

Programming strategically for PK/PD data

Grace Lu, AstraZeneca, Waltham, USA

ABSTRACT

Pharmacokinetic (PK) and pharmacodynamic (PD) data analysis plays an important role in today's clinical research, especially in early phase drug development. Programmers receive many requests not only to support PK analysis for regular clinical studies, but also data integrations and derivations to support PK/PD modeling and simulation. Commonly, because of the urgency and complexity even an experienced programmer can see a significant increase in resource demand and pressures. This is often due to lack of understanding the nature of the data and workload, insufficient resource planning, and not effectively utilizing team input on programming specifications. In today's programming environment it's not enough to just be a highly skilled technical programmer, we also need to think strategically, work collaboratively, and act proactively. This paper will discuss the basic concept of PK/PD, the data flow, interactions and touch points between programmers and other functions, and share the best practices of programming with PK/PD data. The target audience will be the entry to intermediate level programmers, and the aim is for them to gain confidence on dealing with PK/PD data, and to be successful in this area.

INTRODUCTION

The high failure rate in Phase II and Phase III clinical trials increases costs of drug development dramatically in pharmaceutical industry(1). One of the causes for the high attrition rate is inadequate information from early phase studies to support go/no-go decisions or to design later-phase clinical studies correctly. To address the issue, it is essential to have a good understanding on what the drug does to the body (PK) and what the body reacts to the drug (PD). However, only limited information can be provided if PK and PD analysis are performed separately (2). PK/PD modeling and simulation, which links the dose-concentration relationships and the concentration-effect relationships, provides quantitative predictions of the full course of drug effects following different candidate dose regimens. Therefore the integrated PK/PD modeling and simulation has been increasingly used in today's clinical research.

This has led to an increased need in data collection, process, analysis, and reporting around PK/PD modeling and simulation, and thus more comprehensive programming work is required for the clinical programmers. In order to provide analysis or modeling ready PK/PD dataset, programmers will need to combine variables from different datasets, pool data from different sources, derive certain variables, and organize the data into a specific format depending on the modeling software/system requirements. Due to the exploratory nature of PK/PD analysis, the data integration and derivation can be quite complex.

PK/PD MODELING AND SIMULATION

As we know that pharmacokinetics addresses what the body does to the drug, and pharmacodynamics is dealing with what the drug does to the body. For very long time these two areas in pharmacology were considered as separate disciplines, and the information gained from the isolated areas are obviously offered with limitation (2). Sometimes the PD response to drug concentration (PK) can be divergent due to factors like PK process, signal transduction, secondary post-receptor modifications, or variability within study population. This led to the development of new area of PK/PD modeling, which builds the bridge between the two classical areas of pharmacology, integrates and correlates between concentration(PK) and effect (PD)(2, 3).

The importance of PK/PD modeling and simulation in clinical drug development has been widely recognized across industry, academia, and health authorities. The use of modeling and simulation technique to aid drug development in various stages has been greatly increased in recent years. PK/PD modeling and simulation based on data from early dose escalation clinical studies can provide information for rational design of all subsequent trials in identifying effective and safe dose regimens. It also can help to make go-no go decisions earlier, especially if models include comparators data (4). In late phase of drug development, population PK/PD modeling is applied to further examine dose-concentration-effect relationship in patients, and is utilized to explore dose requirements in subpopulations of patients(2). In the case of confirmed favorable efficacy provided from one large Phase III data, PK/PD modeling and simulation would be in obviating the need for a second large trial. The use of exposure-response information in combination with a single pivotal clinical trial data can provide sufficient evidence of effectiveness, which was stated in FDA Modernisation Act of 1977, and Guidance for Industry Population Pharmacokinetics (4, 8). Therefore, PK/PD modeling and simulation plays a important role for both scientific and strategic decision making in drug development, leading to optimized trial design, reducing development cost and time (2, 11).

