

ADCM Macro Design and Implementation

Chengxin Li, Regeneron Pharmaceuticals Inc.

Tingting Tian, Regeneron Pharmaceuticals Inc.

Toshio Kimura, Regeneron Pharmaceuticals Inc.

Abstract

Previous work explained in detail a set of macros that automated the development of safety ADaM datasets. As an extension to that previous work, this paper delves into the design and implementation of a macro that will generate the ADCM dataset. This paper will broaden the focus to include the end-to-end streamlined data processing perspective from data collection to analysis and thus will discuss some of the upstream and downstream components including data collection and SDTM considerations as well as the analysis need that ultimately drives the ADCM requirements. Since the ADCM macro incorporates analysis need, the resulting ADCM will be analysis-ready in supporting common concomitant medication analyses while allowing for traceability back to SDTM.

The GPS (Global, Project, Study) method was applied to the macro development. Global variables are derived without user input, whereas project and study specific variables are controlled by macro parameters. The macro employs logically structured sequential modules with high prioritization on user experience. The macro parameters and core concepts are described in detail in the paper.

Introduction

Clinical Data Interchange Standards Consortium (CDISC) standards have been widely deployed in the pharmaceutical industry in the past decade. Under the CDISC umbrella, a concerted effort has been made to achieve end-to-end streamlined data processing to ensure traceability, transparency, ease of review, quality and efficiency. By leveraging the CDISC end-to-end streamlined data flow, a reusable solution to automate data processing can be developed. As an extension of the safety Analysis Data Model (ADaM) automation framework¹, this paper introduces the macro design and implementation for Concomitant Medications Analysis Dataset (ADCM). The basic end-to-end processing flow is described in the diagram in Figure 1.

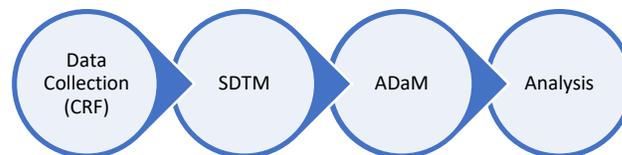


Figure 1: End-to-end streamlined data processing

From the statistical programming perspective, the end-to-end streamlined data processing is initiated from data collection via the case report forms (CRFs). The Clinical Data Acquisition Standards Harmonization (CDASH), a data collection standard harmonized with the Study Data Tabulation Model (SDTM), provides the best practice recommendations on CRF design. SDTM standardizes the collected data in tabulations, required to fully reflect the collected data, whereas any imputations are not permitted. ADaM datasets are then derived from SDTM. The ADaM datasets will fully support the analyses defined in the statistical analysis plan (SAP). Traceability and analysis-ready concepts are the two core features of ADaM development. With the ADaM datasets being analysis-ready, the appropriate data from the ADaM datasets can be selected and the SAP defined statistical procedures can be invoked to perform the analyses.

¹ C. Li, T. Tian, T. Kimura, Safety ADaM Automation from Design to Implementation, PhUSE US Connect 2022

The paper will in general mirror the described flow; however, since the focus of the paper is on ADCM development, the analysis itself will be out of scope. Nevertheless, the common analysis need will be incorporated into the discussion about ADCM design since analysis need must be prospectively considered in order for ADCM to be analysis-ready. The concepts described in this paper are based on CDASHIG 2.1 for CRF data collection, SDTMIG 3.2 for data tabulation, and ADaM OCCDS 1.0 for analysis dataset, specifically, ADCM.

CRF and SDTM Considerations

Data collected on CRFs are mapped to SDTM domains. For concomitant medication (CM), the key information is the CRF collected verbatim medication names (CMTRT). CMTRT is coded using the World Health Organization Drug Dictionary (WHO-DD) standardized hierarchy of terms and codes. The CMTRT standard terms and codes are mapped to the CM and SUPPCM SDTM domains.

The SDTM CM and SUPPCM domains along with ADSL are the source datasets for deriving the ADCM dataset. Since concomitant medication data is an occurrence data type, ADCM will use the OCCDS structure. There are several SDTM considerations that must be taken into account in developing a robust ADCM SAS® macro:

1. **Mapping of hierarchical terms and codes:** Based on the guidance provided by the SDTMIG, hierarchical coding data such as CMDECOD and CMCLAS/CMCLASCD are mapped to the CM domain, and the ATCy (ATC Level y Text) and ATCyCD (ATC Level y Code) are mapped to the SUPPCM domain. However, CDASHIG also provides its own mapping instructions using the naming convention of CMATCy (ATC Level y Description) and CMATCyCD (ATC Level y Code) which are also mapped to the SUPPCM domain. Since both mapping solutions are allowed, both mapping solutions will be supported by the ADCM macro. If both are mapped in SUPPCM, ATCy/ATCyCD will be selected. If only CMATCy/CMATCyCD is mapped, ATCy/ATCyCD ('CM' compressed) will be used; conversely, if only ATCy/ATCyCD is mapped in SUPPCM, then ATCy/ATCyCD will be used.
2. **Fully or partially missing dates:** The dates as captured in CMSTDTDC for the start date or CMENDTC for the end date may be fully or partially missing. In imputing missing dates, the PRIOR or ONGOING status must also be considered as described below.

