

## Unravel the mystery around NONMEM data sets for Statistical Programmers

Sabarinath Sundaram, Seagen Inc., Bothell WA  
Lyma Faroz, Seagen Inc., Bothell WA

### ABSTRACT

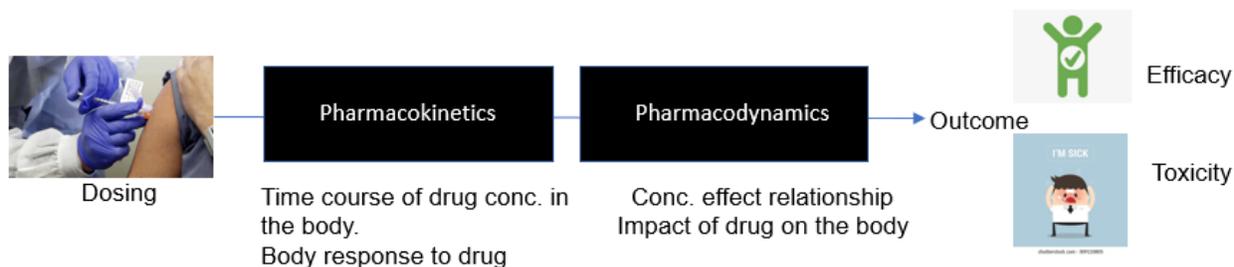
Having a deeper knowledge of the relationships between pharmacokinetics (PK)/ pharmacodynamics (PD) and exposure-response effects is crucial to product development and drug registration. Nonlinear mixed-effects analysis is typically adopted for any repeated measurements or longitudinal data analysis such as PK and PKPD using a software called Nonlinear Mixed Effects Modeling (NONMEM®). A high-quality NONMEM data set is essential for this important decision-making modeling. Statistical programmers must have a solid understanding of PK and PD data associated with the study drug components and the study design. For a NONMEM data set, statistical programmers need to input dosing information, scheduled timepoints, PK data, PD data (labs, response, biomarker, and adverse events) as assigned by the PK scientist, as well as covariant information like demographics, baseline laboratory results, and vital signs and disease characteristics among others. To perform data-driven and model-based updates, a thorough understanding of the PK/PD data is required along with a high degree of collaboration with PK scientists. This paper will discuss the basics for creating NONMEM data sets and some pointers on using the information from various sources to create a NONMEM data set that is different from traditional CDISC-compliant data sets. Additionally, this paper introduces details about standard variables that are expected in such data sets.

### INTRODUCTION

Clinical trials are conducted to assess the treatment efficacy and toxicity. Once the patient gets dosed, something happens, and we get results. The two main black boxes (Figure 1) need to be dissected in detail to understand more about the drug and endpoint relationship.

**Pharmacokinetics (PK)** is how the body responds to a drug, and it describes the drug's exposure by characterizing absorption, distribution, bioavailability, metabolism, and excretion as a function of time. To reduce drug attrition, PK analysis is essential to fully investigate the pharmacokinetics of a new drug.

**Pharmacodynamics (PD)** is the impact of the drug on the body, and PD analysis allows us to quantify the relationship between the drug dose and the pharmacologic or toxicologic effect it has on patients. It describes drug response in terms of biochemical or molecular interactions. Typically, a pharmacometrics model will encompass the results from pharmacokinetic-pharmacodynamic (PKPD) modeling to make inferences on optimal dosing for clinical trials practice. PK and PD behavior is expected to vary from patient to patient based on the physiological properties (covariates) such as Body Mass Index (BMI), age, sex, disease states, organ function (such as renal, hepatic, or cardiac), and genetic polymorphism.



