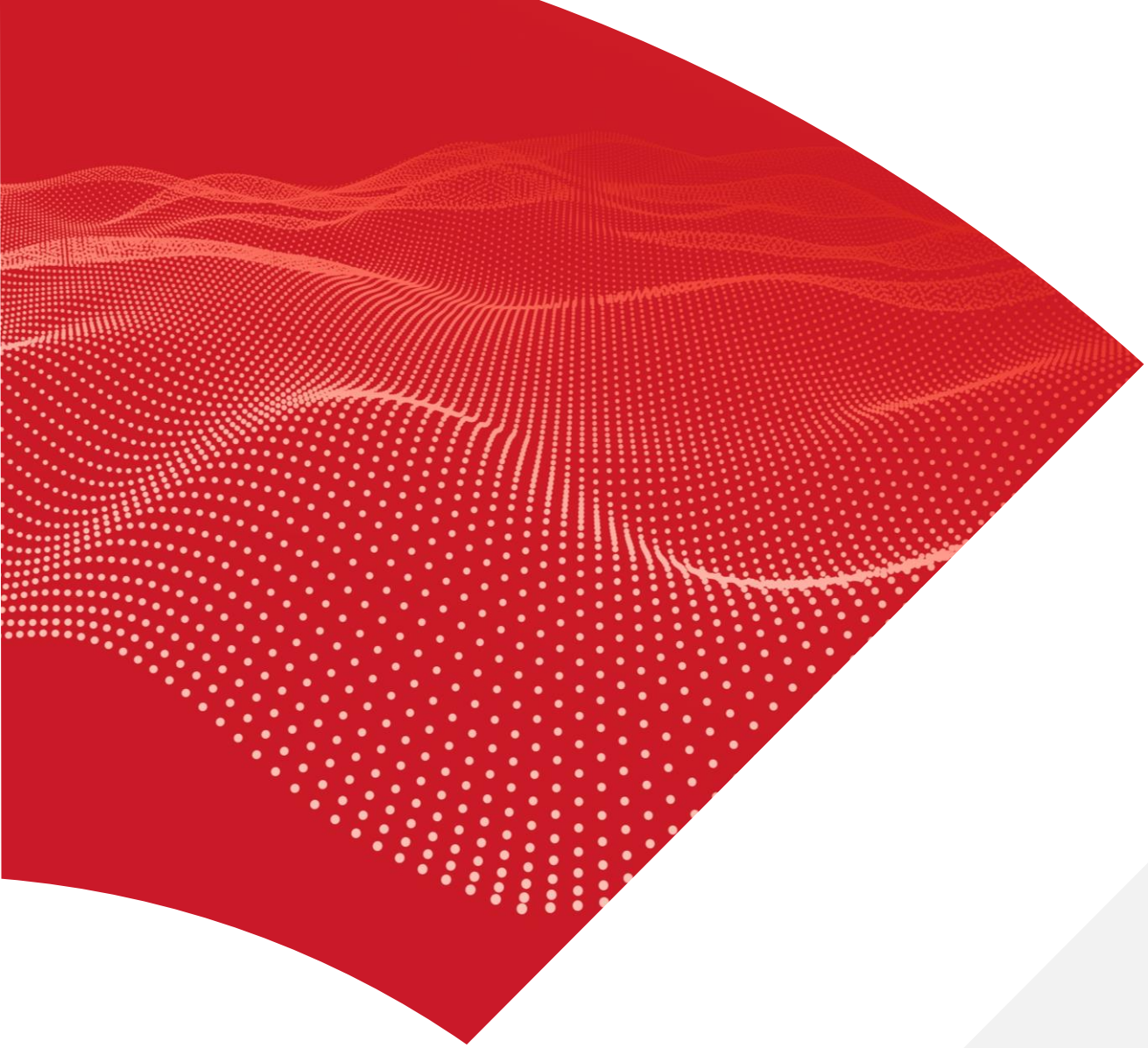


Frank Yang, CIMS GLOBAL

2025 China Pharma RUG Meeting

Leverage Open-Source Knowledge into Statistical Validity with Validation

Telephone
908-557-9366

Address
285 Davidson Ave.
Suite 504
Somerset, NJ 08873
USA

Website
www.cims-global.com

Bio

fyang@cims-global.com

- Frank Yang, Clinical Data Scientist, CIMS Global
- MS in Biostatistics, National Yang Ming Chiao Tung University
- One year of experience in R, with research interests in clinical data processing and related R package development.
- Member of the R/Pharma APAC Track Committee