Start date: The Prior Concomitant Meds (CMPRIOR) field, indicating that the concomitant medication is given or taken prior to the study reference date, is optionally collected when the start date is not provided. When CMPRIOR=Y, a controlled terminology (CT) value of "BEFORE" is populated either in CMSTRTPT (Start Relative to Reference Time Point) with the corresponding start reference time point specified in CMSTTPT (Start Reference Time Point) or in CM.MSTRF (Start Relative to Reference Period) with the corresponding start reference time point specified with DM.RFSTDTDC.

End date: The Ongoing Concomitant Meds (CMONGO) field, indicating that the concomitant medication is ongoing, is often collected as a conditionally required variable if the end date is not provided. When CMONGO=Y, a controlled terminology value of "ONGOING" is populated either in CMENRTPT (End Relative to Reference Time Point) with corresponding end reference time point specified in CMENPT (End Reference Time Point) or in CMENRF (End Relative to Reference Period) with the corresponding end reference time point specified with DM.RFENDTC.

The implementation of the SAP defined imputation rule will use the mapped values in SDTM depending on the SDTM mapping solution that was implemented.

Analysis Need

There are multiple Concomitant Medication analyses that will need to be performed. The analyses that are most often requested include analysis to determine the number and percentage of subjects who received Prior or Concomitant Medications. There are no industry standard definitions for how Prior or Concomitant are defined. For example, a concomitant medication that starts before study treatment but is continued after the initiation of study treatment might only be considered as a Prior Medication and not a Concomitant Medication in one SAP whereas it may be considered as both a Prior Medication and a Concomitant Medication in another SAP. Therefore, the SAP must clearly define the rules that will be implemented. There may also be a special set of medication such as rescue medications or prohibited medications for which separate analyses would be of interest. Furthermore, an analysis examining the duration during which subjects took the concomitant medication may also be requested. These common analysis needs are described in Table 1 and will be proactively incorporated into the ADCM design and implementation.

Table 1: Common Concomitant Medication Analysis Needs

Analysis Needs	Description
Population set	Select the Analysis Population Set to be used in the analysis. This is generally the safety set for CM analysis.
Treatment group	Select the treatment type to be used in the analysis. This is generally the actual treatment arm for CM analysis.
Subgroup	The option to perform subgroups analysis.
Prior/concomitant meds taken during different phases	The ability to differentiate between medication that were taken prior to the initiation of study treatment, medications taken at the same time as the study treatment, or started after the end of study treatment.
Rescue medication	Identify CM that are considered to be a rescue medication
Prohibited medication	Identify CM that are considered to be a prohibited medication
Duration	Perform analyze the duration of the CM
Analysis variables: ATCy and CMDECOD/CMTRT	Perform analysis at different WHO-DD ATC levels

ADCM Macro Design

Similar to adverse events, concomitant medication data are collected when a patient takes a concomitant medication. While the information may be entered during a clinic visit, the concomitant medication can be initiated and stopped at any time. This type of data under the CDISC framework is an occurrence data. As such, the ADCM will use the Occurrence Data Structure (OCCDS). As noted under analysis need, ADCM, consistent with other OCCDS datasets, will support frequency analyses including analysis such as prior/concomitant medications taken prior to study treatment initiation or concomitant with study treatment.

1. ADaM Principles

ADaM design is an art, not a science. This statement also applies to ADaM macro development. Nevertheless, in addition to being ADaM standards compliant, there are core ADaM principles that should be followed, which consists of being able to fully support the SAP, being analysis-ready, and providing traceability to facilitate review.

To allow for traceability and ease of review, the produced ADCM dataset includes most SDTM CM variables along with additional treatment and date variables from ADSL. The datapoint traceability variables such as SRCDOM, SRCVAR and SRCSEQ are not included since there is only a single SDTM CM domain as the primary data source. Furthermore, ASEQ is not derived, since the SDTM variable CMSEQ is kept in the ADCM dataset.