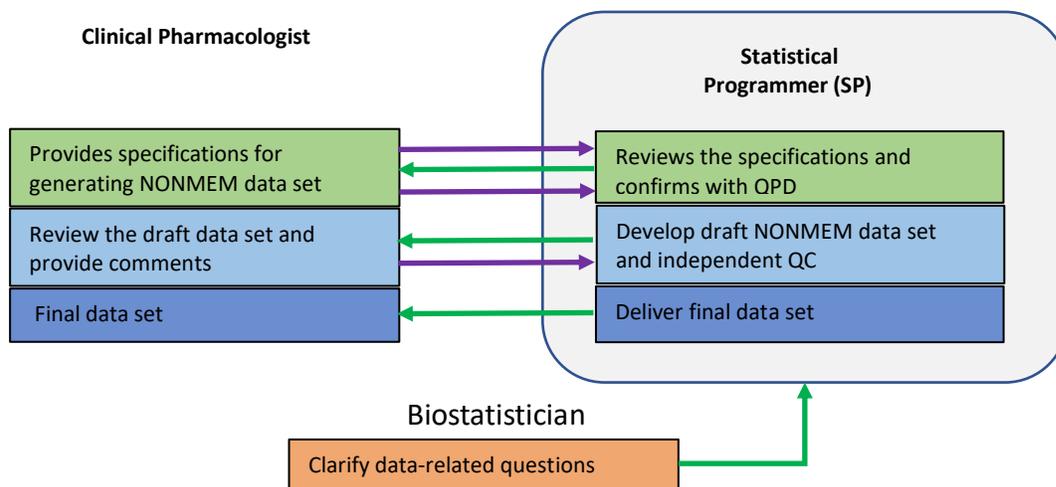
## Figure 1: Overview of the PK and PD in a Clinical Trial

The efficacy and toxicity profiles of many drugs (particularly new drugs) are mixed, complicating decision-making. A population-based PK/PD analysis enhances the decision-making capabilities of clinical trials. It can provide insight into how patients respond to new drugs, as well as help identify drug candidates that may not meet target clinical therapeutics goals early in development. The ability to progress timely and well-informed development translates to major cost savings and gets drugs in the hands of patients in need faster!

Understanding the PK, PKPD, and exposure-response effect relationships (PKPD) is key to the development and approval of every drug registration. The characterization of PK and the evaluation of potential impact of intrinsic/extrinsic factors, patient disease, and co-medication characteristics on PK via population PK analysis help to identify potential factors for dose adjustments. The characterization of PD and PKPD aids the understanding of patient selection and optimal dose selection, e.g., ensure target engagement. Apart from PK and PKPD, exposure-response analysis on efficacy and safety parameters is performed to optimize dose selection. Nonlinear mixed-effects analysis is typically adopted for any repeated measurements or longitudinal data analysis such as PK and PKPD using software called nonlinear mixed effects modeling (NONMEM®). The population PK approach combined with pharmacodynamics modeling allows integrated analysis, interpretation, and prediction of the drug's safety, efficacy, dose-concentration relationship, and dosing strategy. Population PK analysis is expected for drug registration FDA guidance on population PK (<https://www.fda.gov/media/128793/download>). Creating the data set structure required by the NONMEM® software is often challenging.