PhUSE 2015

PK/PD modeling is a mathematical description of the relationship between pharmacokinetics (PK) and pharmacodynamics (PD)(3). The relationship between the administered dose of a drug and the response outcome may be complex, because it's not always in parallel drug concentrations. The technique of modeling and simulation, which is a data (pharmacokinetic and pharmacodynamic) driven exploratory analysis based on a mathematical or statistical model of be non-linear or mixed, allows the characterization and evaluation of this relationship and its change as a function of drug intake and other clinically relevant variables(covariates). It provides quantitative information on complete time course of desired and/or undesired effects in response to a dose regimen in consideration of underlying physiological processes such as disease, age, gender, etc , and even more, and also predicts beyond the existing data (2, 3, 4, 5). Therefore it is a highly integrated analysis established on real data and simulated with given conditions.

In the past several decades, much progress has been made in PK/PD modeling and simulation due to innovation of methodology and availability of computer technology. It becomes increasingly interested among academia, industry, and health authorities (6, 7). This has led to a dedicated professional area of pharmacometrics - a multi disciplinary field of mathematics, statistics, and pharmacology.

Conducted by pharmacometricians (experts at pharmacometrics), a general model building process can be described as following (3, 9):

- Analysis plan development to define hypothesis
- Data collection and data processing for modeling ready format
- Data analysis and exploration to suggest/confirm tentative models
- Model selection and model fitting to estimate the model parameters and precision of the parameters estimates
- Generating analysis and model validation
- Result interpretation and presentation to assist design and decision making
- Model maintenance and update for future use (when more data is available)

There are many computer software and tools available for PK/PD modeling - some may benefit from less programming time required, and some may benefit from building certain types of models. NONMEM has been widely used. Specified by a control file, which is processed to produce Fortran code that is compiled and linked to other objective files to create executive files, NONMEM does not require sophisticated programming skills to build the model(9). This allows pharmacometricians to concentrate on analysis and interpreting modeling/simulation results. However, all of the modeling software or tools will require data to be preprocessed into certain format and structure, as well as require some variable derivation before it can be used for modeling. In addition, unclear or unsatisfied data will need to be handled before in certain ways modeling. Very often PK/PD models are built by using real time data to provide insight into drug exposure safety evaluations and drug-drug interactions. However, when working with data from on-going studies, there might be various data issues (missing values, outliers, errors, etc.). To minimize the potential statistical bias, these data will need to be handled appropriately before using them in modeling (8). And in other cases, modeling and simulation sometimes will be based on large pooling data from different studies or different data sources. Processing data to be modeling ready format is a fundamental and complex part in PK/PD modeling. Assistance from a skilled SAS programmer to build PK/PD data is urgently needed in recent years.

DATA PROCESS AND INTEGRATION FOR MODELING READY

Since clinical data are collected and organized based on CRF modules and clinical studies, the required data for modeling will need to be pulled from different clinical modules and different clinical studies, then to be integrated and re-organized into specific format that required by modeling software. Certain calculation/derivation for new variables and issue data handling are also required. The following categories of data are normally to be integrated for modeling:

1. The main elements: PK concentration and sampling time, PD data and assessment/event time (such as efficacy response data, safety data, biomarkers, some lab tests, etc), dosing information (given, time, intake rate)
2. The covariates: demographic data, concomitant meds, disease characters, interventions, lab values etc
3. The timings: timings of PK/PD tests relative to dosing will need to be calculated accurately

Dealing with these data piece by piece might very straight forward, whereas to integrate all information together and to re-organize them into the specific structure is often not a simple task. Especially when work with multiple data sources (CRF modules, SDTM data, or data from different organization), or with data from ongoing studies which may have missing or error values, the programming work can be very challenging.

