Basic Data Structure for Population Pharmacokinetic (popPK) Analysis

ISoP Data Standards
10-Nov-2020

1. Introduction

Pharmacokinetics (PK) is the study of the effect of the body on a drug, the time course of absorption, distribution and elimination processes of a drug in the body. Population PK (popPK) is a model-based representation of PK processes with a statistical component, enabling identification of the sources of inter- and intra-individual variability. This analysis approach is well suited for analysis of large heterogeneous PK datasets generated as part of standard multi-study clinical programs and are often used as a basis for simulations to inform dose selection and other milestones of drug development. Such analyses are performed regularly throughout the drug development cycle, and represent important components of regulatory submission dossier.
PopPK data are longitudinal, with a degree of complexity which accounts for data items corresponding to time, drug concentration measurements, dosing schedule of the drug, and anonymized individual identifiers. Specifically, the input analysis data need to associate individual-level subject drug concentrations with study drug dosing and specific timing variables to relate concentrations of drug (PK) to the time from dose, and include individual physiological and demographic characteristics. Such datasets should also support exclusion of specific records to facilitate model-based sensitivity analyses. Nonetheless, such data lend themselves to standardization as population pharmacokinetic data and its models need to be typically constructed and used with software packages requiring defined (PK analysis data as its) input. A popPK data standard within the established standardized data tabulations model (SDTM) framework which is compliant with ADaM standards will improve quality of popPK analyses and their interpretation within the context of clinical trial research, facilitating improvements across scientific and regulatory interactions.

The purpose of this document is to present the Basic Data Structure (BDS) and required extensions for popPK. The ADaM Implementation Guide (ADaM IG) supports many of the variables needed for popPK analysis and provides general naming conventions that can be leveraged in the definition of other variables. The ADaM popPK Implementation Guide provides specific guidance for all common variables and may be viewed as the ADaM BDS model plus additional popPK variables. In this document, the ADaM popPK dataset is referred to as an ADPPK dataset. This does not imply a required naming convention. The popPK dataset should be named ADPPK following the ADaM standard naming convention, as described in the Analysis Data Model, Version 2.1 (referred to in this document as ADaM v2.1).

2. Points to Consider in this Document

In reviewing and applying the ADPPK dataset specifications, the following points should be considered:

- **Analysis-ready:** ADPPK datasets should be created with the objective of being "analysis-ready," and contain the variables needed for the intended use of popPK analysis. The ADPPK dataset may be used to create tables, listings, and figures for observed concentration-time data, summary of subject covariates included in the analysis or reporting and diagnostic purposes consistent with the objectives of an analysis but it is not its primary purpose. In addition to required variables such as subject identifiers, treatment variables, and PK sample variables, and other critical variables included in the analysis dataset may be considered "study-specific" or "analysis-specific" as they may depend on the specific nature of the disease / indication / analysis objective. Some variables may be derived and/or populated by the analysts (pharmacometricians) during the process of popPK analysis to document analysis-specific information such as outlier observations. Variables needed to accomplish this are referenced in the metadata. It is out of scope of the ADPPK to define all the possible variables that might be needed to achieve the goals for all the specific analyses. The term "analysis-specific" also refers to specific requirements of a software platform for nonlinear mixed effect modeling (e.g., NONMEM, Monolix, Phoenix NLME, R). Those requirements usually involve particular naming of variables, which can easily be achieved by minimal manipulation of ADPPK. It is beyond the scope of this dataset to cover all the specific details of any commercial software. Therefore, ADPPK is popPK data standard and not NONMEM or any other software data standard.

- **Identification of source dataset:** When identifying the source dataset for a variable, the immediate predecessor variable is used, as described in the latest stable version of ADaM [1]. The dosing and subject level datasets, among others, are considered appropriate predecessor datasets for many ADPPK variables. Dosing datasets may include SDTM EX or EC or a derived exposure datasets, utilizing ADaM standards. If ADaM Data Subject Level Structure (ADSL) variable is not available or a viable option for the purposes of ADPPK generation, it is recommended that applicable SDTM variables as described in the table below be used. Multiple CDISC source datasets may be used to populate ADPPK based on analysis need. Outside of the SDTM PC domain, ADaM sources are expected to be the most common but may not be the only sources used to create ADPPK dataset. All data sources used are to be indicated and provided as supporting information.

- **Rationale for Requiring Optional ADSL Variables in the ADPPK Dataset:** It should be noted that select variables in the ADPPK specification are derived from fields listed in ADSL as 'optional' but are listed as 'required' for ADPPK. Permissible ADSL variables required in ADPPK are necessary for creation of an analysis-ready popPK dataset.

- **Ordering of variables:** The ADaM v2.1 [1] states that the ordering of the variables in the analysis dataset should follow a logical ordering (not simply alphabetic). As such, the specifications of the ADPPK dataset are ordered in a way that is sensible to the

intended use of the dataset such as identification and event variables, time variables, covariates, etc. Within this document, however, no specific ordering of variables within the illustrated datasets is applied, as the tables shown only contain variables relevant to the example. Within this document, the author of each example table applied his or her own analysis-specific ordering.