To ensure that ADCM is analysis-ready and can fully support the SAP, the ADCM macro includes parameters to enable the derivation of the required analysis variables. With those analysis variables derived, the dataset based on a simple WHERE statement for data selection can be passed into SAS procedures to perform the required analysis. Table 2 lists ADCM variables derived to support the corresponding analyses by the macro.

Table 2: ADCM Variables Derived to Support SAP (Analysis Ready)

Analysis Needs	ADCM Implementation
Population set	The analysis population set variable(s) to be copied from ADSL Default: SAFFL
Treatment group	TRTA/TRTAN to be derived For traceability, TRTxX/TRTxXAN and TRxxSDT/TRxxEDT are copied from ADSL
Subgroup	Subgroup variables to be copied from ADSL
Prior/concomitant medications taken during different phases	Discussion regarding prior/concomitant medication are explained in the section above under SDTM considerations.
Rescue medication	A new variable, ACATy, to be derived based on CMTRT/CMDECOD for rescue medication(s)
Prohibited medication	A new variable, ACATy, to be derived based on CMTRT/CMDECOD for prohibited medication(s)
Duration	The ADURN variable will be derived based on the start and stop date. To derive ADURN variable, CMSTDTC and CMENDTC must be imputed.
Analysis variables: ATCy and CMDECOD/CMTRT	CM analysis are generally performed on ATCy and CMDECOD/CMTRT. The ATCy variables will be carried over from SUPPCM.

2. GPS Navigation Method

Similar to the other safety ADaM macros, the GPS (**G**lobal, **P**roject, **S**tudy) navigation method is applied to ADCM macro development. Variables that can be populated or derived without user input are considered as global level variables. Global variables include all variables that are collected on the CRF, coded from WHO-DD, or copied from ADSL or other SDTM domains such as the DM domain. Variables that are conditionally populated or dynamically derived depending on the SAP are considered as project or study level variables. The difference between a project and study level depends on the level at which the parameter is defined (with the project level being higher than the study level). An example use of project level variables would be derivations that are defined at the therapeutic area or product level. Project and study level variables are controlled through macro parameters driven by requirements defined in the SAP.

The following is an example of the GPS concept in ADCM. In ADCM, most variables are populated with CRF collected data or WHO-DD hierarchy code and decode stored in the CM/SUPPCM. This also includes variables copied from ADSL such as SITEID, SEX, RACE, SAFFL, TRTSDT, TRTEDT, etc. Since these are common across all studies, they can be categorized as global variables and derived without any user input. Other variables, however, depend on the derivations defined in the SAP. For example, ASTDT might use a defined imputation method. Since the derivation for these variables might be varied across studies, they are considered project or study level variables. The range of derivation methods in defining these derivations may be more limited; hence, the macro provides the user with a choice of pre-defined methods or some limited parameterization, so that the user does not have to fully parameterize the derivation from scratch.

Ultimately, the production of ADCM dataset is performed for a given study. The macro call would simultaneously utilize all 3 variable levels (global, project, and study) in order to derive the ADCM dataset.

3. User Experience

User experience is another important general consideration. The statistical programmer is the primary user with the statistician being another key stakeholder. The macro should be easy to understand and easy to use. With that in mind, Table 3 summarizes the factors along with the corresponding method that attempts to improve the user experience.

Table 3: Factors Considered in Enhancing the User Experience

Factor	Method
Platform	Multiple platforms are supported, currently supports both PC and EG SAS
Parameter Name	Straightforward and intuitive to understand
Parameter input rules or conventions	Use of consistent input patterns across parameters such as using the “#” to denote multiple items. Macro will catch ERRORS (such as invalid input parameters) to avoid unpredictable SAS messages.
SAS LOG	Saved as an external file and replayed back to the SAS LOG Window
Execution performance	Optimize performance. This includes use of faster processing methods such as PROC SQL or PROC COPY to download or upload datasets as opposed to using the DATA STEP
Format	Use of consistent formats across all date variables, e.g., all date in YYMMDD10. format
Output	ADCM dataset stored in SAS and XPT formats. Specifications are dynamically produced based on the macro call instance. Specifications layout will facilitate the use of define-xml tool.
Facilitate Review	Execution Summary Report generated in separate PDF and Excel files

Execution Summary Report captures the outcome of data checking and execution per parameter. Any data issues which impact the derivations are collected in Execution Summary Report. These issues are also displayed in the SAS log as WARNING. The Execution Summary Report is structured as table with the following columns: Category, Parameter, Outcome, Findings, and Notes. The Notes column includes the definition or objective of each individual parameter in the macro call instance.