Figure 1 shows a general process of generating modeling ready data. As soon as the analysis plan is in place, programmer should be involved to review the plan to obtain first hand information. The analysis plan may not have details on what programming need to be done to process the data. It will very helpful to request a data specification from pharmacometrician with details on:

PhUSE 2015

- Studies to be included (in house studies, outsourced studies, on-going, or data from other source like published trials, academic data, etc)
- Data to be integrated (PK, DOSE, efficacy, safety, etc.)
- Variables to be derived (flags, calculations), and definitions or formulas of derivations
- General rules to be applied on data handling (for outliers, error values, missing values etc.)
- Data structure and format (vertical or horizontal, excel or SAS, etc.)
- Expected delivery date

With the data specification on hand, it is very important for SAS programmer to investigate the source data before starting programming. Spend time to study the specification and explore the source data, ask and find answers of the following:

- Are the data accessible? Will you need a data transfer from CRO? If data is not yet available (like concentration data before database lock), is that possible to have a dummy data for programming purpose? If data is from other organization (public domain or academic), who is responsible to request data?
- Are the variables in the right formats (numeric, character)? If need to convert from character to numeric, how to deal with the character signs associated with a number like <, >, or LOQ?
- Are definitions clear for the calculation/derivations?
- Is the data in good quality in general (too many missing, or some extreme values?)
- Is there is a rule should apply on missing data?
- Does the data collected provides good information (if the assessment time is not collected, there is no way to calculate time to dosing, or too much missing with poor quality)?
- Is the data requested from a single study or from multiple studies? How long time or how many programmers will be needed for the task?