Table of contents

CIMS

Introduction

1. **The rise of open-source technology in recent years and its increasing influence**
    1. Tidyverse
    2. Rmarkdown + quarto
    3. Shiny
    4. Comprehensive tools and platforms:
        - Licensed :Posit Workbench, Posit Connect, Posit Package Manager
        - Open-source: Rstudio Server, Shiny Server, Posit Public Package Manager
    5. AI: ChatGPT, DeepSeek, {ellmer}
2. **The impact of open-source technology in the pharmaceutical industry (e.g., {pharmaverse}, {NEST})**
    1. Data Generation: {sdtm.oak}, {metatools}, {admiral} series
    2. Results Reporting: {formatters}, {rtables}, {rlistings}, {tern}, {gtsummary}
    3. Interactive Result: {teal}, {golem}
    4. Utility: {quarto}, {officer}, {flextable}

CIMS

Developing R Packages in Pharma

1. **R Packages: Empowering R for Robust & Reproducible Analysis**
   - Encourages modularity and reusability
   - Supports validation and compliance
   - Long term maintain

2. **Good Development Practices (Based on OpenStatsGuide)**
   - **Documentation & Reproducibility:**
     Clear documentation & version control for traceability
   - **Testing & Quality Assurance:**
     Unit and functional tests to ensure accuracy & stability
   - **Lifecycle Management:**
     Minimize dependencies & follow structured validation

