

R for Submission


and CDISC Open-Source Alliance

Michael Stackhouse | Chief Innovation Officer





 GPT-3.5

 GPT-4 

ChatGPT

Brainstorm edge cases

for a function with birthdate as input, horoscope as ou...

Write a thank-you note

to a guest speaker for my class

Help me pick

a gift for my dad who loves fishing

Plan an itinerary


for a literary tour of England, visiting famous authors' ...

Send a message





 GPT-3.5

 GPT-4 

ChatGPT

Give me ideas

for what to do with my kids' art

Recommend a dish

to impress a date who's a picky eater

Suggest fun activities

to do indoors with my high-energy dog

Create a workout plan

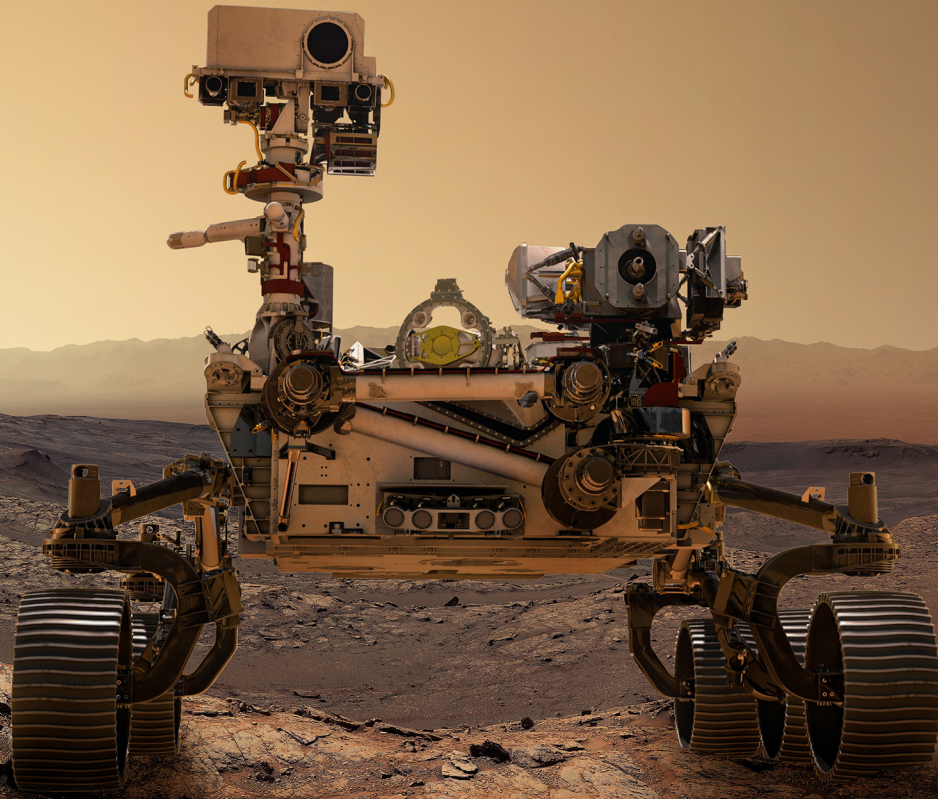
for resistance training

Our Work Is Important

High Risk, High Reward

Real Impact

Hardened Technology





Why Bother Changing?

Current processes
are optimized

What's the ROI of
new technology?



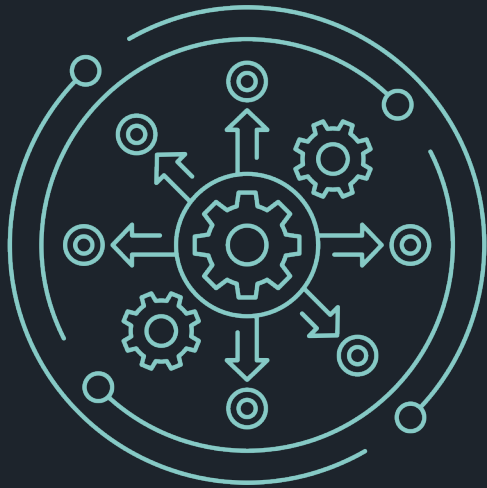
Why Bother Changing?

Current processes
are optimized

What's the ROI of
new technology?

