PharmaSUG 2015 - Paper SP07

Growing Needs in Drug Industry for NONMEM Programmers Using SAS®

Sharmeen Reza, Cytel Inc., Cambridge, MA

ABSTRACT

Nonlinear Mixed Effects Modeling (NONMEM) is a type of population pharmacokinetics/pharmacodynamics (pop-PK/PD) analysis that is used in Clinical Pharmacology research. At different stages of drug development the population PK approach, coupled with PD modeling, allows integrated analysis, interpretation, and prediction about drug safety, efficacy, dose-concentration relationship, and dosing strategy. Analysis reports help regulatory agencies evaluate new drug submission, review safety/effectiveness of a drug, and guide drug labeling. Pharmacologists work with NONMEM® software which requires pop-PK/PD data in text (ASCII/CSV) format, whereas agencies request data be submitted in SAS transport files. SAS is used to create, maintain, update, and recreate data sets required for modeling purposes, and facilitate the creation of regulatory format as well as the text file. Utilizing SAS also assists in maintaining data integrity, handling large data, tracking data manipulation and derivation through log. The use of SAS for the creation of pop-PK/PD analysis data sets that are consumed by NONMEM® software has led to a greater demand for specialized 'NONMEM' programmers. These professionals are equipped to pool data from multiple clinical studies, manipulate data coming in diverse formats, combine and validate records in a single SAS output data set. This paper explains a NONMEM data set structure, some core variables, group interaction, and a programmer's tasks and challenges involved.

INTRODUCTION

Nonlinear Mixed Effects Modeling (NONMEM) is a statistical approach commonly used for population pharmacokinetics/pharmacodynamics (pop-PK/PD) analysis by Clinical Pharmacologists to assess drug safety, efficacy, and dose-concentration relationship in target populations (CHMP, 2006). PK samples and dosing records are the key components of NONMEM and therefore included in the pop-PK/PD analysis data set (hereafter also referred to as NONMEM data set); often lab measurements, PD endpoints are added to obtain exposure-safety or exposure-efficacy information. Subject characteristics and/or treatment-specific factors can be potential determinants of variability in PK/PD model parameters. Thus covariates such as body weight, age, sex and certain health conditions in populations that impact outcomes and help determine dosing strategy are added in a NONMEM data set too.

Pop-PK/PD reports in recent years have been an integral part of submissions to regulatory agencies and results are used in evaluating new drugs, decision making, drug labeling, etc. (HHS, 1999). Agencies like Food and Drug Administration (FDA) require data in transport files generated from SAS data sets (Pang, 2005). Since population analyses are usually performed on large volume of data, preparing data sets using SAS has become a norm because it is efficient in data handling, processing and memory management (Crevar, 2007). Formalities in submission such as traceability of data, reproducibility of outputs, and quality documentation are well served using SAS.

Over the years SAS clinical programmers have been focusing and specializing in pop-PK/PD data set creation, understanding the exact nature of analyses, pooling input data sources, keeping track of group dynamics and challenges involved. For the purpose of data set development, input data might come in different formats. A NONMEM programmer has to collaborate with numerous groups, create programming specs, write code, derive variables using rules and algorithms provided by Clinical Pharmacology. Challenges arise when rules do not coincide with data scenarios, thereby causing frequent spec updates. Data refreshes also make reproduction, debugging, and revalidation time consuming especially when multiple studies or input sources are combined in a single NONMEM data set (Bonate et al., 2012).

The industry's rapid growth and capitalization of population based analysis has led to the need for SAS programmers who understand the complexity and requirements of data sets used in NONMEM analysis (NONMEM Programmers). Being part of the clinical environment, these professionals work with population PK and PD parameters. They are also referred to as Pop-PK/PD programmers which distinguish them from individuals using NONMEM® software.

CONSUMERS OF NONMEM DATA SET

In the clinical research industry, NONMEM data sets are used at different stages of the drug development process. Mainly the Clinical Pharmacology department, also known by Quantitative Pharmacology (QP) or Pharmacokinetics and Drug Metabolism (PKDM) in some organizations/companies, use NONMEM data sets for modeling purposes to evaluate dosing scheme and its impact of variability in the population of interest. In PKDM, in addition to PK scientists there might be pharmacometricians or QP analysts playing key roles in the analysis and model development. It is

worthwhile to note that the above roles are sometimes interchangeable, meaning 'analyst' refers to all roles, and tasks might be performed by multiple individuals.

