PharmaSUG 2025 - Paper DS-256

What Comes First, the Chicken (ADSL) or the Egg (ADNCA)? Modularize Your Covariate Creation to Support Flexible Analysis Dataset Implementation!

David C. Izard, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Good Programming Practice dictates that you should create a variable once and then share it to other locations where it is needed. In the context of analysis datasets following the CDISC ADaM standard, this typically involves creating your covariates during the implementation of the ADSL dataset and then merging them on to other analysis datasets at subsequent points in the development sequence. ADNCA, the ADaM BDS dataset designed to support non-compartmental PK analysis, would utilize these covariates from ADSL but is usually needed and implemented before ADSL is available. This paper explores modular covariate creation and how it supports both the subtle details of covariate differences between ADSL and ADNCA and good programming practices.

INTRODUCTION

Pharmacokinetic (PK) Non-Compartmental Analysis (NCA) is a method used to calculate PK parameters by directly using the observed drug concentration measurements at specific time points. It does not use predictive modeling to fit a curve for Area under the Curve (AUC), rather it deploys simple techniques such as linear interpolation to achieve its goals. As a result, for NCA to be as accurate as possible, you require a significant number of timepoints across participants to achieve successful characterization of a molecule, particularly if drug absorption is swift [1].

LEGACY PROCESSING OF PHARMCOKINETIC DATA

To support NCA in a pharmaceutical setting, leveraging your collected data optimistically captured based on CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standards and organized based on CDISC Study Data Tabulation (SDTM) standards, you identify the SDTM domains needed to support this analysis. This typically includes, but is not limited to, the following:

- Demographics (DM)
- Study Medication Exposure (EC/EX)
- Adverse Events (AE)
- Disposition (DS)
- Protocol Deviations (DV)
- Pharmacokinetic Data [actual drug concentration at actual collected time points] (PC)

Other domains may play into this, but this is what is typically consumed by a PK scientist to take it to the next level, which is integrating this data into a specific format in order to be successfully consumed by a software package such as Phoenix WinNonlin™ from Certara, Inc. These transformations typically involve the following steps:

- Identifying the participants in the PK analysis
- Computing necessary covariates and baseline characteristics
- Organizing the PK concentration data sequentially within a participant
- Standardizing / harmonizing PK concentration data if not already done so
- Calculating a number of timepoint values tied to the original dose within a cycle

- Determining if a timepoint / full set of data for a participant is eligible for consideration in the analysis

Here is a typical legacy data flow for creating PK parameters leveraging NCA:

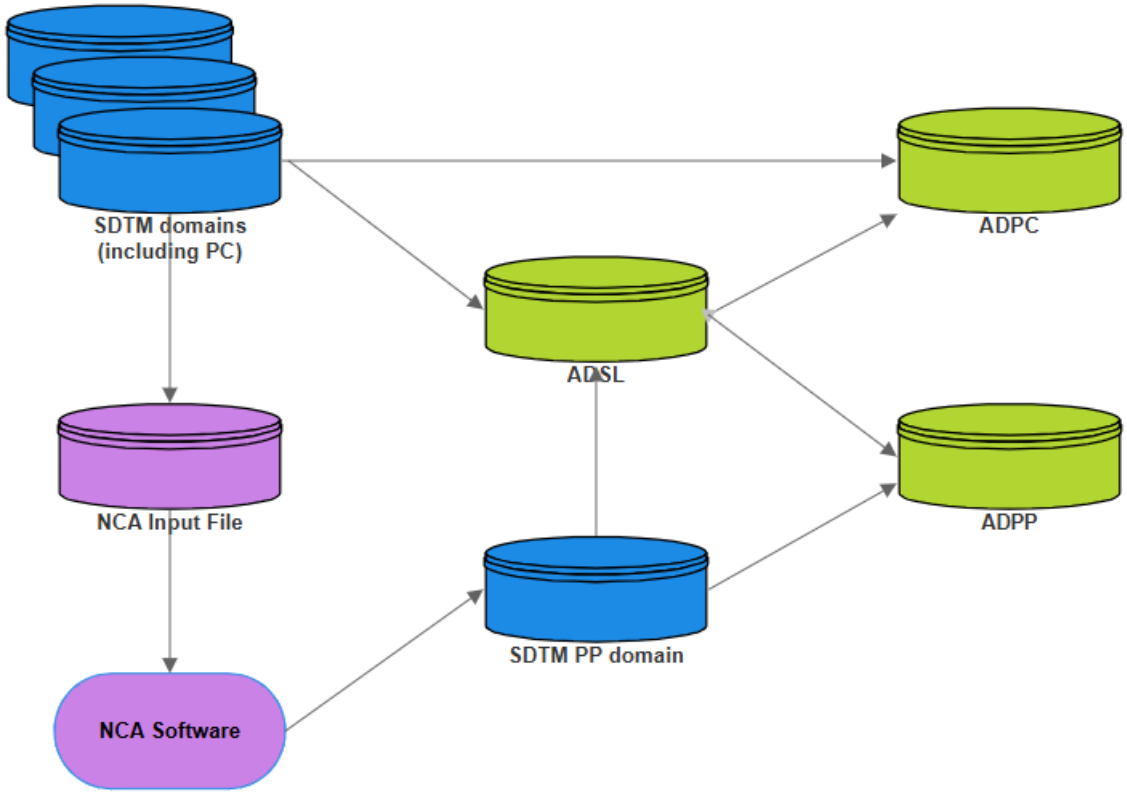

Figure 1. Legacy Data Flow Supporting PK Parameter Calculation

Up until now, the creation of this input dataset has been largely creative and heroic. Phoenix WinNonlin™ specifies specific data that they need but they leave the specifications for individual variables largely to the creator. As a result, even the most sophisticated PK analysis teams may have standardized some aspects of creating the input dataset but typically program certain items each time they endure this exercise, particularly in later stage trials where dosing can be more complex and identifying situations that exclude PK data from analysis due to complex situations in participants with the disease in question.

ADNCA – BENEFITS AND CHALLENGES

So, how do we streamline this process to make it more repeatable and reliable? Standardize the input dataset for the Non-Compartmental Analysis! Enter CDISC and the CDISC Analysis Data Model Team, which have given us the "Analysis Data Model Implementation Guide for Non-compartmental Analysis Input Data" [2]. This IG provides a standard for data organization and variables used leveraging the CDISC Analysis Data Model (ADaM) Basic Data Structure (BDS) dataset class.

CDISC's NCA Implementation Guide details the dataset, variable and value level metadata to support Non-compartmental Analysis. It clearly delineates what covariates are common between ADNCA and the subject-level ADaM analysis dataset, ADSL. It provides specifications and rationales for covariates that might have different definitions between ADSL and ADNCA, specifically the DOSEP, DOSEA, DOSEU

and AVISIT variables. It also includes a number of covariates and baseline characteristics specifically needed for ADNCA.

How else can ADNCA be useful? This dataset can be used interchangeably with the typical ADPC ADaM analysis dataset as the necessary variables are all present to support tables, figures and listings for observed concentration time data ([2], Section 3, 2nd paragraph). As a result, you can have a standard dataset to support NCA and, at the same time, eliminate the cost of producing ADPC.

Another benefit of ADNCA implementation is the reduced cost of feeding your NCA software with data. You now have a target standardized input, so you can establish this standard in that application and simply tweak it as needed for each use based on unique study characteristics.

Notice that, up until now, only mentioned SDTM had been mentioned as a data source, not ADaM, so how did we make this leap? Because this is the reality of the relationship between non-compartmental analysis data and other data within a clinical study. The legacy method of producing the input dataset for NCA mimics the efforts used to create these same values for inclusion in ADSL and ADNCA. It slips under the Good Programming Practice radar as these activities are typically carried out by two different work groups within an organization, potentially even different partner organizations operating in different environments under different SOPs / work practices.

Even CDISC has challenges wrestling with this topic. When it comes to establishing the origin of variables, they go on a long discourse ([2], Section 3, 4th paragraph):

> "When identifying the source dataset for a variable, the immediate predecessor is used. The dosing and subject-level datasets, among others, are common input for many ADNCA variables, if they are available. Dosing datasets may include SDTM exposure domains or a derived exposure dataset utilizing ADaM standards. If the subject-level analysis data (ADSL) is not available or is not a viable option for the purposes of ADNCA generation, use the applicable SDTM variables. Multiple CDISC source datasets may be used to populate ADNCA based on analysis need. Outside of the SDTM PC domain, ADaM sources are expected to be the most common but not the only sources used to create this dataset."