- **Examples are for illustration only:** The examples in this document are only intended as illustrations. In addition, the examples are intended to illustrate content and not appearance; it is understood that there are different ways that data and results can be displayed. This document does not cover display formats or front-end tools.

- **Display of metadata and dataset examples for illustration of content only:** Though the metadata elements have been defined in the ADaM [1], their display is a function of the mechanism used to display the content. Examples of datasets, formatting and presentation styles used in this document are for the purposes of illustration only, and are not intended to imply any type of display standard or requirement.

- **Examples of variables are not meant to be all inclusive:** The examples in this document describe some of the key variables and records that may be included in the ADPPK dataset. They are not intended to illustrate every possible variable that might be included in the analysis dataset, as many variables are study specific. This is particularly the case for covariates, where only some types (e.g., basic demographics, lab assessments) of variables are included.

- **No endorsement of vendors or products:** In an effort to provide illustrations of the ADaM concepts, examples provided may reference specific programming languages. As with other ADaM documents, references of specific vendor products are examples only, and should not be interpreted as an endorsement of these vendors or products.

3. ADaM Metadata

Table 3.1 illustrates a typical Analysis Dataset Metadata specification. Note that the ADPPK data structure adheres to the ADaM BDS, as is denoted in the "Class of Dataset" field. The dataset structure will follow as below noting that parameter refers to analyte, analysis visit refers to a dose event and analysis time point refers to a PK sample (or a dosing event).

Table 3.1 Example of ADPPK Dataset Metadata

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Key Variables of Dataset	Class of Dataset
ADPPK	Data for Population Pharmacokinetic Analysis	adppk.xpt	One record per subject per analyte, per analysis time point per event (dosing or observation)	USUBJID, ARELTM (Actual time since first dose), DVID, EVID (event ID)	BDS

Table 3.2 describes common variables in an ADPPK dataset. The two rightmost columns ("Core" and "CDISC Notes") provide information about the variables to assist the preparation of datasets. These columns are not meant to be metadata included in the data definition file (i.e., define file), such as define.pdf or define.xml. The "Core" column describes whether a variable is required (Req), conditionally required (Cond), or permissible (Perm). The "CDISC Notes" column provides more information about the variable relevant to the ADPPK dataset. In addition, the "Type" column is being used to define whether the variable is character (Char) or numeric value (Num). Variable units

should be specified in the variable. More specific information related to data type will be provided in data definition file (e.g., variable name, label, type, codes, and comments).

Table 3.2 Population Pharmacokinetics Analysis Variables

Variable Name	Variable Label	Type	Codelist/CT	Core	CDISC Notes
PROJID	Project Identifier	Char		Perm	Text representing protocol or compound name
PROJIDN[1]	Project Identifier (N)	Num		Perm	Unique numerical representation of PROJID
STUDYID	Study Identifier	Char	-	Req	PC.STUDYID. Must be identical to the ADSL variable.
STUDYIDN[1]	Study Identifier (N)	Num		Perm	Unique numerical representation of STUDYID
PART	Part of the Study	Num		Cond	As defined per protocol. Required when study has more than one part, e.g., there can be a dose-escalation part (part A) and dose-evaluation part (part B). In SDTM it is mapped in STUDYID by some sponsor companies.
EXTEN	Extension of the Core Study	Num		Cond	As defined per protocol. Required if study has an extension.
SUBJTYP	Subject Type	Char		Perm	For first-in-human studies, the value can be "Healthy volunteers"
SUBJTYPN [1]	Subject Type (N)	Num		Perm	Unique numerical representation of SUBJTYP

USUBJID	Unique Subject Identifier	Char	-	Req	PC.USUBJID. Must be identical to the ADSL variable.
USUBJIDN	Unique Subject Identifier (N)	Num		Req	Unique numerical representation of USUBJID
SUBJID	Subject Identifier for the Study	Char		Perm	DM.SUBJID or ADSL.SUBJID.
SUBJIDN[1]	Subject Identifier for the Study (N)	Num		Perm	Unique numerical representation of SUBJID
SITEID	Study Site Identifier	Char		Perm	DM.SITEID or ADSL.SITEID
SITEIDN[1]	Study Site Identifier (N)	Num		Perm	Unique numerical representation of SITEID
RECSEQS	Record Sequence Number Within a Subject	Num		Perm	The dataset should be ordered by subject, time, event, dependent variable name and then the sequence should be derived for each subject. Sequential values should start with 1 on the first row for each subject and incrementing by 1 for each subsequent row. It is often used for convenience purposes and merging of PK and PD dataset. It can also be useful for the tracking of outlier exclusion, since the variable is preserved from the original to the final dataset.
RECSEQ	Record Sequence	Num		Perm	Derived sequence for the whole dataset. Sequential values should start with 1 on the first non-header row of the data file (i.e., skipping the variable names) and incrementing by 1 for each subsequent row. These are basically row numbers. It is often used for convenience purposes and merging of PK and PD dataset. It can also be useful for the tracking of outlier exclusion, since the