## NONMEM DATA PROGRAMMING

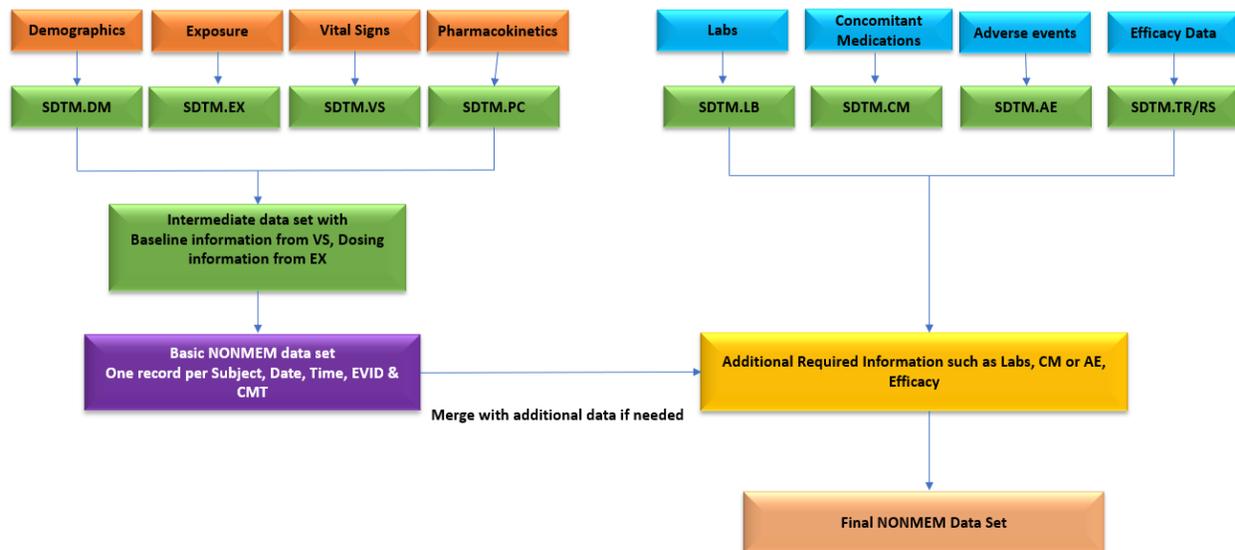
NONMEM data programming is the process of deriving a NONMEM-ready data file based on clinical source data files, CDISC data sets such as SDTM and ADaM. The data definition table (DDT), the specification created by the clinical pharmacologist, should be reviewed thoroughly by the programmer to ensure that the required information is available in SDTM or ADaM data sets. We have standardized this process for efficiency and accuracy of the NONMEM data set. The DDT contains detailed information about how the NONMEM data file shall be derived and consists of a complete list of required variables and their definitions. The DDT serves as the programming specification and to clarify the data file for the reviewers. [Figure 2](#) illustrates collaboration and information exchange between various functional teams.



**Figure 2: Workflow of the Cross-functional Teams in Creating NONMEM Data Sets**

## NONMEM INPUT DATA SET FLOW CHART

The study design and NONMEM-specific analysis requirements result in incorporation of various baseline variables/covariates (demographics, vital signs, exposure, biomarker data, etc.). This helps in creating the basic NONMEM data set. Additional data records such as baseline labs, concomitant medications, adverse events, etc. can also be added if requested by the clinical pharmacologist for additional analysis. The clinical pharmacologist may request a NONMEM data set for an individual study or based on integrating several studies within a product.

