ISCB GMDS 2026

Automated Generation of Clinical Trial Results based on the CDISC Analysis Results Standard (ARS) using SAS (or R)

Hennadiy Olshanyi[a], Astghik Mikayelyan[b], Arman Azaryan[a], Wilmar Igl[b]

[a] Statistical Programming, Veranex, Inc., [b] Biostatistics, Veranex, Inc.,

Veranex

Introduction

Data management and statistical analysis consume considerable resources in clinical trials.

Before 2024, generation of **analysis results from clinical trials** suffered from the following issues:

- no standards for describing & organizing outputs
- not machine-readable
- difficult to automate, reproduce, trace, or reuse

In 2024, CDISC published the **Analysis Results Standard (ARS)** to resolve the above issues by providing [1, 2]:

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

Therefore, the high degree of **standardization of datasets (ie, SDTM, ADaM) and results (ie, ARS) by the CDISC consortium** creates opportunities for automation of these activities.

Here, insights from a **pilot study** for the automated generation of clinical trial results using SAS (or R) will be presented.

Background

Standards by CDISC consortium

- **Tabulation datasets** from clinical studies are highly standardized in the **CDISC/SDTM** (Study Data Tabulation Model) by the CDISC consortium
- **Analysis datasets** from clinical studies are highly standardized in the **CDISC/ADaM** (Analysis Data Model) by the CDISC consortium.
- **Summary results** from clinical trial data have now also been standardized in the **CDISC/ARS** (Analysis Results Standard) [1,2]

TFL Automation

- A TFL automation framework based on **Automatic Coding** (no AI!) as published by Clymb Clinical, Inc. [6]
- A **SAS-based framework** was published by Wachara (2025) [3]
- An **R-based framework** was published by Bosman (2025) [4] using the R/siera and R/tfrmt packages.

Methods

Resources

- Pilot study was conducted by 2x biostatisticians and 2x statistical programmers in Q2 and Q3 2026

Data:

- Public-use ADaM Datasets (ADSL, ADAE, ADCM, ADMH, ADLB) from CDISC Dataset Generator [5]
- 100 simulated patients
- 3 treatment groups (placebo, treatment X1, treatment X2)

Analysis:

- Descriptive statistical analysis
- Automated Coding using SAS-based framework was published by Wachara (2025) [3]

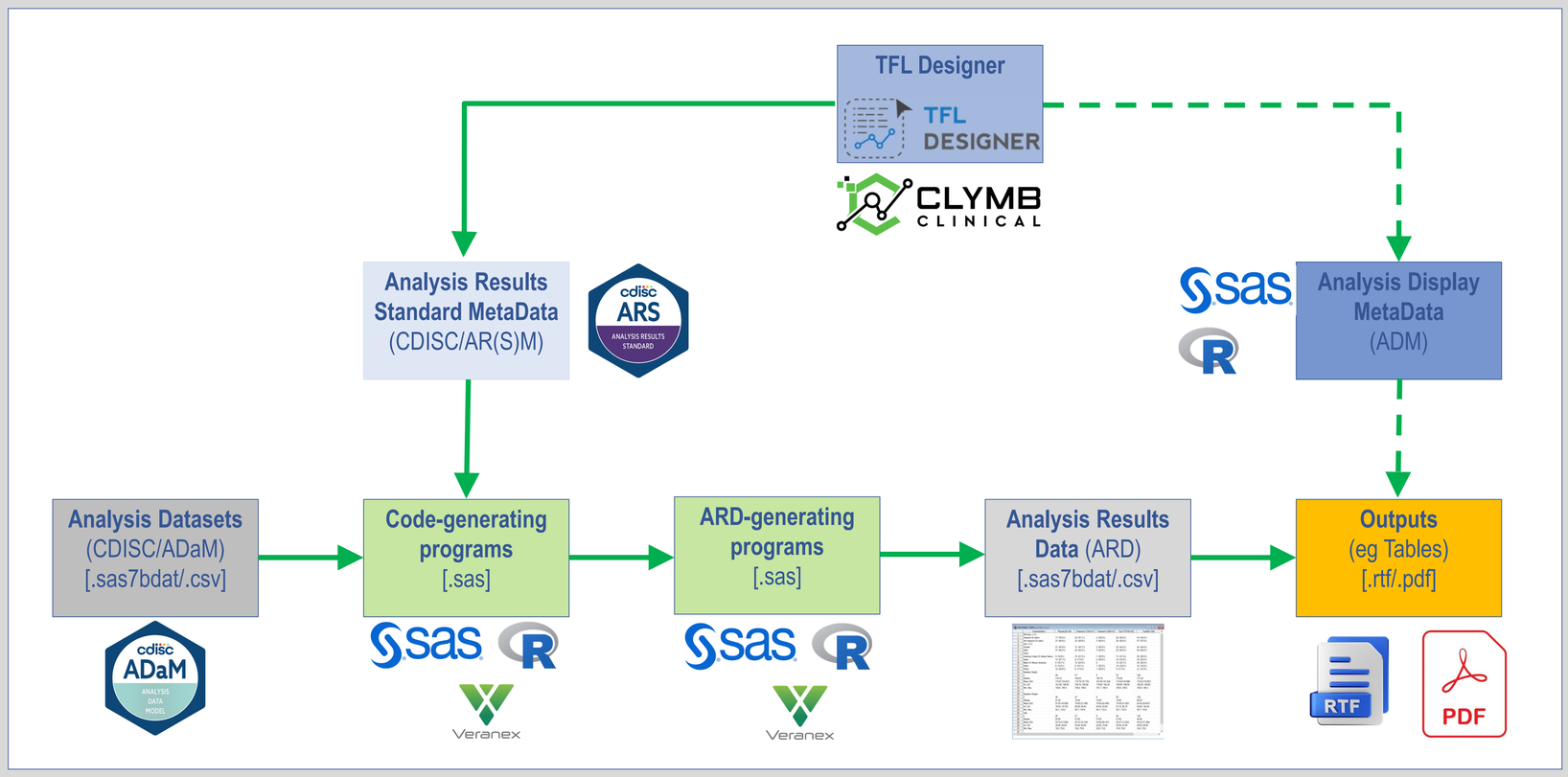
Software:

- Design TFL templates (“mock shells”) with TFL Designer web app [7, 8, 9, 10]
- Software development of the **VRX TFL Automation Engine** as a collection of proprietary SAS macros by Veranex, Inc.

Workflow from datasets to outputs using TFL templates (Fig. 1):

1. **Design a table template** (“mock shell”) for the expected results, ie, without empirical results, with TFL Designer web app.
2. **Export meta-data files** with table structure and layout.
   a) Export the Analysis Results Meta-data (**ARM**) as the abstract description of the table structure in JSON format.
   b) [optional with SAS] Export the Analysis Display Meta-Data (**ADM**) as the concrete description of their graphical layout.
3. **Generate the Analysis Programs** (ie, SAS or R) based on ADaM and ARM specifications.
4. **Generate the Analysis Results Data** (ARD), ie, the empirical summary results, by running the analysis programs.
5. **Generate the Final Outputs**, eg, showing the summary results in human-readable, printable tables in RTF/PDF format by combining the Analysis Results Data (ARD) and the Analysis Display Meta-Data (AMD).

Figure 1: Workflow of table automation from design of table template to table output using SAS (or R) for Automatic Coding



Results

Table 1: Number and title of automated tables in pilot project

Table ID	Table Title
#1/14.1.1.1.1	Summary of Demographics and Baseline Characteristics (Intent to Treatment Population)
#2/14.1.2.1	Summary of Medical History by System Organ Class and Preferred Term (Intent to Treatment Population)
#3/14.1.3.2	Summary of Concomitant Medications by Anatomic Classification & Preferred Term (Intent to Treatment Population)
#4/14.3.1.2.1	Summary of TEAE by System Organ Class and Preferred Term (Safety Population)
#5/14.3.1.2.2	Summary of Serious TEAE by System Organ Class and Preferred Term (Safety Population)
#6/14.3.1.2.3	Summary of Related TEAE by System Organ Class and Preferred Term (Safety Population)
#7/14.3.2.1	Summary of Observed & Change from Baseline by Scheduled Visits Hematology Laboratory Test (Safety Population)
#8/14.3.2.2	Summary of Observed and Change from Baseline by Scheduled Visits - Chemistry Laboratory Test (Safety Population)

- **Eight descriptive tables were successfully generated** using Automatic Coding based on the SAS-based framework (Table 1)
- Total required work time was around **1x employee-month** (with some additional effort during an initial exploratory phase).
- The current VRX TFL Automation Engine includes about **2,500 lines of code across 14 SAS programs**.
- The size of automatically generated SAS macros have between **131 and 160 lines of automated SAS code per table**.
- **Reduction of worktime for a (new type of) descriptive table** from an existing table template to less than 1 hour (instead of multiple hours).

TFL Automation using R

- **Table automation using an R-based framework** (Bosman, 2025) with the R/siera and tfrmt packages (as alternative to SAS) was tested, but not fully implemented for the following reasons:
  – Limitations of the community version of the TFL Designer tool, which does not export the Analysis Display Metadata (ADM)
  – Additional effort for manual programming of the table layout using the R/tfrmt package, thereby, substituting the missing Analysis Display Meta-Data (ADM) with R/tfrmt configurations.