Population modeling and simulation can be done using any or a combination of the following software:

- NONMEM®
- S-Plus
- R (open-source version of S-Plus)
- Phoenix WinNonLin
- SAS (PROC NLMIXED)
- Phoenix NLME
- WinBUGS
- PKBUGS
- JMP
- XPOSE
- MATLAB

Figure 1 below displays NONMEM data generated by a SAS program, are fed to NONMEM® processor outputting summary run and ASCII files (Fisher et al., 2007). Programmers use SAS to create the NONMEM® software-ready files, whereas analysts do further processing for report generation. This setup is common in a clinical environment in preparation for pop-PK/PD analyses (NMuser, 2004). All components in the diagram, except for processors, are part of a regulatory submission.

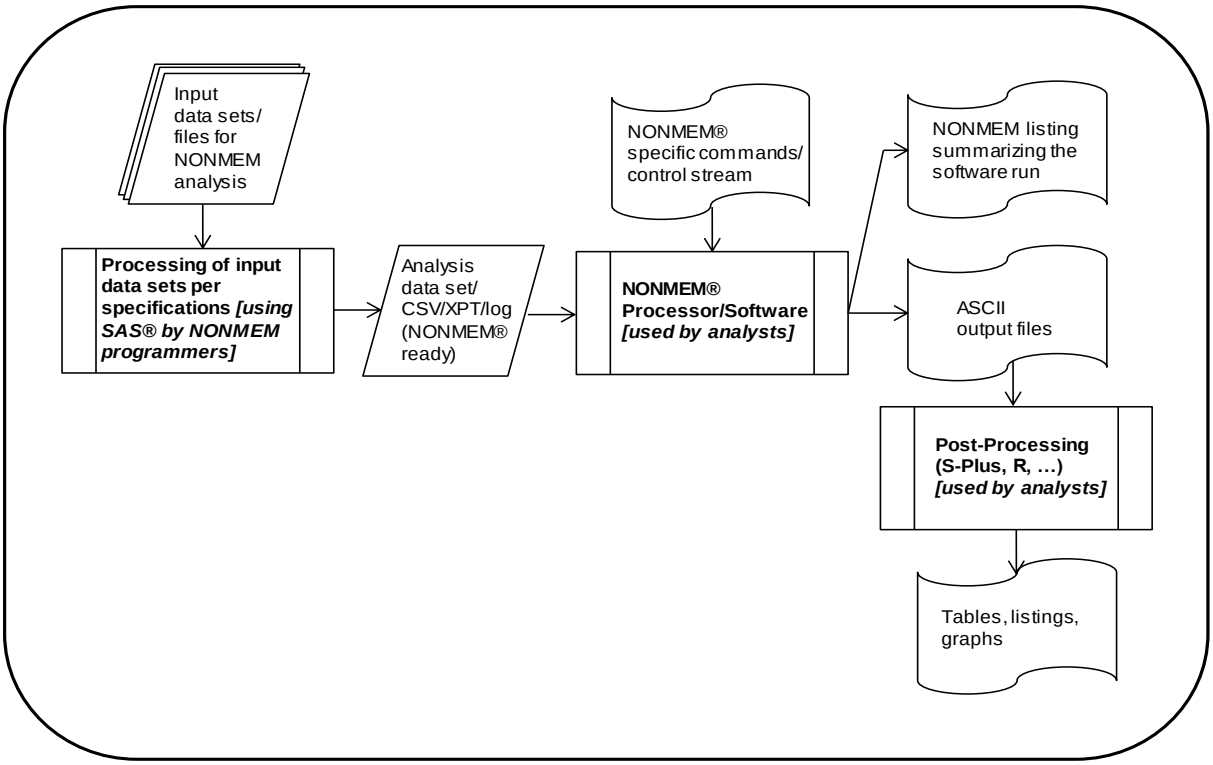

Figure 1. NONMEM analysis

NONMEM DATA SET STRUCTURE, INPUT SOURCES, CORE VARIABLES

Pharmacokinetics (PK) describes drug exposures in a subject over time. It evaluates what the body does to the drug by analyzing drug absorption, distribution, metabolism and elimination. Pharmacodynamics (PD), on the other hand,

checks what the drug does to the subject. In other words PD deals with the relationship between drug concentration and response, whether related to efficacy or adverse safety events. These are two main aspects of the NONMEM data set. Other components of NONMEM data set, like gender, weight, and lab results, help characterize sources of variability in pop-PK/PD analyses (Ette et al., 2013).

Analysis data sets for NONMEM are required to be of a certain structure usually dictated by the type of data available and sorts of models to be investigated. A NONMEM data set primarily contains dosing and concentration records. Lab observations (obs) are regularly added in order to perform complete safety profiles. PD is another component stacked vertically in the data set coming from diverse sources such as efficacy and biomarker measurements. As per modeling requirements, “reset” and “initialization” records are generated and appended as well (Boeckmann et al., 2011). Covariates from demographics, lab, concomitant medications, physical measurements, subject disposition, antibody, and adverse events are incorporated as variables/horizontal fields. Covariates attempt to explain the variability in the target population and can be time varying, time independent, categorical and/or numeric continuous in nature (Shen et al., 2007).