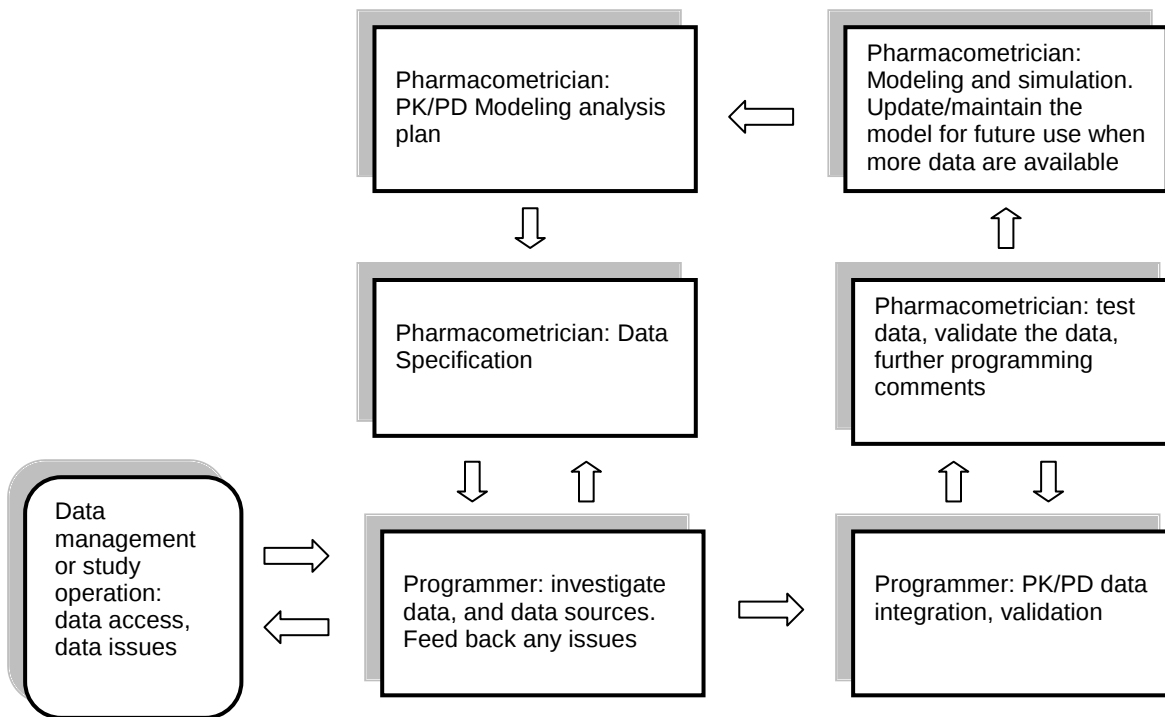


Figure 1 A general process of data integration for PK/PD modeling

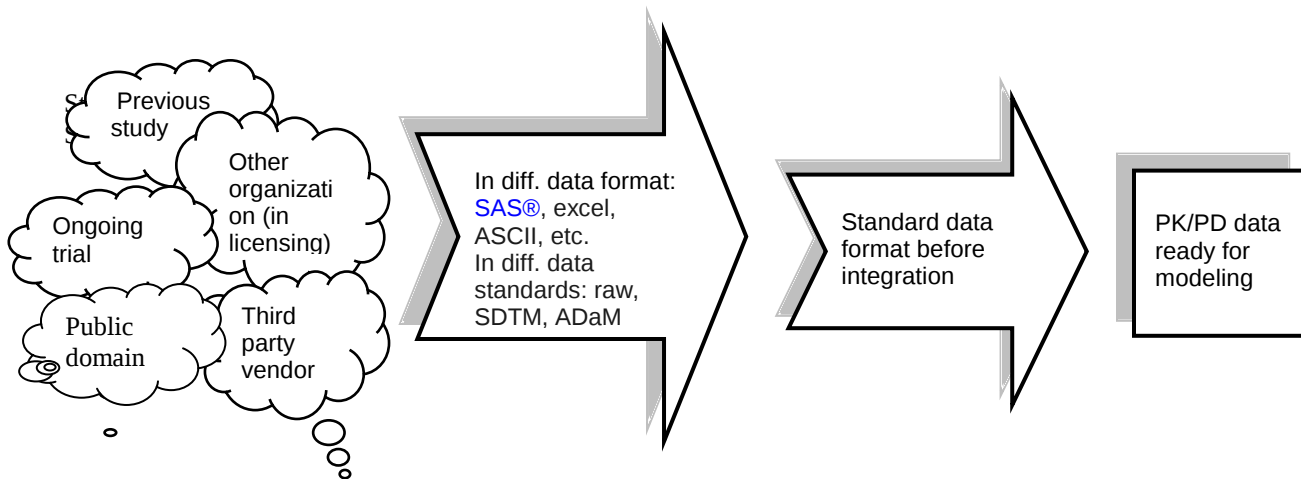


Figure 2 Data flow to generate modeling ready data

After investigating the data and finding answers or questions for the data specification, the SAS programmer should schedule a meeting with pharmacometrician for further discussion aiming on:

- Further clarification on the data specification
- Exam data together on data issues identified (missing, extreme values, non-calculable data, ect) and discuss on resolutions (can these data be eliminated? Or is there any computational rules? What can be done for outliers, and error data? Are the rules for data handling practical/reasonable vs. actual data?)
- Should the data specification be modified/revised based on the actual data situation?
- If the data is not accessible or not available in house, is there any operational process to request data?
- Agree on a reasonable timeline based on the scope of work and resource available
- Identify risk factors

It is critical to have this discussion with the pharmacometrician before start working on programming. Certain rules on data specification may be based on assumptions, which may not be appropriate based on the real data cases. Pay extreme attention on missing values in the data and missing value imputation rules in data specification. With good knowledge on the data, SAS programmer may find that the definition or computational rules may not be clear, or may not be reasonable for certain data cases. This is the opportunity for SAS programmer to get further clarification on data specification, to feedback any concerns or issues from a programming point of view. The data specification can be further refined based on discussion, therefore some revisions could be avoided later. However, revisions on data specification and additional programming are hard to be eliminated completely due to the exploratory nature, which will be discussed later.

When plan the programming work, It is important to know what data is available and accessible to start with, and what data is dependent on other factors (like data transfer schedules, blinded data). Discuss with data management or study operational to request data that is not yet available. At the same time, the SAS programmer should seek appropriate resources to support the completion of the task. High light any risk factors (such as very messy source data, data unavailability, lack of programming resources, other delivery priorities) to the stakeholders (pharmacometrician, translational scientists, pharmacologist, physicians, functional managers) on early stage will allow programmers to gain support from broader team.

