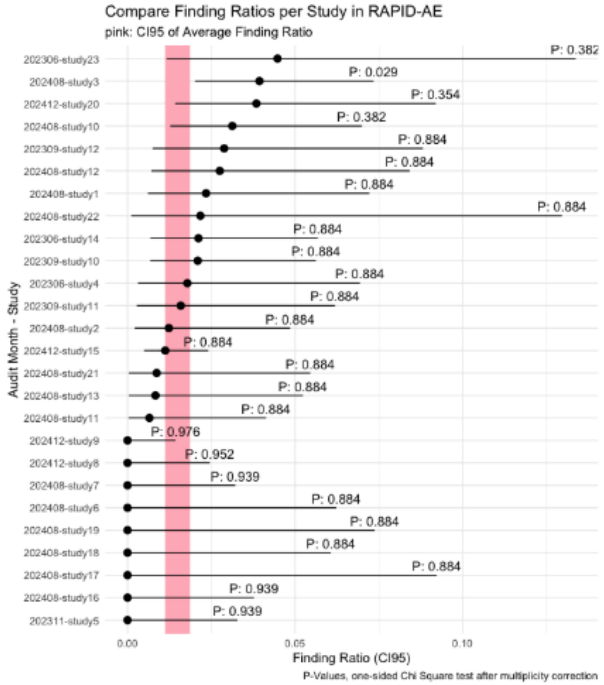
Finding

Positive Affirmation

Studies Audited



Portfolio Oversight



Outlook

- RBQMverse metapackage suite similar to pharmaverse (metapackage for data submission)
- Increase Inter-Operability
- Continued Release of gsm extensions

Thank You

Boehringer Ingelheim

DaVinci: A Modular R/Shiny Framework for Interactive Data Visualization in Clinical Research

Isabel Glauß, Boehringer Ingelheim

Agenda

- **Interactive Data Visualization with R/Shiny**
- **Modular Approach**
- **DaVinci Framework**
- **DaVinci Modules and their Use Cases**
- **How to Set Up a DaVinci App**
- **Our Open-Source Commitment**

Disclaimer

The views and opinions expressed in this presentation are those of the presenter and not necessarily those of Boehringer Ingelheim. Boehringer Ingelheim does not guarantee the accuracy or reliability of the information provided herein.

Interactive Data Visualization with R/Shiny

- Regular data reviews are key in clinical trials

  → Examples:
  data exploration, quality monitoring, safety reviews, etc.

- Traditional data review based on extensive static outputs are cumbersome to handle and may not immediately reveal all insights

- Better solution:
  Interactive R/Shiny apps with real-time access to clinical data

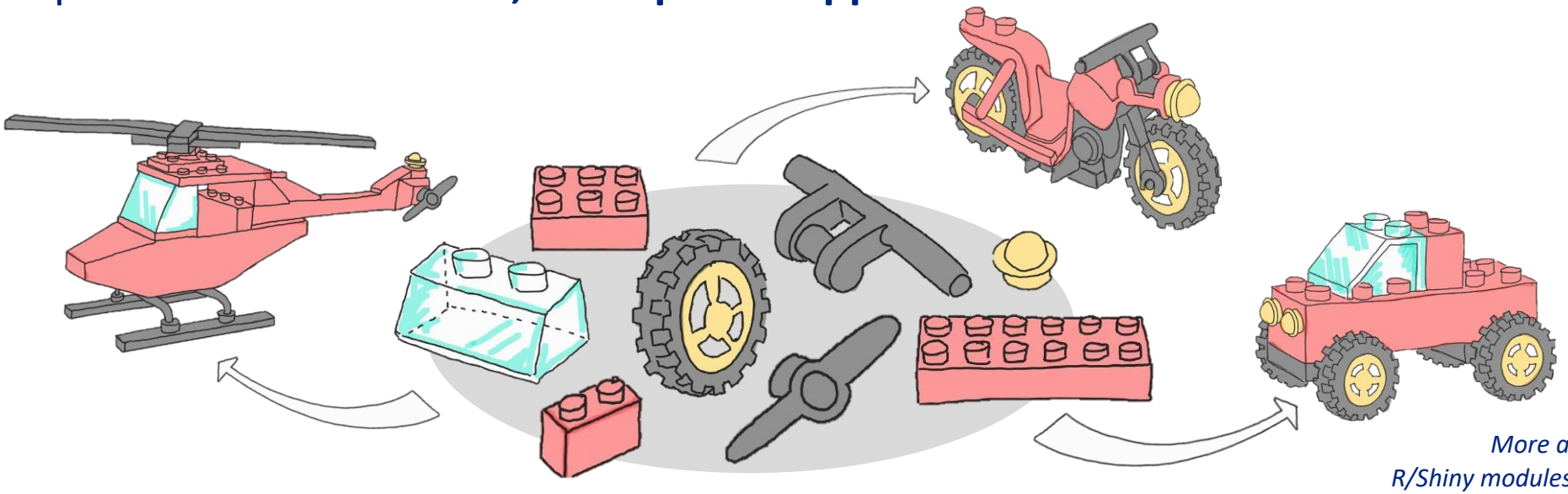
83% of our trial teams using DaVinci apps see a **(strong) positive business impact regarding speed & value** compared to static outputs

