PharmaSUG China 2025 – Paper DS133

From Data to Regulatory: Best Statistical Programming Practices for Population PK, Concentration-QTc and Exposure-Response Analyses

Xiaohan Ye, Lin Liu, Yuyan Wang, and Meng Qu, BeOne Inc.

ABSTRACT

Population Pharmacokinetics (PopPK) and Exposure-Response (E-R) within the scope of Human Pharmacokinetics, as well as Concentration-QT Interval corrected (C-QTc) in Human Pharmacodynamics, are critical components of clinical pharmacology data analysis. However, the role, methodology, and preparation of submissions for these analyses are often unfamiliar to many statistical programmers, with limited guidance available. This paper integrates existing literature and real-world clinical experience to clearly define the specific responsibilities and contributions of statistical programmers in PopPK, E-R, and C-QTc analyses. It also outlines standardized requirements for submission materials, including dataset definitions, the application of required datasets, and compliance with CDISC standards. Through a systematic summary, this paper aims to provide practical guidance and references for statistical programmers, reducing regulatory risks associated with non-compliant submission materials and contributing to cost and time efficiency in drug development.

INTRODUCTION

BACKGROUND

With the development of the biopharmaceutical industry, Model-Informed Drug Development (MIDD) has become a critical tool in new drug development. By simulating the behavior of drugs in the human body through mathematical models, MIDD provides scientific support for dose optimization, efficacy prediction, and safety assessment. MIDD encompasses multi-disciplinary analyses, among which human pharmacokinetic and pharmacodynamic studies related to PK are not widely familiar to many. During the early dose-exploration phase, Population Pharmacokinetics (PopPK) and Exposure-Response (E-R) in Human Pharmacokinetic scope and Concentration-QT Interval corrected (C-QTc) in human pharmacodynamic serve as key focuses for clinical pharmacology. To better facilitate model exploration, collaboration between the Clinical Pharmacologists (CPs) and the Statistical Programmers (SPs) is essential for efficient data processing.

In these analyses, statistical programmers play a vital role in integrating data from multiple sources, ensuring the consistency and traceability of analysis results, and standardizing and streamlining analysis workflows through programming. They must also possess a deep understanding of regulatory requirements to ensure that submitted materials comply with Food and Drug Administration (FDA) or other health authorities (HAs) guidelines, balancing compliance with analytical efficiency—an increasingly critical challenge for the industry.

RESEARCH OBJECTIVES AND SIGNIFICANCE

This paper aims to define the specific responsibilities and contributions of statistical programmers in PopPK, E-R, and C-QTc analyses. It will also summarize the submission standards for PopPK, E-R, and C-QTc analyses, covering dataset definitions, the use of required datasets, and compliance with CDISC standards.

Through this systematic summary, we hope to provide practical guidance and references for statistical programmers in their work, reducing regulatory risks caused by non-compliant submissions, and saving costs and time in drug development. Moreover, this paper seeks to

enhance the industry's recognition and understanding of the role of statistical programmers, fostering further progress in the biopharmaceutical field.

POPULATION PK

During the clinical trials, especially in early research, the Pharmacokinetic (PK) analysis plays a critical role throughout the lifecycle at multiple points in the drug development process, e.g., IND, NDA and BLA, or post-marketing stage. Usually, PK can be divided into two analyses: Individual PK and population PK (PopPK). Unlike individual PK, PopPK analyzes concentration–time data across a population to quantify variability and identify covariate effects. By modeling absorption, distribution, metabolism, and excretion (ADME) processes, PopPK supports dose optimization and informs decision-making for diverse patient groups (Namrata P., 2020).

POPPK INTRODUCTION

PopPK is a model-based representation of PK processes with a statistical component, enabling identification of the sources of inter- and intraindividual variability. It is suited for the analysis of large heterogeneous PK datasets generated as part of standard multi-study clinical programs and often used as a basis for simulations to inform dose selection and other milestones of drug development (Chen WJ, 2017).

POPPK IN CLINICAL INDUSTRY

In clinical trial industry, PopPK performed regularly throughout the drug development cycle and represent important components of a regulatory submission dossier. Since 1972, the concept of "population Pharmacokinetic "was created, it has developed quickly by recent years ((Population Pharmacokinetics | FDA, CDISC Basic Data Structure for ADaM PopPK Implementation Guide v1.0, 2023). For submission guidelines, FDA has also released several requirements of PopPK Datasets in eCTD submissions (FDA eCTD Technical Conformance Guidance).

