

## Considerations in the Submission of Pharmacokinetics (PK) Data in an SDTM-Compliant Format

Fred Wood, Octagon Research Solutions, Wayne, PA  
Peter Schaefer, Certara, Cary, North Carolina  
Richard Lewis, Octagon Research Solutions, Wayne, PA

### ABSTRACT

Pharmacokinetic (PK) data is collected for a variety of purposes in clinical trials. Some of the first-in-human trials provide absorption and elimination data for comparisons with data collected in non-human species. PK data is also used to determine details of the dosing regimen, such as dose amount and frequency; for therapeutic benefits; and for bioequivalence testing for alternative dose formulations.

This paper focuses on some of the challenges associated with managing the collected concentration data and the subsequent calculation of PK parameters. Examples of how the collected data map into the SDTM domains representing pharmacokinetic concentrations and pharmacokinetic parameters datasets, as well as the use of RELREC are presented. In addition, a brief overview of the status of development of CDISC controlled terminology for PK parameters and PK-parameter units is provided.

### INTRODUCTION

#### CDISC Background

The Clinical Data Interchange Standards Consortium (CDISC) is a global not-for-profit organization that actively develops standards to support the acquisition, exchange, submission, and archive of data for protocol-driven medical research. The mission of CDISC is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC develops standards to support the entire drug-development lifecycle. For more information, go to [www.cdisc.org](http://www.cdisc.org).

#### SDTM BASICS

The Study Data Tabulation Model (SDTM) is one of the most well recognized and widely implemented CDISC standards. The purpose of the SDTM is to define the organization, structure, and format of the tabulation data that are to be submitted as part of a product application to a regulatory authority. Tabulation datasets focus on the data collected during a clinical trial, but the SDTM and the SDTM Implementation Guide (SDTMIG) include a few derived variables, and the SDTMIG describes the PK parameters dataset, which is derived from the PK concentration data.

In the SDTM, observations collected about subjects who participated in a clinical study are organized into series of domains. Each domain is identified by a two-letter abbreviation and is defined as a collection of variables that share a common topic. Examples include Demographics (DM), Laboratory Test Results (LB), and Pharmacokinetic Concentrations (PC). The data submitted in a domain can be collected on one or more case report forms (CRFs), and conversely, data collected on one CRF may be submitted in more than one domain. In addition, data can come from sources other than data collection forms or EDC systems. Lab data is generally obtained from an electronic data transfer (eDT) file from a central lab, and PK concentration data are usually provided by a sponsor's or an external bioanalytical group. It is important to recognize that data are placed into domains according to their topic or meaning and not according to their source.

There are three general observation classes where the majority of subject-related data collected during a study are submitted. These classes are as follows:

- **Interventions:** investigational treatments, therapeutic treatments, and surgical procedures administered to the subject or animal. One record per constant dosing/treatment interval.
- **Events:** occurrences or incidents independent of planned study evaluations occurring during the trial (e.g., adverse events) or prior to the trial (e.g., medical history). One record per event.
- **Findings:** observations resulting from planned evaluations (e.g. lab tests, ECGs, microscopic findings). One record per finding result or measurement.

In addition to these general domain classes, the SDTM defines several special-purpose domains that do not use the general-observation-class models. Examples relevant to this paper include the following:

- The Demographics domain describes the essential characteristics that apply to study subjects such as treatment assignment and study start and stop dates.
- The Supplemental Qualifiers (SUPP--) datasets play an important role in the submission of variables that cannot be mapped into the standard domain variables.

- The Related Records (RELREC) dataset is used to describe collected relationships between records in two (or more) datasets, such as an Event record and an Intervention record, or a Finding record and an Event record. RELREC is also used to describe relationships between datasets.