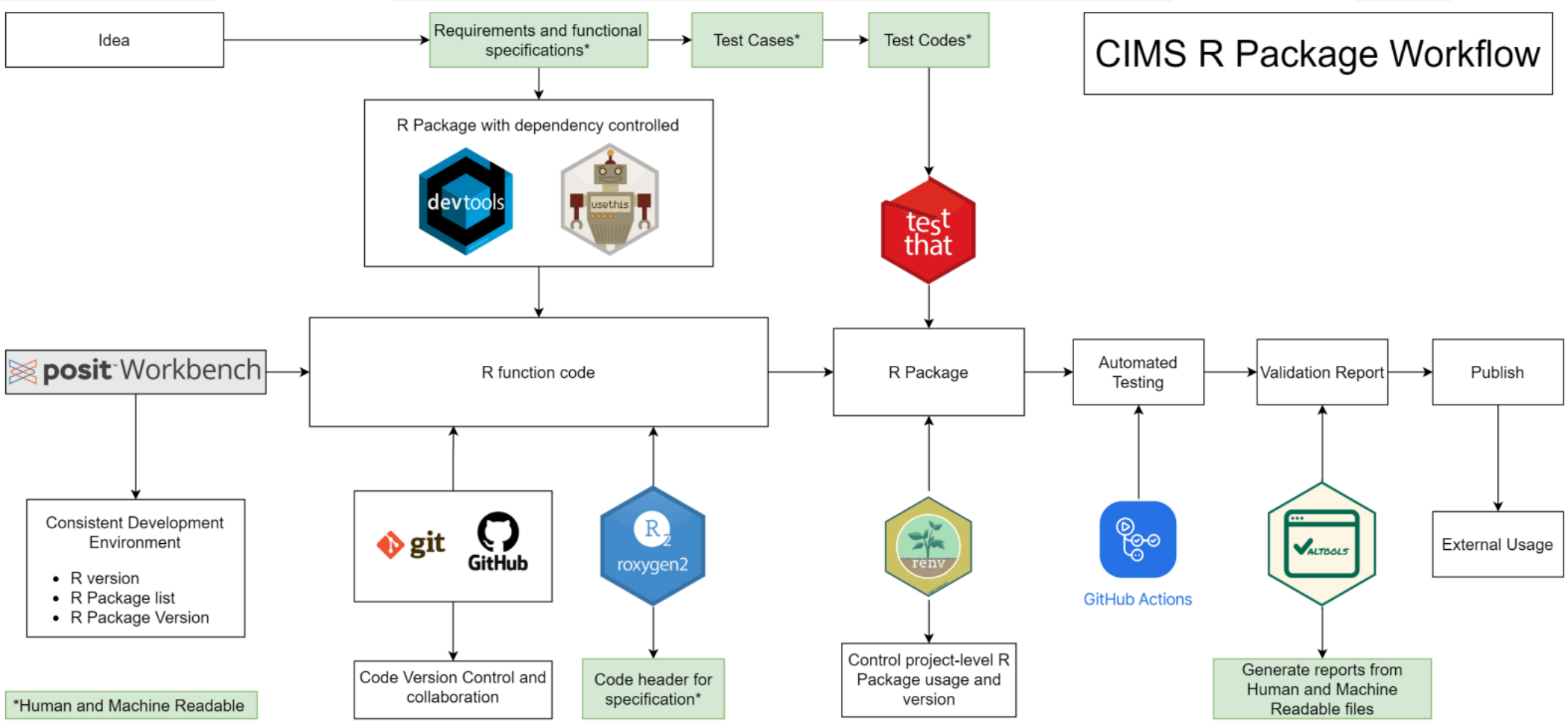
CIMS

Our solution for Key R Package Development

- **Planning & Requirement**
  - Define key questions to resolve and package goals
- **Functional Specifications**
  - Identify functions, inputs, and applicable methodology
- **Development**
  - Use R package structure ({devtools}, {usethis})
  - Maintain code version control (GitHub/Git)
  - Ensure consistent environment ({renv})
- **Documentation & Testing**
  - Function details with {roxygen2}
  - Unit tests with {testthat}
  - Automate with GitHub Actions
- **Validation & Compliance**
  - Generate Validation Report ({valtools})
  - Ensure package version control
  - Compare with SAS results

CIMS

Workflow



Validation Reports

- **PHUSE White paper: R Package Validation Framework**
- **{valtools}**

Validation Framework

{valtools} implements a convention for organizing information referenced during validation of R packages to generate reports. This framework is consistent for all five (5) validation modes.

Validation Modes

- separate from package
- from source code
- at installation
- after installation
- for distribution

Validation Elements

```
- validation/
|   |- requirements/
|   |- test_cases/
|   |- test_code/
|   |- validation.yml
|   |- change_log.md
|
|- report.Rmd
```

vt_use_validation()
vt_use_req()
vt_use_test_case()
vt_use_test_code()
vt_use_change_log()
vt_use_report()

Structure of Validation Reports

- **Description (vt_use_validation()) :**
  - Defines the validation scope and purpose.
- **Requirement (vt_use_req()) :**
  - Specifies expected functionality and behavior.
- **Functional Specification:**
  - Provides technical details on inputs, outputs, and algorithms.
- **Traceability:**
  - Ensures test cases align with requirements.
  - Links Requirements → Test Cases → Test Code.
- **Test Case (vt_use_test_case()) :**
  - Defines specific test scenarios based on requirements.
- **Test Result (vt_use_test_code()) :**
  - Records pass/fail outcomes and outputs.

CIMS

Example: {cimstfl}

- **Purpose: Generate TFLs from ADaM datasets to meet internal requirements.**
- **Process (Following the CIMS R Package Development Workflow):**
  1. Design the shell.
  2. Use packages such as {rtables}, {rlisting}, and {tern} to convert ADaM into TFL.
  3. Use {flextable} and {officer} for formatting DOCX outputs.
  4. Perform unit testing using some test data in the ARDs format.
  5. Generate a complete validation report.
  6. Further works: Compare with SAS results.
- **User-Configurable Options:**
  1. Corresponding variables in ADaM.
  2. Title, column width, footnotes, and other elements in the TFL output.
  3. The output of ARDs or not.

