PharmaSUG 2026 - Paper 272

An End-to-End R-Based Pharmacokinetics (PK) Workflow for Regulatory Submission: The INAVO120 Study

Qi Liu, Shweta Vadhavkar, Clinical Pharmacology, Genentech

ABSTRACT

The INAVO120 clinical study supported the development and regulatory approval of inavolisib and represented Roche/Genentech's first end-to-end regulatory submission executed entirely in R using the OCEAN (One CEntralised Analytics), an open-source, cloud-enabled analytics environment. This work emphasizes modern programming practices for clinical pharmacology analyses, including pharmacokinetics (PK), pharmacodynamics (PD), and model-informed drug development.

The Clinical Pharmacology, Modeling and Simulation Analyst (MSA) team, in collaboration with Pharmaceutical Development Data Sciences (PDDS) and OCEAN platform teams, redesigned the clinical study reporting process by transitioning PK analysis datasets, Analysis Data Model (ADaM) deliverables, Tables, Listings, and Graphs (TLGs), and Modeling & Simulation outputs from SAS to R. PK ADaMs were programmed in alignment with the latest CDISC standards for Non-compartmental Analysis (NCA), enabling standardized, submission-ready outputs and early PK data access.

The Inavolisib New Drug Application (NDA), submitted in March 2024 and approved by the FDA in October 2024, served as the inaugural use case for this approach. Despite challenges associated with adopting new tools, automation, and GitLab-based collaborative workflows, the MSA team successfully delivered CSR-ready PK and M&S datasets. This case study demonstrates how open-source tools (R), automation, and CDISC-aligned standards can be integrated to support scalable, regulatory-ready PK/PD and model-informed drug development workflows.

INTRODUCTION

Clinical pharmacology analyses are essential for understanding drug exposure, safety, and efficacy relationships during drug development. Pharmacokinetic (PK) analyses support dose selection, exposure-response modeling, and regulatory decision-making. Historically, these workflows have relied heavily on proprietary statistical software such as SAS. However, increasing interest in open-source tools has motivated pharmaceutical organizations to explore alternative analytics environments.

The OCEAN (One CEntralised Analytics) platform initiative represents Roche/Genentech's next-generation analytics platform designed to integrate open-source tools, cloud computing infrastructure, and modern software development practices. Within this environment, the INAVO120 clinical study served as an early adopter, implementing a fully R-based programming workflow for regulatory submission.

This paper describes the technical implementation of PK dataset development within the OCEAN platform for the INAVO120 study. The focus is on the programming workflow used to construct regulatory-ready PK ADaM datasets in R, including early PK data ingestion, structured data transfer from upstream systems, standardized ADaM derivation pipelines, and reproducible programming frameworks used to generate submission-ready datasets within OCEAN.

OCEAN PLATFORM ARCHITECTURE

OCEAN is an internal analytics platform developed as a strategic project within Roche/Genentech to modernize clinical data programming and reporting workflows. The platform hosts both next-generation and legacy analytics tools in a unified environment and leverages cloud-based infrastructure and open-source programming languages to support scalable, collaborative, and reproducible clinical data analysis.

Within Roche/Genentech, the Clinical Pharmacology Modeling and Simulation Analyst (MSA) team collaborated closely with the PD Data Sciences (PDDS) group and OCEAN platform development teams to onboard the INAVO120 study. The goal was to transition PK dataset programming and clinical reporting deliverables from traditional SAS-based workflows to standardized R-based programming frameworks implemented within the Roche/Genentech OCEAN environment.

The INAVO120 study began using OCEAN in June 2023 and completed the regulatory submission in March 2024, representing the first end-to-end R-based regulatory filing implemented within the Roche/Genentech OCEAN platform.

PK DATA WORKFLOW AND EARLY PK DATA ACCESS

A key capability implemented in the INAVO120 study was early PK data access within the OCEAN environment, which enabled the programming team to begin dataset development well before the final database lock (DBL). Traditionally, PK dataset programming starts after DBL when finalized clinical data are available. In contrast, the OCEAN workflow allowed preliminary PK data to be securely accessed earlier in the study lifecycle, enabling the MSA programming team to design, implement, and test the ADaM datasets derivation pipeline in advance.

Within OCEAN, early PK access primarily supported dataset creation activities, including development of ADaM structures, validation of derivation logic, and testing of R programming modules used to generate regulatory datasets. Dummy and preliminary PK datasets were used to simulate the final data structure expected at DBL. This approach allowed programmers to verify variable mappings, implement CDISC-compliant dataset specifications, and ensure compatibility with downstream reporting pipelines.

