Analysis Concepts Definition: Initial Perspectives from the CDISC Working Group

Bhavin Busa, Clymb Clinical
Brian Harris, AstraZeneca

Meet the Speakers

Bhavin Busa

Title: Principal and Co-founder

Organization: Clymb Clinical

Bhavin Busa is the Principal & Co-founder of Clymb Clinical. He is recognized as a thought leader in the areas of data standards, programming, and regulatory submission. He serves on the CDISC Open-Source Alliance (COSA) Board. Additionally, Bhavin serves as a CDISC ARS Product Owner/Co-lead and an active contributor to the PHUSE Community.

Brian Harris

Title: Standards Developer Senior Director

Organization: AstraZeneca

Over 25 years of industry experience working as a biostatistician and, most recently, as a standards developer for data collection, cleaning, analysis, and reporting. Over 12 years of volunteer experience on the CDISC ADaM team supporting ADaM conformance, questionnaire supplements, and recent updates to the implementation guide. Served as ADaM team lead in 2022 & 2023.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *The information presented in this session is for educational and informational purposes only. The presenter and affiliated organizations are not liable for any actions taken based on the information provided in this presentation.*

Agenda

1. CDISC Standards - Current State
2. Protocol to CSR: Standards-Driven Conceptual Mapping
3. Analysis Concepts: Building Blocks for the Digital SAP
4. Approaches Considered
5. Expanding into Analysis / Derivation Concepts (AC/DC)
6. Next Steps

CDISC Standards (Oct 2025)

Study Setup & Design

- USDM → Study structure
- SoA → Timing of data collection
- BC → Meaning of data

Data Representation

- CDASH & CRF Portal → collection instruments
- SDTM → Presentation of collected data
- ADaM → Analysis-ready datasets

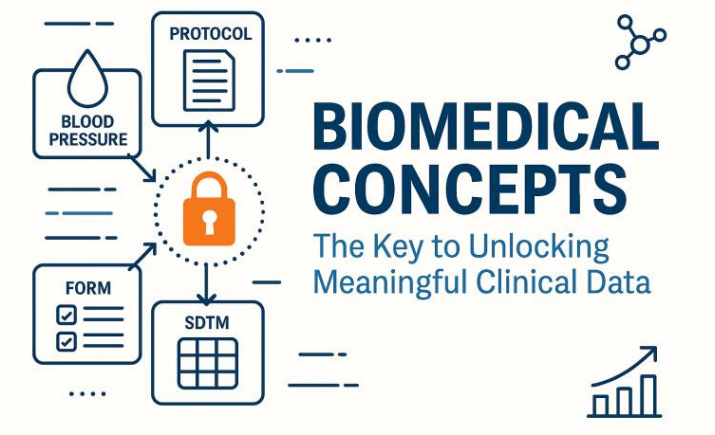
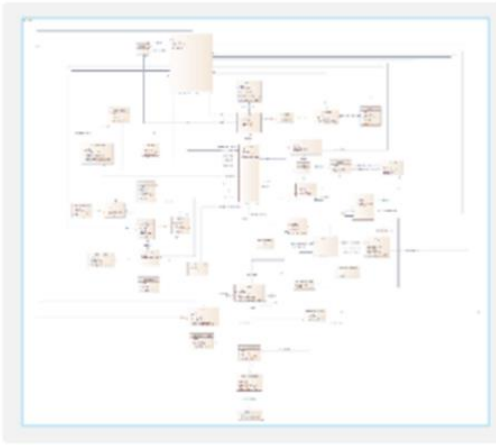


Image Reference: BioMedical Concepts – The Key to Unlocking Meaningful Clinical Data

Analysis & Results

- ARS / ARD → Standardized representation of results
- eTFL Portal → ready-to-use, ARS-compliant reporting templates

USDM Content



Controlled Terms

Study, Identifiers, Amendments

Estimands

Populations

Inclusion & Exclusion

Interventions & Indications

Objectives & Endpoints

Unstructured Content

Study Designs, Arms, Epochs

Detailed Study Logic, Encounters

Procedures, Biomedical Concepts

cdisc

CDISC Analysis Result Standards – Released April 19, 2024!

