phuse·virtual

Single Day Event

Uday Preetham Palukuru
Ruchitbhai Patel
Christine Teng
Yilong Zhang,
Merck & Co., Inc.,
Kenilworth, NJ, USA

Concurrent Use of SAS® and R in Clinical Trial Reporting

Uday Preetham Palukuru
Ruchitbhai Patel
Christine Teng
Yilong Zhang

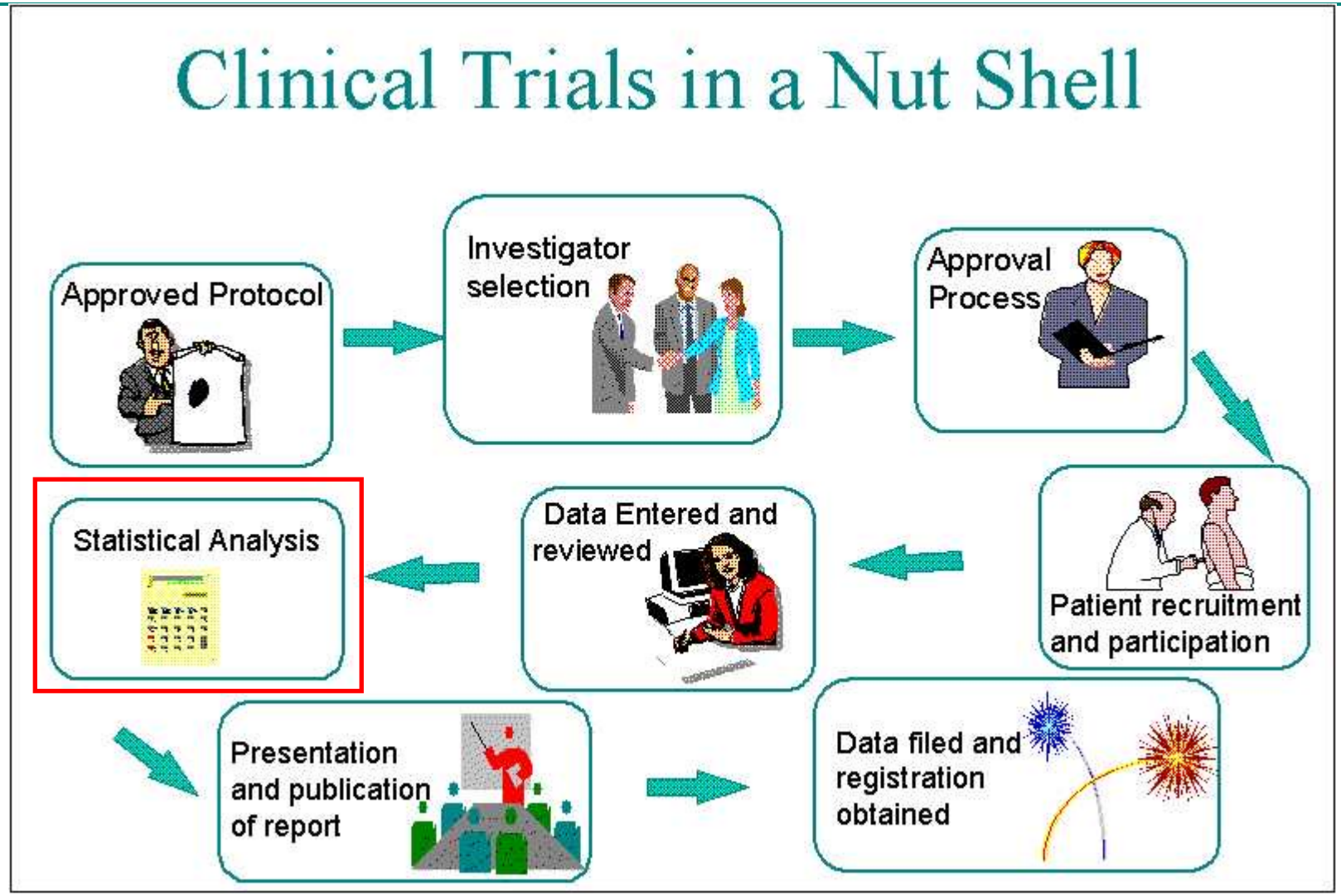
PHUSE Americas Autumn SDE

October 14, 2021

Table of Contents

- Introduction
  - Clinical Trials Lifecycle
  - Analysis & Reporting (A&R) Programming SDLC
  - SAS
  - SAS usage in A&R
  - R
  - R usage in A&R
- Motivation
- Concurrent SAS and R usage
  - Case 1
  - Case 2
- Conclusion

Introduction: Clinical Trials Lifecycle



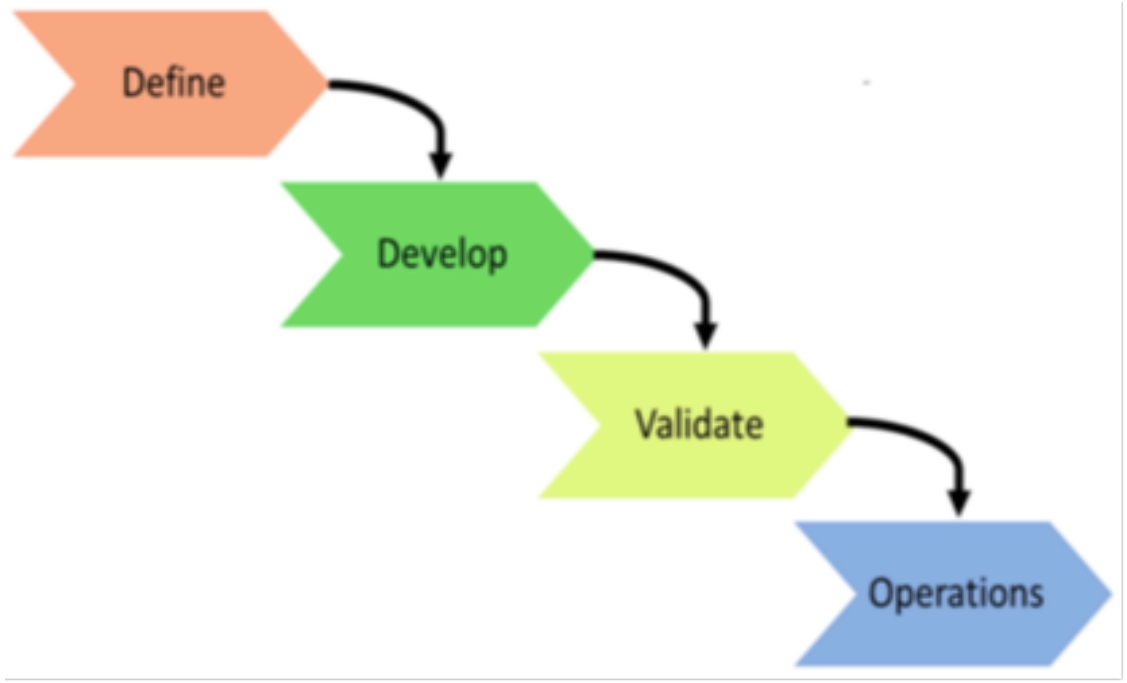
http://rxpharma.blogspot.com/2010/11/clinical-trials.html