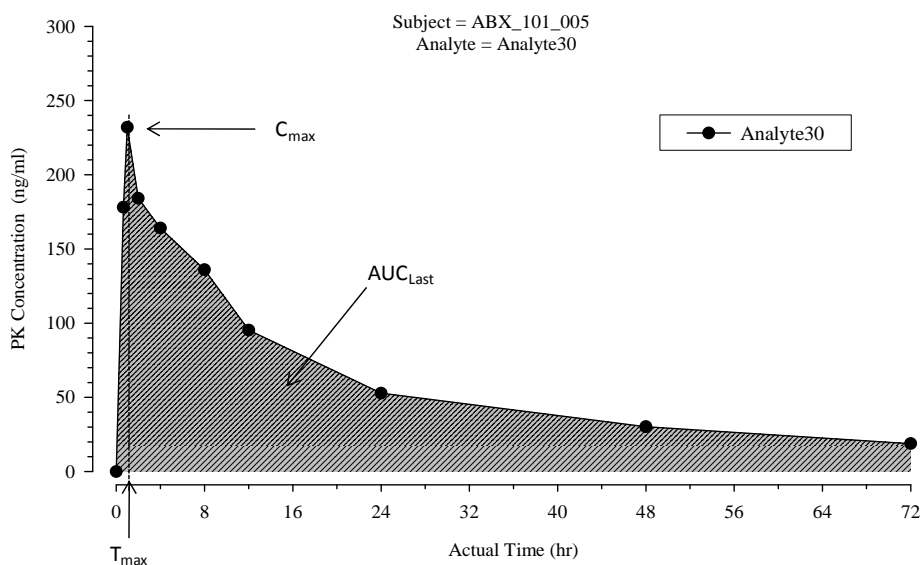
The SDTMIG defines that each domain is represented by a single ‘flat file’ dataset with rows representing observations and columns representing variables. The detailed summary of the SDTM and the SDTM Implementation Guide (SDTMIG) has been described previously (1, 2). Although the SDTM IG does not explicitly specify the data format, it has many provisions to ensure compatibility with SAS® Version 5 Transport files, which the FDA currently expects (3).

## REPRESENTATION OF PHARMACOKINETIC DATA IN SDTM DATASETS

One important part of medical research studies is to analyze and understand the pharmacokinetic and pharmacodynamic characteristics of the drug under development. After a drug is administered to study subjects, concentrations of the drug (and often metabolites) in blood and/or urine are analyzed at various times. These data would be submitted in the SDTM-based Pharmacokinetic Concentrations (PC) domain. This domain is a ‘Findings’ domain with a structure of one record per time-point concentration or sample characteristic per reference dose. This data structure means that a time-concentration profile is stored in multiple records.

The time-concentration profiles are analyzed for the pharmacokinetic characteristics of the drug and related analytes, resulting in multiple pharmacokinetic parameters for each profile, such as  $C_{max}$ ,  $t_{max}$ , half-life, and various AUC values (see Figure re 1). These are submitted in the Pharmacokinetic Parameters (PP) domain. The PP domain is also a Findings domain, and its structure is one record per PK parameter per time-concentration profile per subject. As mentioned above, the PP domain data are not really observed data, but are derived from the time-concentration (PC) data. As such, PP data are not true tabulation data; however, the submission of PK Parameters fits well into a Findings general-observation-class model, and most FDA reviewers are interested in seeing PP data represented in a familiar format of an SDTM-based dataset.

**Figure 1. Illustration of Common PK parameters  $C_{max}$ ,  $T_{max}$ , and AUC for a Time-Concentration Profile for a Drug Following Single-Dose Oral Administration**



Because each set of parameters is related to a specific set of concentrations (a curve), representing this relationship is critical. It should be noted that the SDTMIG v3.1.2 states the following:

*“It is a requirement that sponsors document the concentrations used to calculate each parameter. For many sponsors, this need is currently met via the analysis metadata. As a result of feedback received from many sponsors on the draft version of this document, sponsors may continue to document the concentrations used to calculate each parameter via the analysis datasets.”*

The following section provides additional advice to sponsors wishing to document these relationships using the RELREC dataset. The SDTM uses the Related Records (RELREC) dataset for this purpose. This dataset provides one record per related record, group of records, or dataset.

## IMPLEMENTATION OF RELREC IN THE SUBMISSION OF THE PK DOMAINS

The foundation for relating PK parameters to PK concentration curve, as described in the SDTMIG, is ensuring that two key values are the same in both domains:

- The date of the reference time point (usually a dose), submitted in the variables PCRFTDTC and PPRFTDTC
- The analyte, which is submitted in PCTEST and PPCAT (since PptestCD/Pptest variables contain the names of the PK parameters).

The SDTMIG describes various scenarios for representing which concentrations are used in the calculation of each parameter or group of parameters. All of these use the Related Records (RELREC) dataset.