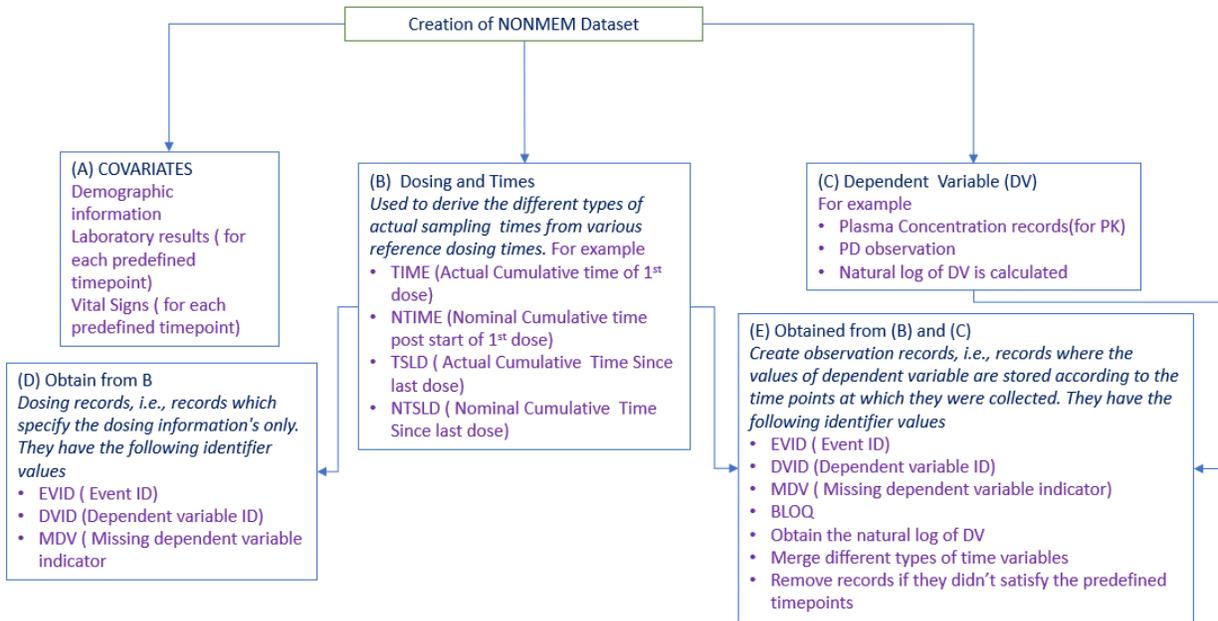


**Figure 3: NONMEM Input Data Set Flowchart**

## DATA SET STRUCTURE

The NONMEM input data set typically has one row per study, per subject, date, time, event ID, and compartment flag. Additional observation types, e.g., PD, will be added as rows with TYPE variable as identifier while covariates of interest will be included as variables (columns). Three basic components are dosing records, PK/PD results, and covariates (PD baseline and/or demographics covariates) as detailed in [Figure 4](#):

- 1) Dosing information such as planned and actual dose along with actual and nominal times of dosing.
- 2) PK data (dependent variables) such as serum or plasma investigational product concentrations, metabolite concentrations and target concentrations with actual and nominal times of measurement and an indicator if measures are below the quantifiable limit (BQL).
- 3) Baseline weight information and other demographic information such as age, race, gender, ethnicity.
- 4) Baseline PD measures (biomarkers and/or efficacy) may also be appropriate.



**Figure 4: Creation of NONMEM Data Set**

Case 1 is a typical example from the NONMEM PK/PD data set where one can see PK parameters (like ADC PK concentration, MMAE PK concentration), PD parameters from both lab data (like Glucose concentration, Absolute Neutrophils concentration) and response data (like tumor size).

	NTIME	TSLD	NTSLD	EVID	CMT	TYPE	DV	MDV	D/INM	UNIT	AMT	DUR	RATE	DOSE	COHT	BLQ
1	0	0	0	0	0	1 ADC PK concentrations	0.25	0	0.347947112	ng/mL	.	.	.	1.0 mg/kg	Dose escalation	1
2	0	0	0	0	0	3 MMAE PK concentrations	5	0	0.0069589422	pg/mL	.	.	.	1.0 mg/kg	Dose escalation	1
3	0	0	0	0	0	4 Glucose concentration	236	0	.	mg/dL	.	.	.	1.0 mg/kg	Dose escalation	0
4	0	0	0	0	0	5 Absolute Neutrophils concentrati.	3.77	0	.	10 <sup>3</sup> /L	.	.	.	1.0 mg/kg	Dose escalation	0
5	0	0	0	0	0	6 Total Antibody PK concentrations	10	0	0.0648508431	ng/mL	.	.	.	1.0 mg/kg	Dose escalation	1
6	0	0	0	0	0	7 Longitudinal tumor size	73	0	.	mm	.	.	.	1.0 mg/kg	Dose escalation	0
7	0	0	0	1	1	Dosing records	.	1	.	mg/kg	95	0.55	172.727	1.0 mg/kg	Dose escalation	0

**Case 1: PK/PD Data Emphasizing the TYPE Variable**

**Case 2** is a typical example from the NONMEM PD data set where one can see only PD parameters from the adverse event dataset. This also shows the TIME variable should be arranged in chronological order. To calculate the TIME variable for AEs as a PD parameter, we use only the date part (TIME=AE start date – dosing date).