					variable is preserved from the original to the final dataset.
AVISIT	Analysis Visit	Char		Perm	The analysis visit description; required if an analysis is done by nominal, assigned or analysis visit. AVISIT may contain the visit names as observed (i.e., from SDTM VISIT), derived visit names, time window names, conceptual descriptions (such as Average, Endpoint, etc.), or a combination of any of these. AVISIT is a derived field and does not have to map to VISIT from the SDTM. AVISIT represents the analysis visit of the record, but it does not mean that the record was analyzed. There are often multiple records for the same subject and parameter that have the same value of AVISIT. ANLzzFL and other variables may be needed to identify the records selected for any given analysis. See Section 3.3.8 of the ADaM Implementation Guide for information about flag variables. AVISIT should be unique for a given analysis visit window. In the event, that a record does not fall within any predefined analysis timepoint window, AVISIT can be populated in any way that the producer chooses to indicate this fact (i.e., blank or "Not Windowed"). The way that AVISIT is calculated, including the variables used in its derivation, should be indicated in the variable metadata for AVISIT. The values and the rules for deriving AVISIT may be different for different parameters within the same dataset. Values of AVISIT are producer-defined and are often directly usable in Clinical Study Report displays.
AVISITN	Analysis Visit (N)	Num		Perm	A numeric representation of AVISIT. Since study visits are usually defined by certain timepoints, defining AVISITN so that it represents the timepoint associated with the visit can facilitate plotting and interpretation of the values. Alternatively, AVISITN may be a protocol visit

					number, a cycle number, an analysis visit number, or any other number logically related to AVISIT or useful for sorting that is needed for analysis. Within a parameter, there is a one-to-one mapping between AVISITN and AVISIT so that AVISITN has the same value for each distinct AVISIT. (Best practice would dictate that the mapping would be one-to-one within a study, but that is not an ADaM requirement.) In the event that a record does not fall within any predefined analysis timepoint window, AVISITN can be populated in any way that the producer chooses to indicate this fact (e.g., may be null). Values of AVISITN are producer-defined.)
VISIT	Visit	Char		Perm	Clinic visit captured in STDM domains EX or PC and variable VISIT
VISITNUM	Visit (N)	Num		Perm	Clinic visit captured in STDM domains EX or PC and variable VISITNUM
AFRELTM	Actual Rel Time From First Dose	Num		Req	Actual elapsed time (for sample point or start of sampling interval) from first dose to study treatment. Could be negative. We recommend that variables ending in RELTM be excluded from the TM timing format. Derived from (ADTM of the current event/record of the subject) - (ADTM of the first dosing event/record of the subject)
TPTREF	Time Point Reference	Char		Perm	Description of reference dose, e.g. first dose in the period, last dose in the period or anything in between
APRELTM	Actual Rel Time From Previous Dose	Num		Perm	Derived from (ADTM of the current event/record of the subject) - (ADTM of the previous dosing event/record of the subject)
NFRELTM	Nominal Rel Time From First Dose	Num		Perm	Planned elapsed time (for sample point or start of sampling interval) from first exposure to study

					treatment. We recommend that variables ending in RELTM be excluded from the TM timing format. For PK it will be in PC domain
NPRELTM	Nominal Rel Time From Previous Dose	Num		Perm	For PK it can be derived from NRRELTM or equivant variable that has nominal time in PC domain
ATPT	Planned Time Point Name	Char		Perm	Text description of the planned/protocol time for the specimen collection. Taken from PC.PCTPT
ATPTN	Planned Time Point Number	Num		Perm	PC.PCTPTNUM, numerical representation for PCTPT
ADT	Analysis Date	YYMMDD10.		Perm	The event date associated with AVAL[2] and/or AVALC .or the dose. When AENDT is populated then this will represent the start date of that dose interval. PC.ADT or EX.EXSTDT
ADY	Analysis Relative Day	Num		Perm	Study day, relative to the start of the study. It can be taken directly from some of the ADaM domains (different ADY variables in different ADaM datasets) or as specified by the modeler
ATM	Analysis Time	B8601TM.		Perm	Analysis time associated with AVAL and/or AVALC. When AENTM is populated then this will represent the start date of that interval as would be expected in urinalysis. Analysis time of event or censoring associated with AVAL in numeric format. PC.ATM
ADTM	Analysis Date/Time	E8601DT.		Perm	Analysis date/ and time associated with AVAL and/or AVALC. When AENDTM is populated then this will represent the start date of that dose intervalis.

AENDT	Analysis End Date	YYMMDD10.		Perm	The end date associated with AVAL and/or AVALC. See also ADT.
AENTM	Analysis End Time	B8601TM.		Perm	The end time associated with AVAL and/or AVALC. See also ATM.
AENDTM	Analysis End Date/Time	E8601DT.		Perm	The end date and time associated with AVAL and/or AVALC..
OCC	Occasion	Num		Perm	The implementation is taken directly from the specs/protocol and defined by the modeler
EXCLFC	Comment For The Record Exclusion	Char		Cond	It can be captured prior to the modeling or after some analysis, with possible iterations and adjustments. Some possible reasons: BLOQ, biological implausibility, outlier based on variability metrics, incorrect dosing information, day 1 pre-dose sample. This can also be coded through numbers that are connected with different reasons. Based on data specification. Could come from multiple sources, e.g., SDTM or ADaM source, CRITyFN variable
EXCLF	Record Exclusion	Num		Cond	It can be captured prior to the modeling or after some analysis, with possible iterations and adjustments. Some possible reasons: BLOQ, biological implausibility, IWER>threshold, incorrect dosing information. There can be several reasons for flagging data during the data set creation e.g., day 1 pre dose samples, missing sample information, deviation of actual time from nominal time > threshold, etc.
FLAG	Flagged records	Num		Perm	Captures exclusion flags plus any non-exclusion flags like dose time imputation flags etc. It's a sequential number starting with 1 and increments with each flag.