### Simple Methods for Relating PC and PP Records

The simplest and easiest method for relating PP parameters to PC concentrations can be used when all concentrations are used to calculate all parameters. This may be by design, or may occur when a sponsor uses a commercial program that doesn't provide specifics regarding its algorithms for the exclusion of specific concentration values from the calculation of certain parameters.

Since the structure of the RELREC table allows only one variable (IDVAR) from each domain to be used in a dataset-to-dataset relationship, and the above bullets show that two are needed, the examples in the SDTMIG show the use of the --GRPID variables. Each grouping consists of a combination of the reference time point and the analyte, as shown in the following table.

PCGRPID Value	Description
DY1_DRGX	All concentrations for Drug X dosed on Study Day 1
DY1DRGX	All parameters for Drug X dosed on Study Day 1
DY11_DRGX	All concentrations for Drug X dosed on Study Day 11
DY11DRGX	All parameters for Drug X dosed on Study Day 11

Once the values of --GRPID have been established, the appropriate records in each dataset for each subject can be related. The table below is one from the SDTMIG, and shows the RELREC records for such a many-to-many (sets of concentrations to sets of parameters) relationship. The RELREC table uses the same value of RELID to show that the respective groups of records are related.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC-123	PC	ABC-123-0001	PCGRPID	DY1_DRGX	MANY	1
ABC-123	PP	ABC-123-0001	PPGRPID	DY1DRGX	MANY	1
ABC-123	PC	ABC-123-0001	PCGRPID	DY11_DRGX	MANY	2
ABC-123	PP	ABC-123-0001	PPGRPID	DY11DRGX	MANY	2

The table above is just for one subject. The content of the RELREC table for a study could be simplified if the same value was used for PCGRPID and PPGRPID, and if all concentrations were used for the calculation of all parameters for all subjects. The result would be a two-record RELREC for the study. This representation can be interpreted as follows: for all subjects, when the values of PCGRPID and PPGRPID are the same, there is a relationship between the sets of records.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC-123	PC		PCGRPID		MANY	1
ABC-123	PP		PPGRPID		MANY	1

### More Complex Methods for Relating PC and PP Records

In some cases, not all concentrations may be used the calculation of all PK parameters, and different sets of concentrations may be used to calculate different parameters. The SDTMIG offers examples for different situations with different methods to create the required RELREC records. This section will describe a few of these.

#### One-to-One Method for Relating Records

In this method, the PCSEQ and PPSEQ variables are used to identify the records to be related in the respective domains. In this structure, each record in the PP domain (i.e., each PK parameter) is related to the records from the PC domain that were used to calculate its value. Because this method requires the inclusion of each related record in the RELREC dataset, it will result in a relatively large number of records. On the other hand, the implementation of this method is a rather straightforward two-step process for each subject:

1. Create a RELREC record for each PP record.
2. Go through the list of PC records and create a RELREC record with the same RELID for each record that was used to calculate the PK parameter in the PP record.

The fact that this does need to be done for each subject may take some time and effort on the sponsor's part.

Many-to-Many Method for Relating Records

This method results in fewer RELREC records, but would likely require more time and effort than the one-to-one method described above. The reason for this, and the difference between the two methods, is that either PC records, PP records, or both can be grouped together using --GRPID, with the value being unique for each grouping of records.

Creating a table such as Table 1a below, with a checkmark used to indicate whether a certain PC record is used in the calculation of each PP record, is a useful first step. This example shows that the PK parameters represented in Columns 1 and 5 used all PC records except those in rows B and E; the PCGRPID is given a value of BE (values for --GRPID are sponsor defined). The creation of PPGRPID values can be done once all the PCGRPID values have been identified, as shown in Table 1b.

**Table 1a. Example to Illustrate the Complexity of Many-to-Many Method for RELREC**

		PP Records							
		1	2	3	4	5	6	7	8
PC Records	A	✓	✓	✓	✓	✓	✓	✓	✓
	B	x	x	✓	✓	x	✓	✓	x
	C	✓	✓	✓	✓	✓	✓	✓	✓
	D	✓	✓	✓	✓	✓	✓	✓	✓
	E	x	✓	x	x	x	✓	x	✓
	F	✓	✓	✓	✓	✓	✓	✓	✓
	G	✓	✓	✓	✓	✓	✓	✓	✓
	H	✓	x	✓	✓	✓	x	✓	✓
	I	✓	✓	✓	✓	✓	✓	✓	✓
	J	✓	✓	✓	✓	✓	✓	✓	✓
	K	✓	x	✓	✓	✓	x	✓	✓

PCGRPID	BE	BHK	E	E	BE	HK	E	B
---------	----	-----	---	---	----	----	---	---

**Table 1b. Summary of PC Concentration Use by Parameters**

PCGRPID	PK Parameters	PPGRPID
BE	1, 5	GRP1
E	3, 4, 7	GRP2
HK	6	GRP3
BHK	2	GRP4
B	8	GRP5

With the values shown in Tables 1a and 1b, the many-to-many RELREC dataset would look like that in Table 1c.