cdisc
Clear data. Clear impact

Analysis Results Standard (ARS) v1.0

cdisc ARS ANALYSIS RESULTS STANDARD

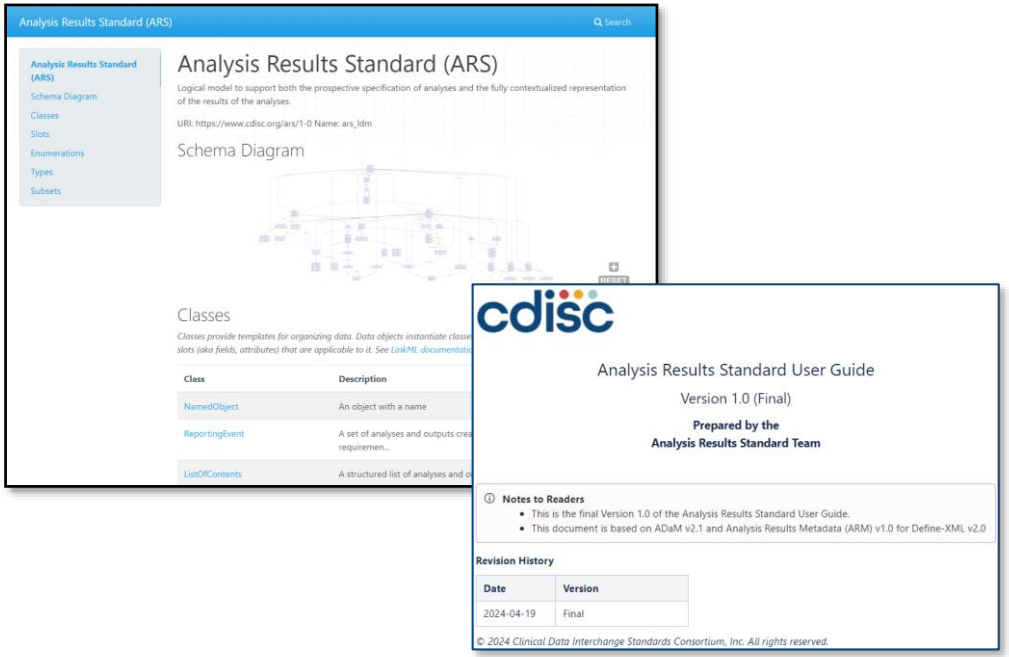
Large trials generate many analysis results in the form of tables, figures, and written reports, yet these results are rarely output in a form that is machine-readable. Previously, there has been no standard way of describing and organizing these results, making it difficult to automate their generation, make them reproducible, trace their origin, or enable them to be reused in other outputs.

To address these inefficiencies, CDISC has developed the Analysis Results Standard (ARS), which aim to facilitate automation, reproducibility, reusability, and traceability of analysis results data.

Features of ARS v1.0

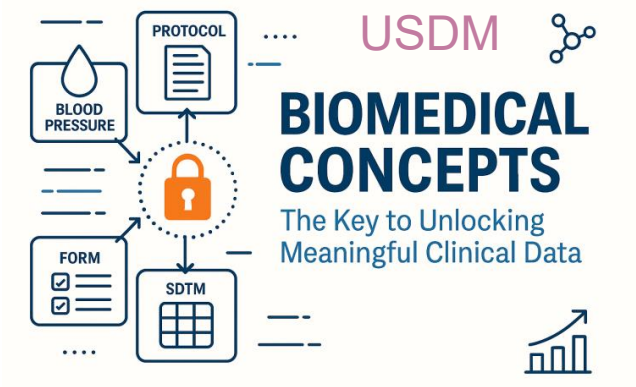
- A Logical Data Model that describes analysis results and associated metadata.
- A User Guide to illustrate and exercise the model with common safety displays.

https://cdisc-org.github.io/analysis-results-standard/



https://wiki.cdisc.org/display/ARSP/Analysis+Results+Standard+User+Guide+v1.0

Protocol to CSR: Standards-Driven Conceptual Mapping



What needs to be derived to set for analysis? [Derivation Concept]

What analysis is needed to answer the question? [Analysis Concept]

Protocol → CDASH (Collection) → SDTM (Tabulation) → ADaM (Analysis Data) → ARD (Analysis Results) → TFL (Analysis Outputs) → CSR

User Stories

Use Case 1: Accelerating Study Start-Up

As a study statistician, **I want to** select from a library of Analysis Concepts **so that I can** quickly draft the digital Statistical Analysis Plan (SAP) and identify needed datasets and variables.