The early PK data access timeline included:

- June 2023: Dummy PK datasets made available within OCEAN for workflow development and testing of dataset derivation scripts
- October 30, 2023: Early PK datasets available restricted OCEAN area to support validation of ADaM programming and dataset structure
- November 23, 2023: Database lock (DBL), when final clinical PK data became available for generation of submission-ready datasets

Within OCEAN, early PK access enabled implementation of the PK dataset creation workflow used for the PK activity. The workflow was structured to support stepwise dataset creation and data exchange between OCEAN and downstream PK calculation tools while maintaining traceability of intermediate datasets.

The operational PK dataset workflow included the following steps:

1. Clinical data preparation in OCEAN
Clinical non-treatment assignment information (non-TAI) datasets were first made available in the ADS (Analytical Data Science) activity space. The ADS team, part of PD Data Sciences within Roche/Genentech, prepares and standardizes upstream clinical datasets (e.g., masking, SDTM alignment, and ADaM-ready structure preparation) for downstream PK ADaM derivation in OCEAN. Masked SDTM and supporting ADaM datasets were shared internally through automated storage scripts within OCEAN to support downstream dataset derivation.

2. PK raw data ingestion
Raw PK datasets (e.g., PKC, dummy DM domain) were ingested into the PK activity within OCEAN by the independent Statistical/Scientific Programming Analyst (iSPA) team, who are responsible for blinded PK data handling and mapping to standardized structures prior to downstream ADaM derivation. These datasets served as the source for PK analysis dataset development. From these inputs, programmers constructed the ADPC dataset, which contains concentration-time observations and derived analysis variables required for PK evaluation.

3. Export of ADPC for NCA calculations
Once ADPC was created and validated within OCEAN, the dataset was exported to the Improve system to support non-compartmental analysis (NCA). This step enabled calculation of PK parameters using standardized NCA tools while maintaining the dataset derivation logic within the OCEAN programming environment.

4. Generation of pre-PP datasets
Following NCA processing, intermediate outputs (pre-PP datasets) were generated. These datasets included PK parameter results that were prepared for integration into the ADPP dataset structure.

5. Integration of PK parameters back into OCEAN
The NCA output generated by Improve (e.g., pre-PP CSV files) was returned to the OCEAN environment. Upon return of NCA outputs, the ADS team mapped the pre-PP files to the SDTM PP domain, and the MSA programming team subsequently constructed ADPP for regulatory reporting.

OCEAN served as the central storage location for intermediate and final datasets, including pre-PP files and derived ADaM datasets. This centralized architecture ensured that all datasets required for PK reporting, modeling, and regulatory submission were accessible within a single controlled environment.

Through this staged workflow, OCEAN supported early development of dataset derivation scripts, validation of dataset structures, and testing of data exchange processes with downstream tools. Because the majority of the programming framework and dataset derivation logic had already been implemented prior to DBL, final PK ADaM datasets could be generated rapidly once the locked database became available, significantly reducing post-DBL programming time.

R-BASED PK PROGRAMMING FRAMEWORK

The INAVO120 PK workflow relied heavily on the pharmaverse ecosystem, a collection of open-source R packages designed for pharmaceutical data analysis.

ADAM DATASET PROGRAMMING

PK analysis datasets were programmed according to CDISC Analysis Data Model (ADaM) standards for Non-compartmental Analysis (NCA). Key datasets included:

- ADPC: Pharmacokinetic concentration dataset
- ADPP: Pharmacokinetic parameter dataset

The following R packages were used to automate dataset development:

- admiral (Bury et al., pharmaverse) – provides modular functions for ADaM dataset creation
- admiralroche (Roche/Genentech internal extension) – Roche-specific extension supporting internal workflows and regulatory requirements

These tools streamlined dataset programming while ensuring compliance with CDISC standards.

MODELING AND SIMULATION DATASET PROGRAMING

In addition to ADaM PK datasets, OCEAN supported the creation of datasets required for population pharmacokinetic (PopPK) and exposure-response (E-R) analyses. These datasets were derived from the ADPC and ADPP datasets and were structured to support downstream modeling activities conducted by internal modeling teams and external vendors.

Within the OCEAN programming environment, the MSA team developed standardized R workflows to construct modeling-ready datasets. Key derivation steps included:

- Integration of PK concentration data from ADPC with subject-level covariate
- Derivation of time-varying exposure variables and dosing records
- Preparation of structured datasets suitable for NONMEM or other population modeling tools

For exposure-response analyses, additional datasets were created by combining PK exposure metrics with clinical endpoints such as efficacy outcomes and safety variables. These datasets enabled evaluation of exposure–efficacy and exposure–safety relationships as part of the model-informed drug development strategy.