Once having data access and obtaining a refined data specification, the programming work can be started (Figure 2). If data integration requested from various data source, the more practical way could be having all data in different format to be standardized first. The benefits are:

- Having standardized data will enable data pooling easier
- Modeling will mostly utilize the most current data, and the new studies may already use SDTM standards or alike. If SDTM data to be pooled with other format, it is more reasonable to align all data to be the CDISC SDTM standards
- The data standardization part can be broken down into individual studies, or data domains. Therefore multiple programmers can work on different studies or data domains at the same time
- It is easier to validate the data on standard before it is integrated into modeling ready data
- By using standardized data, the modeling integration program can focus on directly selecting standard variables, re-organize them into specific structures, and derive new variables.

PhUSE 2015

- By using standard data as starting point, the program for PK/PD data will have chance to be reused

Whereas data standardization is fundamental part for modeling ready data, the data integration and derivation is the core and full of challenging. The SAS programmer will need to have good understanding on the data specification, to follow the definition and rules that defined, and to use smart programming skills (that is to say, use good logic flow and neat programming style) for the core task. The best practices and programming tips will be discussed later. Due to the complexity of data integration/derivation, data validation before providing it for modeling is necessary to ensure quality.

During data integration, it may necessary to have multiple touch points with the pharmacometrician for any emergent issues further identified during programming. In this way, you may have the issues handled before finalizing the program and validation. However, due to the exploratory nature of PK/PD modeling, pharmacometricians will most likely to come back with requests for further programming. This is because that, the modeling is based on certain assumptions, and the data processing on missing values (either excluding or imputing), will likely to add another layer of assumption to the model. Including outliers may lead to disproportionate effect on estimates, but what defines a data outlier needs to be carefully considered. All of these need to be tested, evaluated, and adjusted in the initial PK/PD analysis and model fitting stage. There will be more data issues discovered, more parameters need to be included, or update/adjustment on existing data specification, thus additional programming work is demanded. Therefore, programming for PK/PD data is not just a one way work simply from data specification to modeling ready dataset. It is rather a more complex loop of programming-validation-additional programming-revalidation.

DATA SPECIFICATION AND DOCUMENTATION

To avoid misunderstanding and gain clarity on programming requirement, a good document of data specification is essential. It should include specific information that a programmer need for completion of the task. It may be organized in a way for easy access to any piece of information. An excel work book can be used to organize all such information as it allows to flexibly expand to multiple work sheets and columns to still sum all information in one file. For example:

Sheet 1: General information - studies to be included, general rules, data transfer schedule, timelines, etc.

Sheet 2: Data specification - this is the main part. It can be organized into columns and rows. Example can be seen in table 1. Although the specifications are provided by pharmacometricians, it is very helpful for SAS programmer to add a row of interpretation by indicating the relevant SDTM variables or programming conditions/logics.

Sheet 3: sample data. It will be very helpful if pharmacometrician provide a sample data. SAS programmers will have a better picture on naming convention, value format, and data structure that needs to be generated (See Table 2).

Sheet 4: Documentation. Since the data specification might be changed dynamically, and further programming work will be requested commonly after PK/PD data is delivered, managing the loop of programming-validation-additional programming-revalidation can be very challenging. To avoid confusion and to streamline the programming process, a well maintained documentation is extremely important. Excel might be an easy tool to clearly record all items by rows, and document each item by name, date, status, etc. by different columns. Such log file will be helpful for both programmer and pharmacometrician on how and when the issues/questions are resolved or answered.

Additional work sheets can be added for anything related to the task so that all information can be saved in the same file.