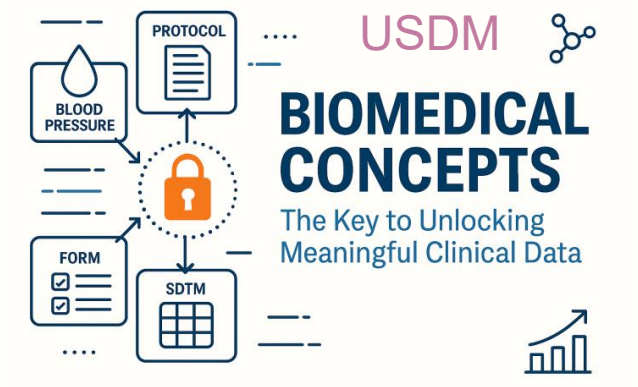
Use Case 2: Automating Standard Analyses Across Studies

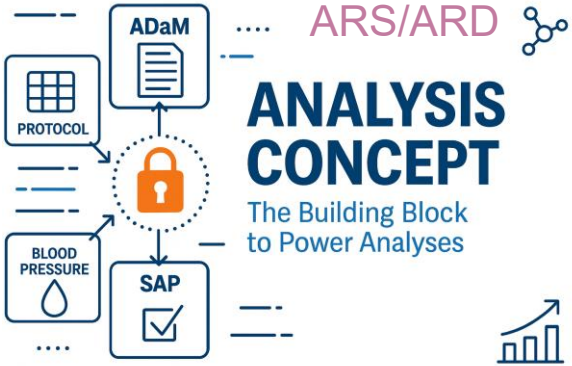
As a statistical programmer, **I want to** use pre-defined Analysis Concepts **so that I can** automate the generation of standard TFLs across studies without recreating logic each time.

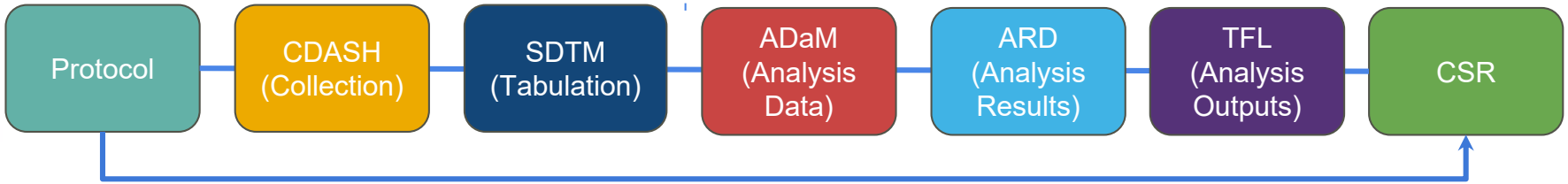
Use Case 3: Enabling Metadata-Driven Automation

As a clinical data standards expert, **I want to** build a structured repository of Analysis Concepts **so that** downstream automation tools (e.g., TFL code generators, validation scripts) can use them as inputs.

Protocol to CSR: Standards-Driven Conceptual Mapping







Analysis Concepts: Building Blocks for the Digital SAP

What AC Brings

- Standardized, configurable elements & methods
- Encodes clinical questions and expected outputs
- Enables automation and consistency across studies

Why This Matters for SAP

- SAPs today are static documents (Word/PDF) → hard to reuse or automate
- AC provides a digital representation of analysis intent
- Bridges the gap between *what is collected* and *what is analyzed*