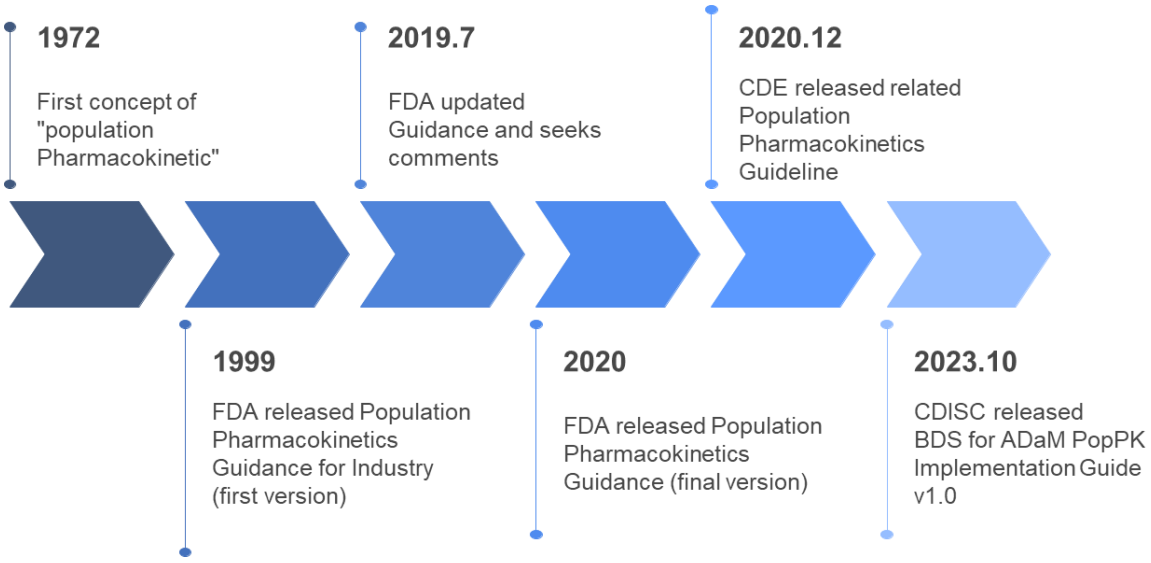


Figure 1 Timeline of PopPK Development

POPPK ON NONMEM

PopPK analysis usually needs cross functions to cooperate, which includes Clinical Pharmacology and Statistical programming. Once the PopPK dataset is ready, it will be used as the input in Non-Linear Mixed Effects Modeling (NONMEM) software by Pharmacology (Raghu K. 2014). It is common to use these tools like NONMEM, Monlix, Phoenix NLME to analyze. They

are all powerful software tools designed for the analysis of pharmacokinetic (PK) and pharmacodynamic (PD) data, with requirements of a very specific pre-defined format and includes a mixture of data components like dosing records, PK and/or PD observations, time independent covariates, time dependent covariates, lab covariates and more (Rebecca H, Mitali G. 2024).

DATA SOURCE AND DATA FLOW

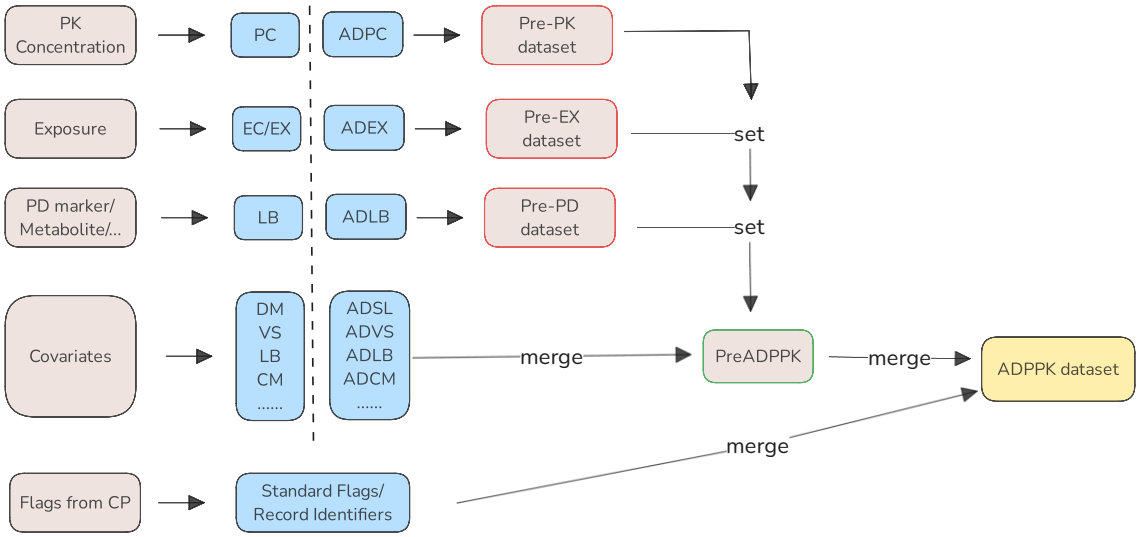
For understanding the PopPK dataset, we will display a table based on the data structure as follows:

- Dataset Metadata: The ADaM Implementation Guide (ADaMIG) supports many of the variables needed for PopPK analysis and provides general naming conventions that can be leveraged in the definition of other variables, by naming the new dataset as “ADPPK”.

Data Structure Name	Data Structure Description	Class of Dataset	Subclass of Dataset	CDISC Notes
ADPPK	Basic Data Structure Population Pharmacokinetic Analysis	BASIC DATA STRUCTURE	POPULATION PHARMACOKINETIC ANALYSIS	Dataset designed to support PPK. Sourced from SDTM (e.g., PC, DM, EC/EX, LB) and ADaM (e.g., ADSL and ADEX datasets).

Table 1. ADPPK Data Structure

- Data Source: ADPPK dataset include different data sources, such as basic demographics, lab assessments, and so on. Based on SDTM or ADaM datasets, we need to follow different rules in a specification.
- Data Flow: This figure illustrates the whole programming process from data to final ADPPK dataset. It includes different SDTM and ADaM data to set or merge following different rules.