What is R/Shiny?

- R is an open-source programming language mostly used for data analysis, statistical computing and graphics
- {shiny} is an R package for creating web apps using R code
- *shiny.posit.co*

The Modular Approach

- One comprehensive stand-alone app that satisfies the needs of as many studies as possible quickly becomes unwieldy with trial-specific adjustments being impossible
- Therefore: DaVinci provides R/Shiny modules which serve as functional building blocks that can be composed into **customized, trial-specific apps**

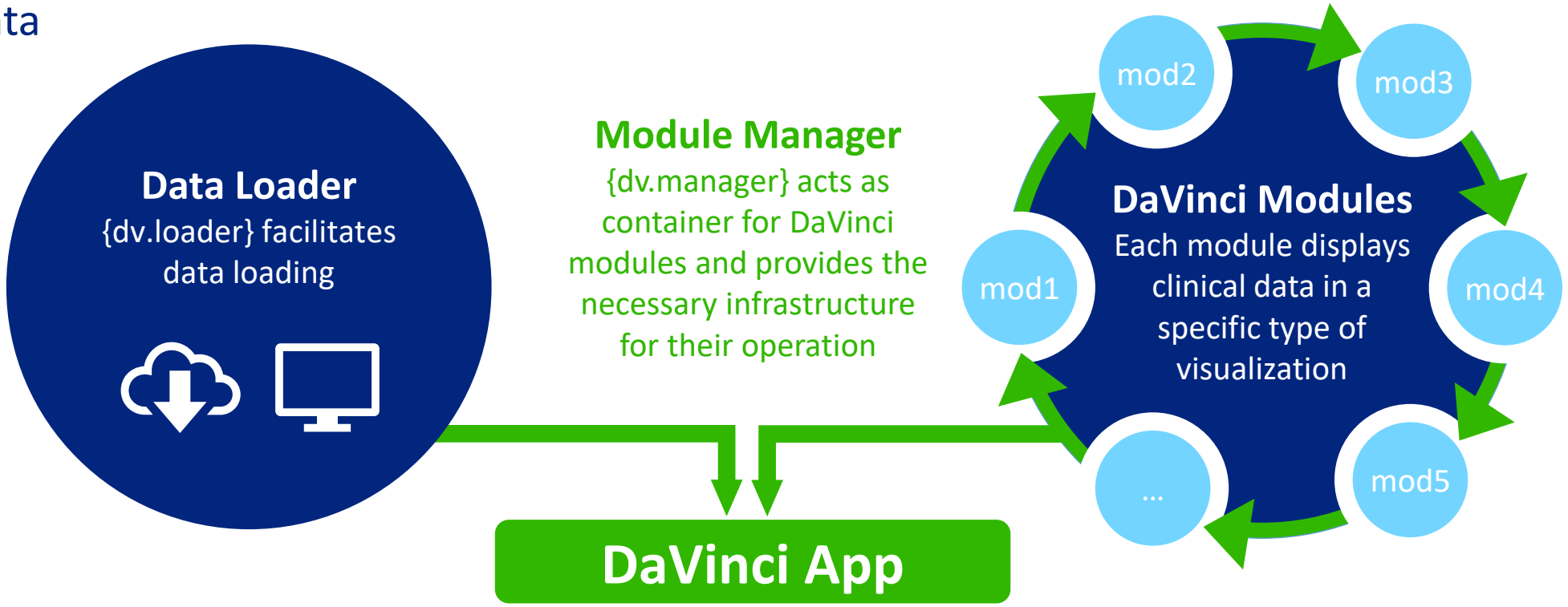


More about the concept of R/Shiny modules can be found here: mastering-shiny.org/scaling-modules