If we can't make what
we have any better,
***make something
different***

Progress and Collaboration



Automation



Traceability



Interactivity

pharmaverse



pharmaverse:
*Breaking boundaries through
open source collaboration*

R/Pharma 2022



1

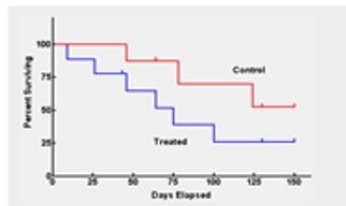
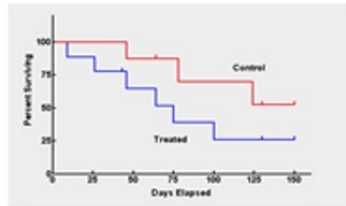
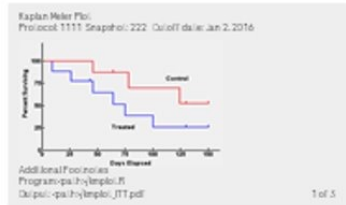


teal



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**

Example App with teal.modules.clinical.modules SPA

Study Information Data Table Variable Browser Demographic Table Forest Plot Kaplan Meier Plot Response Table Time To Event Table Cross Table Cox Reg

Logistic Reg MLSE Binary Response ANCOVA Support precision

Reporter

Encodings

Select Endpoint

Default: **OS**

Filter by

OS Overall Survival

Analysis Variable

Select: **ADSL**

Censor Variable

Select: **DATA**

Facet Plots by

Default: **ADSL**

Default: **ADSL**

Default: **ADSL**

Default: **ADSL**

Nothing selected

Select Treatment variable

Active Filter Summary

Class	Subjects
ADSL	400-400
ADTIE	2000-2000

Active Filter Variables

ADSL

ADTIE

Add Filter Variables

Add **ADSL** filter

Select variable to show

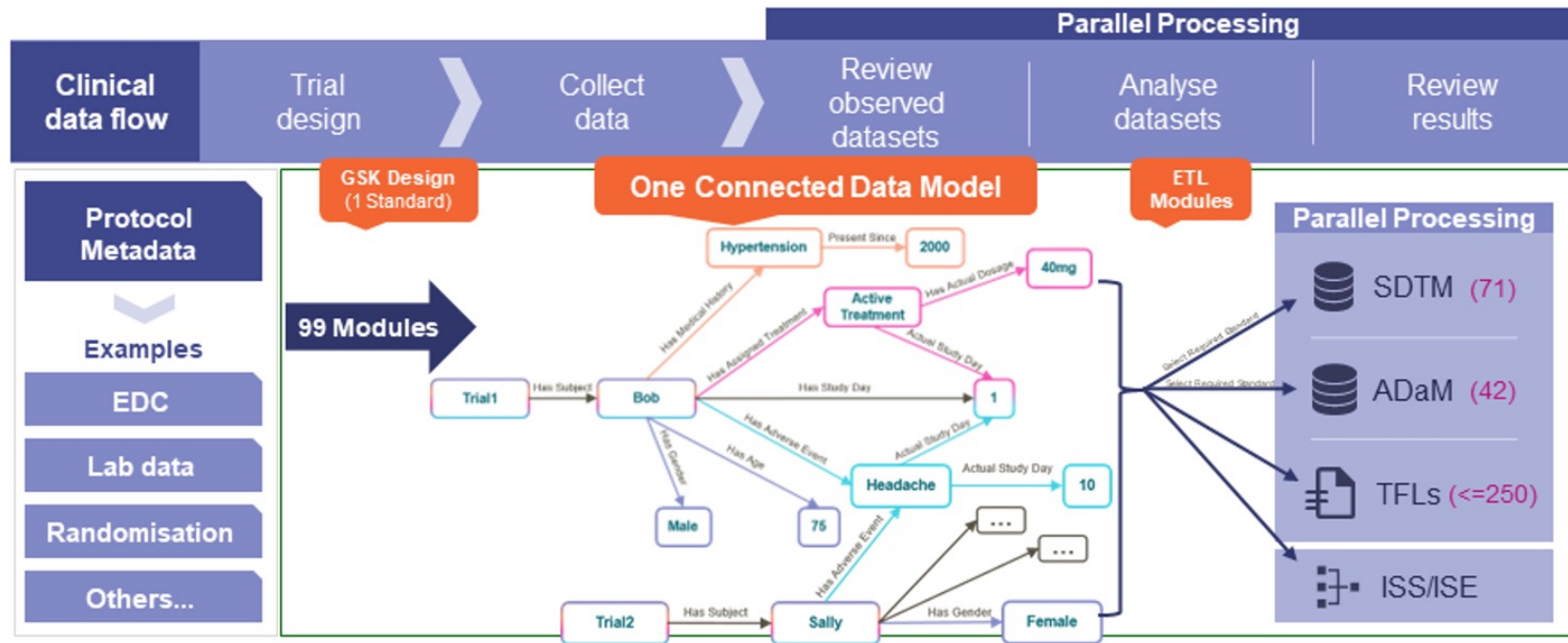
Building scalable exploratory R-shiny web-apps