**Set: stack datasets vertically.*

**Merge: combine datasets horizontally based on a common key.*

Figure 2. ADPPK Data Flow

ADPPK VARIABLES

As PopPK dataset includes different variables from different data sources, and the general variables can be classified into the three categories:

a. **ADPPK required variables:** They are common variables in an ADPPK dataset, including the subject-level (ADSL) variables like *STUDYID, USUBJID, USUBJIDN* and some general required variables of dosing variable like *AMT* or timing variable like *AFRLT*.

b. **NONMEM required variables:** These are some software specific need variables, which help to identify the records between pk and dosing. There are also some variables which both belonged to CDISC required and NONMEM required, like *EVID, DV, MDV.*

c. **Additional variables:** There still need more variables in real study, here we also want to introduce two groups: CP Standard Flags and Record Identifiers and Covariates variables:

Type		Variable Name	Variable Label	Algorithm
ADPPK Required	CDISC required	STUDYID & USUBJID etc..	Study Identifier & Unique Subject Identifier	Derive from ADSL.
	Time required	UDTC	Date and Time of the Event	Derive variables of datetime, need impute if partial date
		NFRLT/NPRLT	Nominal Rel Time From First/Previous Dose	Planned elapsed time from the first/last previous exposure to study treatment per protocol.
		AFRLT/APRLT	Actual Rel Time From First/Previous Dose	Set to sample/dose datetime - the first/last previous dosing start datetime.

	Dosing required	AMT	Actual Amount of Dose Received	Only populated on dosing records
	Standardize type	EVID	Event ID	EVID=1 is Dosing Event, EVID=0 is observation.
	PK required	DV, MDV	Dependent Variable Result, Missing Dependent Variable Result	Numeric result of dependent variable applicable to observation events. Missing result for DV
NONMEM Required		C	Comment	Flag for NONMEM to exclude records from the dataset
		EVID	Event ID	Same as above
		DV, MDV	Dependent Variable Result, Missing Dependent Variable Result	Same as above
		DVID/DVIDN	Dependent Variable Name (N)	Flag indicates what is in the DV column
		CMT	Compartment	Derived based on the specifications from modeler, based on different software
		II	Dosing Interval	Describes the dosing frequency for multiple doses
		ADDL	Number Of Additional Doses	Number of additional doses like the current dosing event until the next captured dose
Additional variables	CP Record Identifiers	EXCLF/EXCLFCOM	Record Exclusion, Comment for Record Exclusion	Record Exclusion/Comment for the Record Exclusion, can be captured prior to the modeling or after some analysis,
	CP Standard Flags	FLGREAS/FLGREASC	Identification of Data Issue Reason (Character)	Identification of Data Issue Reason, like data issues
	Baseline Covariates	WT, WTBL, HTBL, BMIBL, BSABL, ECOGBL	Weight, Baseline of Weight/Height/Body Mass Index/Body Surface Area/ECOG	Derivation and formulas based on study protocol/SAP from ADSL, ADVS, ADBASE
	Lab test Covariates	TBIBL, ASTBL, ALTBL	Baseline Total Bilirubin/ Aspartate transaminase/ Alanine transaminase	Derivation and formulas based on study protocol/SAP from ADSL, ADLB
	ADA test Covariates	ADA Titer, ADA Onset Day, ADA Duration	Treatment-emergent ADA Titer/Onset/Duration	Derivation and formulas based on ADA biomarker test
	Others	EGFRBL, CRCLBL	Baseline eGFR/ Creatinine Clearance Derivation	Derivation and formulas based on study protocol/SAP
		HEPARMP, RENAIMP	Hepatic Impairment Grade/ Renal Impairment Grade	Derivation and formulas based on study protocol/SAP

Note: These variables all depend on the ADPPK Specification.

Table 2. Summary of PopPK Variables

SPECIAL CONSIDERATION AND REQUEST

As the PopPK dataset is a highly pre-defined dataset, it not only consists of several rules derives some of its structure from the NONMEM analysis software system, but also needs to follow the Analysis Data Mode standards.

a. Dataset format rules

- As followed by CDISC, ADPPK data must be present in a text (i.e., ASCII) format flat file with appropriate delimiters.
- The output dataset must not contain null values. Any null values should be replaced with dot (".").
- NONMEM cannot accept character with space, "-", "," in non-null values, please replace such characters with "_" in all non-null values.
- Both the continuous and categorical covariates should not contain null values or dots. If the information is missing, set it to 99999. (specified by the Clinical Pharmacologist)

b. Missing data imputation rules

- Imputations should be avoided as much as possible. It is important to have CRFs that collect the required information. Besides, traceability must be maintained by documenting in dataset specifications, SAP, and/or PAP. (CDISC ADPPK IG, Section 5 Handling Missing Values)

- Usually, the missing data can be grouped into several sections, such as time or dose missing. We summarize the imputation strategy overview in the following tables:

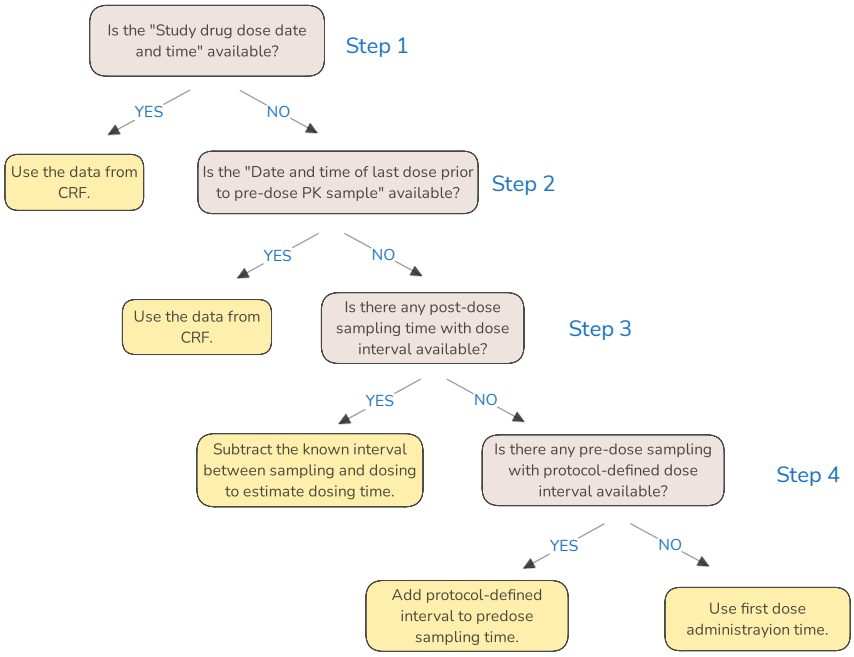
Category	Imputation Approach	Notes
Missing Covariates	Use method specified in PAP (e.g., LOCF)	Clearly document in SAP/PAP.
Missing PK Sample Date/Time	Leave missing and flag, or impute using nominal time or lab time within the PK window	Imputed values must be flagged.
Dose Clock-Time Imputation	Use trough sample + 5 min, or back-calculate from first post-dose sample, or use next/previous dose time	Adjust for BID/TID regimens.
Infusion Clock-Time Imputation	Derive from stop time + protocol-defined duration, or pre-dose sample	If no data, use nearby dose times recursively.
Interval Doses	Impute from visit/lab/PK date or dosing intervals	Special rules for missing start or stop dates.
Infusion Duration	Impute to protocol-defined duration if deviation >100%	Flag the dose and affected PK samples.
Infusion Rate	Calculate as AMT/DOSEDUR	Ensure consistency across records.
Missing Dose Amount	Impute by LOCF if nearby consistent, or use flat dosing value from treatment	Flag imputed dose amounts.

Note: The above summary is based on CDISC ADPPK IG, Section 5 Handling Missing Values.

Table 3: Imputation Strategy Overview

- Here are the scenarios that variable UDTC is missing: For general, UDTC is equal to <PCDTC for PC records>/<EXSTDTC for dosing records>/ <LBDTC for

PD records>. If the drug is taken orally, we need to impute the dosing time according to the following steps based on different conditions:



Note: All the imputation rules and special case rules need to discuss with Pharmacologist.
Figure 3. Imputation Strategy

c. Special case in examples

- In real practice, we find it complex to handle different drugs, so we display a summary table to show how to handle.

Differ/Drug	Periodical drug	Daily drug
Mode of Medication	Patients have an interval dose, usually like infusion drug, e.g. have a dose every 6 weeks.	Patients have drug daily, usually like oral drug, e.g. have a dose twice a day (a.m and p.m)
Drug Exposure	CRF will collect the dose date and time in a record according to the protocol.	CRF only collect a period time with date information, the exact time information is not collected.
Interval Calculate	Not recommended to describe dose interval and additional doses for infusion drug, only need to set as '.'	Need to consider the dosing frequency to calculate number of additional doses (e.g. ADDL in oral drug).

Table 4. Differences Between Periodical Drug and Daily Drug

CONCENTRATION - QT INTERVAL CORRECTED

Since QT/QTc interval prolongation on Electrocardiogram (ECG) is directly linked to increased susceptibility to cardiac arrhythmias, an adequate pre-marketing investigation of the safety of a new pharmaceutical agent should include a rigorous characterization of its effects on the QT/QTc interval. (Food, Drug Administration, HHS 2005.)

C-QTC INTRODUCTION

Concentration-QTc (C-QTc) refers to a method used to assess the relationship between drug plasma concentration and the corrected QT interval.

The International Council for Harmonization (ICH) E14 Questions and Answers document was revised in December 2015 to allow for C-QTc modeling to be used as the primary analysis for assessing the QTc interval prolongation risk of new drugs.

The C-QTc modeling approach utilizes data from all doses and time points, a reliable assessment of QTc prolongation can be based on smaller-than-usual TQT trials or be obtained during trials in early development not specifically targeted at QT effects, which has the advantage of saving time and economic cost.

DATA SOURCE

In Therapeutic Area Data Standards User Guide for QT Studies that CDISC released in December 2014 to give guidance about QT/QTc data collection rules in clinical trials (QT Studies Therapeutic Area User Guide v1.0, 2014). QTc is one of many safety parameters assessed in early Phase 1 Single Ascending Dose (SAD) and Multiple Ascending Dose (MAD) safety and tolerability studies. These studies collect and analyze ECG, PK, and PD data from multiple time points both before and after study treatment. Below is the flow chart of ECG and PK assessment and their data collection order.