Table 1 identifies required and optional NONMEM data set components.

Components in NONMEM data set	**Description**
Dosing	Required. Vertical obs.
PK concentration sample	Required. Vertical obs.
Lab	Often added. Vertical obs.
PD (biomarker, efficacy)	Often added. Vertical obs.
Reset	Often added. Vertical obs.
Initialization	Often added. Vertical obs.
Demographics	Often added. Horizontal obs as covariates. Examples: Time varying: Age, Weight Time independent: Sex Categorical: Sex, Race Numeric: Age, Weight, Height
Lab	Often added. Horizontal fields as covariates. Example: Fasting, Baseline Creatinine Clearance
Concomitant medications	Often added. Horizontal fields as covariates.
Physical measurements	Often added. Horizontal fields as covariates. Example: Height
Antibody	Often added. Horizontal fields as covariates. Example: Neutralizing Antibodies
Subject disposition	Often added. Horizontal fields as covariates. Example: Baseline age calculation from enrollment and birth date.
Adverse events	Often added. Horizontal fields as covariates.

Table 1. NONMEM data set components

A NONMEM data set might consist of multiple studies/trials ranging from phase 1 to post-marketing surveillance (phase 4) of clinical drug development (Shen et al., 2007). Thus, the volume of data, i.e., number of observations could be huge, requiring long processing time. SAS is well suited to handle large data while utilizing simple programming steps (Fraeman, 2008).

A NONMEM programmer receives input data sets in various formats. The following lists some examples:

- Raw data set from Clinical Data Management (CDM): ipadmin, pk, demog, conmeds, lab, physmeas, ab, ae
- Raw excel file from biomarker (lab) specialist: Biomarker data

- Case Report Tabulation (CRT) from Study Programming Team: CORE, EXPOSURE, PC, DEMOG, CONMEDS, LAB, PHYSMEAS, DISPOSIT, AB, AE
- Study Data Tabulation Model (SDTM) from Study Programming Team: EX, PC, DM, CM, LB, PE, VS, DS, AB, AE
- Analysis Data Model (ADaM) from Study Programming Team: ADEX, ADPC, ADDM, ADCM, ADLB, ADPE, ADVS, ADSL, ADAB, ADAE
- Combination

Based on therapeutic area and study design, the analyst includes study-specific variables in the PK specs. Primary core variables in a NONMEM data set are given in Table 2 (Bonate et al., 2012; Boeckmann et al., 2011):

Name	Description	Name	Description
ID	Subject Identification Number	UNIT	Unit of Dependent Variable
POP	Study Population	CMT	Compartment Flag
DATE	Blood Sampling Date	BQL	BQL Data Records?
TIME	Blood Sampling Time	COHORT	Study Cohort
ATAFD	Actual Time Elapsed from First Dose	ARM	Study Arm
ATALD	Actual Time Elapsed from Last Dose	PERIOD	Treatment Period
EVID	Event ID for Records	TRT	Treatment
ADDL	Additional Dose	IPASITE	Dose Administration Site
FAST	Fasted Condition	PKSITE	PK Sample Collection Site
FORM	Formulation of Investigational Product Administration (IPA)	CONMED	Any Con Meds?
ROUTE	Route of IPA	RATE	Rate of Infusion
VISIT	Planned Visit	INFDURN	Duration of Infusion
NTAFD	Nominal Time after First Dose	SS	Steady-state Dose Flag
NTALD	Nominal Time after Last Dose	II	Interdose Interval
AMT	Dose Amount	CRCL	Baseline <characteristics>, example: Creatinine Clearance
DOSENO	Dose Number	SEX	Sex
DV	Dependent Variable	RACE	Race
MDV	Missing Dependant Variable	ALB	Baseline <lab>, example: Albumin

Table 2. Core NONMEM data set variables

Study-specific variables usually replicate efficacy, biomarker or antibody data, whereas core variables define safety parameters of a drug.

STAKEHOLDERS IN NONMEM DATA SET CREATION

A number of groups are involved in the NONMEM data set creation, from support level to production when NONMEM processing is ongoing; each group's contributions must be assessed and inputs integrated. As per study protocol's projection, analysts in PKDM assist with Statistical Analysis Plan (SAP), lays out NONMEM data requirements. Along the way clinicians and the Clinical Safety group are consulted for selection of clinical parameters for pop-PK/PD modeling. Analysts then generate PK specs, and perform modeling once the NONMEM data set is available.