Introduction: Analysis & Reporting (A&R) Programming SDLC

A&R in Late Stage Clinical Trials:

- Analysis datasets (ADaM) to facilitate analysis
- Reporting of data by developing Tables, Listings and Figures (TLF)
- A&R package: ADaM datasets and TLFs along with associated programs
- Multiple packages might be necessary for each clinical trial
- Input(SDTM/Raw) and Output(ADaM and TLFs) data quality is a critical part of the A&R process

A&R Programming Workflow

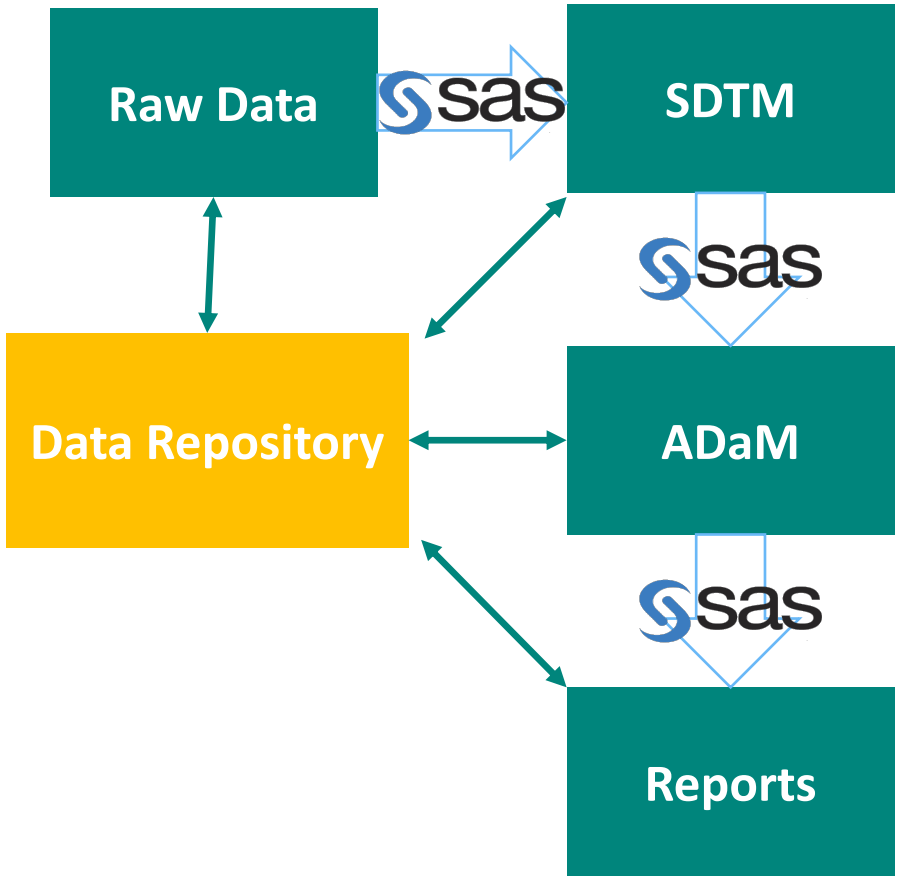


Introduction: SAS

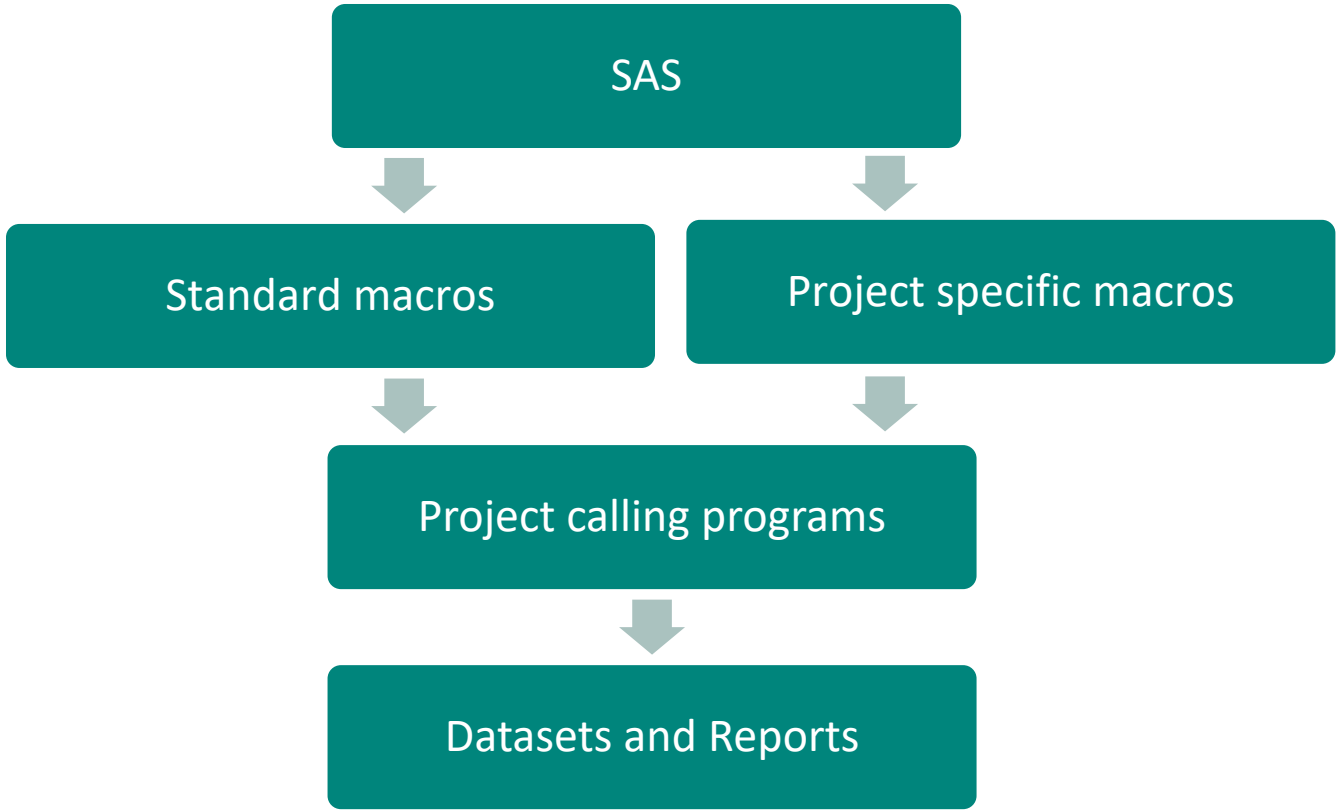
- SAS is a software widely used in the pharmaceutical industry for statistical A&R
- Well established and validated software packages available for use in highly regulated clinical trials environment
  - e.g. SAS/STAT, SAS/GRAPH, SAS/MACRO
- Transformation, analysis, reporting and visualization of data can be performed through graphical user interface and programs
- Widespread Industry use for clinical trials data A&R