What confounds this even more is the consideration that the PK parameter determination process often produces covariates that are ultimately incorporated into ADSL, such as PCFL (included in the PK concentration data population) and PPFL/PKFL (included in the PK parameter data population). Historically this has not been an issue as the source of these flags has been the PP SDTM domain, which is created prior to ADSL in the legacy data flow (See Figure 1). So creating the ADNCA dataset is a prerequisite to completing ADSL if these flags are to be included. As a result, you will continue to create the majority of these covariates twice, once in support of ADNCA and then again in support of ADSL.

This diagram is from a paper earlier paper on the use of ADNCA in the analysis of clinical trial data [3]:

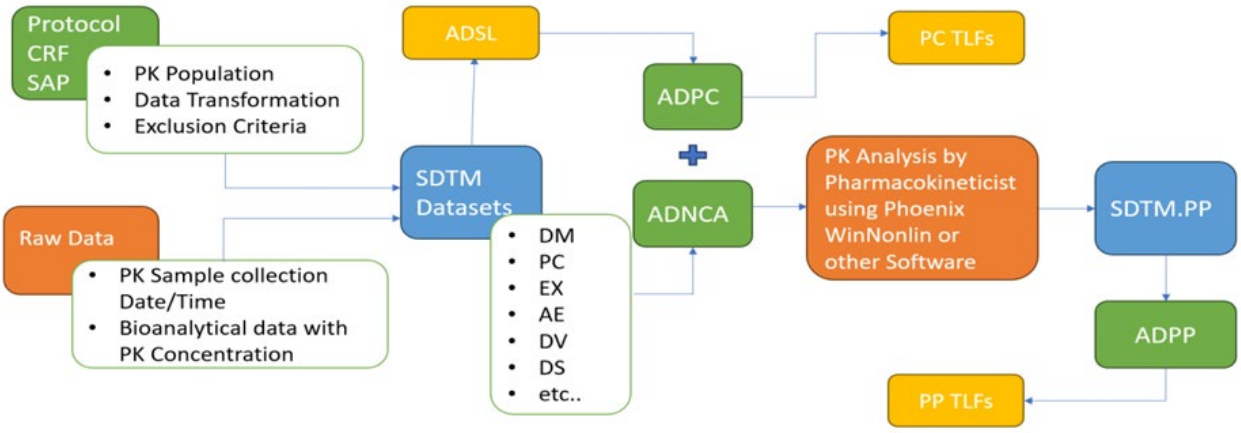

Figure 2. An Early Approach to ADSL with ADNCA Implementation

The flow chart clearly shows that ADSL and ADNCA follow their own paths for creation and do not try to leverage any solution that would have them eliminate duplicate work and potential programming differences between dataset implementations. It also dictates that you still develop ADPC and use it as a source for PK concentration tables, listings and figures, which the CDISC NCA Guidance states can be achieved using the ADNCA analysis dataset ([2], Section 3, 2nd paragraph).

How do you avoid this situation?

SOLUTION! MODULAR COVARIATE CREATION

Creating your covariates using a modular approach will allow you to leverage harmonized, optimistically standardized, programming to produce these variables / metadata / values that will satisfy both ADSL and ADNCA variable requirements.

Here is an updated flow chart emphasizing the use of common code to determine the value of covariates that are shared between ADSL and ADNCA:

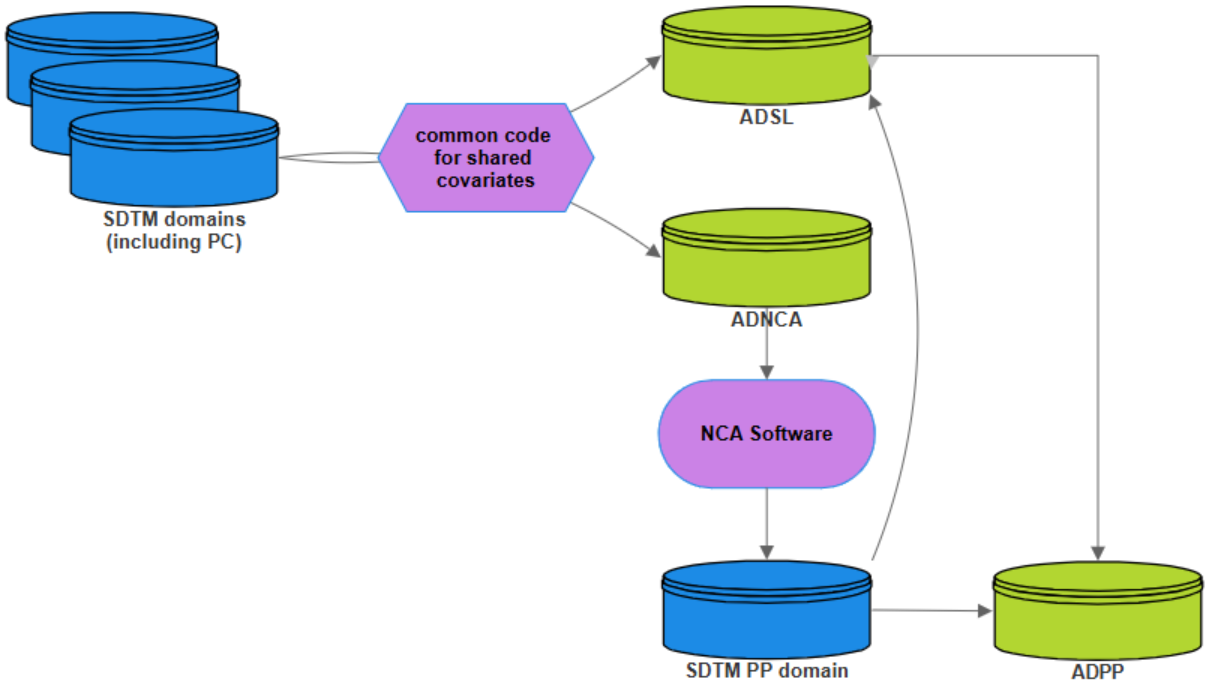

Figure 3. ADSL and ADNCA Produced Leveraging Modular Covariate Creation

Driver programs for ADSL and ADNCA might look like this:

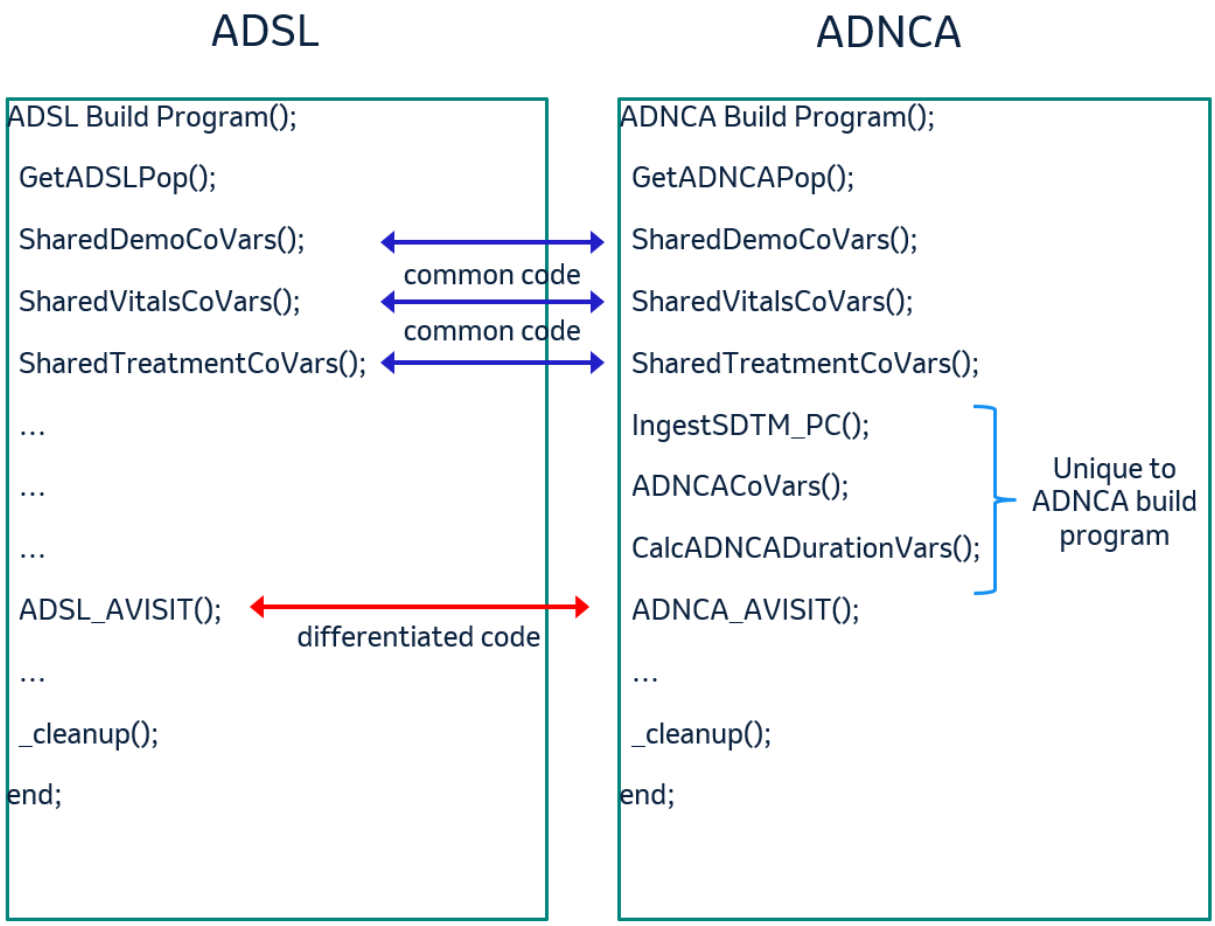

Figure 4. ADSL and ADNCA – High Level Program Flow

There are a number of benefits to this approach:

- Good Programming Practice is preserved! You develop and maintain a single set of programming statements that drive the determination of covariate values that are common across both datasets.
- The participants included in ADSL will typically be a superset of the participants in ADNCA, as ADNCA will only include subjects with available PK concentration data. Calling modular code segments allows you to determine covariate values for your population of interest dynamically.
- A standard driver program can be produced with calls to code to produce the new standard non-compartmental analysis variables that have been defined by the CDISC NCA guidance ([2], Table 4.2.1).
- In places where there are differences between ADSL and ADNCA for variables with the same name, such as the DOSEP, DOSEA, DOSEU and AVISIT variables ([2] Table 4.2.2), you have the opportunity to differentiate the programming needed to produce the variables in a controlled, traceable way.

CONCLUSION

The concept of modularizing covariate creation is fine in theory but definitely requires a healthy appetite for change if you were to consider a full implementation. We have been working for over a decade in a

model where PC plus other domains generates PP and each of these domains generates the companion ADPC and ADPP. Moving to ADNCA involves implementing a dataset that both drives non-compartmental analysis and supports PK concentration TLF generation; at larger organizations, these tasks might be performed by completely separate work groups. At minimum, the timeline to produce ADNCA is typically earlier than the full ADaM dataset production, yet these groups will need to work together to ensure that covariates are calculated correctly on an earlier schedule. And we should not forget that this ADNCA will need to be incorporated into the ADaM analysis dataset submission data package, therefore PK scientist will be directly contributing to the development of this deliverable where they might not have been before.

Despite these considerations, the benefits of modularizing the generation of covariates far outweigh these potential drawbacks as you will have consistent covariate values produced using a single set of code, thereby preserving Good Programming Practice and variable traceability in the eyes of regulators and in the quality of the results you produce.

REFERENCES

[1] Quantics Biostatistics, "Non-Compartmental Analysis (NCA): The Basics of PK/PD Statistics". Accessed March 19, 2025. Available at https://www.quantics.co.uk/blog/nca/#:~:text=Non-compartmental%20analysis%20%28NCA%29%20computes%20key%20pharmacokinetic%20parameters%E2%80%94such%20as,AUC%20%28area%20under%20the%20concentration-time%20curve%29%2C%20and%20half-life.

[2] CDISC Analysis Data Model Team. "Analysis Data Model Implementation Guide for Non-compartmental Analysis Input Data, v1.0". Published November 29, 2021. Available at https://www.cdisc.org/standards/foundational/adam/adamig-non-compartmental-analysis-input-data-v1-0.

[3] Tripathi, Saumilkumar. "Unveiling the Potential of ADNCA by CDISC ADaM IG: Revolutionizing Pharmacokinetic Non-Compartmental Analysis Input Data Standardization", *Proceedings of PHUSE US Connect 2024*, Bethesda, MD, USA. Available at https://www.lexjansen.com/phuse-us/2024/si/PAP_SI04.pdf.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Merck and specifically to Janet Low, Leader of the Submission Consultants at Merck and my fellow Submission Consultants for supporting this paper.

RECOMMENDED READING

- *Analysis Data Model Implementation Guide for Non-compartmental Analysis Input Data, v1.0 – available via CDISC website, direct link in References above.*

CONTACT INFORMATION

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

David C. Izard
Principal Submission Consultant
Biostatistics and Research Decision Sciences (BARDS)
Merck & Co., Inc., Rahway, NJ, USA
david.izard@merck.com
https://www.linkedin.com/in/david-izard-4105625/