

Open-Source

Advancements and Opportunities

Michael Stackhouse | Co-lead Data Visualization and Open-Source Technology



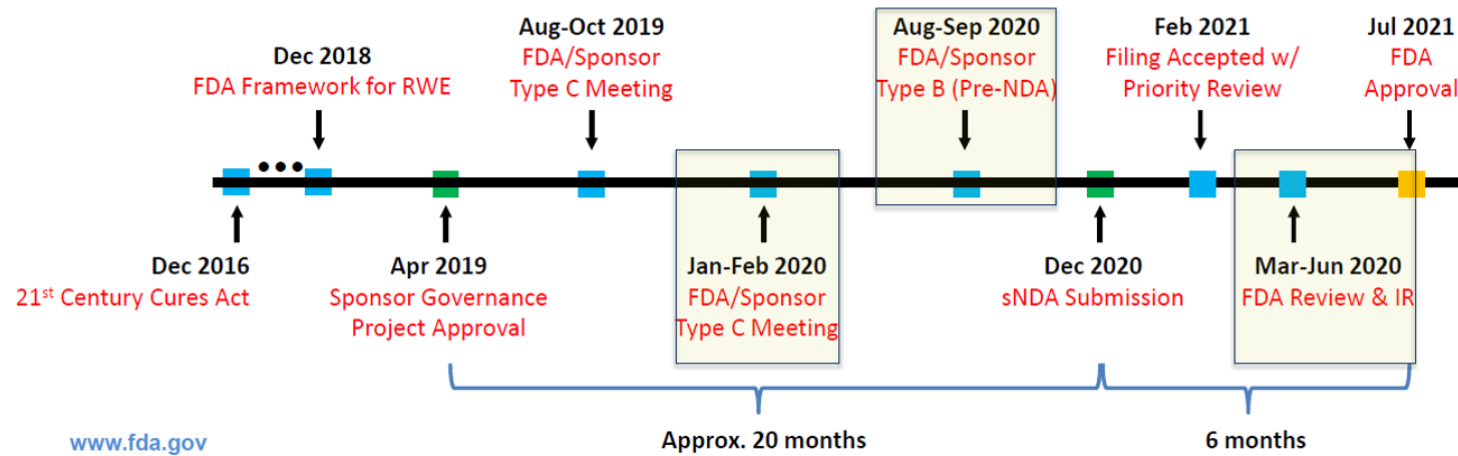
Real-World Data

New Indication for Tacrolimus (Prograf®)



CDER's **first acceptance** of an “**observational study**” as an **adequate and well-controlled study** providing the **primary support** for a finding of **substantial evidence of effectiveness**

Source: Concato, John, and Jacqueline Corrigan-Curay. "Real-World Evidence-Where Are We Now?." *The New England journal of medicine* 386.18 (2022): 1680-1682.



www.fda.gov

5

E2E Open-Source Collaboration



End-to-End Open-source Collaboration Guidance

🔍 ⬇️ 🔗

End-to-End Open-source Collaboration Guidance

Data Visualisation & Open Source Technology Working Group

AUTHOR
PHUSE Working Group

ABSTRACT
This whitepaper provides guidance on the use of Open-Source Software (OSS), as well as collaboration on and creation of open-source projects used by data scientists in clinical reporting workflows

Guidance scope and purpose

The primary aim of this collaboration is to provide guidance within the context of how open source is relevant to PHUSE members, and link out to more information to avoid duplication on more generalisable topics. In this guidance, R packages are referenced as an example OSS project that is a focal point today in clinical reporting, but the principles extend to other libraries in python, julia, javascript, and more. The following topics are covered in this white paper:

Using open source

- Relevance of different licence types
- Watchouts on governance models and assessing risk
- Landscape of tools available for vulnerability detection, validation, qualification, risk and enforcing licence policies, with particular reference to R-specific tools

Table of contents

- Guidance scope and purpose

CVARS

CVARS

Read/Upload Data

Data Source

Default (CDISC Pilot Data)

This option is only available for training and understanding app using CDISC data