Introduction: SAS usage in A&R

SAS usage in Clinical trials A&R



SAS Development Structure

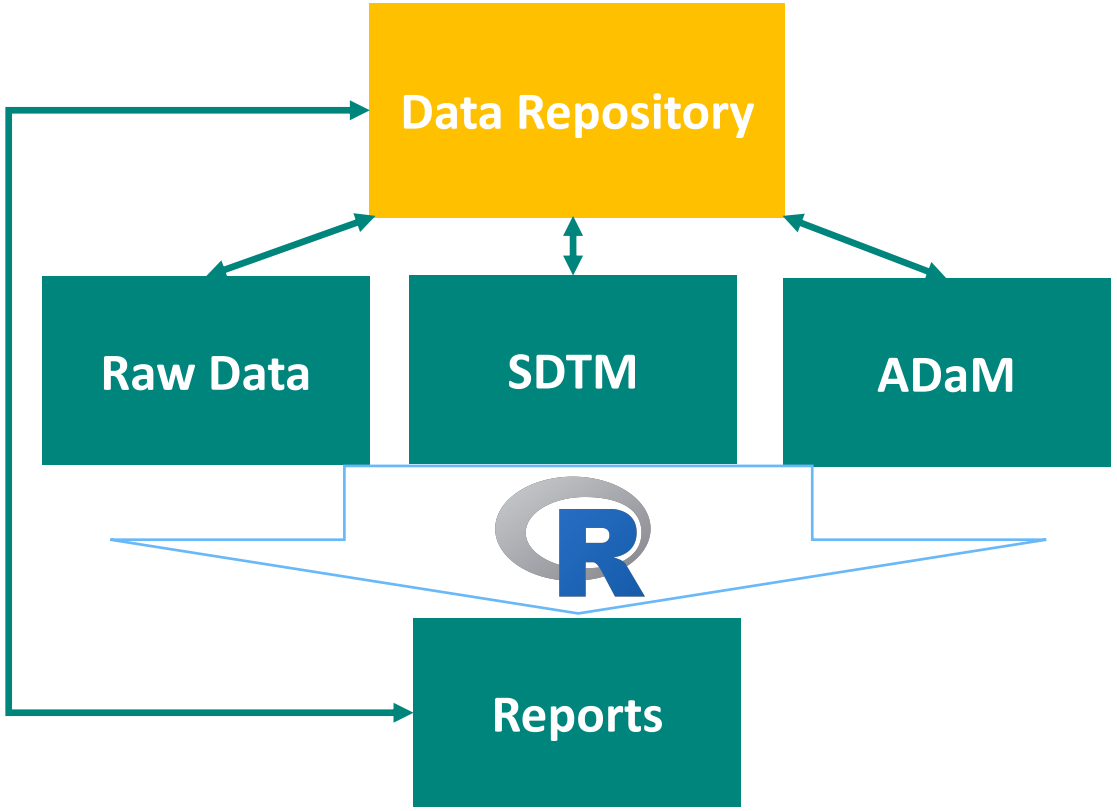


Introduction: R

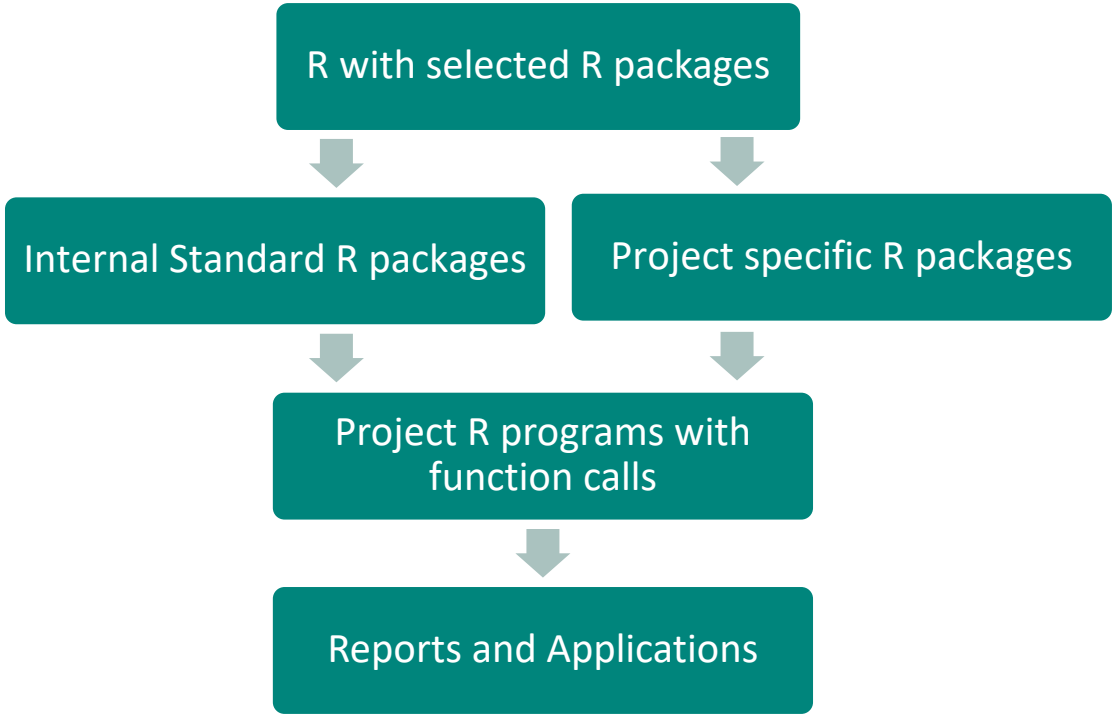
- R is a software also used for statistical computing and graphics (comparable to SAS)
- An R function is a set of R statements organized together for a specific task and can have arguments (comparable to SAS macro)
  - e.g. read.csv() is a function read a csv file into R
- An R package is a collection of R functions
  - e.g. ggplot2 is an R package to create graphs
- Highly interactive Integrated Development Environment(IDE) available from vendors such as RStudio for code development and deployment
- Recent increase in interest in use of R for clinical trials data A&R

Introduction: R usage in A&R

R usage in Clinical trials A&R



R Development Structure



Motivation: Leveraging Individual Strengths

Leveraging Individual Strengths

SAS
Large number of existing validated procedures for A&R use
Established SOPs for use in clinical trial A&R due to widespread industry use
Support Infrastructure and commercial validation in place due to widespread use

R
Large number of open-source packages for A&R use
Highly interactive and customizable data visualization packages
Ability to customize features based on user needs, owing to being open source