FLAGC	Comment for the flagged records	Char		Perm	Captures reason for exclusion flags plus any non-exclusion flags like dose time imputation flags etc.
DVID	Dependent Variable Name	Char		Perm	Analyte/drug name. Most of the time ADaM is not available when we prepare PMX datasets, so we pull from STDM. PC.PCTEST for PK, EX for dose, ADLB for labs
DVIDN[1]	Dependent Variable Name (Num)	Num		Perm	Unique numerical representation of DVID.. Convention is normally 0=dose, 1=for the first observation of interest, etc.
CMT	Compartment	Num		Perm	Derived based on the specifications from modeler.
DV[2]	Analysis Value	Num		Req	Numeric analysis value
DVC	Analysis Value (C)	Char		Perm	Character results/findings in a standard format. The purpose of this column is to capture which records are BLQ or ALQ in the DV column. "PC.PCSTRESC (SDTM) AVALC (ADaM)"
DVL	Log Analysis Value	Num		Perm	Natural log. It's "." (CDISC value for null) if value is 0 and log otherwise
EVID	Event Id	Num		Req	Derived based on the standard codes in the codelist
MDV	Missing DV	Num		Req	Derived based on the standard codes in the codelist
PCULOQ	Upper Limit Of Quantitation	Num		Perm	In assay report
PCLLOQ	Lower Limit Of Quantitation	Num		Perm	In assay report or CSR. PC.PCLLOQ
BLQFN[1]	Blq Flag	Num		Perm	0=No,1=Yes

ALQFN[1]	Alq Flag	Num		Perm	0=No, 1=Yes
DOSEA	Actual Treatment Dose	Num		Req	DOSEA represents the actual treatment dosage associated with the record. This is the actual numeric amount of the dose used for the population analysis and may differ from the EX.EXDOSE. It can be derived from the EX.EXDOSE or based on related dose. Populated on all individual records as carry-forward
DOSEP	Planned Treatment Dose	Num		Perm	DOSEP represents the planned treatment dosage associated with the record. This is the numeric amount of the dose used for the p opPK analysis and may differ from the EX.EXDOSE.
AMT	Actual Amount of Dose Received	Num		Req	Only populated on dosing records
DOSCUMA	Cumulative Amount of Dose Received	Num		Perm	Calculated by adding the doses up as we go
DOSETDD	Total Daily Amount of Dose Received	Num		Perm	Used mostly for b.i.d. dosing schemes
DOSEDUR	Duration Of Treatment Dose	Num		Perm	This record is generally considered to be associated with an infusion dose and is distinct from TRTDURD, TRTDURM, and TRTDURY which references the duration of the entire study rather than the duration of a single treatment event. "EX domain

					1) Derived from EXENDTC - EXSTDTC. 2) Taken from EXDUR. In case we need to extend, additional column could be defined Derived from TRTRSDTM - TRTREDTM." The label should contain the unit of time.
RATE	Infusion Rate	Num		Perm	Units are the units of AMT divided by the units of DOSEDUR. AMT/DOSEDUR or calculated otherwise
II	Dosing Interval	Num		Perm	24 for QD, 12 for BID, etc… If time units are hours. Connected with time units in TIMEUNIT Protocol or some inspiration in EX.EXDOSFRQ
ADDL	Number Of Additional Doses	Num		Perm	Number of additional doses like the current one until the next dose, e.g. if the value is 1 then 1 additional dose, if value is 2 then 2 additional doses. It is commonly used for long lasting studies with frequent dosing in order to not have the dataset extremely large.
SS	Steady State	Num		Perm	A steady-state dose is a dose that is imagined to be the last of a series of implied doses, each exactly like the dose in question, given at a regular interval specified by the II data item and leading to steady-state by the time the steady-state dose is given.
FORM	Formulation	Char		Perm	Type of formulation (e.g., tablet, capsule, aerosole) EX.DOSFRM or Protocol
FORMN[1]	Formulation (N)	Num		Perm	Unique numerical representation of FORMN
ROUTE	Route	Char	(ROUTE)	Perm	Route of treatment delivery. May be derived from the EX.EXROUTE.
ROUTEN[1]	Route (N)	Num		Perm	Derived from ROUTE as one-on-one unique match