**Table 1c. Many-to-Many RELREC Example**

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
1	ABC-123	PP	ABC-123-0001	PPGRPID	GRP1		1
2	ABC-123	PC	ABC-123-0001	PCGRPID	BE		1
3	ABC-123	PC	ABC-123-0001	PPGRPID	GRP2		2
4	ABC-123	PP	ABC-123-0001	PCGRPID	E		2
5	ABC-123	PC	ABC-123-0001	PPGRPID	GRP3		3
6	ABC-123	PC	ABC-123-0001	PCGRPID	HK		3
7	ABC-123	PP	ABC-123-0001	PPGRPID	GRP4		4
8	ABC-123	PC	ABC-123-0001	PCGRPID	BHK		4
9	ABC-123	PC	ABC-123-0001	PPGRPID	GRP5		5
10	ABC-123	PC	ABC-123-0001	PCGRPID	B		5

One-to-Many and Many-to-One Methods for Relating Records

The SDTMIG includes two additional hybrid methods for relating PC and PP domain records. The one-to-many method to express the relationship uses the PCSEQ and the PPGRPID variables (i.e. the PC records are identified individually, while the PP records are identified by a value in --GRPID. The many-to-one method is the reverse of the one-to-many method: It

uses the variable PCGRPID to identify the PC records as a group and the PPSEQ variable to identify the PP records individually.

There are also numerous other methods that might involve the use of some grouping of PC or PP records (using --GRPID) mixed with references to single records (using --SEQ) within each dataset. It is up to each sponsor to determine the best way to represent the relationships between PP and PC records.

## Summary

The SDTMIG describes several methods to relate PC and PP records together with detailed use cases. Clearly, the simplest RELREC dataset is for a study where all PC records are used for calculation of all tabulated PK parameters and therefore the many-to-many method can be implemented without any complicated assignment of values to --GRPID variables. From our experience, however, the real world demands more complexity: the one-to-one method provides a lot of flexibility, and can be implemented easily. The resulting larger RELREC datasets from using the one-to-one method might be seen as a drawback to some, and an implementation that would take advantage of grouping in either of the domains via one of the hybrid methods may be seen by some to be desirable. Hybrid methods add complexity, however, and their use may provide no tangible benefits.

Other than the case where all PC records are related to all PP records, it can be challenging to create the RELREC dataset. Software tools that are typically used for PK analysis don't often provide the necessary information in a way that can be used to automatically create the RELREC records. Tool vendors need to incorporate this into their roadmaps and should offer solutions quickly to further support the adaptation of the SDTM.

## CONTROLLED TERMINOLOGY FOR PK PARAMETERS AND UNITS

Most of the variables used in the domain models of Version 3.1.2 of the SDTM have referenced controlled terminology (CT) in that version. Since the publication of that version in 2008, CDISC has been active in creating CT for variables that did not have it. In 2009, a CT subteam for PK parameters was formed. This team consisted of members from CDISC, NCI EVS, and industry experts from pharmaceutical companies, vendors, and service providers. They were charged with examining the referenced UNIT codelist to determine which terms might be missing, and to develop CT for PK parameters.

### PK Parameter Units

The PK CT subteam team concluded that a separate list for PK units was needed, and held biweekly teleconferences to create that list. As a result, the PKUNIT codelist was created. It contains more than 190 values for PK parameter units. This list, like most CDISC CT lists, is extensible, which means that sponsors can add new terms, as long as they are indeed new terms, and not synonyms of existing terms. For example, the unit of "kg/L" is not in the PKUNIT codelist as CDISC Submission value; however, one can find a submission value of "g/mL" (i.e., Gram per Milliliter), which has the "kg/L" as a synonym. Therefore, the value of g/mL would be the value submitted.