CIMS

Unit testing

- **Suggest using ARD/ARS from CDISC Standard for validation:**
  - One Column for numeric values, other columns for value's property
- **Example: AE table**

7. ARDS: Adverse Event by System Organ Class and Preferred Term Table

group1	group1_level	group2	group2_level	trt_var	trt	stat_name	result
				TRT01P	Treatment Group	N	
				TRT01P	Control Group	N	
				TRT01P	Total	N	
			At least one AE	TRT01P	Treatment Group	n	
			At least one AE	TRT01P	Treatment Group	pct	
			At least one AE	TRT01P	Control Group	n	
			At least one AE	TRT01P	Control Group	pct	
			At least one AE	TRT01P	Total	n	
			At least one AE	TRT01P	Total	pct	
AESOC	Primary SOC 1			TRT01P	Treatment Group	n	
AESOC	Primary SOC 1			TRT01P	Treatment Group	pct	
AESOC	Primary SOC 1			TRT01P	Control Group	n	
AESOC	Primary SOC 1			TRT01P	Control Group	pct	
AESOC	Primary SOC 1			TRT01P	Total	n	
AESOC	Primary SOC 1			TRT01P	Total	pct	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Treatment Group	n	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Treatment Group	pct	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Control Group	n	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Control Group	pct	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Total	n	
AESOC	Primary SOC 1	AEBODSYS	Preferred Term 1	TRT01P	Total	pct	

System Organ Class Preferred Term	Treatment Group (N=xx)	Control Group (N=xx)	Total (N=xx)
At least one AE	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Primary SOC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 3	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Primary SOC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Preferred Term 3	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

Validation Report

Validation Report for cimstfl

Peng Zhang, Frank Yang, Nina Han, Vivian Chang, Jade Lee

2025-03-14

Contents

• User Requirement Specification

01.04 : Medical History by System Organ Class Table

The table will include the medical history of the study population, categorized by different preferred terms under each primary System Organ Class (SOC), as coded using MedDRA. It will provide a descriptive summary that includes both counts and percentages. It will include three columns for active treatment, control treatment and total group. For different types of the variables, the rows will include:

- number of patients with any primary system organ class,
- summary of primary SOC,
- summary of preferred term within each SOC.

The table can be sorted by system organ class incidence in (specified descending or increasing) order and by preferred term incidence in order within each system organ class. The ordering is according to the specified column. The user can specify the target population and the threshold for filtering values for the specified column, showing only the rows larger than the specified value.

• Test Cases

01.04.1:

Use the function `t_b_mh_01()` to generate a medical history by system organ class Table. Supply the following input values to the function:

- planned treatment, which groups the data by treatment arms, distinguishing between the control group and treatment group,
- `pop_var = 'SAFFL'` to filter the safety population as our dataset,
- primary system organ class,
- reported term for the medical history.

Verify that the function correctly calculates the medical history table and ARD based on the provided data. The table should include a total of 268 subjects, with 134 in the Drug X group and 134 in the Placebo group. Only subjects within the safety population should be included, incorporating primary system class and under each primary system class, showing the reported preferred term, providing count and percentage, both categorized using MedDRA. In this trial, a total of 245 patients (91.4%) reported any primary system organ class, with 122 patients (91.0%) in the Drug X group and 123 patients (91.8%) in the Placebo group. Specific classifications such as class A, B, C, and D, and the associated preferred terms within each classification should be detailed. Finally, the ARD should transform the information in the table into a data frame, which holds the summary results and can be used for quality control checks.

To further test the robustness of the function, apply test cases 01.04.2 through 01.04.6 with the following variables, which are of the same type as the original target variables but with different values, and compare the results with those obtained from the code written by a secondary developer:

- `arm_var = 'TRT01P','TRT01A','TRT02P',TRT02A'`,
- `soc_var = 'MHSOC','MHBODSYS','MHDISTAT','EOSSTT','STRATA1'`,
- `pt_var = 'MHTERM','MHDECOD','EOTSTT','STRATA1','STRATA2'`,
- `pop_var = 'ITTFL','SAFFL','BMEASIFL'`.

Also, test case 01.04.7 will apply Study 1 data to verify the function works on a different data set.