CDISC ADaM Data

ADAE

Run Report

Download Report

☒ html

☐ pdf

☐ docx

Save Report

Report Selection

Report base domain

adae

Report Name

adae_r001

Therapeutic Area

Any

Report Number

2.1

Report Type

Table

Report Description

Summary of Adverse Events by System Organ Class and Preferred Term

Treatment & Population Selection

Treatment Variable

TRTA

Treatment Sort Variable

TRTA

Population Filter

Overall Population

Subsetting Conditions

Analysis Subset Condition

USUBJID!="

Denominator Subset Condition

STUDYID!="

Percentage Denominator

Treatment

AE Summary Table Inputs

Treatment Big N

☐ Y ☒ N

Total treatment

☐ Y ☒ N

Add Missing Grouping Variable

☐ Y ☒ N

Calculate Risk

☐ Y ☒ N

Generic Inputs

Adverse Event Filter(s)

Period

← → ↺

CAMIS

[CAMIS](#) [Home](#) [About](#) [Contributions](#) [Meeting Minutes](#) [Conferences](#)

CAMIS

Introduction

Several discrepancies have been discovered in statistical analysis results between different programming languages, even in fully qualified statistical computing environments. Subtle differences exist between the fundamental approaches implemented by each language, yielding differences in results which are each correct in their own right. The fact that these differences exist causes unease on the behalf of sponsor companies when submitting to a regulatory agency, as it is uncertain if the agency will view these differences as problematic. In its Statistical Software Clarifying Statement, the US Food and Drug Administration (FDA) states that it “FDA does not require use of any specific software for statistical analyses” and that “the computer software used for data management and statistical analysis should be reliable.” Observing differences across languages can reduce the analyst’s confidence in reliability and, by understanding the source of any discrepancies, one can reinstate confidence in reliability.

On this page

[Introduction](#)

[Motivation](#)

[<](#) [>](#) [↺](#)

pharmaverse



pharmaverse:
*Breaking boundaries through
open source collaboration*

R/Pharma 2022

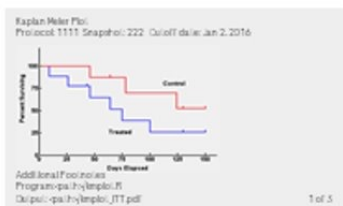


1

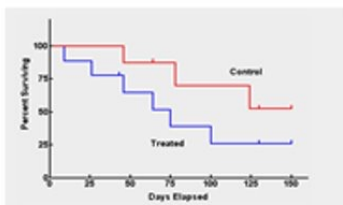


Improve efficiency in the way we work

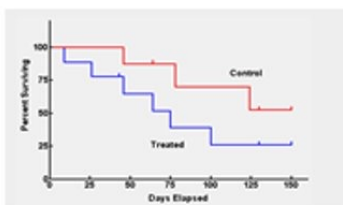
Analyzing clinical trial data requires multiple ways of presenting and interacting with our data



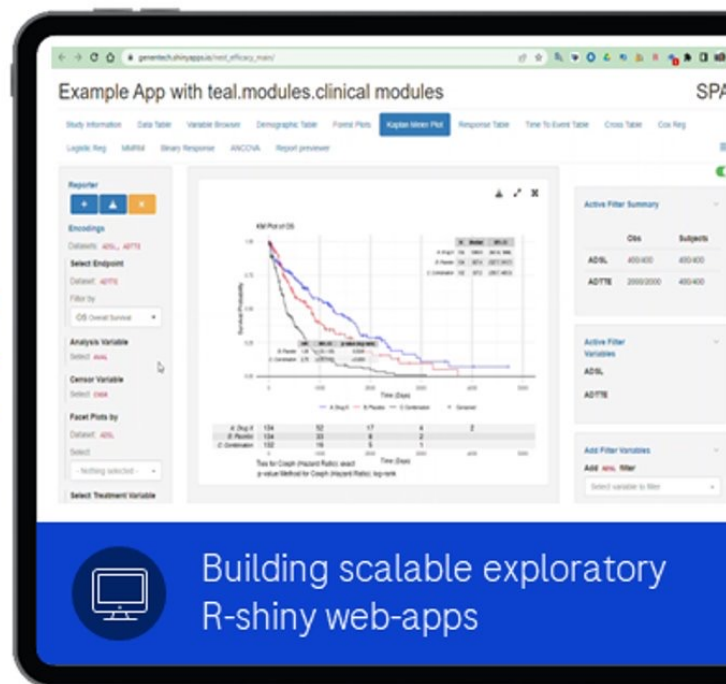
Per-SAP static output
on **OS**



Ad-hoc analysis
on **PFS**



Ad-hoc analysis on **PFS**
in **subpopulation**



CDISC - CORE

Why is CDISC doing CORE?