### PK Parameter Tests and Codes

The PK subteam also worked to develop CT for PK parameter (PPTTESTCD and PPTTEST). The most recent version of CT (March 2012) contains more than 200 values. The PK parameter CT sub-team continues to work on adding more required terms and to improve that list. CT Package 10, which should be finalized by Q2 of this year, includes more than 90 additional standard PPTTESTCD and PPTTEST values. The parameter codes and names are based upon those used widely in the industry and in PK analysis programs such as WinNonlin®.

The following points should be helpful in understanding a few of the conventions used in the "PK Parameters Code" codelist:

- Many PK parameters can be calculated using the observed or predicted value of the last non-zero concentration. This dependency is expressed in the code using the letter "O" or "P" in the term. For example, the codes AUCIFP and AUCIFO refer to the  $AUC_{\infty}$  values using the observed or predicted last non-zero concentration value.
- For many PK parameters a 'normalized' term is included in the code list, i.e. there is a distinct term for the original value divided by dose, body mass index, surface area, and weight. The term for the normalized PK parameter is indicated by adding the letter "D", "B", "S", or "W", respectively, to the end of the term for the original term.

There is a small group of PK parameter codes (Submission Values) that require special attention during the implementation of SDTM datasets. First, there are the Area Under Curve (AUC) parameters over an interval given by a start and end time point T1 and T2 as shown in Table 2 below.

**Table 2. AUC parameters over an interval T<sub>1</sub> to T<sub>2</sub>**

<b>Submission Value</b>	<b>CDISC Definition</b>
AUCINT	The area under the curve (AUC) over the interval from T1 to T2.
AUCINTB	The area under the curve (AUC) over the interval from T1 to T2 divided by the body mass index.
AUCINTD	The area under the curve (AUC) over the interval from T1 to T2 divided by the dose.
AUCINTS	The area under the curve (AUC) over the interval from T1 to T2 divided by the surface area.
AUCINTW	The area under the curve (AUC) over the interval from T1 to T2 divided by the weight.
AURCINT	The area under the urinary excretion rate curve (AURC) over the interval from T1 to T2.
AURCINTB	The area under the urinary excretion rate curve (AURC) over the interval from T1 to T2 divided by the body mass index.
AURCINTD	The area under the urinary excretion rate curve (AURC) over the interval from T1 to T2 divided by the dose.
AURCINTS	The area under the urinary excretion rate curve (AURC) over the interval from T1 to T2 divided by the surface area.
AURCINTW	The area under the urinary excretion rate curve (AURC) over the interval from T1 to T2 divided by the weight.

The PK CT team concluded that the start and end point of the interval should be recorded in two specific qualifier variables that should be added to the PP domain, PPSTINT, defined as Time Point of Start of Interval, and PPENINT, defined as Time Point of End of Interval. This normalized approach would be more efficient than defining specific CT values for every possible AUC interval a sponsor might wish to calculate. These variables will be added to a future version of the SDTM; until then, however, they should be submitted in the Supplemental Qualifier dataset for the PP domain.

The same more-normalized approach was used in a second case: The PK CT team decided not to define separate controlled terms for parameters measured under steady state conditions. Instead, PPSSFL (Steady State Flag) was proposed for use as an indicator that a specific PK parameter applies to a steady-state situation. Until this variable is published in a future version of the SDTM, it should be submitted in the Supplemental Qualifier dataset for the PP domain.

## **OBSTACLES IN CONVERTING LEGACY DATA**

### **Source-Data Issues**

When preparing to convert legacy studies to SDTM, sponsors may be faced with some surprises before the conversion can begin. PK studies have often been completed long before the SDTM conversion process is started, and the source datasets have not been evaluated for completeness. This is often the case for drugs which have been purchased or acquired.

#### Lack of Electronic Data

One issue we have seen in our legacy-data conversion work is that the PK concentration and parameter tables were created for the clinical study report but, for whatever reason, the PK datasets were not maintained. Sometimes the CRF collected only portions of the PK data. The collection times, for example, are available, but there are no concentrations or parameters. At this point, many sponsors have decided to enter the data back into SAS®, thereby recreating the original PK data. One common problem with this method is that the tables often do not include all the data necessary to recreate robust PC and PP datasets. While we recognize the importance of submitting PK data in an electronic format, it is recommended that a sponsor contact their review division in such cases to inquire about the necessity of performing this task, because of its time-consuming nature.