• Function Usage Overview

01.04 Medical History by System Organ Class Table

Description

Medical History

Usage

```
t_b_mh_01(
  adam,
  arm_var = "TRT01P",
  soc_var = "MHSOC",
  pt_var = "MHTERM",
  pop_var = NULL,
  decreasing = T,
  sort_col = "Total",
  filter_val_col = NULL,
  filter_val_pct = 0,
  ard = T
)
```

Arguments

- `adam`: The list of data frames, which should include the `adsl` and `admh` datasets.
- `arm_var`: The variable representing the planned treatment.
- `soc_var`: The variable for the primary system organ class.
- `pt_var`: The variable for the preferred term.
- `pop_var`: The variable used to filter the population for analysis.
- `decreasing`: A logical value indicating whether to sort the table in decreasing order. If TRUE, the table will be sorted in descending order based on `sort_col`.
- `sort_col`: The column to be used for sorting the table.
- `filter_val_col`: The column used for filtering values.
- `filter_val_pct`: A numeric value specifying the threshold for filtering in `filter_val_col`. Only values that are greater than this percentage will be retained.
- `ard`: A logical value indicating whether to generate the analysis result data (ARD).

Example: MH table

CIMS

• Functions & Test Case Authors

Function Name	Editor
f_e_km_01	Frank Yang, Vivian Chang
f_s_lb_01	Frank Yang, Vivian Chang
f_s_vs_01	Vivian Chang
l_b_cm_01	Frank Yang, Vivian Chang
l_b_dm_01	Frank Yang, Vivian Chang
l_b_ds_01	Frank Yang, Vivian Chang
l_b_mh_01	Frank Yang, Vivian Chang
l_b_pd_01	Frank Yang, Vivian Chang
l_s_ae_01	Frank Yang, Vivian Chang
l_s_lb_01	Frank Yang, Vivian Chang
l_s_vs_01	Frank Yang, Vivian Chang
t_b_cm_01	Frank Yang, Vivian Chang
t_b_dm_01	Frank Yang, Vivian Chang
t_b_ds_01	Frank Yang, Vivian Chang
t_b_mh_01	Frank Yang, Vivian Chang
t_b_pd_01	Frank Yang, Vivian Chang
t_e_tte_01	Frank Yang, Vivian Chang
t_s_ae_01	Frank Yang, Vivian Chang
t_s_ae_02	Frank Yang, Vivian Chang
t_s_ae_03	Frank Yang, Vivian Chang
t_s_ae_04	Frank Yang, Vivian Chang
t_s_lb_01	Frank Yang, Vivian Chang
t_s_lb_02	Frank Yang, Vivian Chang
t_s_lb_03	Frank Yang, Vivian Chang
t_s_lb_04	Frank Yang, Vivian Chang
t_s_vs_01	Frank Yang, Vivian Chang
t_s_vs_02	Frank Yang, Vivian Chang

test_cases	editor	editDate
01. Baseline test cases	Nina Han	2025-02-07
02. Safety test cases	Nina Han	2025-02-08
03. Efficacy test cases	Nina Han	2025-02-11
04. Listing test cases	Nina Han	2025-02-11

• Traceability

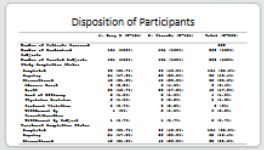
Requirement Name	Requirement ID	Test Case Name	Test Cases
Requirement 01	01.01	01. Baseline test cases	01.01
	01.02		01.02
	01.03		01.03
	01.04		01.04
	01.05		01.05
Requirement 02	02.01	02. Safety test cases	02.01
	02.02		02.02
	02.03		02.03
	02.04		02.04
	02.05		02.05
	02.06		02.06
	02.07		02.07
	02.08		02.08
	02.09		02.09
	02.10		02.10
	02.11		02.11
	02.12		02.12
Requirement 03	03.01	03. Efficacy test cases	03.01
	03.02		03.02
Requirement 04	04.01	04. Listing test cases	04.01
	04.02		04.02
	04.03		04.03
	04.04		04.04
	04.05		04.05
	04.06		04.06
	04.07		04.07
	04.08		04.08