ID	DRN	RATE	RATEMOLHR	TIME	DV	MDV	EVID	CMT	MIT1	MIT2	AEDECOD
10015	0.6666666667	240	1.578	0	.	1	1	1	.	0	
	0.7	218.57142857	1.4371428571	28	.	1	1	1	.	0	
	0.7	217.14285714	1.4285714286	49	.	1	1	1	.	0	
	.	.	.	56	2	0	0	3	.	0	Vision blurred
	.	.	.	70	2	0	0	3	.	0	Blepharitis
	.	.	.	70	2	0	0	3	.	0	Dry eye
	.	.	.	70	2	0	0	3	.	0	Meibomian gla...
	0.6833333333	220.97560976	1.4546341463	79	.	1	1	1	.	0	
	0.65	240	1.5784615385	100	.	1	1	1	.	0	
	.	.	.	121	2	0	0	3	.	0	Conjunctivitis b.
	.	.	.	127	1	0	0	3	.	0	Dry eye
	.	.	.	127	1	0	0	3	.	0	Ectropion
	.	.	.	134	1	0	0	3	.	0	Arcus lipoides
	.	.	.	134	2	0	0	3	.	0	Eye irritation
	.	.	.	134	2	0	0	3	.	0	Eye pain
.	.	.	134	2	0	0	3	.	0	Lacrimation inc...	
.	.	.	134	1	0	0	3	.	0	Photophobia	
.	.	.	134	1	0	0	3	.	0	Pinguecula	
.	.	.	150	1	0	0	3	.	0	Vision blurred	

## Case 2: PD Data with Additional AE Information

Case 3 is a typical example from the NONMEM PK data set where one can see PK parameters only. To calculate the TIME variable for AE as a PK parameter, we use only datetime as evident from the TIME variable values. (TIME=PK blood collection datetime – dosing datetime)

TRT	CYCL	TIME	NTIME	TSLD	NMID	AMT	DUR	RATE	DV	UNIT	DVNM	TYPE	CMT	FLAG	MDV	EVID
2	1	0	0	0					0.25	ng/mL	0.347947...	ADC PK conce...	1	1	0	0
2	1	0	0	0		95	0.55	172.727...		mg/kg		Dosing records	1	0	1	1
2	1	0.0299	0.0417	0.0299					523	ng/mL	727.9053...	ADC PK conce...	1	1	0	0
2	1	0.0993	0.0833	0.0993					471	ng/mL	655.5323...	ADC PK conce...	1	1	0	0
2	1	0.1861	0.125	0.1861					443	ng/mL	616.5622...	ADC PK conce...	1	1	0	0
2	1	0.9639	0.1667	0.9639					258	ng/mL	359.0814...	ADC PK conce...	1	1	0	0
2	1	1.9361	3.2083	1.9361					158	ng/mL	219.9025...	ADC PK conce...	1	1	0	0
2	1	8.0785	7	0					24.3	ng/mL	33.82045...	ADC PK conce...	1	1	0	0
2	1	8.0799	7	8.0799						mg/kg		Dosing records	1	0	1	1
2	1	8.1056	7.0417	0.0257					618	ng/mL	860.1252...	ADC PK conce...	1	1	0	0
2	1	14.0569	14	0					54.2	ng/mL	75.43493...	ADC PK conce...	1	1	0	0
2	1	14.0583	14	14.0583		95	0.55	172.727...		mg/kg		Dosing records	1	0	1	1
2	1	14.0833	14.0417	0.025					648	ng/mL	901.8789...	ADC PK conce...	1	1	0	0
2	1	21.025	21.25	6.9667					51.5	na/mL	71.67710...	ADC PK conce...	1	1	0	0

## Case 3: Records in Chronological Order

### VARIABLES EXPECTED IN NONMEM DATA SET

The following variables are required in a NONMEM data set for modeling purposes which are different than any CDISC variables. There are no standards that govern the use of these variables and hence the discretion lies with the PK team within a company. Depending on the model, they can use different values based on various covariates and the study design. Here we have listed some of the variables that are expected in a NONMEM data set, however it is always advised to confirm the derivation with the end user of this data set, in this case the clinical pharmacologist.