Concurrent SAS and R usage: Case 1

CASE 1, PART A

- Input data review
- Data issues reporting

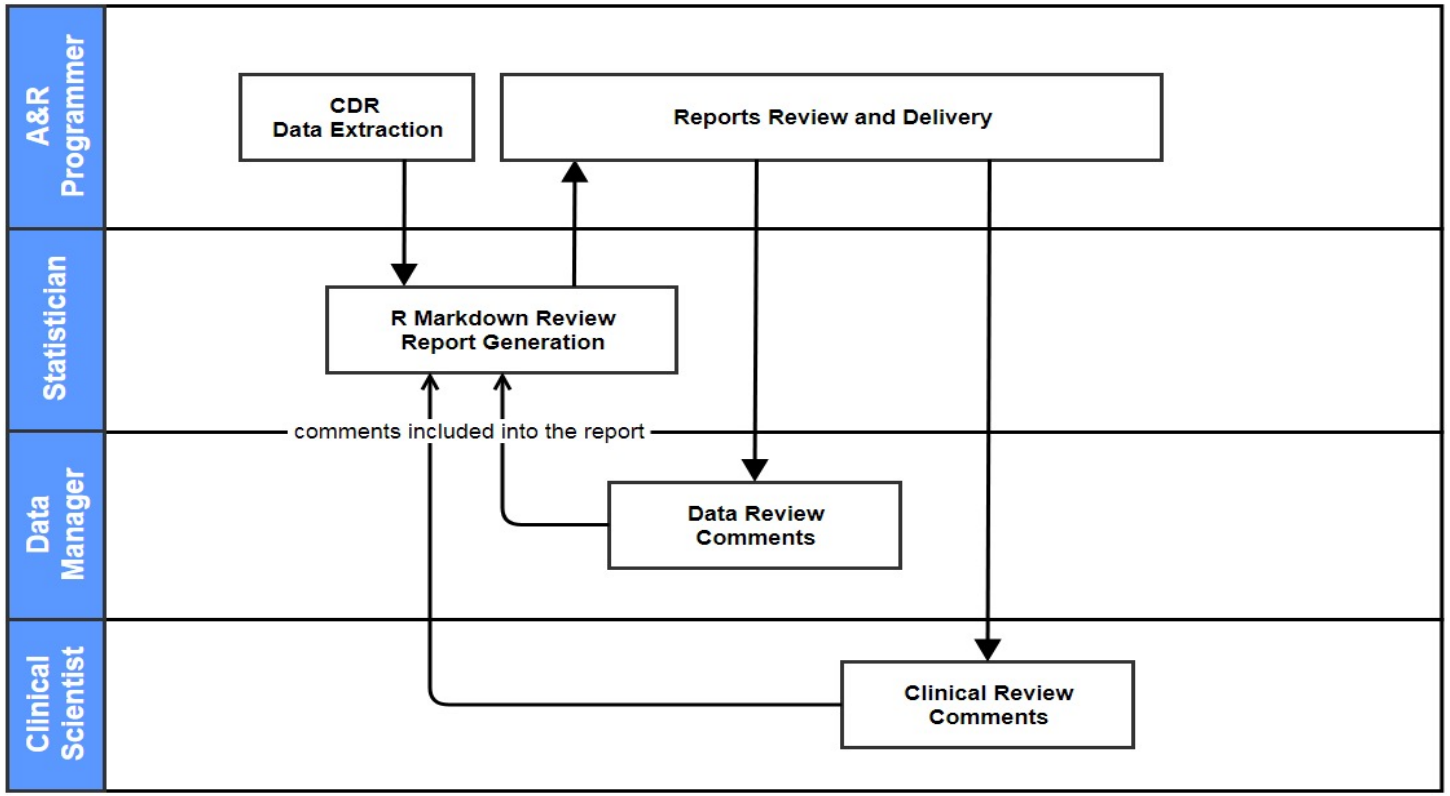
CASE 1, PART B

- A&R Output data review
- A&R Programming issues review

Concurrent SAS and R usage: Case 1 Part A

Clinical Trial Input Data Review Workflow

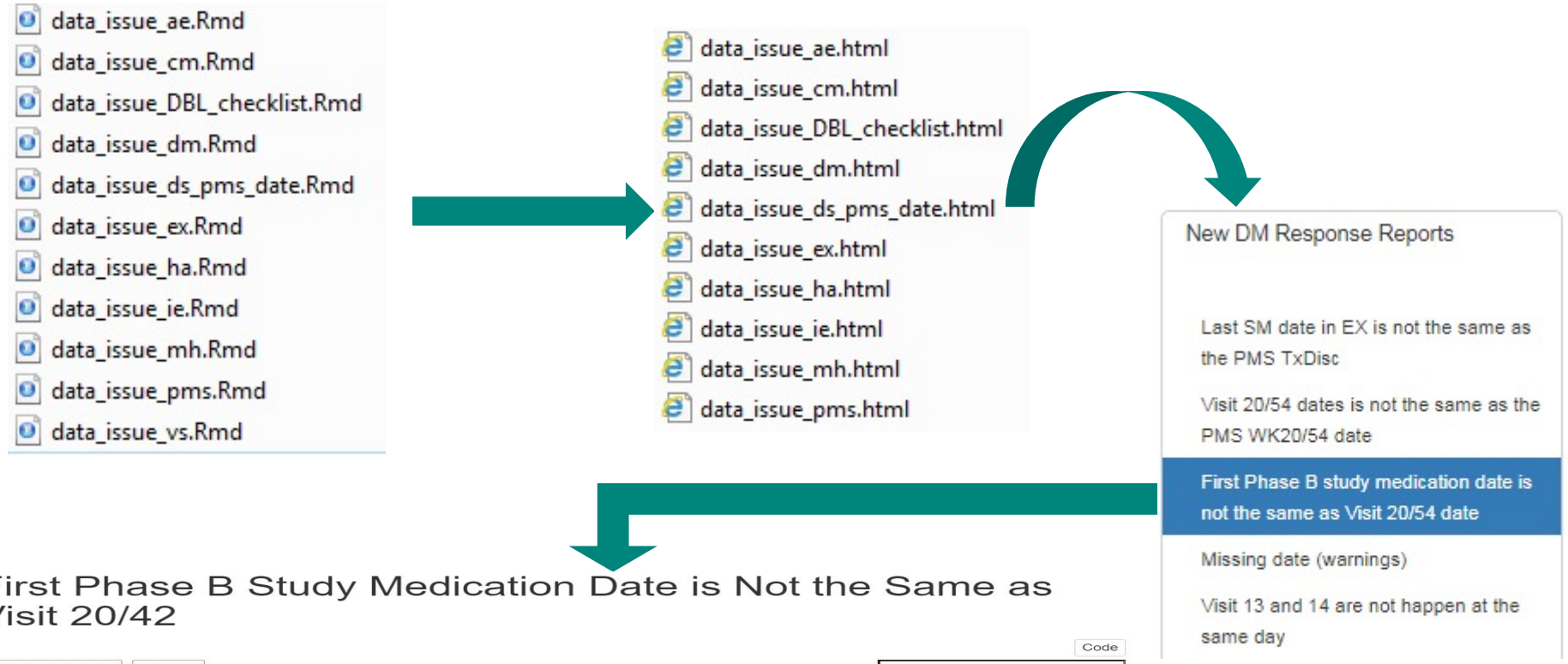
Clinical trial data review is an integral part of clinical data analysis facilitating identification of critical data issues