The DaVinci Framework

Main Components

DaVinci (= **Da**ta **Vi**sualizatio**n** for **C**linical **I**nsights) provides a modular framework with a rich and growing set of R/Shiny modules to review and explore clinical trial data interactively



The DaVinci Framework

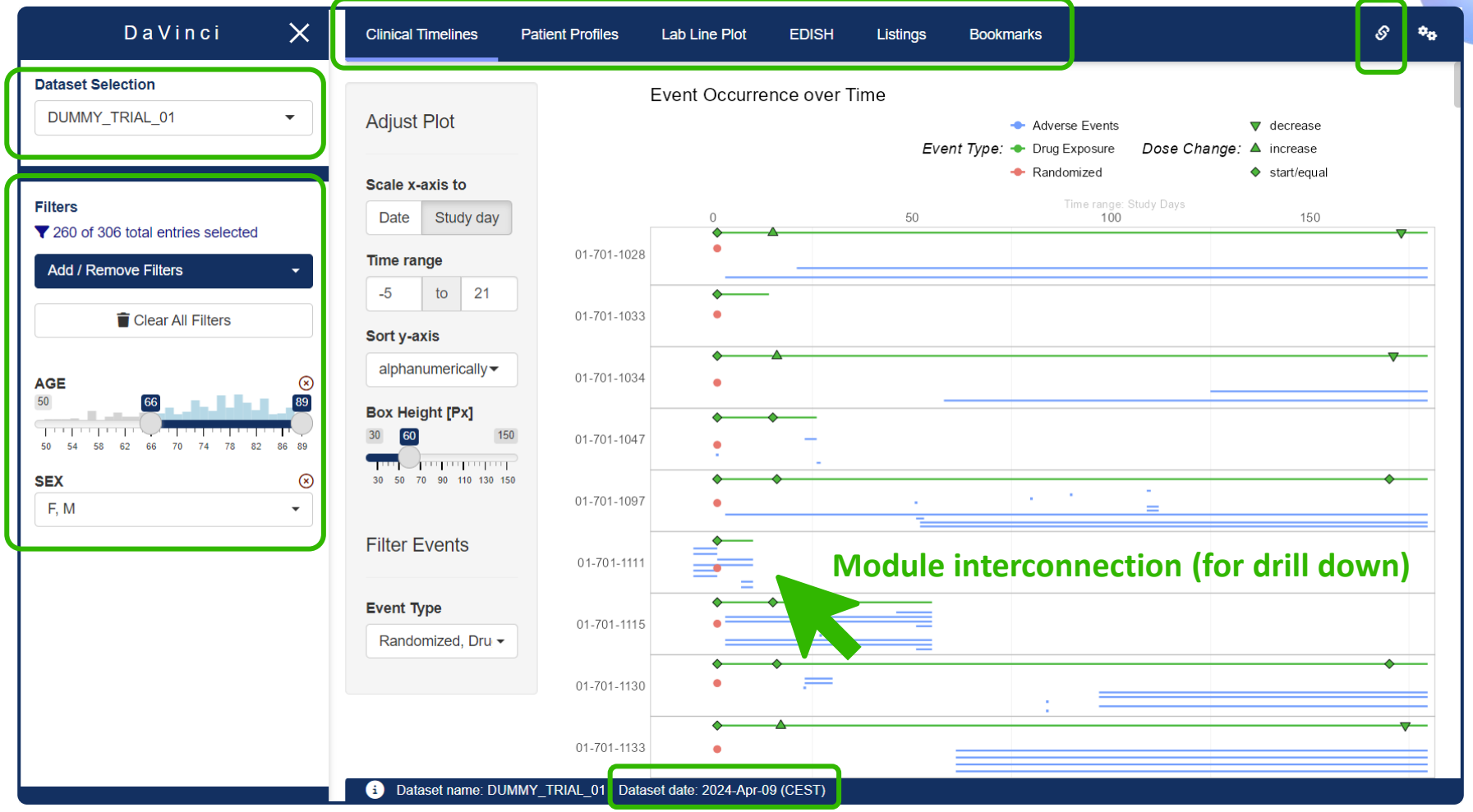
Key Features

List of modules

Bookmarking

Study switch

Global filter



+ Access restriction

DaVinci Module Gallery

… and example use cases

Safety Review, Clinical Quality Monitoring

Patient Profile {dv.papo}	Clinical Timelines {dv.clinlines}	eDISH Plot {dv.edish}	Hierarchy Table {dv.tables}
Lineplot {dv.explorer.parameter}	Boxplot {dv.explorer.parameter}	Correlation Heatmap {dv.explorer.parameter}	Forest Plot {dv.explorer.parameter}
ROC curve {dv.explorer.parameter}	Scatterplot {dv.explorer.parameter}	Matrix of Scatterplots {dv.explorer.parameter}	Waterfall Plus Heatmap {dv.explorer.parameter}
Listings {dv.listings}	Bookmark Manager {dv.bookman}	*Further modules are already in the pipeline…*	

Exploratory Parameter Analysis

Any use case

SCDM Live

The DaVinci Workflow

How to Set Up a DaVinci app

Note:
DaVinci can handle **any data source**, including SDTM and ADaM.
Raw data may require further preprocessing.