Table 1. Example of Data Specification*

		Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8
1	Data ID (specific variable names required)	OBS	STUD	SCODE	SUBJ	ID	SDAY	DATE	TIME
2	Variable description	Unique record identifier	Study identifier	Study code	Subject specific identifier	Unique subject identifier	Study day	Date of event	Time of event
3	Value format	integer	integer	char	integer	integer	integer	mm/dd/yyyy	hh:mm
4	Significant digit				4				
5	Value decimal places								
6	Unit						day		
7	Further notes	A sequential number assigned for every record in	Study 1234 = 1, Study 5678 =2	The actual study code	The 4 digit patient number	Assign the patients in a sequential order, 1, 2, 3,...	(date of assessment/ event – date of first dose+1)	Assess. or event date	Asses. or event time

PhUSE 2015

		the data set							

*This table is adapted from data specification used by AstraZeneca programmers

Table 2. A sample NONMEM data*

OB S	S T U D	SCOD E	SUBJ	I D	S D A Y	DATE	TIME	TSFD	TSLD	AMT	RATE	DV	D V I D	M D V	E V I D
1	1	C1234	0001	1	4	2/29/2012	14:20	-112.67	-112.67	.	.	4.386	2	0	0
2	1	C1234	0001	1	1	3/5/2012	7:30	-0.5	-0.5	.	.	2.299	2	0	0
3	1	C1234	0001	1	1	3/5/2012	8:00	0	0	85	26.000	.	.	1	1
4	1	C1234	0001	1	1	3/5/2012	9:30	1.5	1.5	.	.	5.27334	1	0	0
5	1	C1234	0001	1	1	3/5/2012	11:00	3	3	.	.	8.00570	1	0	0
6	1	C1234	0001	1	1	3/5/2012	11:30	3.5	3.5	.	.	5.74005	1	0	0
7	1	C1234	0001	1	1	3/5/2012	12:00	4	4	.	.	4.79434	1	0	0
8	1	C1234	0001	1	1	3/5/2012	14:00	6	6	.	.	1.07587	1	0	0
9	1	C1234	0001	1	1	3/5/2012	16:00	8	8	.	.	5.39564	1	0	0
10	1	C1234	0001	1	2	3/6/2012	8:00	24	24	.	.	8.01640	1	0	0
11	1	C1234	0001	1	3	3/7/2012	8:00	48	0	.	.	0.00927	1	0	0
12	2	C1234	0001	1	5	3/9/2012	8:00	96	0	.	.	0.00631	1	0	0
13	2	C1234	0001	1	8	3/12/2012	8:00	168	0	.	.	0.00850	1	0	0
14	2	C1234	0001	1	8	3/12/2012	10:45	170.75	2.75	.	.	1.27056	2	0	0

*There are more variables in NONMEM. Only part is showed here due to limited space.

THE BEST PRACTICES /PROGRAMMING TIPS

Processing data for PK/PD modeling ready format is not a simple task. A programmer has to sort out the data from various sources first, convert them into standard format, gather all variables, organize them into specified structure, derive further variables and handle outliers/missingvalues/error data based on specific requirements. It may not require super programming skills to generate some fancy codes, but the entire process is normally very time consuming and resource demanding.

The following best practices and programming tips may help to reduce the burdens and improve programming efficiency.

- Involve programmers as early as possible so that the work can be planned in advance
- Once the data specification is available, derive a good programming plan. Consider working on SDTM data mapping first if the data are not in SDTM format. And this part can be assigned to multiple programmers at the same time if resources are available.
- Plan some extra programming time to deal with addition requests or unexpected situations (model adjustment, data issues, system issues, etc)
- If concentration data are not yet available(before data base lock), consider using dummy data to enable programming to start
- Develop a simple macro to convert DTC(date and time) to be numeric and to calculate time of event/assessment to the first dose, or to the dose last administrated. The macro can be reused in the future
- The program to derive NONMEM PK/PD might be very lengthy. It will be helpful to write clean programming comments before each logic blocks. This will help you later on any revisions needed
- Because the final data will contain many variables from different data or data sources, it will be helpful to derive a flag or note in the comments to indicate the source data. It is important to be able to trace back to the source data to ensure the quality.
- Be proactive. Reach out for help. Because the data and the process can be very complex, it is very important to have a clear understanding on the requirement, and deep knowledge on the data. Data specification may not clear enough for programming, or data issues might affect programming progress. For any questions/issues/concerns, discuss with pharmacometrician, data management, or

PhUSE 2015

other team members. Communication is the key for seeking resolution and improving collaboration, and thus managing expectations.