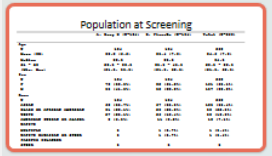
• Validation Results

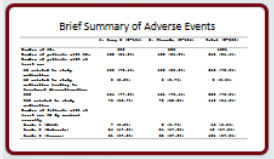
Test	Results	Pass.Fail
01.05.7.2	As expected	Pass
01.05.7.3	As expected	Pass
01.05.7.4	As expected	Pass
01.05.2.1	As expected	Pass
01.05.2.2	As expected	Pass
01.05.3.1	As expected	Pass
01.05.3.2	As expected	Pass
01.05.4.1	As expected	Pass
01.05.4.2	As expected	Pass
01.05.5.1	As expected	Pass
01.05.5.2	As expected	Pass
01.05.6.1	As expected	Pass
01.05.6.2	As expected	Pass
01.05.8.1	As expected	Pass
01.05.8.2	As expected	Pass

Potential connection : {AutoRepo}

- Prepared {quarto} file and {Rmarkdown} file to customize the need based on dummy/dry-run data
- Configure what tables and figures are needed in the slides, using validated gallery from {cimstfl}
- When actual data comes in, one-click on the codes to generate the slides





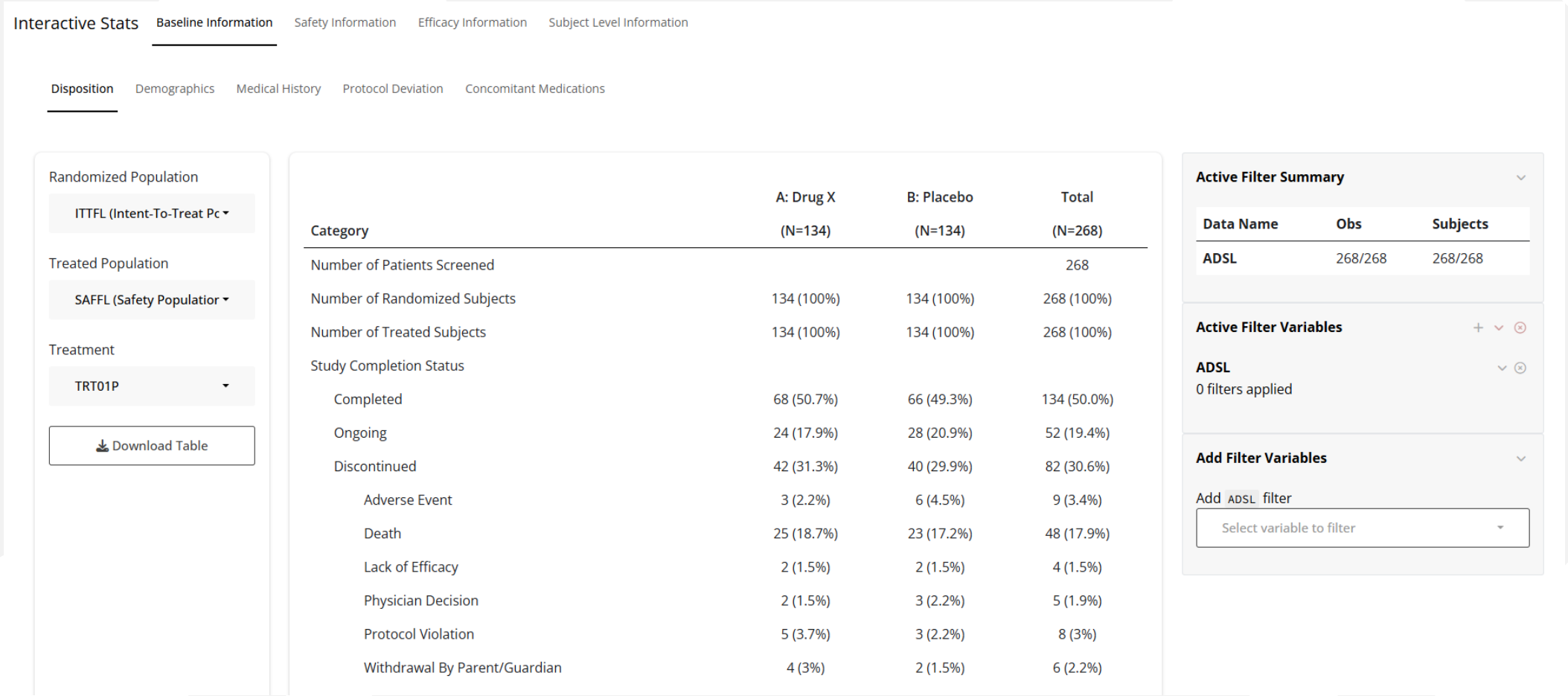


Brief Summary of Adverse Events

	A: Drug X (N=134)	B: Placebo (N=134)	Total (N=268)
Number of AEs	609	622	1231
Number of patients with AEs	122 (91.0%)	123 (91.8%)	245 (91.4%)
Number of patients with at least one			
AE related to study medication	105 (78.4%)	108 (80.6%)	213 (79.5%)
AE related to study medication leading to treatment discontinuation	3 (2.2%)	5 (3.7%)	8 (3.0%)
SAE	104 (77.6%)	101 (75.4%)	205 (76.5%)
SAE related to study medication	76 (56.7%)	70 (52.2%)	146 (54.5%)
Number of Patients with at least one AE by maximal severity			
Grade 1 (Mild)	7 (5.2%)	9 (6.7%)	16 (6.0%)
Grade 2 (Moderate)	24 (17.9%)	24 (17.9%)	48 (17.9%)
Grade 3 (Severe)	91 (67.9%)	90 (67.2%)	181 (67.5%)

{Interactive.stats} – adopting {teal}

- **Resolve current difficulties in DMC practices of Large number of pages, Inefficient static output, Insufficient information, Time-consuming follow-up**