Various groups—PKDM, Pop-PK/PD and phase 1 to 4 Study Programming teams—work together to facilitate production. The Study Programming Team, as the name suggests, is involved in study conduct activities ahead of NONMEM group. For Clinical Study Report (CSR) purposes the Study Lead Programmer creates Study programming specs, called Study Data Definition Table (DDT), taking PK specs and SAP into consideration. Comments on NONMEM DDT (described in the next section) to NONMEM programmers from study team members, particularly

from the Study Lead Programmer and the Study Biostatistician, provide valuable feedback given their knowledge of study design, input raw data, SDTMs and/or ADaMs. Analysts collaborate with study biostatisticians as well in order to ensure accuracy of PK specs.

A NONMEM programmer interacts with other groups depending on the maturity of data. If any issue is found at the raw data level, CDM group will need to clean data before further evaluation. If analyses involve biomarker data, specific biomarker group representatives play a part in establishing data accuracy and clean up. When NONMEM data set creation begins, two programmer roles are usually assigned working in parallel doing independent coding: Source Programmer for development and QC Role for validation purposes.

The timelines and milestones for NONMEM data set production is mostly driven by the Clinician at least for the planned analyses, but ad hoc requests come from different levels. For example, Senior Management might present results to prove potency/efficacy of a drug, and regulatory agencies could query data for safety reasons. A stakeholder map in Figure 2 best illustrates one's contribution.

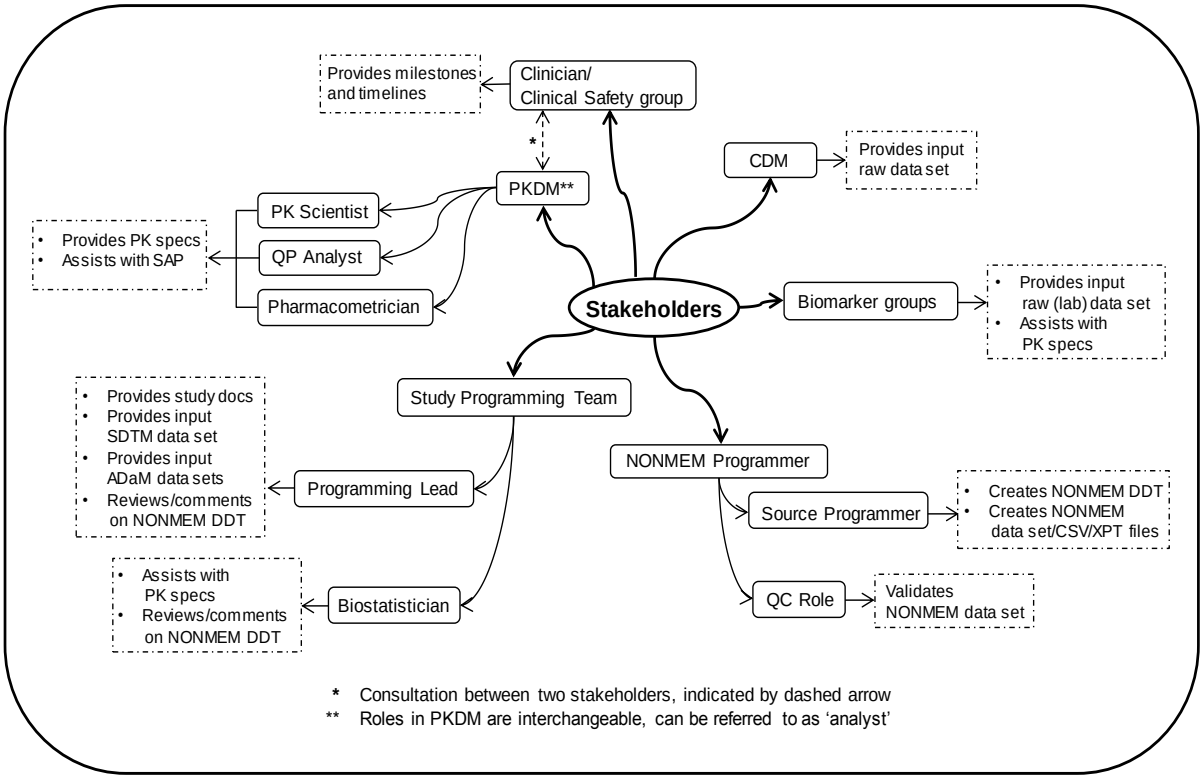

Figure 2. Stakeholder map

NONMEM PROGRAMMER'S TASKS

A NONMEM programmer starts with gathering specifications from the assigned analyst(s). All study documents come from Study Programming Team. Relevant study docs include: Protocol, SAP, Case Report Forms (CRFs), and Study DDT. At project initiation a NONMEM programmer gets acquainted with all study docs, reviews PK specs and provides input to the analysts.