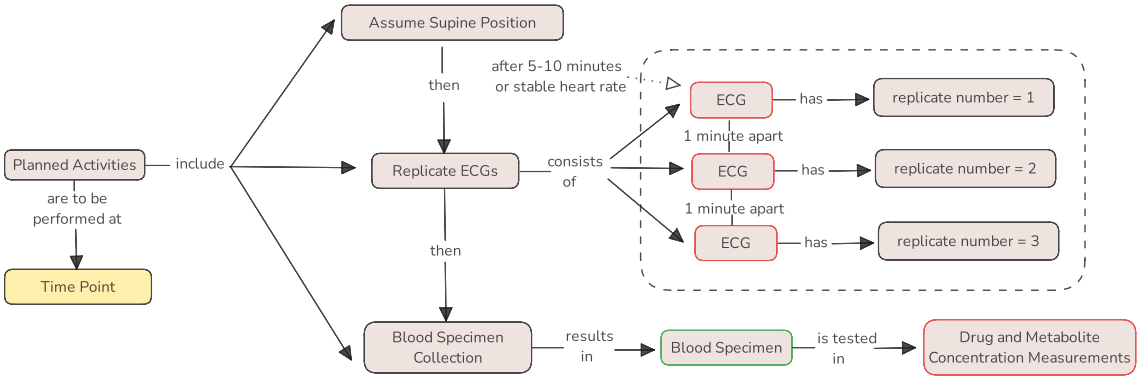


Figure 4. QT/QTc Data Collection Rules

1. ECG assessments:

    a. Quantitative triplicate ECG data (e.g., heart rate, PR interval, QRS duration, absolute QT interval, and QTc) represents one aspect of the ECG data analysis that should be included in the SDTM.

    b. QT Interval Adjusted (Corrected) for Heart Rate. Normally use Bazett's formula (QTcB) or Fridericia's formula (QTcF) from ECGs recorded in a small group of healthy subjects. An alternative sub-approach is to fit the individual subject's data to several preselected mathematical models and use the best mathematical model for each individual subject (QTcI).

2. PK assessment:

    a. An appropriate PK sampling schedule should be designed to adequately characterize the PK/PD concentration/response relationship for QT/QTc prolongation.

b. To investigate potential exposure-response relationships and confirm the availability of the drug, blood samples for PK measurements can be collected at the same time points as the ECG recordings. Blood draws are performed after the ECG recordings to avoid confounding stress.

ADCQTC VARIABLES

For ease of reference, the name ADCQTC was assigned in accordance with ADaMIG's nomenclature standards.

Data Structure Name	Description	Class of Dataset	Data Structure
ADCQTC	C-QTc Analysis Dataset	ADAM OTHER	One record per subject per visit per analysis date per time point.

Table 5. ADCQTC Data Structure

ADCQTC dataset mainly contains demographic, dosing, PK concentration and ECG information. The table below describes seven categories of variables that will be used in ADCQTC.

Type	Variable Name	Variable Label	Algorithm
Identifier	STUDYID, USUBJID etc..	Study Identifier, Unique Subject Identifier	Derive from ADSL.
Dosing	DOSEA, DOSEP	Actual Treatment Dose (unit), Plan Treatment Dose (unit)	Derived from ADEX.ECDOSE/ECPDOSE or based on related dose. Populated on all records by plan dose as carry-forward.
Concentration	PKCONC	Pharmacokinetic Concentration (unit)	Set to ADPC.AVAL. Set to 0 for pre-first dose or post-dose concentration that is lower than LLOQ.
ECG Assessment [a]	EGHRMN, QTCFAG, BQTCF, MBQTCF etc..	ECG Mean Heart Rate, QTcF Interval, Aggregate, Baseline QTcF for Each Subject, Mean of All Subjects Baseline QTcF Etc..	Derive from ADEG.AVAL where DTYPE='AVERAGE' and ADEC.PARAMCD is corresponding parameter. Generally, Heart Rate Unit is beats/min or msec.
Time Related Variables	AVISIT, ATPT	Analysis Visit, Analysis Timepoint	AVISIT and ATPT in ADEG and ADPC should be consistent.
	PCADTM	Analysis Datetime of PK collection	Derive from ADPC.ADTM
	ECGADTM	Analysis Datetime	Set to ADEG.ADTM from the last triplicate ECG (QTcF) for the average record.

	ATMDF	Relative Time between ECG (QTcF) and PK concentration	Set to PCADTM-ECGADTM in *unit*. Leave blank if time of ECG assessment is missing or PK sampling time is missing.
	ATMDFFIL	Time Difference Indicator Flag	Indicate whether the relative time between ECG and PK (ATMDF) is greater than time difference window. If so, those records will be excluded in C-QTc analysis.
Inclusion and Exclusion	ECGPCFL	ECG Matching PK Flag	Set to 'Y' if there is matching PK sampling for the ECG assessment by patient ID, visit and timepoint. ECG performed after PK sample collection will be excluded.
	EXCLF, EXCLFCOM	Record Exclusion Flag, Record Exclusion Comments	Flag and comments for Concentration QT Analysis to exclude record from dataset.
Covariates	Age, Race, ECOGBL, etc..	Baseline Covariates Information	Derive from ADSL, etc..

[a] ECG assessment also include RRAG (RR Interval, Aggregate), QTAG (QT Interval, Aggregate), PRAG (PR Interval, Aggregate), QRSAG (QRS Duration, Aggregate), QTCBAG (QTcB Interval, Aggregate) and QTcI (Individually corrected QT interval) for optional.