By implementing these dataset derivation pipelines within OCEAN, the programming team ensured that PopPK and E-R datasets were generated in a reproducible and standardized manner, facilitating efficient data sharing with modeling teams and supporting regulatory submission deliverables.

QUALITY CONTROL (QC) AND VALIDATION STRATEGY

To ensure the accuracy, reliability, and regulatory acceptability of datasets generated in R within the OCEAN environment, a robust quality control (QC) framework was implemented. A key component of this strategy was the use of independent SAS-based programming to perform double programming and validation of critical PK datasets.

For key deliverables such as ADPC, ADPP, and modeling datasets, parallel QC programs were developed in SAS to independently reproduce dataset derivations implemented in R. This dual-programming approach enabled systematic comparison of outputs between the two programming environments, ensuring:

- Consistency of derived variables and dataset structures
- Accuracy of PK concentration and parameter mappings

- Alignment with CDISC ADaM standards and dataset specifications
- Verification of edge cases and complex derivation logic

Discrepancies identified between R and SAS outputs were investigated through structured QC workflows, including variable-level comparisons, dataset-level reconciliation, and traceability checks back to source data. This process ensured that all differences were resolved and that final datasets met regulatory quality standards.

In addition to accuracy, this QC approach also provided an opportunity to evaluate the efficiency and robustness of R-based workflows relative to traditional SAS programming. The results demonstrated that R-based dataset creation within OCEAN could achieve equivalent accuracy while improving automation, modularity, and reproducibility.

Overall, the integration of SAS-based QC alongside R programming established a high-confidence validation framework, supporting the use of open-source tools for regulatory submissions within Roche/Genentech.

AUTOMATED TABLES, LISTINGS, AND GRAPHS (TLGS)

Generation of Tables, Listings, and Graphs (TLGs) was performed entirely within R using several packages from the pharmaverse ecosystem.

Key packages included:

- tern (Insights Engineering) – creation of clinical trial analysis tables
- chevron (Insights Engineering) – standardized TLG templates for clinical reporting
- citril (Insights Engineering) – workflow tools for generating R Markdown reports
- rtables (Insights Engineering) – flexible table creation framework
- rlistings (Insights Engineering) – generation of regulatory-compliant listings

The MSA team also contributed to the development of a TLG catalog containing reusable templates for PK and ADA outputs. These templates improved standardization and efficiency across clinical study reporting workflows.

PROJECT MANAGEMENT AND REPRODUCIBILITY

All programming code was managed through GitLab repositories. GitLab supported collaborative development, version control, and tracking of programming progress through issue management.
This software engineering approach improved reproducibility and transparency of the PK programming workflow.
Additionally, modeling datasets were securely transferred to external modeling vendors using BOX.
TLG review processes were performed using the entipub tool, and finalized outputs were published in the RIM system as part of the regulatory submission package

CHALLENGES AND LESSONS LEARNED

The transition to OCEAN introduced several challenges, including:

- Working within a new and evolving analytics environment
- Adopting R instead of the traditional SAS programming framework
- Establishing new workflows for sharing deliverables with internal teams and external vendors

Despite these challenges, close collaboration between the Roche/Genentech Clinical Pharmacology MSA team, the Roche/Genentech PD Data Sciences (PDDS) group, and the Roche/Genentech OCEAN development teams enabled successful completion of the regulatory submission.

CONCLUSION

The INAVO120 study demonstrated the feasibility of performing an end-to-end regulatory submission using R within the Roche/Genentech OCEAN analytics platform. The integration of open-source tools, CDISC-compliant data standards, and automated reporting workflows within Roche/Genentech enabled efficient generation of PK analysis outputs and regulatory deliverables.
This case study highlights how modern analytics platforms and open-source programming ecosystems can transform clinical pharmacology workflows, improving reproducibility, scalability, and operational efficiency for future regulatory submissions.

REFERENCES

Bury, J., et al. admiral: An R Package for ADaM Dataset Creation. pharmaverse. https://github.com/pharmaverse/admiral
Insights Engineering. tern: Tables for Clinical Trials. https://github.com/insightsengineering/tern
Insights Engineering. rtables: Reporting Tables for Clinical Trials. https://github.com/insightsengineering/rtables
Insights Engineering. rlistings: Clinical Trial Listings. https://github.com/insightsengineering/rlistings
Insights Engineering. chevron: Clinical Reporting Templates. https://github.com/insightsengineering/chevron
Insights Engineering. citril: R Markdown Workflow Tools. https://github.com/insightsengineering/citril

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Qi Liu
Liuq36@gene.com