#### 1) ID - Unique Patient ID Number

This variable is used to identify information that belongs to a certain patient. For example: if multiple study data sets need to be pooled to create the NONMEM data set.

#### 2) DV - Dependent Variable

Value of an observation in standard units. Depending on what the model needs, this variable value changes. For dosing records or missing data, enter as "."; for other observations, enter the numeric result value.

#### 3) MDV - Missing Dependent Variable

This variable identifies input data set records that do not have values for an observation.

- The MDV value of 0 is assigned for observation records, the dependent variables to be modeled. The DV value of the record is to be used in the analysis.
- The MDV value of 1 is assigned to non-observation records such as dosing records and dummy records. This value would be appropriate to evaluate the prediction value at a non-observation TIME or for the DV values of this record to be excluded from the analysis.

#### 4) EVID - Event ID (Event Identification Data Item)

This variable helps to identify the event described by every record within the data set. For example, categories may include Observation (EVID = 0) and Dosing records (EVID=1). If there are cross-over subjects, then different EVIDs (3, 04, 4) are used depending on the situation.

#### 5) AMT - Amount Dosing Records

The amount of a dose is displayed using this variable. It should be a positive number. The actual total dose administered to a patient in units mentioned in the study SAP (Statistical Analysis Plan)

can be entered. For dosing records, the actual dose value can be used, whereas for other records this is set to missing.

**6) RATE - Rate of Infusion**

For other routes of administration such as oral, IV (intravenous) bolus injection, or subcutaneous injection, set it to 0 or as discussed with the PK scientist. It is displayed only for dosing records and set to missing for other records.

**7) CMT - Compartment Number**

Usually populated for drugs delivered via infusion and not for orally administered drugs. The following example is for drugs delivered via infusion. This value is assigned for different PD/biomarker records, e.g., set to 1 for dosing records, such as PK analyte, else set to 2 for other covariates.

## NONMEM DATA SET TIME VARIABLES

**1) Time - Actual Cumulative Time of First Dose**

This is the actual cumulative time from the first dose day of the study. For missing sample collection time, imputation would be needed by adding the nominal time (NTSLD) to the last dosing time.

**2) NTIME - Nominal Cumulative Time Post the Start of First Dose**

This is the nominal cumulative time from the start of the first dose day of the study. This information comes from the protocol. In general, we will focus only on scheduled visits. In rare cases, we may include NTIME for unscheduled visit where we set NTIME as 999. Confirm with the PK scientist before assigning 999 for this variable.

**3) TSLD – Actual Cumulative Time Since Last Dose in Days**

This is the actual cumulative time from the last dosing time. For records prior to the first dose, set it to 0. This is calculated as (actual datetime of record – datetime of start of infusion of last dose) in days.

**4) NTSLD – Nominal Cumulative Time Since Last Dose in Days**

This is the nominal cumulative time from the last day of the dose of the study. It is a protocol-defined nominal time from the last dosing time.

## HANDLING OF MISSING DATA

If there are any records with missing data, it can cause an issue while analyzing the data and hence programmers should notify the PK representative and get guidance about handling the missing data information. There are various considerations for missing data by data type:

- Missing dosing information such as actual amount or dosing time will not be included for analysis. No imputation would be done. Subsequent PK samples cannot be included for analysis unless steady state is reached again. By default, concatenate a character “c” and NONMEM ID.
- Missing observation records and sampling time will be excluded from the data set. No imputation would be done.
- Missing baseline covariates should be set to -99 (or any other pre-agreed value that would not be representative of the variables).
- Missing time-dependent variables.