Table 6. ADCQTC Required Variables

ADCQTC OUTPUT REQUIREMENTS

1. General Notes

    a. Only subjects with available C-QTc samples were included in the analysis. If a QTcF measurement result was missing, the observation record was excluded.

    b. Required time-matched observed study drug concentrations and QTc for heart rate (HR). Concentration and ECG measurements were matched by nominal time point.

    c. Before merging with the concentration data, the ECG triplicate values of QTc interval were averaged by nominal time to obtain a single value per time point per subject.

    d. When pooling ECG interval data across multiple studies, heterogeneity testing should be performed.

2. Data Handling of Special Cases

Category	Approach
Concentrations Below the limit of quantification (BLQ)	Concentrations below the limit of quantification (BLQ) following study drug administration were set to zero.

Data Editing and Exclusion	All imputed values and data excluded for any reason were documented in data editing and analysis exclusion listings.
Missing Pharmacokinetic Data	Any post baseline time points without observed PK concentrations were excluded, along with ECG data at those time points.
Missing QT/QTc Data	If the date and time of the QTcF observation was missing, the relevant record was excluded. If a subject lacked baseline QTcF measurements, they were excluded from the analysi.
Missing Covariate Data	If the baseline value for any covariate to be treated as stationary was missing, the value was imputed from the value recorded at the screening visit.
Outliers	All values were included in the analysis unless there was a clear reason why the value was erroneous, such as measurement error or preexisted significant cardiac conditions at baseline (i.e., bundle branch block).

Table 7. Data Handling of Special Cases

OTHERS

SPECIAL INSTRUCTIONS

Large, targeted proteins and monoclonal antibodies have a low likelihood of direct ion channel interactions, and a thorough QT/QTc study is not necessary, unless the potential for proarrhythmic risk is suggested by mechanistic considerations or data from clinical or nonclinical studies. (E14 and S7B Clinical and Nonclinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential, 2022).

CLINICAL PHARMACOLOGIST APPLICATION

The evaluation of QTc interval prolongation will follow the ICH E14 guidance (Scientific white paper on concentration-QTc modeling, 2017) and method outlined in the scientific white paper on C-QTc modeling (Garnett C, 2018).

1. Software

Exploratory graphical analysis, model development and simulation will be performed either using R (e.g. packages nlme or lme4), SAS (e.g. PROC MIXED) or NONMEM.

2. Concentration-QTc Model

    a. Results of the exploratory analyses will be used to determine if the assumptions of the model are met.

    b. A pre-specified linear mixed effects model is recommended as the primary analysis to exclude a 10-ms QTc prolongation effect. However, for drugs known to prolong QTc intervals, the objectives of C-QTc analysis shift toward exploratory characterization, potentially rendering the pre-specified model unsuitable in such cases.

EXPOSURE-RESPONSE

E-R INTRODUCTION

Exposure-Response (E-R) Analysis: Includes PK/PD modeling as a special case where the exposure variable is drug concentration, the response is often a clinical endpoint (safety or

efficacy endpoint). A drug can only be determined to be safe and effective when the relationship between beneficial and adverse effects and a defined exposure is understood.
Some oncology development programs employ a seamless design, enabling rapid progression from initial dose-finding trials to registration trials to accelerate drug development. With careful planning, dose optimization can be strategically aligned with this expedited clinical development pathway.

FDA GUIDANCE

Traditional MTD-focused dose-finding trials often overlook critical data (e.g., grade 1-2 toxicities, PK/PD, and exposure-response) when selecting doses for later studies. The FDA advocates a more comprehensive approach that integrates safety, tolerability, and activity data with preliminary exposure-response analyses to optimize dose selection for further clinical evaluation (Exposure-Response, 2003).

SOURCE DATA

E-R dataset contains safety and efficacy information from corresponding ADAMs, including ADSL, ADAE, ADTTE, ADRS and etc. For modeling part, Clinical Pharmacologist will combine PopPK dataset and base on three key questions to establish proper analysis model (Establishing Good Practices for Exposure–Response Analysis of Clinical Endpoints in Drug Development, 22Sep2015).

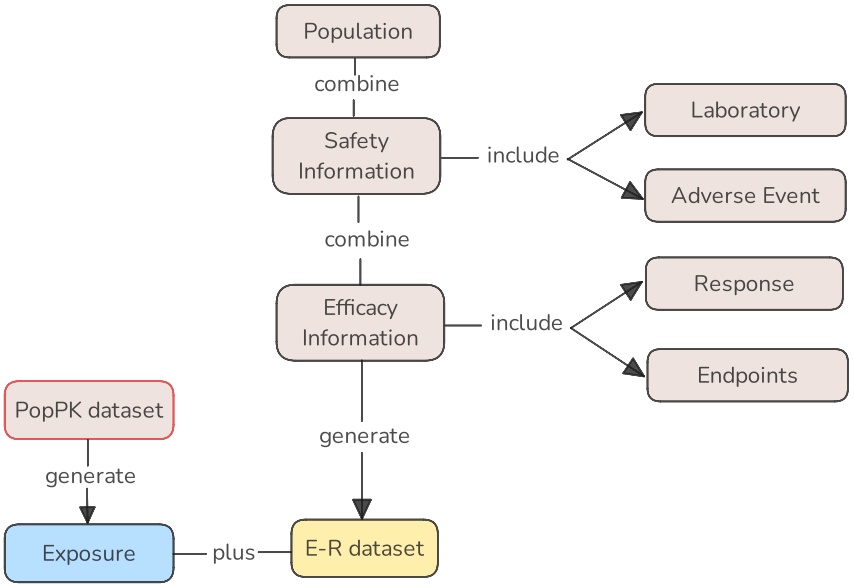


Figure 5. E-R Source Data and Combination with PopPK for Modeling

1. Exposure
    a. Drug exposure information derive from ADPPK by Clinical Pharmacologist per they need.
2. Response
    a. For E-R analysis, the response variable could be continuous, categorical, or time-to-event data.