TRTP	Planned Treatment	Char		Cond	TRTP is a record-level identifier that represents the planned treatment attributed to a record for analysis purposes. Record-level identifier that represents the planned treatment attributed to a record for analysis purposes. This variable should contain both the name of the drug and the dose amount. and may also include information related to delivery of the drug if it is relevant for the analysis.
TRTPN[1]	Planned Treatment (N)	Num		Perm	The numeric code for TRTP. There must be a one-to-one map to TRTP. One-to-one map to TRTP
TRTA	Actual Treatment	Char		Cond	TRTA is a record-level identifier that represents the actual treatment attributed to a record for analysis purposes. TRTA is a record-level identifier that represents the actual treatment attributed to a record for analysis purposes. This variable should contain both the name of the drug and the dose amount and may also include information related to delivery of the drug if it is relevant for the analysis.
TRTAN[1]	Actual Treatment (N)	Num		Perm	The numeric code for TRTA. There must be a one-to-one map to TRTA.
APHASE	Phase	Char		Perm	Higher level categorization than period. APHASE (ADaM IG)
APHASEN[1]	Phase (Num)	Num		Perm	
APERIOD	Period	Num		Perm	Record-level timing variable that represents the analysis period within the study associated with the record for analysis purposes. The value of APERIOD (if populated) must be one of the xx values found in the ADSL TRTxxP

					variables.
APERIODC	Period (C)	Char		Perm	Text characterizing to which analysis period the record belongs. It must be one-to-one mapping within a dataset to APERIOD. It must be identical to the ADSL variable.
ACYCLEC	Analysis Cycle	Char		Perm	Record level identifier that reflects cycle and may be of particular importance for studies that examine concentrations in cancer patients From AVISIT or Protocol
ACYCLEN[1]	Analysis Cycle (N)	Num		Perm	
ARM	Description Of Planned Arm	Char		Perm	Subject level variable. ADSL.ARM
ARMN[1]	Arm (N)	Num		Perm	Derived from ARM. Ensure consistency across studies when pooling
ACTARM	Description Of Actual Arm	Char		Perm	Subject level variable
ACTARMN[1]	Actual Arm (N)	Num		Perm	Derived from ACTARM/ACTARMCD. Ensure consistency across studies when pooling
COHORT	Cohort Subject Enrolled Into	Char		Perm	Subject level variable. Could be some sort of sub-population. Study can have 5 cohorts and 3 arms. DM or ADSL.
COHORTN[1]	Cohort Number Subject Enrolled Into	Num		Perm	Numeric representation of the COHORT variable. There must be a one-to-one mapping to COHORT.

[1] Not required for the initial popPK dataset, but needed in some cases for analysis dataset. It can be easily derived from its corresponding Char pair variable.

[2] DV is a wide-spread used notation in the field of Pharmacometrics since its inception for the value of dependent variable (observation)

Table 3.3 describes common covariates in an ADPPK dataset. As there is often the need to analyze the effect of baseline versus time-varying covariates, they can be distinguished by the suffixes in their name, e.g.,

- *COV* for time-varying covariate, where *COV* is a name of covariate
- *COV*B for baseline covariate, where *COV* is a name of covariate
- *COV*N for numerical version of categorical covariate, where *COV* is a name of covariate and there is one-to-one relationship between *COV* and *COV*N
- COVI can be used to for any imputed covariates.

This table is not exhaustive and does not include every permutation for each covariate and many other covariates may be relevant for a specific analysis.

Table 3.3 Population Pharmacokinetics Analysis Variables - Covariates

Variable Name	Variable Label	Type	Codelist/CT	Core	CDISC Notes
WT	Body Weight	Num		Perm	Variable may be derived from VS.VSTEST and VS.VSORRESU but may also be pulled in from other datasets. Weight is associated with the time of dosing. SDTM, VS or ADVS
WTB	Baseline Body Weight	Num		Perm	Variable may be derived from VS.VSTEST and VS.VSORRESU but may also be pulled in from other datasets. Weight is associated with the time of dosing. Use flags for baseline. SDTM, VS or ADVS
HTB	Baseline Body Height	Num		Perm	Variable may be derived from baseline VS.VSTEST and VS.VSORRESU but may also be pulled in from other

					datasets. SDTM, VS or ADVS
BMIB	Baseline Body Mass Index	Num		Perm	Variable may be derived from baseline VS.VSTEST and VS.VSORRESU but may also be pulled in from other datasets or derived.. ADSL.BMI for baseline and VS.BMI for time-varying or derived
BSAB	Body Surface Area at Baseline	Num		Perm	
AGE	Age	Num		Req	DM.AGE or ADSL.AGE. If analysis needs require a derived age that does not match ADSL.AGE, then AAGE (Analysis Age) must be added.
SEX	Sex	Char	(SEX)	Perm	The sex of the subject is a required variable in ADSL; must be identical to DM.SEX or ADSL.SEX. Single value for all records within a patient. When integrating multiple studies, need to conform the values across studies to show the same convention. For example, the values should be F or FEMALE across the studies but not combination of both the values.
SEXN	Sex (N)	Num		Perm	Numeric version of SEX. From DM or ADSL.
RACE	Race	Char	(RACE)	Perm	The race of the subject is a required variable in ADSL; identical to DM.RACE or ADSL.RACE. May categorize differently if analysis demands.
RACEN	Race (N)	Num		Perm	Numeric version of RACE. Codes used as per the specification.
ETHNIC	Ethnicity	Char		Perm	Derived based on RACE and COUNTRY if not in the SDTM/AdaM datasets