GSK Clinical Reporting Knowledge Graphs

Our Idea

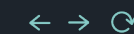
...the Google Translate for our clinical data – helping us translate our complex data landscape to answer important scientific questions



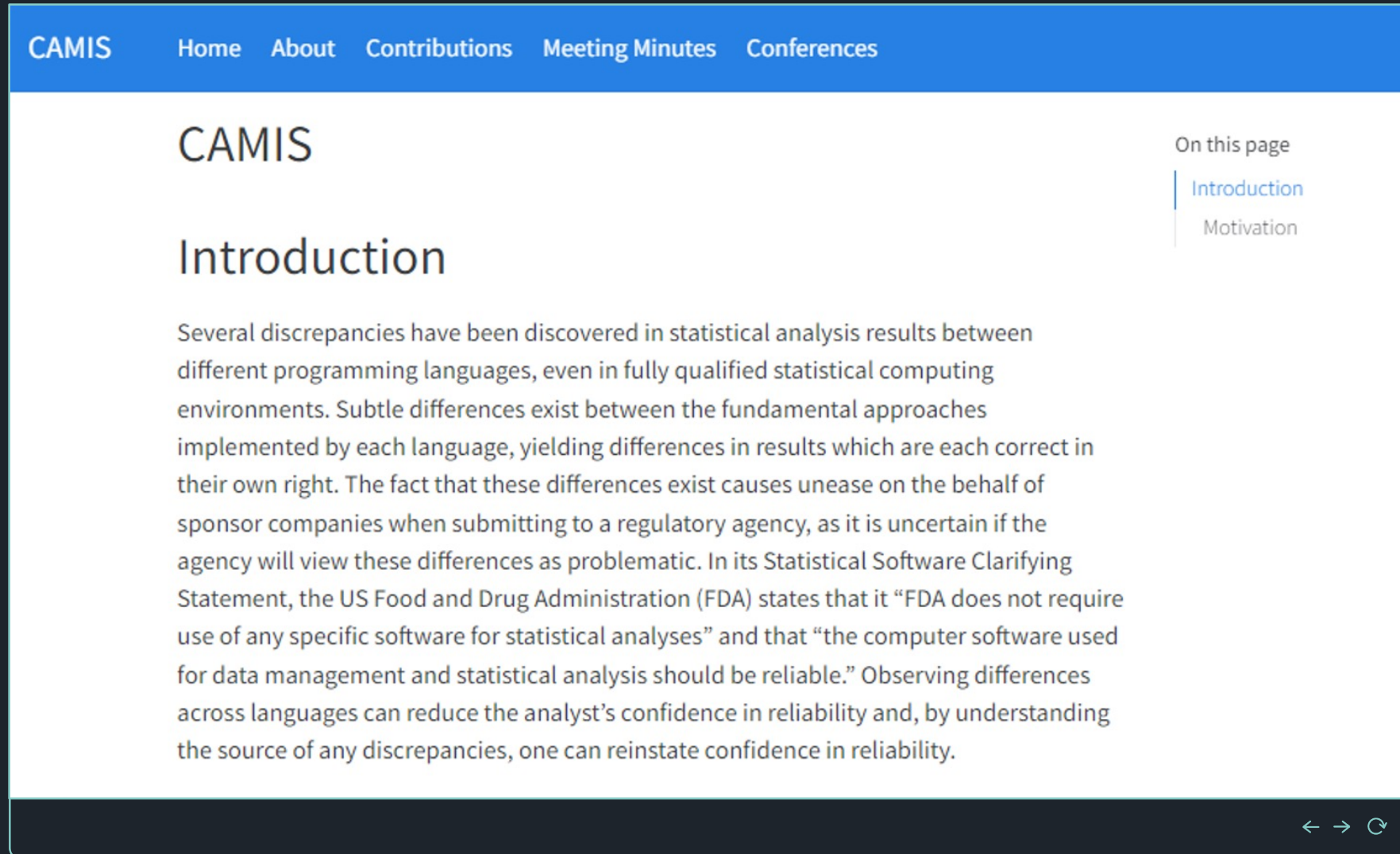
GSK

08 May 2023

5



CAMIS



The screenshot shows a web browser window with a blue header bar. The header contains the text "CAMIS" followed by navigation links: "Home", "About", "Contributions", "Meeting Minutes", and "Conferences". Below the header, the page title "CAMIS" is displayed in a large font. To the right of the title is a "On this page" section with two links: "Introduction" (which is highlighted with a blue vertical bar) and "Motivation". Below the title is a large heading "Introduction". The main content area contains a paragraph of text discussing discrepancies in statistical analysis results between different programming languages. At the bottom right of the page, there are three small navigation icons: a left arrow, a right arrow, and a refresh/circular arrow icon.

CAMIS Home About Contributions Meeting Minutes Conferences

CAMIS

On this page

- Introduction
- Motivation

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.

← → ↻

CDISC - CORE

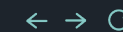
Why is CDISC doing CORE?

- 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

id	Observation	id	Time	id	Population	id	Treatment	id	Parameter	id	Unit	id	Aggr	id	Statistic	id	Analysis	id	Method	
1001	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1002	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1003	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1004	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1005	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1006	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1007	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1008	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1009	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1010	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1011	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1012	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1013	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1014	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1015	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1016	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1017	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1018	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1019	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1020	01	summary	analysis	Treatment A	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1021	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1022	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1023	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1024	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1025	01	summary	analysis	Treatment B	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1026	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1027	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1028	