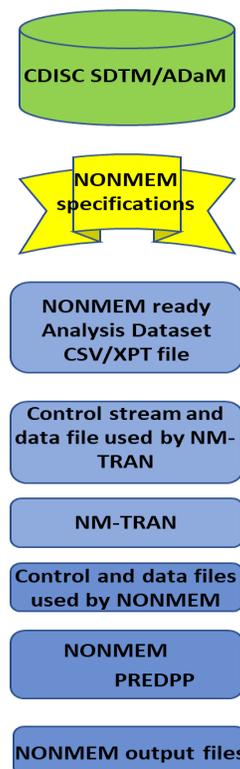
This information should be defined in collaboration with the clinical pharmacologist for the study and documented in the specs. Various imputation methods such as LOCF (Last Observation Carried Forward), WOCF (Worst Observation Carried Forward), FOCB (Forward Observation Carried Backward),

and COA (Closest Observation Algorithm) can be used if necessary.

## POST PROCESSING STEPS AFTER CREATING NONMEM DATA SETS

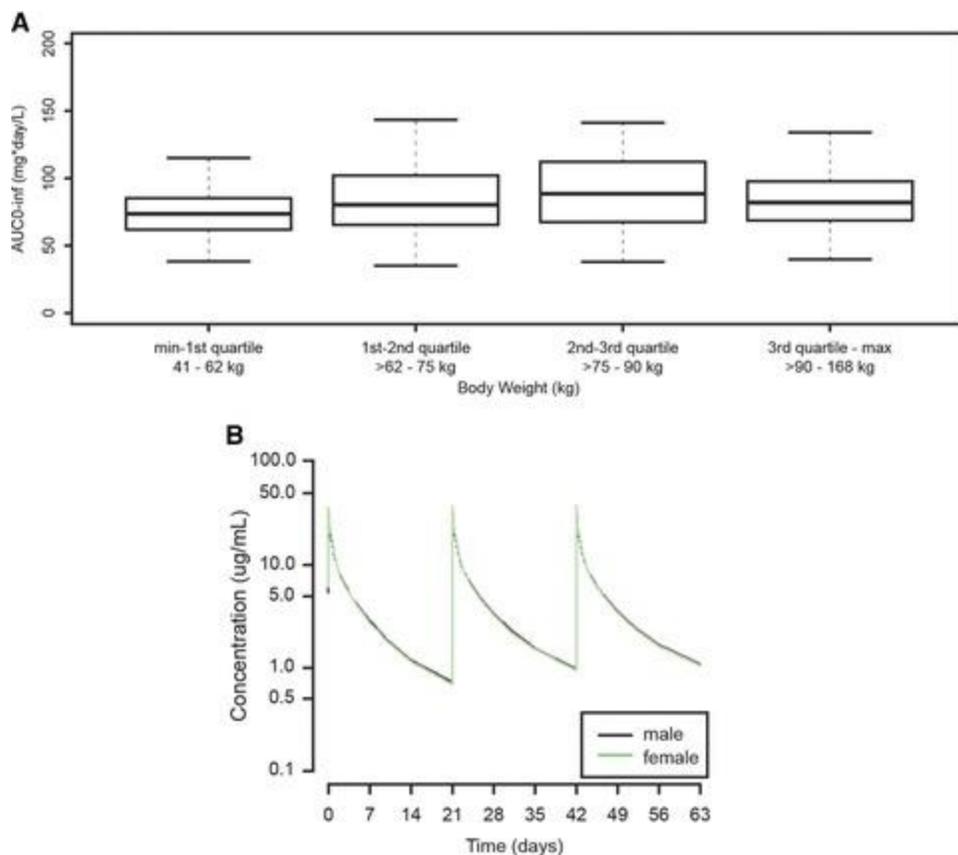
Understanding the post processing steps helps programmers have better knowledge about the NONMEM data set. NONMEM® software used for the PK/PD modeling consist of three components:

- NM-TRAN (NonMem TRANslator) which converts the data file and the control stream into FORTRAN code files for use by NONMEM
- PREDPP (PRED for Population Pharmacokinetics), a very powerful package of subroutines, handling population PK data as well as general linear and nonlinear models, which can free the user from coding standard kinetics type equations while simultaneously allowing complicated patient-type data to be easily analyzed
- NONMEM® (NONlinear Mixed Effects Modelling) uses various statistical methodologies to provide estimates of parameters (and their statistical uncertainties) that result in best fit of the model predictions to the observed data. Using NONMEM, one can optimally fit parameters of a mathematical model to observed patient data. In a PK setting, the mathematical model would be a set of equations the pharmacometrician believed would best describe how the concentrations of drug changes over time in a patient's body.