- Ensure each standard has a set of unambiguous, executable Conformance Rules
- Ensure consistency across Conformance Rule implementations
- Expedite the availability of executable Conformance Rules for new Foundational Standards
- Create executable Conformance Rules vetted by the CDISC standards development teams
- Develop an open-source engine that serves as a Reference Implementation
- Publish the Rules in the CDISC Library and the engine under the CDISC Open Source Alliance (COSA)

➡ *CORE Initiative = Rules + Engine*



<https://www.cdisc.org/core>

CDISC - ARS

Shifting the Paradigm

Table 3.1.1: ADHYPO Analysis Dataset

Row	STUDYID	USUBJID	MIDS	CEDECOD	WASAEYN	ASTDTM
1	XYZ	000001	HYP0 1	Hypoglycemia	Y	07Sep2012 22:29:00
2	XYZ	000001	HYP0 2	Hypoglycemia	N	10Sep2012 09:12:00
3	XYZ	000001	HYP0 3	Hypoglycemia	N	10Sep2012 23:05:00
4	XYZ	000001	HYP0 4	Hypoglycemia	N	11Sep2012 15:24:00
5	XYZ	000001	HYP0 5	Hypoglycemia	N	18Sep2012 11:39:00
6	XYZ	000002	HYP0 1	Hypoglycemia	N	22Oct2012 13:28:00
7	XYZ	000002	HYP0 2	Hypoglycemia	N	25Oct2012 13:59:00
8	XYZ	000002	HYP0 3	Hypoglycemia	N	17Nov2012 05:01:00

ADaM Dataset

dm Observation	dm Table	dm Population	dm Treatment	dm Parameter	dm Unit	dm Aggreg	dm Statistic	analysisResult
1001 dm summary	enrolled	Treatment A	param subjects	ssn ALL	aggrat ALL	stat freq	100	
1002 dm summary	enrolled	Treatment A	param subjects	ssn F	aggrat ALL	stat freq	50	
1003 dm summary	enrolled	Treatment A	param subjects	ssn M	aggrat ALL	stat percent	50	
1004 dm summary	enrolled	Treatment A	param subjects	ssn ALL	aggrat ALL	stat freq	40	
1005 dm summary	enrolled	Treatment A	param subjects	ssn M	aggrat ALL	stat percent	40	
1006 dm summary	enrolled	Treatment B	param subjects	ssn ALL	aggrat ALL	stat freq	50	
1007 dm summary	enrolled	Treatment B	param subjects	ssn F	aggrat ALL	stat freq	30	
1008 dm summary	enrolled	Treatment B	param subjects	ssn M	aggrat ALL	stat percent	30	
1009 dm summary	enrolled	Treatment B	param subjects	ssn ALL	aggrat ALL	stat freq	20	
1010 dm summary	enrolled	Treatment B	param subjects	ssn F	aggrat ALL	stat percent	20	
1011 dm summary	enrolled	Treatment ALL	param subjects	ssn ALL	aggrat ALL	stat freq	180	
1012 dm summary	enrolled	Treatment ALL	param subjects	ssn F	aggrat ALL	stat percent	90	
1013 dm summary	enrolled	Treatment ALL	param subjects	ssn M	aggrat ALL	stat percent	80	
1014 dm summary	enrolled	Treatment ALL	param subjects	ssn ALL	aggrat ALL	stat freq	100	
1015 dm summary	enrolled	Treatment ALL	param subjects	ssn M	aggrat ALL	stat percent	40	
1016 dm summary	enrolled	Treatment A	param age	ssn ALL	aggrat ALL	stat mean	40.7	
1017 dm summary	enrolled	Treatment A	param age	ssn ALL	aggrat ALL	stat mean	10.7	
1018 dm summary	enrolled	Treatment A	param age	ssn ALL	aggrat ALL	stat mean	37.0	
1019 dm summary	enrolled	Treatment A	param age	ssn ALL	aggrat ALL	stat min	21.0	
1020 dm summary	enrolled	Treatment A	param age	ssn ALL	aggrat ALL	stat max	95.0	
1021 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat freq	50	
1022 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat mean	41.2	
1023 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat min	18.0	
1024 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat max	96.0	
1025 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat min	23.0	
1026 dm summary	enrolled	Treatment B	param age	ssn ALL	aggrat ALL	stat max	97.0	
1027 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat freq	180	
1028 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat mean	40.8	
1029 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat min	10.4	
1030 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat max	97.0	
1031 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat min	21.0	
1032 dm summary	enrolled	Treatment ALL	param age	ssn ALL	aggrat ALL	stat max	97.0	