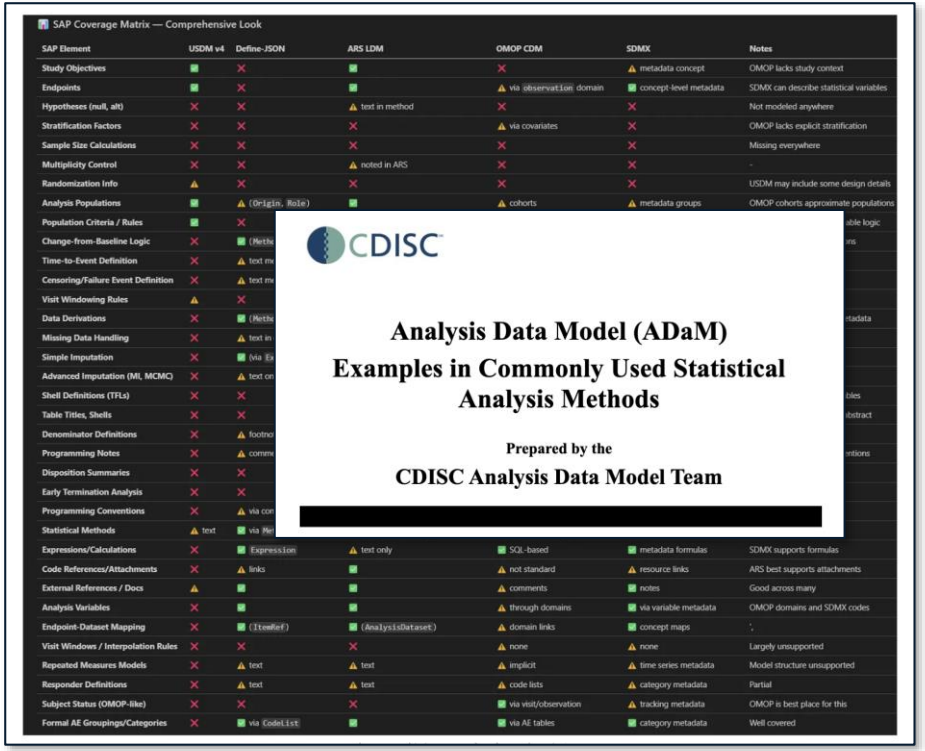
Towards an eSAP Model

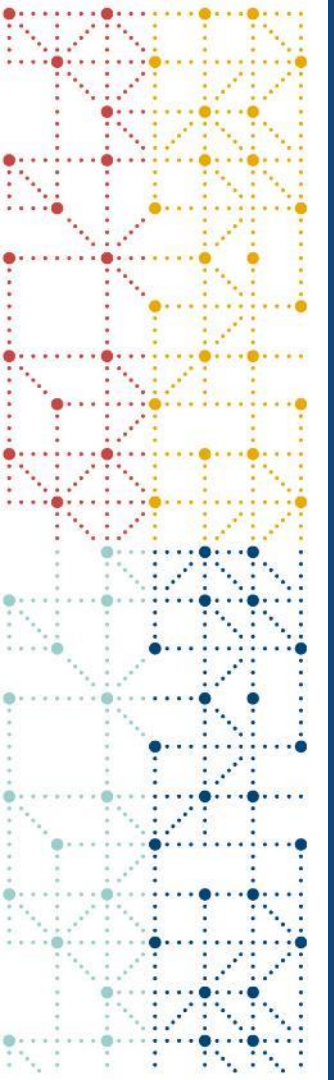
- AC definitions serve as the foundation for a Digital SAP
- Digital SAP → machine-readable, executable, and version-controlled
- Drives automation of analyses, traceability, and regulatory alignment

Approaches Considered

- **Top-Down**
  - What are the element covered across sample SAPs?
  - What element are covered across standards?

- **Bottom-up**
  - Build a picture of an analysis concept using the most common statistical analyses



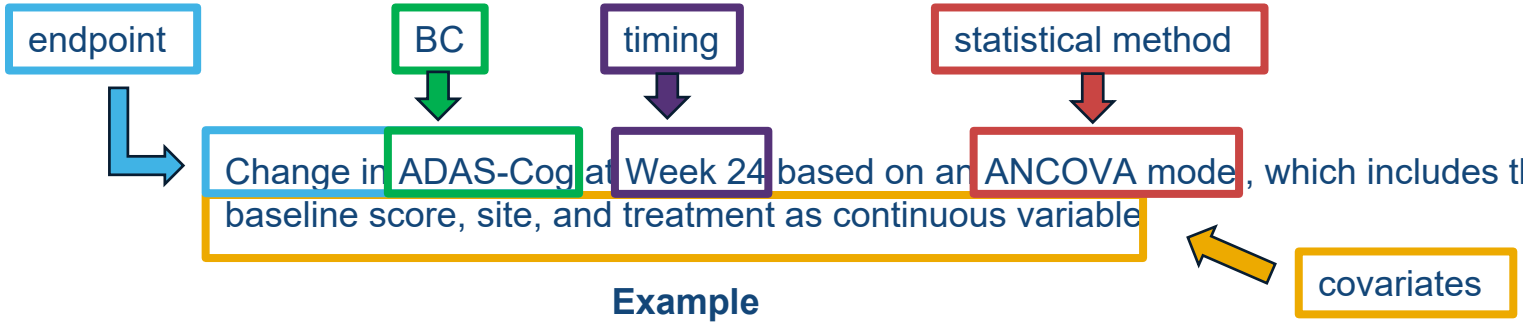


Expanding Into Analysis Concepts

Analysis Concepts

- What is an Analysis Concept?
  - Collection of standardized configurable elements & methods that are needed to describe a final output or answer a clinical question
  - Standard definitions in a digital statistical analysis plan

- What is the purpose of an Analysis Concept?
  - Supports configuration and automation of analysis
  - Informs what needs to be collected to answer the clinical question



Today's specification of analysis

- Written in plain text
  - not machine readable
- Some details are left out
  - often (at best) retrievable from the analysis executable program/define.xml
  - reason why regulators requires SAS programs to be submitted
- Specifications are not software agnostic
  - SAS versus R syntax
  - different set of options