ETHNICN	Ethnicity (N)	Num		Perm	Numeric version of ETHNIC.
REGION	Region	Char		Perm	DM
REGIONN	Region (N)	Num		Perm	DM
COUNTRY	Country	Char		Perm	DM
COUNTRYN	Country (N)	Num		Perm	DM
CREATB	Baseline Creatinine Serum	Num		Perm	LB domain
CRCLB	Baseline Creatinine Clearance	Num		Perm	LB domain
EGFRB	Baseline eGFR	Num		Perm	LB domain
TBILB	Baseline Total Bilirubin	Num		Perm	LB domain
ASTB	Baseline Aspartate transaminase	Num		Perm	LB domain
ALTB	Baseline Alanine transaminase	Num		Perm	LB domain

Table 3.4 Population Pharmacokinetics Analysis Variables - Flags

Variable Name	Variable Label	Type	Codelist/CT	Core	CDISC Notes

ITTFL	Intent-to-treat Population Flag	Char		Perm	ADSL.ITTFL
ITTFN	Intent-to-treat Population Flag (N)	Num		Perm	ADSL.ITTFL with "Yes" or "Y" = 1 and "No" or "N"=0
COMPLFL	Completers Population Flag	Char		Perm	ADSL.COMPLFL
COMPLFN	Completers Population Flag (N)	Num		Perm	ADSL.COMPLFL with "Yes" or "Y" = 1 and "No" or "N"=0
TCOMPLFL	Treatment Completers Population Flag	Char		Perm	ADSL.TCOMPLFL
TCOMPLFN	Treatment Completers Population Flag (N)	Num		Perm	ADSL.TCOMPLFL with "Yes" or "Y" = 1 and "No" or "N"=0
SAFFL	Safety Population Flag	Char		Perm	ADSL.SAFFL
SAFFN	Safety Population Flag (N)	Num		Perm	ADSL.SAFFL with "Yes" or "Y" = 1 and "No" or "N"=0
VOMITFL	Flag for Vomit Related To Dose	Char		Perm	
FOODFL	Food Status	Char		Perm	

4. Examples for derivation of some typical covariates and their baseline values

Height (cm)

- Use the value(s) indicated by the baseline flag in the source dataset (subject level source dataset like adsl)
- If baseline flag is missing use the last result up to and including Day 1, unless otherwise specified in SAP
- If more than one record at baseline and the values differ, consult statistician
- Leave as missing (code to a numeric number if tool does not accept missing, like -99)

Weight (kg)

- Use the value(s) indicated by the baseline flag in the source dataset (subject level source dataset like adsl)
- Use the last result up to and including Day 1, unless otherwise specified in SAP
- If more than one record at baseline and the values differ, consult statistician
- If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Ideal Body Weight (kg) [13]

When a patient's height is over 60 in Male (kg): = 50 + 2.3 (Height (in) - 60) Female (kg) = 45.5 + 2.3 (Height(in) - 60) When a patient's height is 60 in or less, the IBW is 50 kg for male and 45.5 for female. If height in cm, use to convert : 1 cm = 0.3937 in	• Use from source data If already available (subject level source dataset like adsl) • If deriving using the formula provided use baseline height • If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Body Mass Index (kg/m^2)

$WT[kg]/(HT[m])^2$ If height in cm, use to convert : 1 cm = 0.3937 in	• Use from source data If already available (subject level source dataset like adsl) • If deriving using the formula provided use baseline height • If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Body Surface Area (m^2) [14]

$0.007184*WT[kg]^{0.425}* HT[cm]^{0.725}$	• Use from source data If already available (subject level source dataset like adsl) • If deriving using the formula provided use baseline height • If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Age

- Use the value(s) indicated by the baseline flag in the source dataset (subject level source dataset like adsl)
- If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Baseline Lab values

- Use the value(s) indicated by the baseline flag (subject level source dataset like ADSL)
- If baseline flag is missing use last result up to and including Day 1
- If more than one record at baseline and the values differ, consult statistician
- If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

CrCL derivation [15] - This may not be the only way to derive, there are other potential derivations one could use as per analysis requirement

Male: ((140-age in years)*weight in kg) / (72*creat) Female: above value*0.85	Use the subjects creatinine clearance at baseline in mL/min Baseline Values: Age(yrs) ,weight(kg),creat in mg/dL If the population is obese (ie wt>=1.2*ibw then use: Male:(((140-age)*ibw/(72*creat))) Female:(((140-age)*ibw/(72*creat)))*0.85 • Variable Values used in derivation should be prior to the rounding off • If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

GFR [16] - This may not be the only way to derive, there are other potential derivations one could use as per analysis requirement

eGFR (mL/min/1.73 m^2) = **175** x (Scr, std)$^{-1.154}$ x (Age)$^{-0.203}$ x (0.742 if female) x 260 (1.212 if African American) 261 262 Scr, std: serum creatinine measured with a standardized assay

- Variable Values used in derivation should be prior to the rounding off
- If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

Liver dysfunction groups [17]

GROUP A: Normal (total bilirubin ≤ULN and AST ≤ULN
GROUP B: Mild (total bilirubin > 1.0x-1.5xULN or AST > ULN)
GROUP C: Moderate (total bilirubin > 1.5x-3xULN)
GROUP D: Severe (total bilirubin > 3xULN)

- The values of the lab covariates should be considered at baseline.