Pop-PK/PD analyses need to occur at different phases of the drug development process. They are usually defined within a SAP. In some cases, requests for such analyses are completely ad hoc and triggered by safety and regulatory review (e.g., analysts might take a glance at early un-blinded data due to adverse events reporting). Concentration data reveal treatment assignments, thus are highly protected until final database lock when the study becomes un-blinded. A NONMEM programmer seeks to access appropriate data through a formal approval process to maintain data integrity and treatment blinding. Input data can be partial, interim, final, scrambled, restricted, or dummy. For programming development, in order to maintain trial integrity (and the blind of the study), NONMEM programmers may use dummy or scrambled data at the initial code setup and with real data used for the final NONMEM data set creation.

From PK specs and input metadata information NONMEM programmers create programming specs called a NONMEM (Pop-PK/PD) DDT, which is essentially a mapping doc specifying algorithms on how input variables are

laid out to derived fields. A NONMEM DDT is a dynamic document and updates occur as programming continues; when finalized it is passed to the Study Lead Programmer for feedback and confirmation on the mapping. NONMEM DDT is mainly based on PK specs and Study DDT, although other study docs are referenced. Independent coding commences once programmers are assigned as Source Programmer and QCer. Both roles check input data, making sure all required variables are available, and report data issues to the appropriate group (depending on where issues originate). In program development, NONMEM programmers always interface with the analyst especially when specs limitations are found and rules do not cover data scenarios. Once created a NONMEM data set is passed to the analyst for initial review. Comments from the analysts are assessed, after which programmers respond indicating specs' limitation, data issues, and/or fixing programming errors. Every time data is refreshed, a NONMEM programmer is granted data access and a data set is reproduced/revalidated. This allows analysts to work with more observations and thus helping out with modeling task.

A simplified task flow is shown in Figure 3 where study docs are assumed to be final and communication between programmers and other teams are omitted. Spec changes and/or data reload makes a NONMEM data set deliverable go through complex looping, i.e., multiple iterations.