- Increase efficiency, transparency, and organization
- Regular data issue review meetings:
  - Single data review summary email
  - Issue tracking
  - Knowledge sharing
  - Agile responsiveness to changes
- A&R programming review

Concurrent SAS and R usage: Case 1 Part A

A&R Programming Data Issue Reports



First Phase B Study Medication Date is Not the Same as Visit 20/42

Code

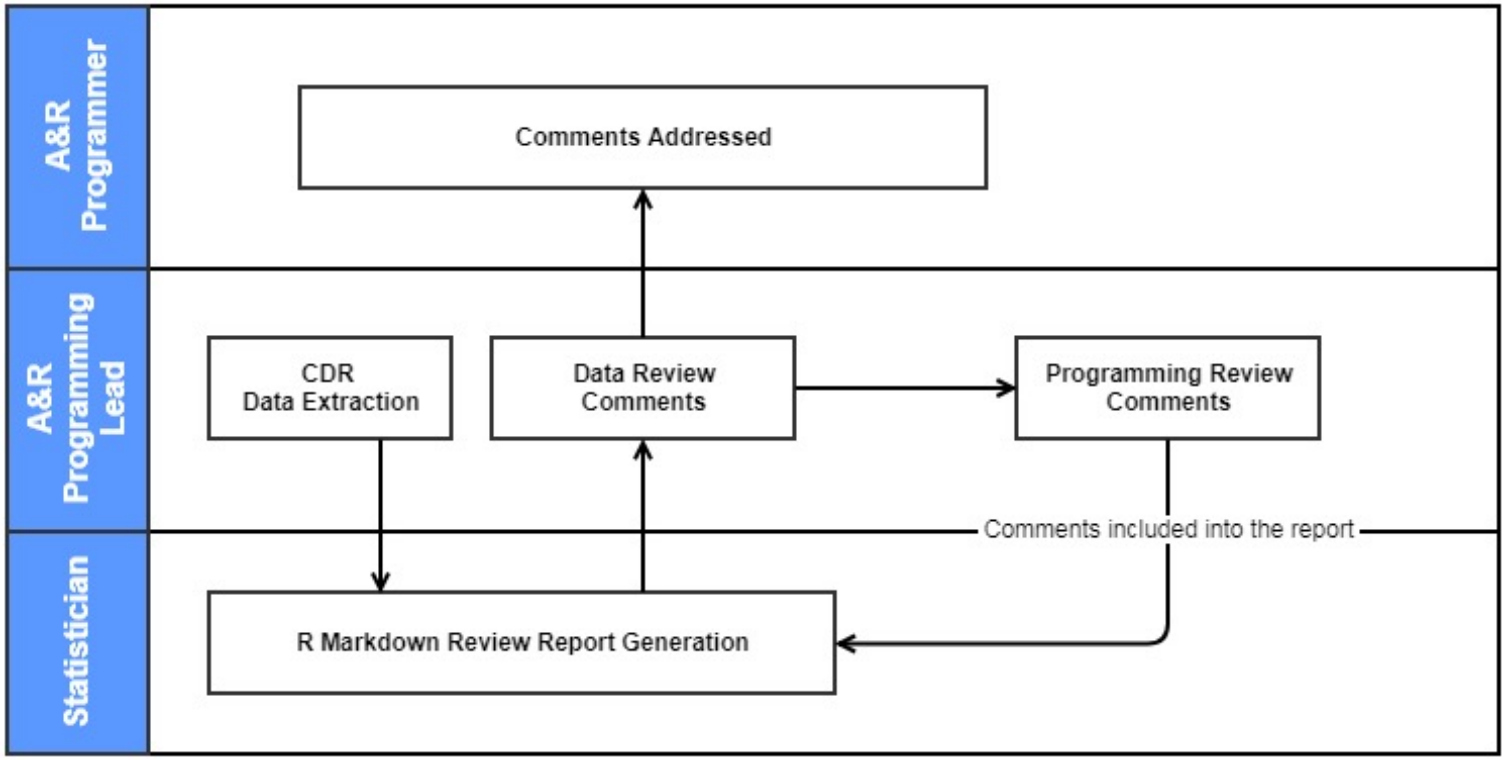
Column visibility | Excel | Search:

	SUBJID	Batch	Study_Status	CRB_Status	CMD_Comments
	All	All	All	All	All
1	053	2	COMP D	3 LOCKED	Comment Providd on xxx Date
2	023	E	COMP D	3 LOCKED	Querried and no futher comment
3	078	F	COMP D	Pending CRA Query	Further Investigation needed
4	012	J	In Phase B	IN PHASE B	

Concurrent SAS and R usage: Case 1 Part B

Clinical Trial A&R Output Data Review Workflow

Clinical trial A&R programming review is essential for maintaining quality and identifying critical programming issues with programming deliverables



The purpose is to:

- reduce the volume of emails
- keep organized solution records
- ensure consistent logic for all studies within a trial
- knowledge transparent (new people onboard)
- agile to adapt change request