CONCLUSION

PK/PD modeling plays an important role in today's drug development. Preparing data for PK/PD modeling has become a more frequent task for clinical programmers. It is essential to understand the exploratory nature of PK/PD modeling work, and be prepared to work on large data integration or to deal with uncleaned data from ongoing studies. Furthermore, different with the normal clinical data (which is CRF module based or domain based like SDTM data), the PK/PD data for modeling is highly integrated from various data and data sources. In this challenging programming environment, it is not just enough to have solid programming skills, it is also very important to think strategically, work collaboratively, and act proactively.

REFERENCES

1. *Ismail Kola and John Landis* (2004). Can the pharmaceutical industry reduce attrition rates? *Nature reviews drug discovery*. Vol. 3 August 2004: 711 - 715
2. *Hartmut Derendorf and Bernd Meibohm* (1999). Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: Concepts and perspectives. *Pharmaceutical research*. Vol 16, No. 2: 176-185
3. *Dheeraj Gop and Gomathi P* (2012). Pharmacokinetics/pharmacodynamic (PK/PD) modeling: an investigational tool for drug development. *International journal of pharmacy and pharmaceutical sciences*. Vol. 4 Suppl 3: 30-37
4. *Iris Rajman* (2008). Pk/PD modeling and simulations: utility in drug development. *Drug discovery today*. Vol. 13 No. 7/8: 341-346
5. *Jose Perez-Urizar et al* (2000). Pharmacokinetic-pharmacodynamic modeling: why? *Archives of medical research* 31: 539-545
6. *Chantal Csajka and Davide Verotta* (2006). Pharmacokinetic-pharmacodynamic modeling: history and perspectives. *J. of pharmacokinetics and pharmacodynamics*. Vol 33, No. 3: 227-279
7. *Douwe D. Breimer* (2008). PK/PD modeling and beyond: impact on drug development. *Pharmaceutical research*. Vol. 25, No. 12: 2720-2722
8. *FDA* (1999). Guidance for industry: population pharmacokinetics.
9. *Ene I. Ette and Paul J. Williams* (2013). *Pharmacometrics: The science of quantitative pharmacology*. Published by John Wiley and Sons.
10. *Jenny Y. Chien, et al* (2005). Pharmacokinetics/pharmacodynamics and the stages of drug development: role of modeling and simulation. *The AAPS journal* 7(3): E544 – E559
11. *Raymond Miller, et al* (2005). How modeling and simulation enhanced decision making in new drug development. *Journal of pharmacokinetics and pharmacodynamics*. Vol. 32. No. 2: 185-197

ACKNOWLEDGMENTS

The author is very thankful for many valuable comments and suggestions from colleagues in AstraZeneca. Special thanks to Dr. Jianguo (James) Li and Dr. Diansong Zhou for guidance and advices in the area of PK/PD modeling. The author also greatly appreciates Mark Burke, Pat Mitchell, Rob Fox, Diansong Zhou and Jianguo (James) Li for their critical review on the article, and thanks Alison Dobson for her all time support.

RECOMMENDED READING

1. *Pharmacometrics: the science of quantitative pharmacology*. Edited by Ene I. Ette and Paul J. Williams (2013). Published by John Wiley and Sons.
2. NONMEM PK/PD dataset programming: make it simpler. Theorem. In PHUSE wiki: http://www.phusewiki.org/docs/Bangalore%202014%20SDE%20Presentations/NONMEM_Presentation.pdf
3. Growing needs in drug industry for NONMEM programmers using SAS®. Sharmeen Reza. *Pharmsug* 2015

CONTACT INFORMATION

Grace Lu
AstraZeneca
35 Gatehouse Dr.
Waltham, MA02451, USA
Work Phone: 1 781 839 4476
Email: grace.lu@astrazeneca.com

Note:

PhUSE 2015

The author is a current employee at AstraZeneca Pharmaceuticals. The view and opinions expressed in this article are from author's own learning and experiences, and do not necessarily represent the views of AstraZeneca.