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1029	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1030	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1031	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1032	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1033	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1034	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1035	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1036	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1037	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1038	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1039	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all
1040	01	summary	analysis	Treatment ALL	para	adhy	all	all	all	all	all	all	all	all	all	all	all	all	all	all

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
RESULT IDENTIFIER	Period, Intention-to-treat Population
PARAM	Treatment difference results (LSMean, confidence interval, p-value)
PARAMCD	HbA1c (%)
ANALYSIS VARIABLE	HBA1C
ANALYSIS REASON	CHG (Change from baseline)
ANALYSIS PURPOSE	SPECIFIED IN SAP
ANALYSIS DATASET	PRIMARY OUTCOME MEASURE 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 (Intention-to-treat) Population	
		Drug A	Drug B
		N=115	N=115
BASISLINE	SE	X.XX (X.XXXX)	X.XX (X.XXXX)

WEEK 4	SE	XXX	XXX
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)	
95% Confidence Interval for difference		(XX.XX, XX.X)	
p-value vs. Drug B			X.XXXX

WEEK 12	SE	X.XX (X.XXXX)	X.XX (X.XXXX)
Change from baseline: Mean (SD)		XXX	XXX
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)	
95% Confidence Interval for difference		(XX.XX, XX.X)	
p-value vs. Drug B			X.XXXX

Display



Digital Data Flow

Future Value Streams for Digital Data Flow

Team will begin to address all three at varying degrees (in priority order)



Complete Protocol Digitization & Regulatory Alignment

Includes collaboration through the Vulcan Working Group between ICH M11 & CDISC

- Complete (100%) digitization of all protocol elements in alignment with M11 and relevant CDISC SDTM domains
- Begins with gap analysis between USDM and ICH M11 content model, CDISC SDTM, and Global Trial Registry Reporting
- Goal to capture "breadth" of ICH M11 completely within USDM, followed by greater "depth" of structured content within model (e.g. structured I/E criteria)



Expand Downstream Connectivity

Includes collaboration with expanding community of tech solution providers across range of clinical solutions

- Further develop USDM to enable downstream connectivity with priority systems, enabling a future state of "write once, read many times"
- Work collaboratively with the vendor ecosystem to better understand existing gaps and development requirements for the USDM



Alignment with Point of Care

Includes collaboration with Vulcan FHIR Accelerator

- Alignment of DDF and FHIR resources for end-to-end enablement of EHR workflow set-up and eSource
- Comparative assessment of USDM and FHIR currently underway

©2023 TRANSCCELERATE BIOPHARMA INC., ALL RIGHTS RESERVED.



What Changes May the Future Bring?

Our purpose is there, the skills will be different

What are you doing to grow?