Concurrent SAS and R usage: Case 1 Part B

A&R Programming Review Reports

Concurrent SAS and R usage: Case 2

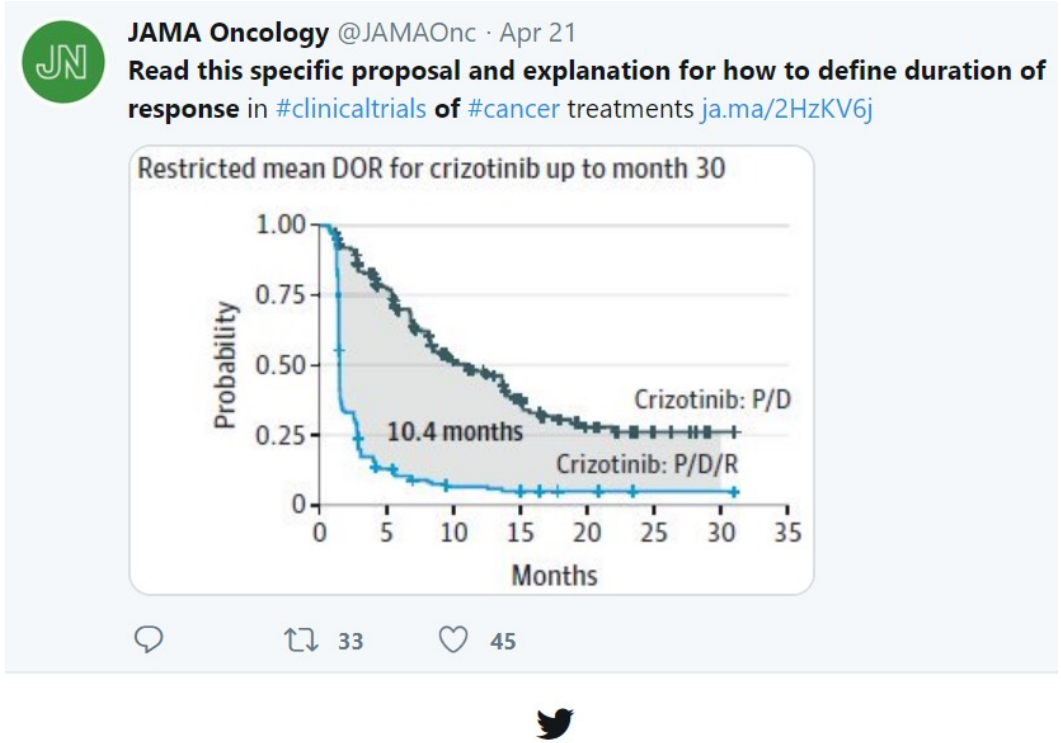
A&R using both SAS and R

- Majority of A&R deliverables are completed in SAS
  - SDTM/ADaM data is created using SAS
  - Most TLFs are generated using standard SAS macros
- Selected TLFs are generated using R
  - Complex statistical methodology and no existing procedure available in SAS
  - e.g. RMST Analysis
- If R is used for an analysis, all associated programs are done using R

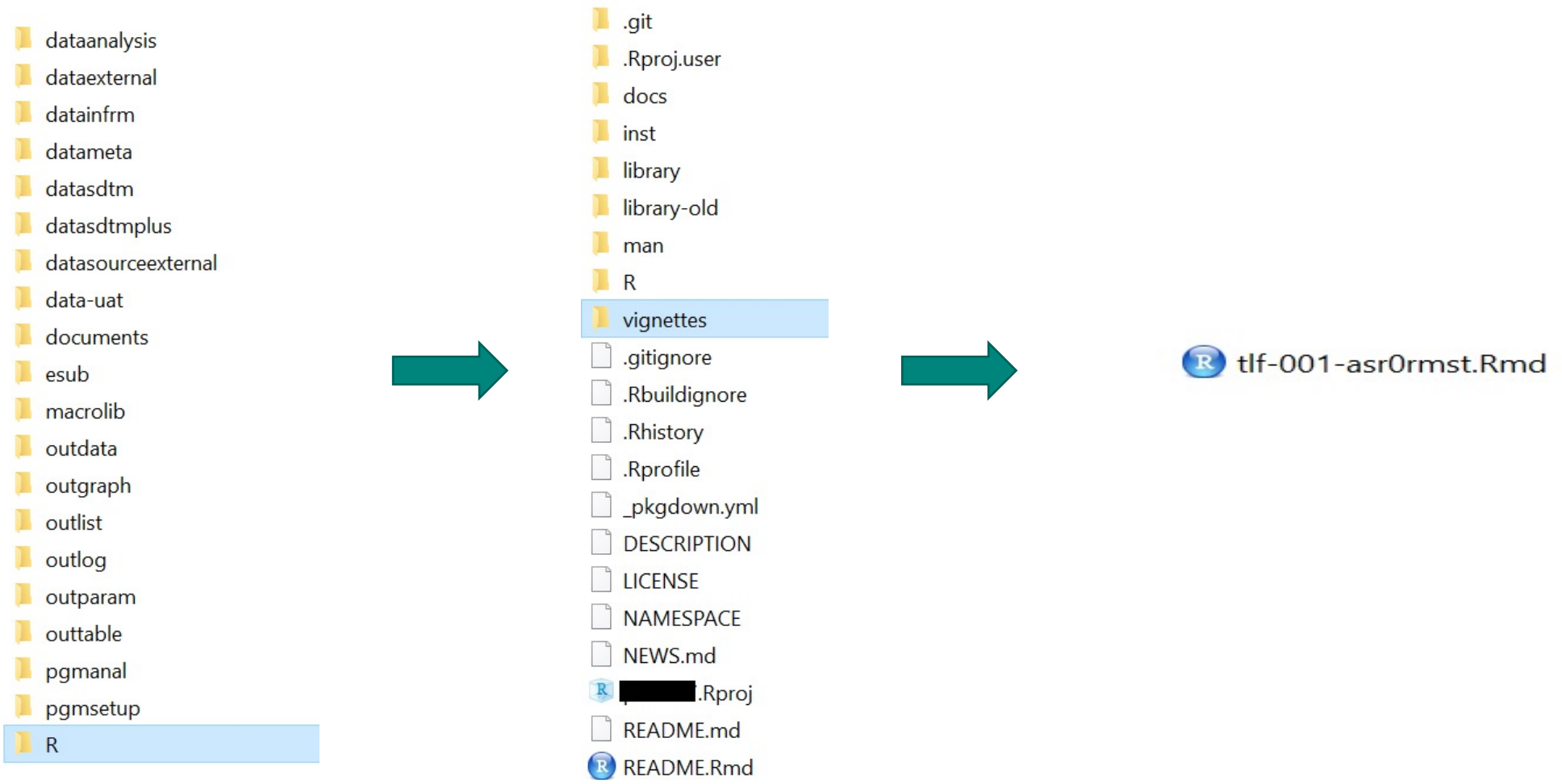
Concurrent SAS and R usage: Case 2

Restricted Mean Survival Time (RMST):

- Proportional hazard (PH) assumption is usually violated in complex analyses. When PH assumption is not met, Hazard ratio (HR) is difficult to interpret.
- RMST based statistics can be more powerful than HR under Non-PH.
- It's the often-used supportive method for non-proportional hazard (usually defined in SAP for Progression Free Survival or Overall Survival analysis) in regulatory filing or publication which is the new 'norm' in immuno-oncology studies
- RMST can use all the data until the last censored or observed time point, but HR can only use the last observed data point.
- It's mathematically easy to calculate and easy to interpret for non-stats audience with graphic output
- Early death (crossing KM curve) could be defined by RMST



Concurrent SAS and R usage: Case 2



R folder has R related analysis code

.Rprofile will direct input path location such as ADaM dataset (ex:dataanalysis\) for the R project folder

Concurrent SAS and R usage: Case 2

MKSURV Package:

This package provides standard tools for time-to-event data analysis including:

- Kaplan-Meier curve (including drug labeling)
- RMST

Input : ADaM datasets in SAS format

```
adsl   <- read_sas(file.path(path$adam, "adsl.sas7bdat"))
adtte <- read_sas(file.path(path$adam, "adtte.sas7bdat"))
```

Functions used for the examples:

`tidy_survdata()`	Standardize ADaM datasets for Time-to-event Analysis
`surv_stat()`	Generate Statistical Summary Table
`rmst_tbl()`	Table for Analysis of Restricted Mean Survival Times
`km_plot_agency()`	Generate kmplot from Standardized Survival Data