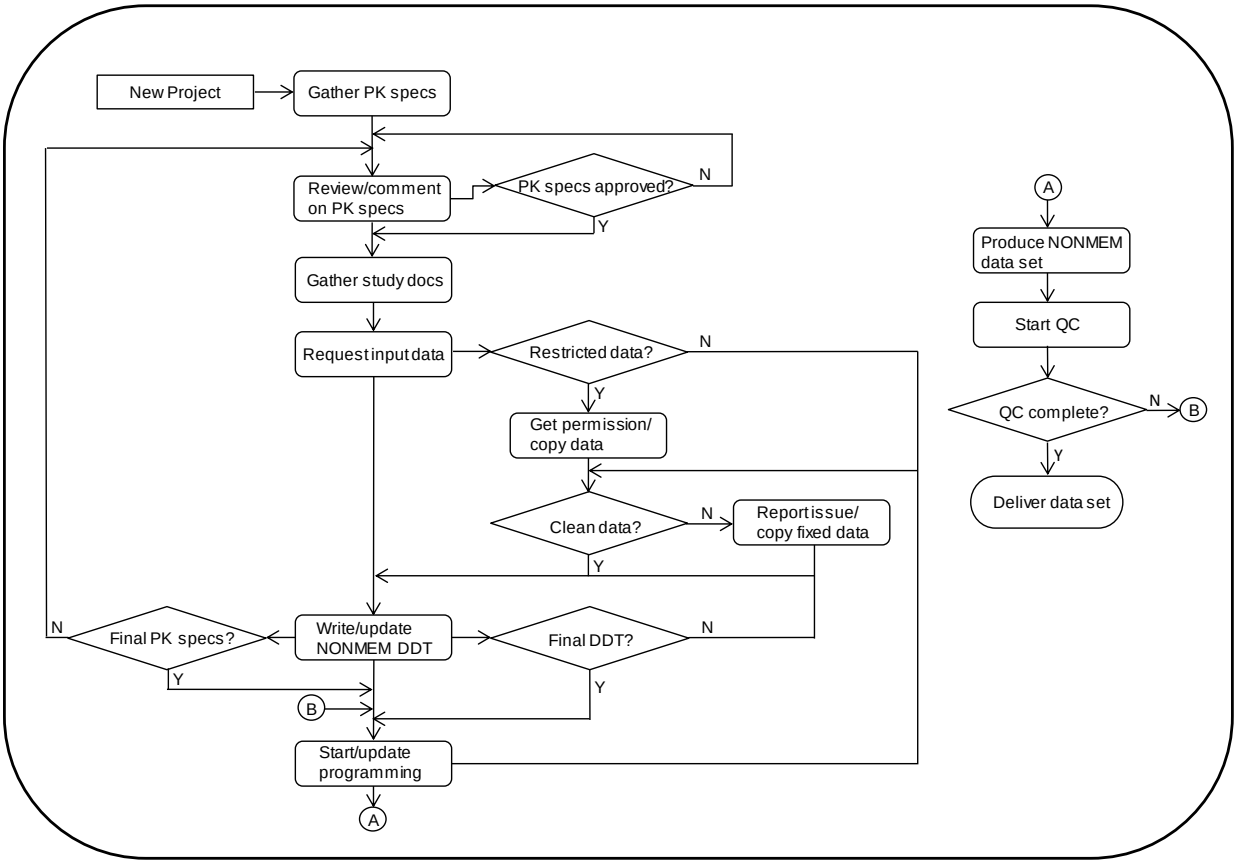

Figure 3. NONMEM Programmer's task flow

PROGRAMMING CHALLENGES

In general, analysts construct PK specifications from study protocols and CRFs before data become available. In other words, variable coding/decoding lists are not accessible to analysts when specs are created. Thus, initial specifications in most cases do not cover all data scenarios in a study, and frequent changes to the specs occur throughout the NONMEM data set creation, making it difficult for NONMEM programmers to meet timelines. At each data refresh more observations come that might expose code-lists not encountered in PK specs, necessitating additional updates. In the process, the NONMEM DDT is impacted and re-review is needed by the Study Programming Team, which could cause additional delays.

The spec changes become more prominent when a new study (with possibly different data capture) is added to an analysis. Timely data arrival to NONMEM programmers is always a challenge since PK concentration data are usually received late in the data set creation process because concentration information reveals treatment assignments. A NONMEM programmer has to prepare code ahead of time from only data structure (without having

real PK data) and be acquainted with actual data quickly when released. As one would expect maturity of input data drive the whole process; data cleanup, edit checks, issue resolution by CDM should be aligned with pop-PK/PD analysis timelines.

As NONMEM DDT has dependency on Study DDT, the readiness/completeness of Study DDT is critical. Else too many rounds with the Study Programming Lead are needed. When raw data is used in NONMEM data set creation, there is no associated Study DDT and NONMEM programmers derive variables by going through each data type and content which can be tedious and error prone. Instead of using analysis-ready variables from Study DDT, derivations from raw variables pose another concern that pop-PK/PD results might differ from CSR. This is due to two distinct groups, NONMEM and Study Programming Team, deriving the same variables causing the pop-PK/PD and study CSR results out of sync.

The specs from analysts are usually study-specific and algorithms defined might be complex depending on study design and therapeutic area involved. Complexity becomes multi-folded when a NONMEM program deals with several studies concurrently (Bonate et al., 2012). Each study can have a set of exclusion criteria, missing data, imputation rules, closest observation identification strategy (if time of measurement is missing). Hence, non-standard specs/programs impose extra stretch in NONMEM data set creation process.

DEBUGGING TECHNIQUES

Like with any analysis data set, validation is crucial for NONMEM data sets which are submitted to regulatory agencies as part of the submission package. Here quality control and maintaining standards are of the utmost importance (Bonate et al., 2012). Common debugging practices are given below:

- Create independent validation code
- Follow Standard Operating Procedures (SOPs)
- Isolate each study
- Isolate type of records by key variables
- Drop covariates and/or study-specific variables in first pass
- Check sort order
- Match number of observations
- Validate common/standard variables
- Validate one set of variables at a time
- Provide feedback to Source Programmer via formal documentation
- Specify details in feedback so that problem record is identifiable

EXPECTATIONS FROM A NONMEM PROGRAMMER

NONMEM programmers have to be familiar with PK concentration data, comfortable with input metadata information and equipped with pooling data from diverse input sources for multiple studies in parallel. Understanding of each study design is essential and requires collaboration with cross-functional teams to facilitate progress. Studies are often distributed in sub-groups given programming resource availability. In program development phase NONMEM programmers are expected to cross-check dosing and concentration data to see if data fit together as intended (e.g., visits are sequential, concentration samples are collected before/after dosing). Calculation of actual clock time and knowledge of acceptable range of nominal time points are valuable qualifications in this situation.

A NONMEM programmer is supposed to work with analysts to ensure quality and accuracy of specs by checking against study docs, providing comments, and/or bringing up intuitive questions for the analysts. Programmers need to have good grasp of data/variable definition as well as all stakeholders' responsibilities and report to desired groups when specs and/or data issues arise. These individuals must be experienced in SAS, preferably knowledgeable in NONMEM, and are expected to gather all pieces to meet production timelines that are non-negotiable in most cases.

CONCLUSION

Although sophisticated programming skills are not required from a NONMEM programmer, a good clinical background, understanding of input data and their formats, adaptability to specs changes, excellent data handling

and manipulation capacity in SAS, expertise in analysis-ready data set creation, ability/flexibility to work in globally dispersed and cross-functional team environment are essential.