Analysis Results Dataset

Table 4.2.2: HbA1c Longitudinal Repeated Measures Analysis Results Metadata

Metadata Field	Metadata
DISPLAY IDENTIFIER	Table 4.2.1/Figure 4.2.1
DISPLAY NAME	Mean Change from Baseline in HbA1c (Percent) Longitudinal Repeated Measures Analysis, Intention-to-treat Population
RESULT IDENTIFIER	Treatment difference results (LSMean, confidence interval, p-value)
PARAM	HbA1c (%)
PARAMCD	HBA1C
ANALYSIS VARIABLE	CHG (Change from baseline)
ANALYSIS REASON	SPECIFIED IN SAP
ANALYSIS PURPOSE	PRIMARY OUTCOME MEASURE
ANALYSIS DATASET	ADHBA1C

ARM v1

ARM Extension Technical Specification

Automation



Reuse
Traceability

Table 4.3.1: HbA1c Longitudinal Repeated Measures Analysis - Table Shell

Page 1 of 2

		HbA1c (%) Longitudinal Repeated Measures Analysis	
		24-Week Short-term Double-Blind Treatment Period	
		Intention-to-treat Population	
		Drug A	Drug B
BASELINE	Mean (SD)	5.15 X.XX (X.XXXX)	5.15 X.XX (X.XXXX)
WEEK 4	SR	5.00	5.00
	Change from baseline: Mean (SD)	X.XX (X.XXXX)	X.XX (X.XXXX)
	Adjusted change from baseline: Mean (SD)	X.XX (X.XXXX)	X.XX (X.XXXX)
	95% Confidence Interval for adjusted mean	(XX.XX, XX.X)	(XX.XX, XX.X)
	Difference vs. Drug B (SD)	XX.XX (X.XXXX)	XX.XX (X.XXXX)
	95% Confidence Interval for difference	(XX.XX, XX.X)	(XX.XX, XX.X)
	p-value vs. Drug B		X.XXXX
...			
WEEK 12	SR	X.XX (X.XXXX)	X.XX (X.XXXX)
	Change from baseline: Mean (SD)	X.XX (X.XXXX)	X.XX (X.XXXX)
	Adjusted change from baseline: Mean (SD)	X.XX (X.XXXX)	X.XX (X.XXXX)
	95% Confidence Interval for adjusted mean	X.XX (X.XXXX)	X.XX (X.XXXX)
	Difference vs. Drug B (SD)	(XX.XX, XX.X)	(XX.XX, XX.X)
	95% Confidence Interval for difference	(XX.XX, XX.X)	(XX.XX, XX.X)
	p-value vs. Drug B		X.XXXX

SR: the number of subjects in the Intention-to-treat (ITT) population.
SD: the number of subjects in the ITT population with non-missing baseline and non-missing Week 4 value.
Reported measure (SR): change / baseline treatment visit visit/treatment
Program Source: `summarize(over(arm) = HbA1c) by week`
CDISC v15.2020

Display



atorus
Your Complete Data Solution