Concurrent SAS and R usage: Case 2

ADTTE BDS structure, listed key variables below

VarName	TypeLen	Label
STUDYID	Char12	Study Identifier
USUBJID	Char30	Unique Subject Identifier
SUBJID	Char10	Subject Identifier for the Study
ITTFL	Char1	Intent-to-treat Flag
TRTFL	Char1	Treated Population Flag
TRT01P	Char40	Planned Treatment for Period 01
TRT01PN	Num8	Planned Treatment for Period 01 (N)
TRT01A	Char40	Actual Treatment for Period 01
TRT01AN	Num8	Actual Treatment for Period 01 (N)
PARAMCD	Char8	Parameter Code
PARAMN	Num8	Parameter (N)
AVAL	Num8	Analysis Value
STARTDT	Num8	Time-to-Event Origin Date for Subject
ADT	Num8	Analysis Date
ADTF	Char1	Analysis Date Imputation Flag
CNSR	Num8	Censor
SRCDOM	Char8	Source Data
SRCVAR	Char8	Source Variable
SRCSEQ	Num8	Source Sequence Number
EVNTDESC	Char100	Event or Censoring Description
CNSDTDSC	Char100	Censor Date Description

Concurrent SAS and R usage: Case 2

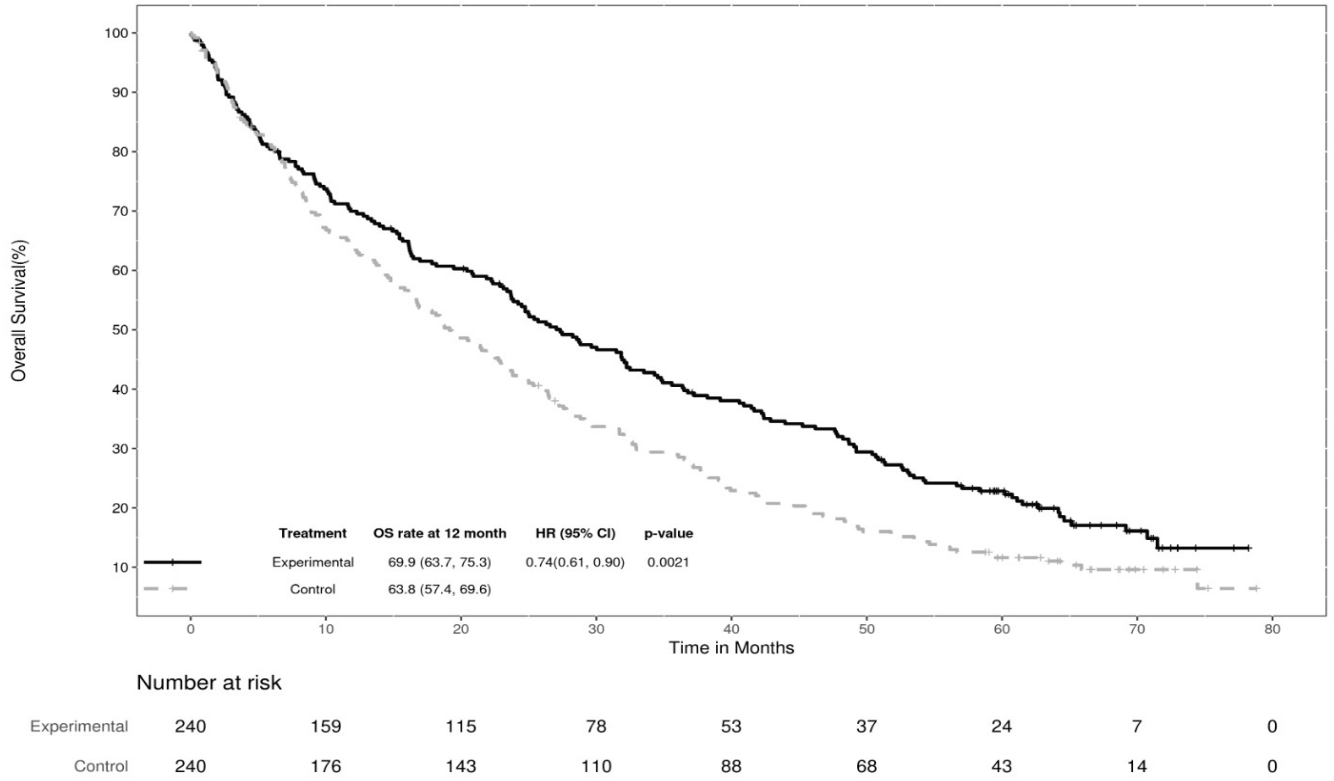
ADaM dataset

tidy_survdata

surv_stat

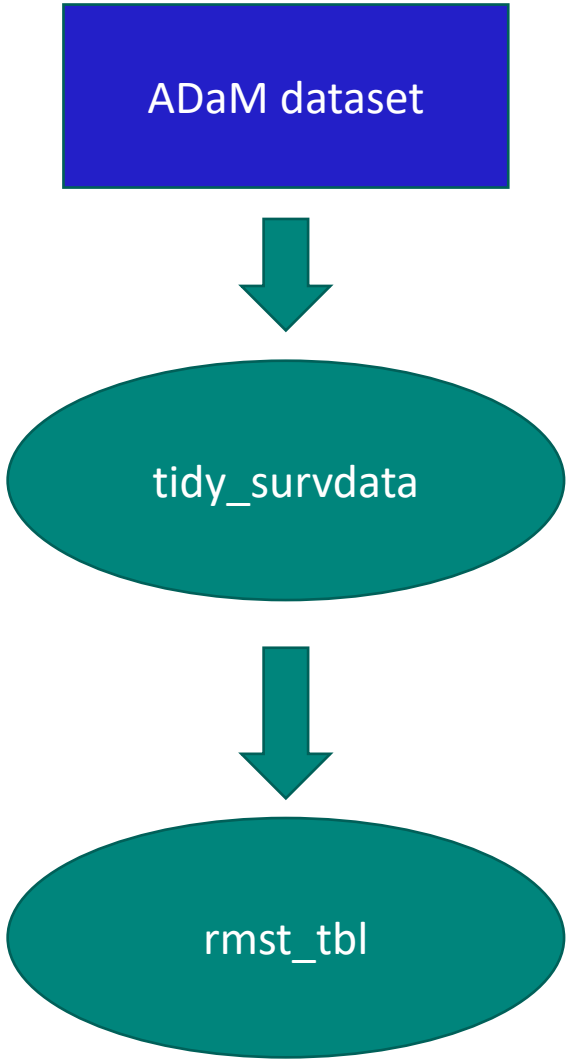
km_plot_agency

KMPLOT



Flexibility of legend position has been provided in the km_plot_agency() function for drug labeling specific requirements.