{stat2csr}

Section 10.1 | t_b_dm_01 | Generate the Section

Select Table Rows

- [x] AGE
- [x] SEX
- [x] RACE
- [x] ETHNIC

	A: Drug X (N=134)	B: Placebo (N=134)	Total (N=268)
Age			
N	134	134	268
Mean (SD)	33.8 (6.6)	35.4 (7.9)	34.6 (7.3)
Median	33.0	35.0	34.0
Q1 - Q3	28.0 - 39.0	30.0 - 40.0	29.0 - 39.0
(Min, Max)	(21.0, 50.0)	(21.0, 62.0)	(21.0, 62.0)

- The mean (SD) Age was {AGE_mean} ({AGE_sd}) with a median of {AGE_median} (range {AGE_min} to {AGE_max}).
- Of {num_participants} participants enrolled, {SEX1_num} ({SEX1_pop}%) participants were {SEX1_name}, {SEX2_num} ({SEX2_pop}%) participants were {SEX2_name}.
- Of {num_participants} participants enrolled, {RACE1_num} ({RACE1_pop}%) participants were

- The mean (SD) Age was 34.6 (7.3) with a median of 34 (range 21 to 62).
- Of 268 participants enrolled, 161 (60.1%) participants were F, 107 (39.9%) participants were M.
- Of 268 participants enrolled, 135 (50.4%) participants were Asian, 59 (22%) participants were Black or African American, 53 (19.8%) participants were White, 19 (7.1%) participants were American Indian or Alaska Native, 1 (0.4%) participants were Multiple, 1 (0.4%) participants were Native Hawaiian or Other Pacific Islander, 0 (0%) participants were Other, 0 (0%) participants were Unknown.
- Of 268 participants enrolled, 33 (12.3%) participants were Hispanic or Latino, 207 (77.2%) participants were Not Hispanic or Latino, 16 (6%) participants were Not Reported, 12 (4.5%) participants were Unknown.

CSR Report

- Once the review and edit is completed, the report and corresponding {quarto} file can be generated
- If one has pre-specified template, it can be read into the code and works for the final report.
- Such report can be editable in the future if necessary.