Conclusion

PROs:

- **Increased Efficiency**
  – Biostatisticians can design TFL templates more effectively using the TFL Designer web app
  – Statistical programmers reduce manual SAS programming by automation and standardization
- **Improved Quality & Consistency**
  – Decreased risk of manual programming errors
  – Decreased variability of programs depending on programmer
  – Increased use of metadata-driven and standardized macros reducing size of SAS programs
- **Alignment with regulatory and industry standards**
- **Reproducibility, traceability and auditability** (ie no AI used!)
- **Additional (freed up) resources** for quality assurance, interpretation, and innovation

Conclusion (cont.)

CONs:

- Pilot project currently limited to **tables with results from descriptive statistical analysis**
- **No automation of figures and listings** currently available.
- **Validation of table outputs by manual statistical programming** still recommended
- **Studies which do not follow CDISC (SDTM/ADaM) standards** may not benefit from TFL automation (eg, medical device studies)
- **Effects of standardization and automation on required resources for Biostatisticians and Statistical Programmers** need to be further evaluated using historical examples (horses vs cars for mobility, coal as energy source for steam engines, cf Jevons Paradox) [11, 12].

Next Steps:

- **Expand SAS macro library** to tables with more complex statistical analysis including inferential statistical analysis
- **Comprehensive validation** of TFL automation systems and processes
- **Update internal infrastructure and processes (eg, SOPs)** of production systems

Acknowledgements

The pilot study was a joint effort between the Departments of **Biostatistics** (HoD Global: Wilmar Igl) and the Departments of **Statistical Programming** (HoD Ukraine: Serhii Voievutskyi, HoD India: Koteshwar Govind) at **Veranex** in collaboration with **Clymb Clinical** (https:/clymbclinical.com).

References

[1] Busa, B. et al. (2023). All You Need to Know about the New CDISC Analysis Result Standards! PharmaSUG 2023, Paper MM-327. https://pharmasug.org/proceedings/2023/MM/PharmaSUG-2023-MM-327.pdf

[2] Bhusa, B. & Dedhia, N. (2025). Harnessing AI & CDISC ARS for Effortless Statistical Reporting & CSR Writing. PharmaSUG 2025, Paper AI-349, https://pharmasug.org/proceedings/2025/AI /PharmaSUG-2025-AI-349.pdf

[3] Wachara, J. (2025). Let the Machine do the Coding: A SAS-based Framework for Automating TFL Generation with ARS Metadata. Phuse US, 2025, https://www.lexjansen.com/phuse-us/2025/sm/PAP_SM13.pdf

[4] Bosman, M. (2025-03-19). Advancing TFL Automation: R-Based Meta-Programming Powered by ARS. Phuse EU, 2025, Paper OS20. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2025/ Connect/ US/Orlando/PAP_OS20.pdf

[5] CDISC Dataset Generator (2026). https://cdiscdataset.com

[6] Clymb Clinical (2025-06-23). Clymb TFL Automation Suite - From Designing Shells to Review of Outputs [Video Tutorials], https://clymb-tfl-automation.carrd.co/

[7] Clymb Clinical (2026-09-14), TFL Designer [Web App], https://www.tfldesigner.org

[8] The Clymb Team (2024-08-15). Webinar: TFL Designer Community Release Version 1.0.0. Video: https://www.youtube.com/watch?v=cZTf9QMslpo

[9] Busa, B. (2022-09-13). Open-Source TFL Designer: Virtual Workshop. https://github.com/bhavinbusa/tfldesigner/blob/main/design-thinking-workshop/TFL-Designer-Virtual-DesignThinking-Workshop-Part1-13Sep2022.pdf

[10] Busa, B. (2022-10-06). Open-Source TFL Designer: Virtual Workshop 2. https://github.com/bhavinbusa/tfldesigner/ blob/main/ design-thinking-workshop/TFL-Designer-Virtual-DesignThinking-Workshop-Part2-06Oct2022.pdf

[11] Hermansson, N. (2026-08-19). Är du kol eller en häst? Så vet du om ditt jobb överlever AI-revolutionen, https://hrnytt.se/aktuellt/ar-du-kol-eller-en-hast-sa-vet-du-om-ditt-jobb-overlever-ai-revolutionen/

[12] Wikipedia (2026). Jevon’s Paradox [Lexical Entry]. https://en. wikipedia.org/wiki/Jevons_paradox

Contact Information

Corresponding author: Wilmar Igl, PhD (Veranex)
Email: wilmar.igl@veranex.com

Website: https://veranex.com

29-P1-04