ADER VARIABLES

The name ADER was also assigned following ADaMIG nomenclature standards for consistent reference.

Data Structure Name	Description	Class of Dataset	Data Structure
ADER	Exposure-Response Analysis Dataset	ADAM OTHER	One record per subject*.

* The data structure also incorporates records-level metadata to support optional analytical methods.

Table 8. ADER Data Structure

Normally, ADER variables are study specific. There are several categories that required most in below table.

Type	Variable Name	Variable Label	Algorithm
Identifier	STUDYID, USUBJID etc..	Study Identifier, Unique Subject Identifier	Derive from ADSL.
Population Flag	SAFFL, EFFFL, TRTEMFL etc..	Safety Population Flag, Efficacy Evaluable Population Flag, Treatment Emergent Analysis Flag, etc..	Derive from ADSL, ADAE.
Safety Information	AESER, AEGR3, AEDOSECHv, AESIxIM[a], etc..	Serious Event, AE of Grade 3 and Above, AE Leading to Dose Modification of Any Grade, Special Interest of Immune-related AEx, etc..	Derive from ADAE.
Efficacy Information	PFSDxxIN[b], OSD, BORCxxIN[b], etc..	Progression Free Survival (days) by Investigator (xxx) [b], Overall Survival (days), Best Overall Response with Confirmation by Investigator (xxxx) [b] etc..	Derive from ADRS, ADTTE and etc..

[a] 'x' could be number, e.g. 1 ,2 3….

[b] 'x' in Variable Name and '(xxx)' in Variable Label are for modality (align with PARAM/CD in ADEF).

Table 9. ADER Variables Categories

ADER OUTPUT REQUIREMENTS

1. Only consider those patients from the full analysis set (FAS) for which exposure data are available. When combine with ADPPK dataset, the population should keep consistence with it.
2. Given the E-R dataset includes more study-level safety and efficacy endpoints, which is quite study-specific. It would be better to clarify whether use subject-level or record-level dataset structure in advance.

OTHERS

ADER COMBINED APPLOCATION WITH ADPPK

Population PK analysis is a mature discipline, and numerous methodological papers on model building, covariate selection, and model diagnostics exist.

E-R analysis, in its broad definition, includes PK/PD modeling as a special case where the exposure variable is drug concentration, but most often the term E-R refers to analyses that differ from PK/PD models in several aspects:

1. The exposure variable is a summary variable such as area under the curve (AUC), rather than the concentration time course. The most obvious summary measure of drug exposure is the area under the concentration–time curve (AUC) in a dosing interval at steady-state, but other measures such as maximal drug concentration (C_{max}) or trough concentration (C_{trough}) may be more appropriate, depending on the indication, the endpoint, and the PK properties of the drug.
2. The response is often a clinical endpoint, typically expressed as the change of response variable from baseline to the end of trial.
3. In many instances, E-R analysis is conducted by simple regression type of analysis, rather than time course models.

For E-R modeling, three key questions to establish proper analysis model see below table.

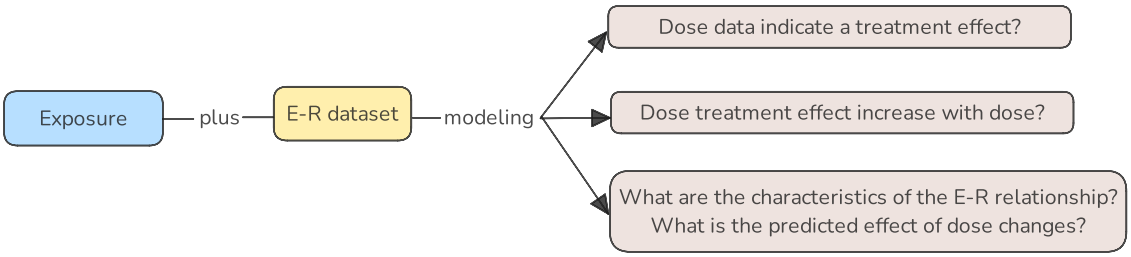


Figure 6. The Principle of Establishing Proper Analysis Model

SUBMISSION PACKAGE AND WORKFLOW

During the data submission phase, due to the absence of clear guidance from HA, a process based on practical submission experience has been summarized to standardize submission workflows and content.

The submission package, assembled by the SP, consists of two key components: the datasets, which include final formatting and alignment with submission standards, and the corresponding define file, which documents the structure and content of the datasets. Separate submission packages must be prepared for each analysis type for PopPK, E-R, and C-QTc.

PREPARATION OF DATASETS

All datasets included in the submission package must be provided in two formats: Comma-Separated Values (CSV) and SAS Transport Format (XPT). If a submission requires compliance with Chinese regulatory standards (CDE), a Chinese version of the dataset must be prepared.