Load data

Pre-process

Specify modules

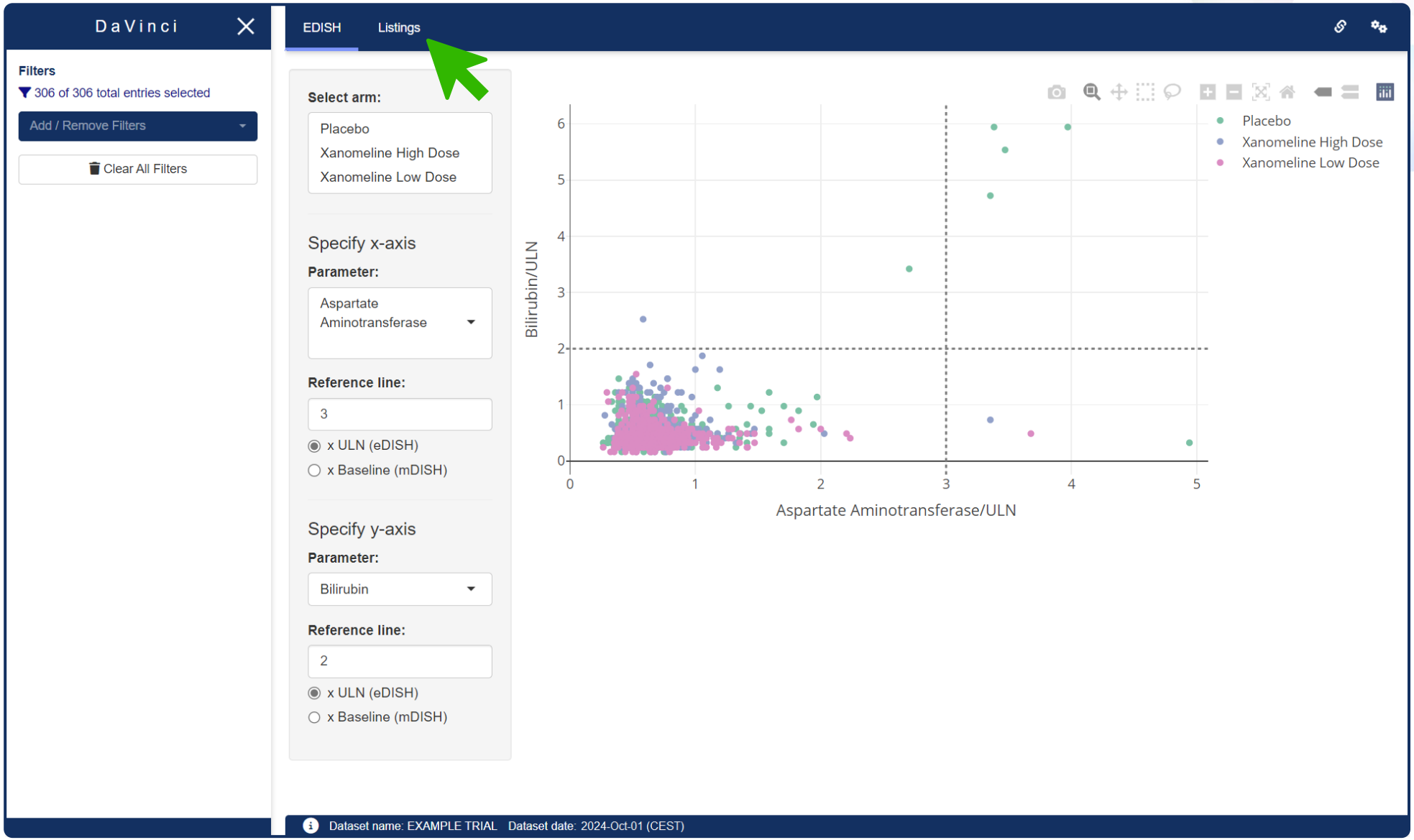
Launch app

```
# I) LOAD DATA ---
data_list <- dv.loader::load_data(
  sub_dir = "path/to/data",
  file_names = c("dm", "ae", "lb")
)

# II) PREPROCESSING ---
data_list$dm <- dv.listings::convert_data(data_list$dm)
data_list$ae <- dv.listings::convert_data(data_list$ae)
data_list$lb <- dv.listings::convert_data(data_list$lb)

# III) SPECIFY MODULES ---
edish <- dv.edish::mod_edish(
  module_id = "mod1",
  subject_level_dataset_name = "dm",
  lab_dataset_name = "lb",
  baseline_visit_val = "SCREENING 1"
)
listings <- dv.listings::mod_listings(
  module_id = "mod2",
  dataset_names = c("dm", "ae", "lb")
)

# IV) LAUNCH APP ---
dv.manager::run_app(
  data = list("EXAMPLE TRIAL" = data_list),
  module_list = list("EDISH" = edish, "Listings" = listings),
  filter_data = "dm"
)
```