Baseline Performance Status

<table>
<tr>
<td>If deriving from KPS [18]:
<table>
<tr><th>KPS</th><th>ECOG</th></tr>
<tr><td>100</td><td>0</td></tr>
<tr><td>90
80</td><td>1</td></tr>
<tr><td>70
60</td><td>2</td></tr>
<tr><td>50
40</td><td>3</td></tr>
<tr><td>30
20
10</td><td>4</td></tr>
</table>
</td>
<td>

Use the ECOG/KPS value(s) indicated by the baseline flag
If baseline flag is missing use the value(s) on the nearest subsequent visit record
If more than one record at baseline and the values differ, consult statistician
If missing leave as missing (code to a numeric number if tool does not accept missing, like -99)

</td>
</tr>
</table>

5. Examples for handling of missing values

Imputation of missing covariates can be done by the pharmacometrician based on the analysis needs.

Time Variant Covariates:

- For deriving time variant covariates use the time varying results for each of the covariates like weight, age. For time varying IBW, BMI, BSA, CRCL, GFR use the time varying results of height, weight and other elements in the equation on the same date/day.
- If missing use last observation carried forward (LOCF)

Missing date and time records: Imputations should be avoided as much as possible. It may be required in some cases where complete dosing history is not available. It's important to have CRFs collect the required information.

PK sample date and Time Imputation

If date and/or time missing or partially missing

- Leave as missing and flag the records
- Impute based on nominal time relative to the dose after reviewing the profile of the subject within that occasion
- Impute based on other lab date and time in the same window of PK
- The imputed date/time records should be flagged

Dose clock time imputation for CRF designs where every single dose date and time is captured

If a dose date is available but time is missing:

Note: Account for missing doses/dose interruptions based on the number of tablets or comment in text like variable

- If a trough sample was taken on the same day, the trough sampling time is used as the dosing time. One can add 5 min to the sample time to impute dose time.
- If dose time is missing at the date when post dose PK samples are available, use the first post dose sample to back impute the dosing time (eg, the first post dose sample is 2 hour post dose, subtract 2 hours from the actual sampling time of the 2 hour post dose sample to obtain the dosing time)
- If there is no PK sample associated with the missing dose time, it will be imputed by using next or previous available dosing time or an arbitrary nominal/expected time. For BID/TID adjust the imputation based on frequency
- If day 1 dose has no time and no associated PK samples, baseline lab/pharmacodynamic (PD) time will be used
- The imputed clock time records should be flagged

Dose clock time imputation For IV Doses, if dose date is available but time is missing:

- If infusion stop time is available but infusion start time is missing, the protocol defined duration (e.g. 1 hour or 30 min) is used to determine the start of infusion and vice versa if stop time is not available.
- If both infusion stop and start times are missing on any day other than day 1
  a. If a trough sample was taken on the same day, use the trough sample time as the dosing time (start of infusion).
  b. If there is no trough but an EOI sample is taken on the same day then use EOI sample time minus duration of infusion to impute start time of the dose (start of infusion).
- If both infusion stop and start times are missing on day 1 then use pre-dose sample time or end of infusion sample time on the same day along with nominal infusion duration to determine the start of infusion time.

- If there are no concentrations associated with the missing dose time, then the dose time on the previous (or next if it is the first dose) occasion's dose will be used as the current dosing time. This rule will be applied recursively if the dose time is missing for multiple dosing occasions.
- The imputed clock time records should be flagged

Where CRF is designed to capture interval doses with start and stop dates recorded and only dose times relative to PK sample are recorded

If a dose date is missing:

Note: Account for missing doses/dose interruptions based on the number of tablets or comment in text like variable

- Impute using visit date
- Impute using lab date or PK sample date with in the window
- Flag as missing

If a start or stop dose date is missing:

- If start dose date is missing , impute using previous stop date plus dosing interval
- If start date of the first dosing interval is missing use dose date missing rules
- if stop date is missing impute using start date from next interval minus dosing interval
- if stop date of the last dosing interval is missing impute using below:
    - if there is PK/lab on last dose day then use last PK/lab record date
    - impute using last visit date
    - impute using unscheduled visit date as applicable
    - impute using SAP

If dose date is available and time is missing:

- Impute the non-recorded doses with ADDL (number of additional doses) and II (inter dose interval). Account for any dose interruptions.
- If dose date/time is missing on the day of a trough PK sample, it will be imputed as trough date/time

- If dose time is missing at the date when post dose PK samples are available, use the first post dose sample to back impute the dosing time (eg, the first post dose sample is 2 hour post dose, subtract 2 hours from the actual sampling time of the 2 hour post dose sample to obtain the dosing time)
- If there is no PK sample associated with the missing dose time, it will be imputed by using next or previous available dosing time
- If day 1 dose has no time and no associated PK samples, baseline lab/pharmacodynamic (PD) time will be used

- The imputed date/time records should be flagged

Infusion Duration

- If infusion duration is >= +/-100% of protocol defined duration, then impute the duration to protocol defined duration (e.g., 1 hour or 30 min) and flag the dose record and following PK samples
- If infusion stop time is available but infusion start time is missing, the protocol defined duration (e.g. 1 hour or 30 min) is used to determine the start of infusion and vice versa if stop time is not available.
- If both infusion start and stop times are missing on day 1, pre-dose sample time or end of infusion sample time is used along with nominal infusion duration to determine the start of infusion time
- The imputed date/time records should be flagged

Infusion Rate

Rate of Infusion: AMT/DOSEDUR

Dose Amount

If dose information is missing for one or a few occasions of a subject, but treatment information (x mg/kg) is available:

- If missing for one occasion and everything around it is consistent then impute missing dose by using last observation carried forward approach.
- If flat dosing then use the information from the treatment variable
- Flag all records that have imputed amount.

