Paper SM05

Automated QC Framework for Reliable ADNCA Dataset Generation with {admiral}

Valentin Legras, Roche, Basel, Switzerland

Lucy Aspridis, Roche, Basel, Switzerland

ABSTRACT

The generation of robust analysis datasets for Non-Compartmental Analysis (NCA), such as ADNCA, is a critical and time-intensive task for PK/PD data scientists. Manual quality control (QC) is often inefficient and prone to structural inconsistencies. We present a standardized, automated QC framework developed in R. This framework systematically validates foundational SDTM data, leverages the {admiral} package for transparent dataset construction, and performs final verification on the resulting ADNCA. This process ensures a reliable, analysis-ready dataset, which seamlessly integrates with analysis tools like {aNCA} for streamlined NCA and reporting.

INTRODUCTION: THE DATA SCIENTIST'S PRIMARY CHALLENGE

Creating analysis datasets for PK/PD is one of the most rigorous and time-consuming tasks in the data analysis pipeline. The quality control (QC) of source data, particularly from core SDTM domains, is paramount. A dataset may pass standard checks yet contain structural issues that invalidate an NCA, such as dosing or timing inconsistencies. This framework addresses this challenge by automating and standardizing the entire workflow, from raw data integrity checks to the final creation of the Analysis Dataset for NCA (ADNCA), leveraging the {admiral} package for a streamlined and reproducible process.

1. METHODS: A MULTI-STAGE QC FRAMEWORK

Our solution is a four-phased approach to ensure data integrity at every step of the dataset generation process.

STEP 1: SETUP

The script sets up the R environment. It begins by checking the R version and then loads all required packages, installing any that are missing. After loading packages, it sources all custom functions from the /functions directory. Finally, it loads the SDTM (and ADaM if available). As a last step, it filters all loaded data to retain only the subjects belonging to the PK population, which is defined by subjects found in the PC domain.

Then we validate the data and define the study type. Its first task is to assert that all required columns exist in their respective SDTM domains. Next, it derives the key study attributes by scanning the data for specific keywords. Finally, the script generates summary outputs, including a heatmap of PK sample availability and an Excel-based CDISC summary.

The g_study_design object is a central list of TRUE/FALSE flags. This key object is populated by scanning the data once to set logical flags like g_study_design$single_dose, g_study_design$period, g_study_design$IV, g_study_design$urine, and g_study_design$FE (Fed/Fasted). This approach simplifies all downstream scripts, as they no longer need to re-query the raw data. Instead, they can just check a pre-defined flag.

STEP 2: QUALITY CONTROL OF EVERY DOMAIN

Before any analysis dataset is built, we perform a battery of automated checks on the core SDTM domains. This QC identifies and flags potential issues at the source.

Key domains and critical checks include:

- PC (Pharmacokinetics Concentration):
    - Duplicate Timepoints: No identical timepoints for the same subject.
    - Date/Time Integrity: Checks the correct format of PCDTC and PCENDTC.
    - PK Sample Inversions: Ensures PK samples are correctly ordered.
    - Visit Order: Validates the sequence of nominal visits (VISITDY).
    - Alignment between Planned and Actual: Checks for discrepancies between planned and actual sample (PC) times, visualized via a plot.
- EX (Exposure):
    - Dose Integrity: Check for duplicate doses.
    - Zero-Duration Infusions: Flags infusions where start and end times are identical.
    - Dose/Visit Alignment: Compares actual dose times (EXSTDTC) to nominal visit information (VISITDY), flagging discrepancies.
    - Interrupted & Multiple Doses: Flags complex dosing scenarios.
- VS (Vital Signs):
    - Unique Baseline Flag: Ensures each subject/parameter combination has one unique baseline, using visualizations for quick review.

STEP 3: BUILDING THE ADNCA DATASET WITH {ADMIRAL}

Once the source SDTM data passes all QC checks, we proceed to create the ADNCA dataset. We utilize the {admiral} package, part of the R-Pharma ecosystem, to ensure our analysis dataset is robust, transparent, and compliant with ADaM standards.

Key derivations include:

- Merging core domains (EX and PC).
- Deriving critical time variables relative to dosing (e.g., AVISIT, AFRLT, NFRLT).
- Deriving analysis-ready parameters and flags (e.g., AVAL, AVALCAT, ATPTREF).

STEP 4: FINAL CHECKS

The newly created ADNCA dataset is subjected to a final set of verification checks to guarantee all derivations are accurate and no artifacts were introduced.

- Time Variables:
    - Nominal vs. Actual Time: Scatter plots compare nominal (NRRLT, NFRLT) vs. actual (ARRLT, AFRLT) times to visually inspect for deviations and identify outliers.
- Dose Numbering:
    - Planned vs. Actual Dose Number: Tabular summaries compare planned (DOSNOA) and actual (DOSNOP) dose numbers.
    - Skipped Doses: An algorithm compares the observed dose sequence against the planned sequence within each treatment arm to detect and flag potential skipped doses.
- Dataset Integrity:
    - Audit & Comparison: The final ADNCA is programmatically compared against previous versions of the dataset to identify any changes in values, rows, or columns.

2. RESULTS: RUNNING THE NCA WITH {ANCA}

The output of this framework is a fully validated and quality-assured ADNCA dataset. This dataset serves as the direct input for the {aNCA} (Analysis of Non-Compartmental Data) R Shiny application.

{aNCA} is a streamlined, interactive tool that performs the non-compartmental analysis, calculating key PK parameters (e.g., Cmax, AUC) and deriving critical plot data. The application features:

- Interactive Data Exploration: Allows scientists to visually inspect individual concentration-time profiles and explore the results of the NCA.
- Run the NCA with multiple settings options
- Submission-Ready TLG Generation: Provides automated, reproducible Tables, Listings, and Graphs (TLGs) that are formatted according to industry standards for inclusion in clinical study reports.

CONCLUSION

A robust, automated QC process is the bedrock of reliable Non-Compartmental Analysis. By systematically validating data at the SDTM source, using {admiral} for transparent derivation, and performing final verification on the ADNCA dataset, we eliminate a major source of error in PK analyses. This structured R-based approach transforms a traditionally error-prone task into a transparent and dependable workflow.

REFERENCES

CDISC SDTM & ADaM Implementation Guides: cdisc.org

Admiral github.com/pharmaverse/admiral

Pharmaverse pharmaverse.org

aNCA github.com/pharmaverse/aNCA

CONTACT INFORMATION

Valentin Legras, valentin.legras@roche.com, linkedin.com/in/valentin-legras/

Lucy Aspridis, lucy.aspridis@gmail.com, linkedin.com/in/lucy-aspridis/