#### Lack of Metadata

In some cases, legacy PK dataset(s) are available, but metadata describing the variables in the datasets does not exist. This presents a challenge, since PK datasets we receive tend to contain rather cryptic variable names, no labels, and no external documentation. This problem is further compounded by the fact that there is often no one available at the sponsor company to answer any questions. If, for example, there are parameters present in the dataset that were not listed in the PK report, it can be difficult to determine exactly what is being reported. Similarly for the PC dataset, determining what “analyt\_a” and “analyt\_b” represent, in the absence of a study report where one could check the listings, can be difficult if similar results are expected for the two analytes.

#### Inconsistent Data Structures or Format

Octagon has converted PK datasets where the structure of the main source PK dataset was one record per subject per time-concentration curve. This dataset, which contained columns for twenty-five nominal-time concentrations and twelve PK parameters, needed to be split to separate the concentration data from the parameter data. The two datasets resulting from the split needed to be transformed to create the SDTM-based PK Concentrations (PC) and PK Parameters (PP) datasets. The actual date/times for sampling were in a separate dataset, which then needed to be merged onto the PC dataset.

Octagon has also encountered PK data that were supplied by the vendor as one Excel spreadsheet per time-concentration curve per subject, requiring that all the individual-spreadsheet data be combined to create a study-level domain before one could even begin data conversion.

### Common SDTM Mapping Mistakes

#### Using --CAT or SCAT for Arm, Dose, Frequency, and Fed/Fasted State

The category variables, PCCAT, PCSCAT, and PPSCAT, are often used to submit data that belong in the EX or DM domains. Because PPCAT has a clear, predetermined purpose, it is often exempt from this practice. The former variables are often “hijacked” for the purposes of holding the DM variable ARM, or the EX variables of EXTRT, EXDOSE, EXDOSU, and/or EXDOSFRM. Many companies feel that they “need” this data in the PC or PP domains in order to be able to perform their analysis, and they see the category variables as the perfect location. Sponsors should realize that the exposure data is easily located in the EX domain via the SDTM-based PC and PP Timing variables, PCTPTREF, PCRFTDTC, and PPRFTDTC, and that the treatment Arm is easily merged in from the DM domain, using USUBJID as the merge key. The fasting status of subject when data was collected should be submitted in PCFAST.

#### Standardizing BLQ Values

In Version 3.1.1 of the SDTMIG, the SDS Team was not sure how to best represent results that included “<” in the standard variables. The proposed method was for the sponsor to define the rules for standardizing these variables. When the draft PC and PP domains were released, it was shown that the BLQ values should be standardized to “0” in PCTRESC and PCSTRESN. For example a PCORRES of “<2.0” would have a PCSTRESC and PCSTRESN of “0”. CDER provided feedback that this approach was unacceptable, so a change was made (which applies to all domains): results that originally contained “<” or “>” results should maintain the symbol in the --STRESC variable, and that--STRESN should be null. Although this is clearly indicated in v3.1.2, published more than three years ago, the practice of standardizing these values to 0 has not yet faded from common practice.

#### Using Actual Collection Times for Time-Point Variables and Elapsed Times

Another frequent error is to use the actual collection time in the planned time point, PCTPT, and/or the planned elapsed time in PCELTM. In the case of PCTPT, Table 3 shows two examples of the incorrect numbering of time points (PCTPTNUM).

**Table 3. Example to Illustrate Incorrect PCTPT/PCTPTNUM Use**

PCTPT	PCTPTNUM	PCTPT	PCTPTNUM
1.48 HOURS POST	1.47	1.47 HOURS POST	1.5
1.49 HOURS POST	1.48	1.48 HOURS POST	1.5
1.52 HOURS POST	1.52	1.52 HOURS POST	1.5
1.54 HOURS POST	1.53	1.53 HOURS POST	1.5
1.55 HOURS POST	1.55	1.55 HOURS POST	1.5

As can be seen above in the left-hand section of the table, one approach was to create an actual PCTPT and PCTPTNUM for each time point that occurred. On the right-hand section of the table, the PCTPTNUM is the actual planned time point number, but the PCTPT is the actual time the collection occurred. Both of these examples are incorrect because they are depicting actual time points and not planned time points. A similar practice is often observed with PCELTM, populating it with the actual elapsed time instead of the planned elapsed time. All of these values should reflect the planned elapsed time (1.5 HOURS POST) and the time-point number). There is no place in the PC domain for the actual elapsed time; this can be derived from the difference between PCDTC (date/time of sample collection) and PCRFTDTC (date/time of reference time point, which is usually a dose).