6. References

This is a list of references for the guidance on population pharmacokinetic dataset and analysis practices.

1. Analysis Data Model (ADaM) version 2.1: http://www.cdisc.org/adam
2. N Thanneer, A Roy, P Sukumar, J Bandaru, E Carleen. 2014. 'Best practices for preparation of pharmacometric analysis data sets', ACoP, M-058.
3. Bonate, P. L., A. Strougo, A. Desai, M. Roy, A. Yassen, J. S. van der Walt, A. Kaibara, and S. Tannenbaum. 2012. 'Guidelines for the quality control of population pharmacokinetic-pharmacodynamic analyses: an industry perspective', *AAPS J*, 14: 749-58.
4. Byon, W., K. Sweeney, C. Frost, and R. A. Boyd. 2017. 'Population Pharmacokinetics, Pharmacodynamics, and Exploratory Exposure-Response Analyses of Apixaban in Subjects Treated for Venous Thromboembolism', *CPT Pharmacometrics Syst Pharmacol*, 6: 340-49.
5. Dykstra, K., N. Mehrotra, C. W. Tornoe, H. Kastrissios, B. Patel, N. Al-Huniti, P. Jadhav, Y. Wang, and W. Byon. 2015. 'Reporting guidelines for population pharmacokinetic analyses', *J Clin Pharmacol*, 55: 875-87.
6. Gieschke, R., and J. L. Steimer. 2000. 'Pharmacometrics: modelling and simulation tools to improve decision making in clinical drug development', *Eur J Drug Metab Pharmacokinet*, 25: 49-58.
7. 'Guidance for industry on Population Pharmacokinetics; availability. Food and Drug Administration, HHS. Notice'. 1999. *Fed Regist*, 64: 6663-4.
8. Mould, D. R., and R. N. Upton. 2012. 'Basic concepts in population modeling, simulation, and model-based drug development', *CPT Pharmacometrics Syst Pharmacol*, 1: e6.
9. Radivojevic, A., B. Corrigan, N. Downie, R. Fox, J. Fiedler-Kelly, H. Liu, M. Melhem, D. Radke, P. Schaefer, J. Su, M. J. Swat, N. S. Teuscher, N. Thanneer, A. Zong, and J. J. Wilkins. 2018. 'Data standards for model-informed drug development: an ISoP initiative', *J Pharmacokinet Pharmacodyn*, 45: 659-61.
10. Sun, H., E. O. Fadiran, C. D. Jones, L. Lesko, S. M. Huang, K. Higgins, C. Hu, S. Machado, S. Maldonado, R. Williams, M. Hossain, and E. I. Ette. 1999. 'Population pharmacokinetics. A regulatory perspective', *Clin Pharmacokinet*, 37: 41-58.
11. Wade, J. R., M. Edholm, and T. Salmonson. 2005. 'A guide for reporting the results of population pharmacokinetic analyses: a Swedish perspective', *AAPS J*, 7: 45.
12. Williams, P. J., and E. I. Ette. 2000. 'The role of population pharmacokinetics in drug development in light of the Food and Drug Administration's 'Guidance for Industry: population pharmacokinetics'', *Clin Pharmacokinet*, 39: 385-95.
13. Devine BJ. Gentamicin therapy. Drug Intell Clin Pharm 1974;8:650–5.
14. Du Bois D, Du Bois EF (Jun 1916). "A formula to estimate the approximate surface area if height and weight be known". Archives of Internal Medicine. 17 (6): 863–71.
15. Cockcroft, D.W. and M.H. Gault. Prediction of creatinine clearance from serum creatinine. Nephron. 1976. 16(1):31-41.
16. Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, Kusek JW, Van Lente F; Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med. 2006 Aug 15;145(4):247-54. doi: 10.7326/0003-4819-145-4-200608150-00004. Erratum in: Ann Intern Med. 2008 Oct 7;149(7):519. PMID: 16908915.
17. Ramalingam SS, Kummar S, Sarantopoulos J, et al. Phase I Study of vorinostat in patients with advanced solid tumors and hepatic dysfunction: A National Cancer Institute Organ Dysfunction Working Group Study. J Clin Oncol 2010;28:4507-12.
18. Ma C, Bandukwala S, Burman D, Bryson J, Seccareccia D, Banerjee S, Myers J, Rodin G, Dudgeon D, Zimmermann C. Interconversion of three measures of performance status: an empirical analysis. Eur J Cancer. 2010 Dec;46(18):3175-83. doi: 10.1016/j.ejca.2010.06.126. Epub 2010 Jul 30. PMID: 20674334.

APPENDIX

Discussion points with ADaM team:

- Naming of the variables: CDISC vs. pharmacometrics-specific