Importance of pop-PK/PD analysis in drug development process, numbers of drugs in pipeline in a drug company, exposure-response analyses involved, timing/frequency of output needed, numbers of submissions to regulatory agencies have all led to an increase in demand for NONMEM programmers, thus leading to creation of an exclusive Pop-PK/PD programming group.

APPENDIX - NONMEM DATA SET EXAMPLE

Study setup: Multiple dose intravenous infusion study.
Dose = 140 mg QID (4 times a day), where alternate doses are recorded.
II (Interdose Interval) is every 6 hours.

Obs # | 24 hour clock | Not dose unit | In hours | Baseline covariate

#	C	ID	DATE	TIME	EVID	CMT	DV	MDV	VISIT	AMT	UNIT	ADDL	ATAFD	BQL	RATE	SS	II	CRCL
1	.	1	2005-03-02	08:00	3	1	.	1	.	.	.	0	-248.08	0	0	0	0	56.96
2	.	1	2005-03-03	08:00	1	7	.	1	.	1	.	0	-240.08	0	0	0	0	56.96
3	.	1	2005-03-03	08:00	1	14	.	1	.	1	.	0	-240.08	0	0	0	0	56.96
4	.	1	2005-04-03	08:31	0	7	3.02	1	D1PRE	.	ng/mL	0	-1.48	1	0	0	0	56.96
5	.	1	2005-04-03	09:00	0	14	45.90	0	D1PRE	.	mmol/L	0	-1.00	0	0	0	0	56.96
6	.	1	2005-04-03	09:40	0	1	197.29	0	D1PRE	.	ng/mL	0	-0.33	0	0	0	0	56.96
7	.	1	2005-04-03	10:00	1	3	.	1	D1	140	.	3	0.00	0	15.4	0	6	56.96
8	.	1	2005-04-03	22:10	1	3	.	1	D1	140	.	1	12.17	0	15.4	1	6	56.96
9	.	1	2005-04-04	10:30	0	1	2034.53	0	D2	.	ng/mL	0	24.50	0	0	0	0	56.96
10	.	2	2005-03-02	08:00	3	1	.	1	.	.	.	0	-248.07	0	0	0	0	80.34
11	.	2	2005-03-03	08:00	1	7	.	0	.	1	.	0	-240.07	0	0	0	0	80.34
12	.	2	2005-03-03	08:00	1	14	.	1	.	1	.	0	-240.07	0	0	0	0	80.34
13	.	2	2005-04-03	08:25	0	7	5.79	0	D1PRE	.	ng/mL	0	-1.33	0	0	0	0	80.34
14	.	2	2005-04-03	08:55	0	14	31.40	0	D1PRE	.	mmol/L	0	-0.83	0	0	0	0	80.34
15	1	2	2005-04-03	09:20	0	1	17.29	1	D1PRE	.	ng/mL	0	-0.42	1	0	0	0	80.34
16	.	2	2005-04-03	09:45	1	3	.	1	D1	140	.	3	0.00	0	15.4	0	6	80.34
17	.	2	2005-04-03	22:25	1	3	.	1	D1	140	.	1	12.67	0	15.4	0	6	80.34
18	.	2	2005-04-04	09:50	0	1	6784.34	0	D2		ng/mL	0	24.08	0	0	0	0	80.34

Table 3. NONMEM data set sorted by ID DATE TIME EVID CMT

Record identification (obs #) by CMT and EVID:

Reset (#1, 10)	CMT = 1 and EVID = 3
Initialization (#2, 3, 11, 12)	CMT in (7, 14) and EVID = 1
Lab (#4, 5, 13, 14)	CMT in (7, 14) and EVID = 0
PK concentration (#6, 9, 15, 18)	CMT = 1 and EVID = 0
Dosing (#7, 8, 16, 17)	CMT = 3 and EVID = 1

Assumptions (should be derived as per PK specs):

C, VISIT, UNIT	Missing values are shown as '.' although these are character variables.
CRCL	Decimal precision is 2.
Obs #1, 2, 3, 10, 11, 12	TIME = 8:00 for reset and initializations records.
Obs #2, 3, 11, 12	AMT = 1 for initialization records.
Obs #4	BQL = 1 when result contains '<', where original lab value is '<3.02'.
Obs #15	BQL = 1 when concentration is less than 50 ng/mL.
Obs #15	C = 1 for PK concentration records when BQL = 1.
Obs #4, 15	MDV = 1 when BQL = 1.
Obs #7, 8, 16, 17	MDV = 1 for dosing records.
Obs #8	SS (steady-state) = 1.