Concurrent SAS and R usage: Case 2



Analysis of Restricted Mean Survival Times (RMST) of First Event
(ITT Population)

RMST based on Follow-up duration (Weeks)	Xanomeline Low Dose (N=84)		Placebo (N=86)		Difference vs. Placebo (95% CI)
	n	RMST (95% CI)	n	RMST (95% CI)	
3	24	2.64 (2.47, 2.81)	8	2.82 (2.68, 2.96)	-0.18 (-0.40, 0.03)
6	44	4.28 (3.86, 4.69)	16	5.37 (5.04, 5.70)	-1.09 (-1.63, -0.56)
9	52	5.38 (4.70, 6.05)	19	7.71 (7.14, 8.28)	-2.34 (-3.23, -1.45)
12	56	6.20 (5.27, 7.14)	25	9.91 (9.09, 10.74)	-3.71 (-4.95, -2.47)
15	58	6.89 (5.71, 8.06)	27	11.92 (10.84, 13.01)	-5.04 (-6.64, -3.43)
18	62	7.39 (6.00, 8.77)	28	13.86 (12.49, 15.23)	-6.48 (-8.43, -4.53)
21	62	7.76 (6.20, 9.33)	28	15.79 (14.13, 17.46)	-8.03 (-10.32, -5.74)
24	62	8.14 (6.38, 9.91)	28	17.72 (15.75, 19.70)	-9.58 (-12.23, -6.93)
27	62	8.52 (6.54, 10.50)	29	19.63 (17.34, 21.91)	-11.11 (-14.13, -8.08)

N = number of participants. n = number of events.
RMST: Restricted mean survival time.

Source: [Study MK9999P001: adam-adsl adtte]

RMST Table

Concurrent SAS and R usage: Case 2

- Create table using `adsl` and `adtte` datasets.
- In an actual study A&R process, `df1` shall be saved in the `outdata` folder.

```
# Define surv_data
surv_data <- tidy_survdata(population_from   = mksurv_adsl,
                           population_where  = "ITTFL=='Y'",
                           observation_from  = mksurv_adtte,
                           observation_where = "PARAMCD =='OS'",
                           start_date_var    = "STARTDT",
                           time_unit         = "months",
                           treatment_order   = c("Xanomeline High Dose", "Placebo"))
```

```
df1 <- surv_data %>% rmst_tbl(trunc_time = c(2, 3, 4, 5, 6),
                              title      = c("Analysis of Restricted Mean Survival Times (RMST) of
                                             "(ITT Population)"),
                              footnote   = c("N = number of participants. n = number of events.",
                                             "{^a} RMST: Restricted mean survival time."),
                              datasource = "Source:  [Study MK9999P001: adam-adsl, adtte]")
df1 %>% knitr::kable()
```

Conclusion

Conclusion:

- SAS and R can be used concurrently in clinical trial A&R activities for variety of tasks
- Each software environment's individual strengths can be leveraged to increase efficiency of A&R SDLC

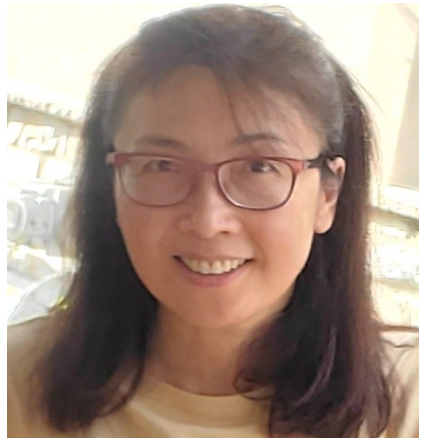
Future Work:

- Concurrent use of SAS & R in validating A&R deliverables is being evaluated

Awesome Team!

Uday Preetham Palukuru

Ruchitbhai Patel



Christine Teng

Yilong Zhang

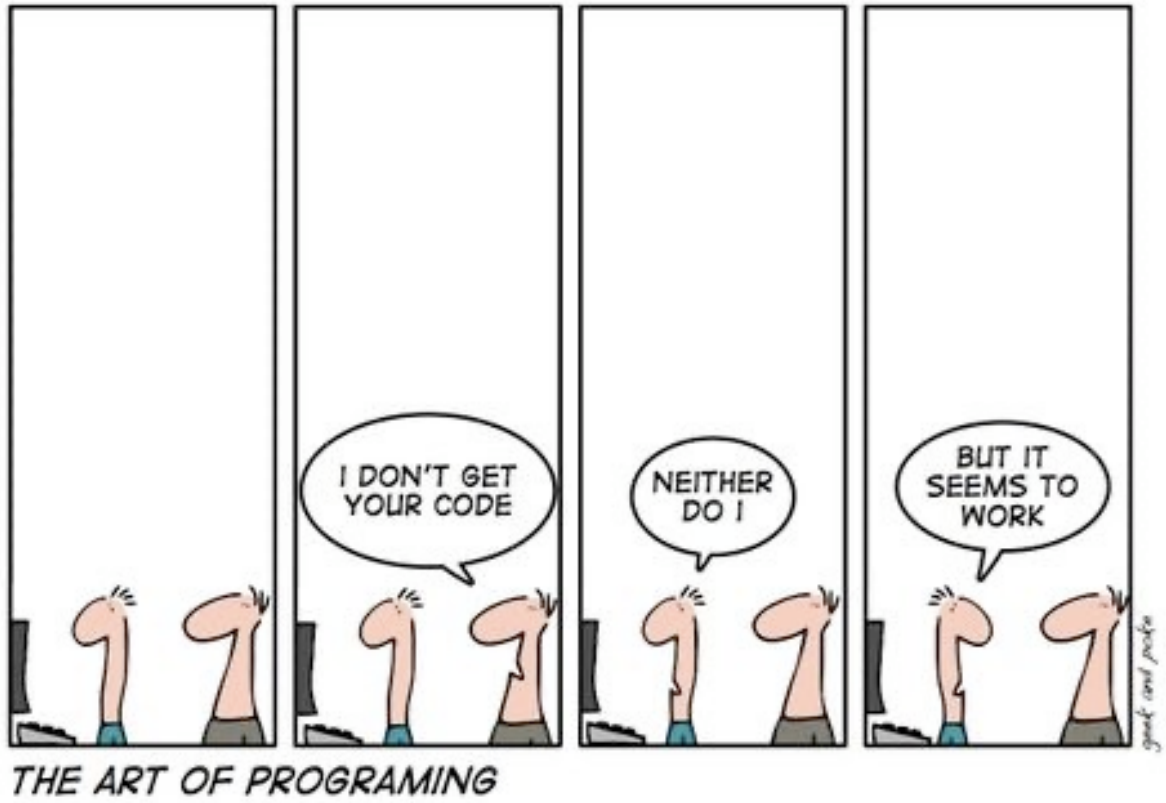
Thank you

- **Uday Preetham Palukuru**

Tel: +1732-594-6626
E-mail: preetham.palukuru@merck.com

- **Yilong Zhang**

Tel: +1732-594-6196
E-mail: yilong.zhang@merck.com



The Global Healthcare Data Science Community

Contact Channels
UK +44 1843 609600
USA +1 609 514 5105
office@phuse.global
phuse.global

Social Media
@phusetwitta
/phusebook
/phusetube
/company/phuse