**Figure 5: Workflow of NONMEM Operating Environment**

The Figure 6 is an example output from population pharmacokinetic analysis, which was conducted to characterize the PK of ADC and unconjugated MMAE in patients with CD30-expressing hematologic malignancies to evaluate the effects of covariates on PK of the ADC. Figure 6(A) refers to the ADC exposure across weight quartiles. Figure 6(B) illustrates the model-predicted typical patient ADC concentration-time profiles for male and female patients.



**Figure 6:**

**(A) Model-predicted ADC exposure (AUC<sub>0-inf</sub>) for patients (n = 314) receiving brentuximab vedotin administered intravenously at 1.8 mg/kg every 3 weeks with the dose capped at 180 mg. The box-and-whisker plots indicate minimum and maximum values (lower and upper whiskers), median (heavy band), and first and third quartiles (bottom and top of box). (Adapted from Li et al., 2017)**

**(B) Model-predicted typical patient ADC concentration–time profiles for male (black line) and female (green line) patients receiving brentuximab vedotin administered intravenously at 1.8 mg/kg every 3 weeks. ADC, antibody-drug conjugate. (Adapted from Li et al., 2017)**

## CONCLUSION

Statistics programmers create data sets that support analyses which help interpret the outcome of clinical trials on investigational products. The quality and accuracy of the NONMEM data is very critical. Basic knowledge about Pharmacokinetics and Pharmacodynamics plus understanding the clinical trial design and background knowledge in handling lab data are always an added advantage for the statistical programmers working on creating NONMEM data. Here one can visualize the complete patient profile from data. In addition, extensive cross-functional collaboration between Statistical Programmers, Biostatisticians, Clinical Pharmacologists, and Medical Monitors is needed to understand the modeling purpose, which helps provide the required data in NONMEM data set format. As we have seen throughout the paper that there are no standards yet for creating this data from CDISC or FDA, however it is in our best interest to use data in standard formats such as SDTM or ADaM for creating this NONMEM data set. For example, if a study needs frequent updates to NONMEM data sets at an early stage, then that can be incorporated in shorter time span by using SDTM data sets. If modeling requires derived data, ADaM data sets can be used which as well can help maintain traceability and improve efficiency-

In our company, we are standardizing the flow of PK/PD (NONMEM)-related deliverables. Standard data sets are used to create the NONMEM data set by following the Cross functional DDT specification, followed by QC of the generated NONMEM data set through independent programming. We create a traceability document containing information about data source (from SDTM or ADaM) to NONMEM, conduct review by a subject matter expert within Programming or Biostatistics, and cross-check the structure of the NONMEM data set before delivering it to Clinical Pharmacology. We plan to share our experience in a future PharmaSUG!

## REFERENCES

**NONMEM® Tutorial Part I: Description of Commands and Options, With Simple Examples of Population Analysis:** <https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.12404>

**NONMEM® – A Programmer point of view:** <https://www.lexjansen.com/phuse/2014/pp/PP12.pdf>

**Li et al., Population pharmacokinetics of brentuximab vedotin in patients with cd30-expressing hematologic malignancies:** [J CLIN PHARMACOL](#), 2017 SEP; 57(9): 1148–1158.

## ACKNOWLEDGMENTS

We would like to thank Boxun Zhang, Bala Pitchuka, and Shefalica Chand, and Michiel Hagendoorn for their valuable feedback and constant support and guidance.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Sabarinath Sundaram  
Seagen Inc.  
21823 - 30th Drive S.E.  
Bothell, WA 98021  
[ssundaram@seagen.com](mailto:ssundaram@seagen.com)

Lyma Faroz  
Seagen Inc.  
21823 - 30th Drive S.E.  
Bothell, WA 98021  
[lfaroz@seagen.com](mailto:lfaroz@seagen.com)