REFERENCES

- U.S. Department of Health and Human Services (HHS), Food and Drug Administration (FDA), 1999. Guidance for Industry. Population Pharmacokinetics.
- Committee for Medical Products for Human Use (CHMP), European Medical Agency (EMEA), 2006. Guideline on reporting the results of population pharmacokinetic analyses.
- Wade J.R., Edholm M., Salmonson, T., 2005. A Guide for Reporting the Results of Population Pharmacokinetic Analyses: A Swedish Perspective. AAPS Journal, v.7(2); 2005 Jun, PMC2750982.
- Pang, H., 2005. Some Statistical Programming Considerations for e-Submission.
- Shen, D., Lu, Z., 2007. Population Pharmacokinetics Studies with Nonlinear Mixed Effects Modeling.
- Fraeman, K.H., 2008. Common Sense Tips and Clever Tricks for Programming with Extremely Large SAS® Data Sets.
- Crevar, M., 2007. How to Maintain Happy SAS® Users.
- NMuser, 2004. http://www.cognigencorp.com/nonmem/nm/99dec072004.html
- Ette, E.I., Williams, P.J., 2013. Pharmacometrics: The Science of Quantitative Pharmacology.
- Bonate, P.L., Strougo, A., Desai, A., Roy, M., Yassen, A., Walt, J.S., Kaibara, A., Tannenbaum, S., 2012. Guidelines for the Quality Control of Population Pharmacokinetic–Pharmacodynamic Analyses: an Industry Perspective. AAPS Journal, v.14(4); 2012 Dec, PMC3475847.
- Fisher, D., Shafer, S., 2007. Fisher/Shafer NONMEM Workshop Pharmacokinetic and Pharmacodynamic Analysis with NONMEM. 2007.
- Boeckmann, A.J., Sheiner, L.B., Beal, S.L., 2011. NONMEM Users Guide - Part V.

GLOSSARY

Code-list: A list of possible values of a variable in SAS data set.

Derivation of a NONMEM variable: Variables are derived only for the purposes of population PK/PD analysis like imputation of time varying covariates based on the rules specified in PK specs, or lab results using Last Observation Carried Forward (LOCF).

Input data consistency check: NONMEM programmers check date/time of dosing and concentration data, see if PK samples are collected and doses have been administered as per Protocol mentioned time points. Any measurement taken not within an acceptable range of nominal time point would be a data issue and should be run by CDM. Code-lists defined for variables need to be consistent across studies, for example for VISIT - D1PRE and D2 might not be the same as D1HR0 and D1HR24, respectively.

Input metadata: Characteristics of input data defined by its structure, variable type and content.

Imputation of missing data: NONMEM programmers incorporate imputation rules provided in PK specs. Example: for missing covariates imputation method can be use of median or mean. If time of a measurement is missing, it is imputed from the closest previous or next measurement.

Exclusion criteria: PK specs specify exclusion criteria that must be taken into account when creating NONMEM data sets. Criteria for unusable records might be but not limited to: duplicate observations, screen failure subjects, missing dose or blood sampling time, no associated concentration with dosing or vice versa.

NONMEM data set: NONMEM® software-ready analysis data set prepared by Pop-PK/PD programmers, also called Pop-PK/PD data set.

NONMEM DDT: Data Definition Table/programming specs created by NONMEM programmers from PK specs and Study DDT. Other study documents (Protocol, CRFs, SAP, …) are referenced as well.

NONMEM Programmer: Individuals who create NONMEM DDT and associated data set. These programmers work with analysts and Study Programming Team to gather specs and report issues if any. They are also referred to as Pop-PK/PD programmers which distinguish them from analysts using NONMEM® software.

Observations: Vertical rows in a SAS data set, also called records.

PK Specs: PK/PD specifications that are provided by analysts. Study biostatistician assist with creating this

document. It is also referred to as NONMEM Specifications or PKDM Requirement Document or Population PK/PD Data Specifications.

Study DDT: Data Definition Table/programming specs created by Study Lead Programmer providing data structure and variable information, this document is used for generating analysis-ready data sets. It is derived from SAP mainly and other study documents (Protocol, CRFs, …) are referenced as well.

Study Programming Team: Group assigned for clinical study projects and responsible for all expected deliverables for CSR.

Variables: Horizontal fields/columns in a SAS data set.

ACKNOWLEDGEMENTS

The author would like to acknowledge Cytel Inc. for providing the opportunity to work on this paper. The author thanks the following colleagues/managers from present and previous workplaces for thoughtful comments.

Baker, Jim (Senior Vice President of Clinical Research Services at Cytel Inc.)

Eschenberg, Michael (Director Biostatistics at Amgen)

Melhem, Murad (Principal Scientist at Amgen)

Xiao, Jim (Associate Director Clinical Pharmacology at Clovis Oncology)

CONTACT INFORMATION

Sharmeen Reza
Cytel Inc.
Work Phone: 269-743-7221
E-mail: Sharmeen.Reza@cytel.com

In this paper NONMEM with the registered trademark refers to the software.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.