### Incorrect Matching of PCRFTDTC and PPRFTDTC

The --RFTDTC variables contain the date/time of the reference time point, usually the date of the relevant dose. Parameters calculated from set of concentrations should have the same value in the PCRFTDTC and PPRFTDTC variables, as noted in the RELREC section above. There have been cases where this information has not been provided in the legacy data. Sometimes these dates can be found in the exposure data.

One of the biggest challenges is presented when a sponsor provides multiple legacy PK concentration and PK parameter datasets from the same study, and no information regarding which ones belong together. If one is lucky, there will be tables, listings, or figures in the clinical study report that will lead to the identification of the related legacy datasets. Equally as challenging are cases where PCTPTREF and PPTPTREF have been populated with values such as “MOST RECENT DOSE” or “STUDY DOSING.” The use of more-precise terms, such as “DAY 1 DOSING” or “DAY 14 DOSING” would facilitate data conversions as well as internal quality checks.

## Using Draft (v0.92) Version Variables

In February of 2006, draft versions (v0.92) of PC and PP were released for public comment. Included in the proposed PC domain were two variables, PCBATID and PCLOQ. These were not part of the SDTM, and thus were not valid variables at the time, but were merely proposed variables which were being considered at the time. They were never added to the SDTM or the modeled PK domains (although --LLOQ is a valid variable in 3.1.2); however, some sponsors implemented them immediately. While those variables can still be found in various "SDTM" PC domains, the increased use automated validation tools is quickly making it clear to sponsors that these variables are not valid SDTM variables.

## Improper Population of PCTESTCD and PCTEST

There has been some tendency to populate PCCAT with the name of the analyte, presumably because PPCAT is used this way. A sponsor who does this then finds themselves needing to create a value for PCTESTCD and PCTEST, and often uses a value such as "Concentration." As noted in previous sections, PCTESTCD and PCTEST should contain the name of the analyte or specimen characteristic, and PCCAT can be used to indicate "ANALYTE" or "SPECIMEN PROPERTY".

## Matching Specimen-Collection Times and Concentrations

In many studies, specimen-collection times are collected on the CRF, while the concentrations are obtained through electronic data transfer (eDT). Octagon has seen cases where the only timing information in the specimen-collection dataset was the sample collection date and the only timing in the concentration dataset was the nominal collection time. Without a common merge key, ensuring that the appropriate results are merged with the appropriate specimen may either require considerable detective work and consume costly resources, or may not be possible at all.

## CONCLUSION

PK data is critical in understanding a drug's safety and determining its dosing frequency. Yet the processes for collection and submission of this data result in challenges to its representation in an SDTM format.

One of the biggest challenges lies in representing the relationship between sets of PK parameters and the concentrations that were used in their calculation. The underlying concepts have been discussed, and options to express these relationships using the Related Record (RELREC) dataset were presented. When all concentrations are used for the collection of all parameters, RELREC can be as simple as two records. When only certain concentrations are used in the calculation of one or more parameters, there are several methods that can be used to represent the relationships using RELREC. From current experience, the authors see the one-to-one method as the superior option, because it combines flexibility with respect to the relationships that need to be tabulated with a rather easy implementation procedure. The relatively large resulting RELREC datasets can be considered a minor drawback.

This paper provided a summary of CDISC Controlled Terminology that has been developed for the PP domain, specifically the PK parameter code and unit variables. Sponsors should be aware that this terminology exists, and, as with all CDISC CT, should have a plan for its implementation.

## REFERENCES

1. Wood, F., and Guinter, T. (2008) Evolution and Implementation of the CDISC Study Data Tabulation Model (SDTM). *Pharmaceutical Programming* 1 (1): 20-27.
2. Wood, F. (2008) The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics. *PharmaSUG Proceedings*, June 2008.
3. Study Data Specifications, Version 1.6 (2011). <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM199759.pdf>

## CONTACT INFORMATION

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

Fred Wood  
Vice President, Data Standards Consulting  
Octagon Research Solutions, Inc.  
585 East Swedesford Road  
Wayne, PA 19087  
484-881-2297  
fwood@octagonresearch.com

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.