DaVinci
EDISH
Listings
Filters
306 of 306 total entries selected
Add / Remove Filters
Clear All Filters
Select arm:
Placebo
Xanomeline High Dose
Xanomeline Low Dose
Specify x-axis
Parameter:
Aspartate Aminotransferase
Reference line:
3
x ULN (eDISH)
x Baseline (mDISH)
Specify y-axis
Parameter:
Bilirubin
Reference line:
2
x ULN (eDISH)
x Baseline (mDISH)
Bilirubin/ULN
Aspartate Aminotransferase/ULN
Placebo
Xanomeline High Dose
Xanomeline Low Dose
Dataset name: EXAMPLE TRIAL Dataset date: 2024-Oct-01 (CEST)

DaVinci

EDISH | Listings

Filters

306 of 306 total entries selected

Add / Remove Filters

Clear All Filters

Click to see inputs

Download

Reset Rows Order

Search:

	STUDYID [Study Identifier]	USUBJID [Unique Subject Identifier]	SUBJID [Subject Identifier for the Study]	RFSTDTC [Subject Reference Start Date/Time]	RFENDTC [Subject Reference End Date/Time]	RFXSTDTC [Date/Time of First Study Treatment]
	All	All	All	All	All	All
1	CDISCPILOT01	01-701-1015	1015	2014-01-02	2014-07-02	2014-01-02
2	CDISCPILOT01	01-701-1023	1023	2012-08-05	2012-09-02	2012-08-05
3	CDISCPILOT01	01-701-1028	1028	2013-07-19	2014-01-14	2013-07-19
4	CDISCPILOT01	01-701-1033	1033	2014-03-18	2014-04-14	2014-03-18
5	CDISCPILOT01	01-701-1034	1034	2014-07-01	2014-12-30	2014-07-01
6	CDISCPILOT01	01-701-1047	1047	2013-02-12	2013-03-29	2013-02-12
7	CDISCPILOT01	01-701-1057	1057			
8	CDISCPILOT01	01-701-1097	1097	2014-01-01	2014-07-09	2014-01-01
9	CDISCPILOT01	01-701-1111	1111	2012-09-07	2012-09-17	2012-09-07
10	CDISCPILOT01	01-701-1115	1115	2012-11-30	2013-01-23	2012-11-30
11	CDISCPILOT01	01-701-1118	1118	2014-03-12	2014-09-09	2014-03-12
12	CDISCPILOT01	01-701-1130	1130	2014-02-15	2014-08-16	2014-02-15
13	CDISCPILOT01	01-701-1133	1133	2012-10-28	2013-04-29	2012-10-28
14	CDISCPILOT01	01-701-1145	1145			
15	CDISCPILOT01	01-701-1146	1146	2013-05-20	2013-06-30	2013-05-20

306 records

Dataset name: EXAMPLE TRIAL Dataset date: 2024-Oct-01 (CEST)

Our Open-Source Commitment

- After several years of in-house user research and development, DaVinci is now **open-source and available on GitHub!**

  - DaVinci homepage: https://boehringer-ingelheim.github.io/davinci
  - DaVinci repositories: https://github.com/orgs/Boehringer-Ingelheim/repositories?q=dv.

- Benefits of open-sourcing DaVinci:

  - Makes our packages accessible to others
  - Encourages enhancements and direct contributions from the community

Homepage

Repositories

Thank you!

Any questions?

In case of questions please contact:

Isabel Glauß
isabel.glauss@boehringer-Ingelheim.com

Boehringer Ingelheim