*know the document is old, but very likely the same is seen in newer documents

Primary analysis from CDISC pilot study*	**Issue**
The primary analysis of the ADAS-Cog (11) at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24.	define.xml revealed that it was Change from baseline (CHG) that was analysed.
*An ANCOVA model will be used with the **baseline score**, **site**, and **treatment** included as **independent variables**.* ***Treatment** will be included as a **continuous** variable, and results for a **test of dose response** will be produced.* ***Interaction terms will not be investigated**.*	The define.xml specifies how the analysis is done in SAS, bot how to translate to R or other programming language?
*If the test for **dose response is statistically significant**, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.*	Pairwise estimates made – but test for dose response was significant ($p < 0.05$)

Why Develop Analysis Concepts?

- Reduces ambiguity
  - Analysis Concepts with standardized metadata structure enforce precision in specifying analysis settings and assumptions
- Enables machine-readable analysis plans
  - Analysis Concepts structured as metadata rather than narrative text, can be directly linked to statistical programming code
- Supports traceability
  - Analysis Concepts help maintain clear linkage between protocol objectives, endpoints, and the specific analytical methods applied
- Streamlining collaboration
  - Analysis Concepts provide a common language between statisticians, clinicians, data managers, and other stakeholders

Working Towards a Logical Model

endpoint

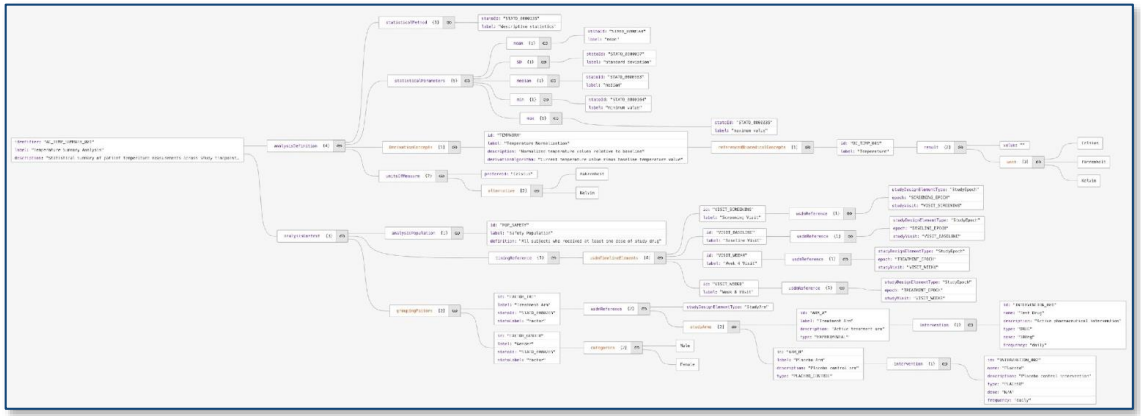
BC

timing

statistical method

Change in ADAS-Cog at Week 24 based on an ANCOVA model, which includes the baseline score, site, and treatment as continuous variable

covariates



AC DC

	Analysis Concept (AC)	Derivation Concept (DC)
Purpose	Examines existing data or information to draw conclusions, identify patterns, or test hypotheses	Generates new derived values from raw or derived data
Direction	Typically works with completed measurements or observations to extract meaning	Transforms or processes data to create new representations
Process	Involves applying statistical methods, critical thinking, and interpretative frameworks to understand data	Uses mathematical operations, formulas, or algorithms to calculate new quantities
Outcome	Produces insights, conclusions, or evaluations based on the data - aggregated data (not subject-level)	Creates derived data (subject level) that serve as inputs for subsequent analysis or derivations
Example 1	The p-value (from Type III Sums of Squares for treatment dose), based on linear model analysis of CHG for dose response; using randomized dose and site group in model.	CHG: Change from Baseline to Week 24 in ADAS Cog (11). Use LOCF is missing value at week 24.
Example 2	Mean value of CHG by visit	CHG: Change from BASELINE in ADAS Cog (11) by visit BASELINE = 'Y' if ADAS Cog (11) at visit 2

Next Steps

- Develop Data Model
- Defining the boundaries between USDM, Analysis Results Standard, Biomedical Concepts, and Analysis Concepts
- Determine how to leverage existing ontologies such as STATO
- Determine how to make this relatable to the CDISC community?
  - Creation of an eSAP?

Thank You!

For Analysis Concepts Working Group

- Bhavin Busa, Clymb Clinical
- Brian Harris, AstraZeneca