Clinical Study Report

9.1 Disposition of Participants

	A: Drug X (N=134)	B: Placebo (N=134)	Total (N=268)
Number of Patients Screened			268
Number of Randomized Subjects	134 (100%)	134 (100%)	268 (100%)
Number of Treated Subjects	134 (100%)	134 (100%)	268 (100%)
Study Completion Status			
COMPLETED	68 (50.7%)	66 (49.3%)	134 (50.0%)
ONGOING	24 (17.9%)	28 (20.9%)	52 (19.4%)
DISCONTINUED	42 (31.3%)	40 (29.9%)	82 (30.6%)
ADVERSE EVENT	3 (2.2%)	6 (4.5%)	9 (3.4%)
DEATH	25 (18.7%)	23 (17.2%)	48 (17.9%)
LACK OF EFFICACY	2 (1.5%)	2 (1.5%)	4 (1.5%)
PHYSICIAN DECISION	2 (1.5%)	3 (2.2%)	5 (1.9%)
PROTOCOL VIOLATION	5 (3.7%)	3 (2.2%)	8 (3%)
WITHDRAWAL BY PARENT/GUARDIAN	4 (3%)	2 (1.5%)	6 (2.2%)
WITHDRAWAL BY SUBJECT	1 (0.7%)	1 (0.7%)	2 (0.7%)
Treatment Completion Status			
COMPLETED	68 (50.7%)	66 (49.3%)	134 (50.0%)
ONGOING	24 (17.9%)	28 (20.9%)	52 (19.4%)
DISCONTINUED	42 (31.3%)	40 (29.9%)	82 (30.6%)

A total of 268 participants were initially screened for inclusion in the study. Of these, 268 participants were randomized into the study, and 268 participants received medication. The participants were divided into two groups: the A: Drug X group (134) and the B: Placebo group (134). 134 subjects had completed the study, while 82 subjects discontinued prematurely, and 52 subjects were ongoing.

10.2.1 Population at Screening

	A: Drug X (N=134)	B: Placebo (N=134)	Total (N=268)
Age			
N	134	134	268
Mean (SD)	33.8 (6.6)	35.4 (7.9)	34.6 (7.3)
Median	33.0	35.0	34.0
Q1 - Q3	28.0 - 39.0	30.0 - 40.0	29.0 - 39.0
(Min, Max)	(21.0, 50.0)	(21.0, 62.0)	(21.0, 62.0)
Sex			
N	134	134	268
F	79 (59.0%)	82 (61.2%)	161 (60.1%)
M	55 (41.0%)	52 (38.8%)	107 (39.9%)
Race			
N	134	134	268
ASIAN	68 (50.7%)	67 (50.0%)	135 (50.4%)
BLACK OR AFRICAN AMERICAN	31 (23.1%)	28 (20.9%)	59 (22.0%)
WHITE	27 (20.1%)	26 (19.4%)	53 (19.8%)
AMERICAN INDIAN OR ALASKA NATIVE	8 (6.0%)	11 (8.2%)	19 (7.1%)
MULTIPLE	0	1 (0.7%)	1 (0.4%)
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	0	1 (0.7%)	1 (0.4%)
OTHER	0	0	0
UNKNOWN	0	0	0
Ethnicity			
N	134	134	268
HISPANIC OR LATINO	15 (11.2%)	18 (13.4%)	33 (12.3%)
NOT HISPANIC OR LATINO	104 (77.6%)	103 (76.9%)	207 (77.2%)

Discussion

- **Advantages of Using Open-Source:**
  - Shared Solution: People customize it for their own usage based on this foundation.
  - Collaboration: Supports a long-term development cycle with collaborative effort.
  - Advanced Features: Enables tools like Shiny and RMarkdown.
  - Scalability: More efficient tools continue to emerge.

- **On the Other Hand:**
  - Justify Reliability: Open-source is not always reliable—build confidence in statistical validity through validation (e.g., {valtools}).
  - Figure Out What You Want to Build: Carefully plan and design.
  - Develop with Good Practices.
  - Maintain & Consider User Feedback.

Thank you.

We welcome any questions or thoughts that you may have.

Tai Xie, CEO, CIMS Global
tomx@cims-global.com

Peng Zhang, Associate Director, CIMS Global
pzhang@cims-global.com

www.cims-global.com

CIMS