The preparation of the PopPK dataset requires special attention to the following steps:

1. **Handling of Variable C**:
   As described in adppk variables section, certain variable C in the PopPK dataset must be completed by the CP. After SP generates the initial ADPPK dataset in CSV format, it should be shared with the CP, who will use their expertise to populate Variable C. The CP will also provide the rules for populating this variable, which SP will reference when preparing the define file.

2. **Final Submission Version**:
   Once the CP completes populating the variables, the updated CSV dataset is returned to the SP as the finalized version. The SP is then responsible for converting this finalized dataset into XPT format.

3. **Additional Components**:

   - **Model Analysis Data**: If necessary, model analysis data from the CP should also be included in the dataset. The related define should also be provided.

PREPARATION OF DEFINE FILES

The define file in the submission package can be a simplified version compared to the Clinical Study Report (CSR)-level define file. The format and content requirements are as follows:

1. **Content**:
   The define file should include only three elements:

   - **Variable Name**
   - **Label**
   - **Data Source and Algorithm**: For Variable C, the algorithm should be provided by the CP.

2. **Format**:
   The define file should be submitted in Microsoft Word format (.docx) and does not need to include hyperlinks.

PACKAGING AND UPLOAD

Once all components are prepared, the package should be compressed into a ZIP file. The ZIP file must include:

- Datasets in both CSV and XPT formats
- The define file

The CP is responsible for uploading the finalized ZIP file to the designated submission location. The upload path of PopPK, E-R and C-QTc is shown in Figure 7.

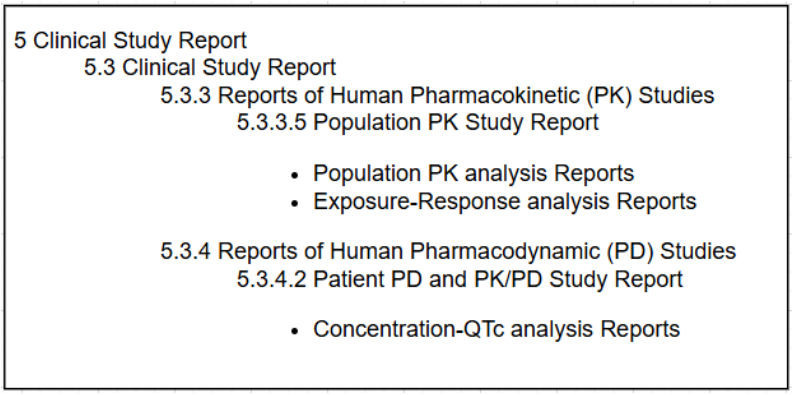


Figure 7. The Upload Path of PopPK, E-R and C-QTc

CONCLUSION

Population Pharmacokinetics (PopPK), Exposure-Response (E-R), and Concentration-QT Interval corrected (C-QTc) analyses play a pivotal role in clinical pharmacology, particularly during early-phase drug development. However, the lack of clear guidelines and limited familiarity with these analyses among statistical programmers often create challenges in data processing and submission preparation.

This paper has clarified the specific responsibilities of statistical programmers in PopPK, C-QTc and E-R analyses, emphasizing their contributions to data integration, variable population, and alignment with CDISC standards. It also provides a structured approach for preparing datasets and define files to meet regulatory requirements. By systematizing these processes, statistical programmers can ensure high-quality submissions, reduce regulatory risks, and streamline drug development timelines.

By addressing these challenges and providing actionable guidance, this work aims to serve as a valuable resource for statistical programmers, fostering collaboration between clinical pharmacologists and programmers while enhancing the overall efficiency and compliance of clinical pharmacology analyses.

REFERENCES

Chen WJ, 2017, Population pharmacokinetics and its application in new drug research.

Namrata P. 2020. Population PK Data Sets – Basic Concepts and Challenges from the Perspective of a SAS Programmer.

Population Pharmacokinetics Guidance for Industry https://www.fda.gov/regulatory-information/search-fda-guidance-documents/population-pharmacokinetics

CDISC Basic Data Structure for ADaM PopPK Implementation Guide v1.0, 2023

FDA eCTD Technical Conformance Guidance. https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-common-technical-document-ectd

Basic Data Structure for ADaM PopPK Implementation Guide v1.0

Raghu K. 2014. NONMEM® – A Programmer point of view.

Rebecca H, Mitali G. 2024. An End User's Perspective of Clinical Dosing Data Used for Population PK / PD Analysis with Nonlinear Mixed Effects Modeling.

Food, Drug Administration HHS: International Conference on Harmonisation; guidance on E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs; availability. Notice. Fed Regist 2005.

QT Studies Therapeutic Area User Guide v1.0, 08Dec2014.

E14 and S7B Clinical and Nonclinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential--Questions and Answers. 2022-18515 (87 FR 52786).

Scientific white paper on concentration-QTc modeling,21Jun2017.

Garnett C, Bonate PL, Dang Q, Ferber G, Huang D, Liu J, et al. Scientific white paper on concentration-QTc modeling. J Pharmacokinet Pharmacodyn. 2018 Jun;45(3):383-397.

Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications, Apr2003

CONTACT INFORMATION

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

Xiaohan Ye
BeOne, Inc.
xiaohan.ye@beigene.com

Lin Liu
BeOne, Inc.
lin2.liu@beigene.com

Yuyan Wang
BeOne, Inc.
yuyan.wang@beigene.com

Meng Qu
BeOne, Inc.
meng.qu@beigene.com