4. Date Imputations

As previously described, the start and end dates for the concomitant medication may be required to meet the analysis need as defined in the SAP. If the start and/or end dates are missing, CMSTDTC (start date) and/or CMENDTC (end date) variables may need to be imputed. While there is no standard method for these imputations, there are commonly used imputations that are often used which will be available as an option in the macro. One approach is to use a conservative and simple approach. Specifically, the imputation will favor a start date that begins during study treatment and an end date that results in the longest duration. Table 4 depicts the imputation methods implemented in the ADCM macro.

Table 4: Imputation Methods for Missing CMSTDTC/CMENDTC Dates

DTC (ISO8601) ³	Missing Type	Method
CMSTDTC	Full Missing	1. If PRIOR is collected in CMSTRF/CMSTRPT as 'BEFORE' or 'PRIOR' ¹ , then impute with RFICDTC; else 2. Impute with TRTSDT
	Partial Missing	1. If non-missing date component is the same as TRTSDT, then impute the missing date component with the corresponding counterpart of TRTSDT; else 2. Impute the missing date component the first calendar of the missing date component
	Alignment	No later than CM end date or last participation date
CMENDTC	Full Missing	1. If ONGOING is collected (CMENRF/CMENRTPT='ONGOING'), then impute with RFPENDTC ² (CMENRF), or with specified reference point in CMENTPT; else 2. Impute with RFPENDTC
	Partial Missing	Impute the missing part with the last calendar part
	Alignment	No earlier than CM start date or later than last participation date

Note: 1. BEFORE is the CT for PRIOR. The Macro also accepts PRIOR as an acceptable CT value.
2. Due to the ambiguity of DM.RFENDTC, the macro takes RFPENDTC as an alternative while outputting a WARNING to LOG for the case of CMENRF mapping.
3. Since CMSTDTC/CMENDTC are expressed using ISO8601, the partial missing or full missing patterns can be easily identified. Details are not included here.

Macro Parameters

Table 5 describes the parameters in the ADCM macro. Global variables are derived without user input. Additionally, the macro is capable of providing a platform enabling project and study variables to be derived through the available macro parameters.

Table 5: Macro Parameters of ADCM Macro

Parameter	Core	Definition	Notes/Example
SDTMLIB	Req	SDTM source libname	Example: %str(SDTM)
ADaMLIB	Req	ADaM source libname Macro will use this libname to access ADSL Macro will output final dataset to this libname	Example: %str(ADaM)
POPU	Opt	Analysis population set to be copied from ADSL	Default: SAFFL

ADSL_VAR	Opt	Additional variables to be copied from ADSL This includes variables used for subgroup analysis	Default: SITEID
STDT_IMPOT	Opt	Refer to section on SDTM considerations for missing dates, and Tables 1 and 2 for analysis need and implementation	Valid values: Y or N Default: N
ENDT_IMPOT	Opt	Refer to section on SDTM considerations for missing dates, and Tables 1 and 2 for analysis need and implementation	Valid values: Y or N Default: N If Y then STDT_IMPOT=Y is required
ACAT	Opt	Categorize records as a rescue medication or prohibited medication	Parameter uses the following structure: srcVAR/tgtVAR/format=\$xxx Example: CMDECOD/ACAT1/format=\$rescue. # CMDECOD/ACAT2/format=\$prohibit.
AVALFL	Opt	Flag records as a prior medication or concomitant medication	Parameter uses the following structure: tgtVARNM/tgtVARLBL/WHERE Example: PRIORFL/Prior Concom Med Flag/ASTDT<TRTSDT # CONCOMFL/ Concom Med Flag/ASTDT>=TRTSDT
SUPPQ	Opt	List the additional QNAMS to be merged into ADCM from SUPPCM	Example: %str(CMDOSFQ) Default: ATCy and ATCyCD from SUPPCM will be included
WHODRUG	Req	WHO-DD version	Example: %str(March 2021)
SPECS	Opt	Supporting document	Valid values: Y or N Default: N

ADCM Macro Implementation

1. Structured Programming

The ADCM macro leverages structured design with sequential modules, which organizes codes into sections that achieve specific objectives. Modular coding is easier to maintain, since any issues that may appear can be isolated to a particular module. It is also easier to reuse, since the modules can be swapped in and out depending on the macro being developed. Table 6 summarizes the five sequential programming modules used in ADCM macro development.

Table 6: Sequential Programming Modules Used in the Macro

Module	Objective	Description
1	Initiation	1) Set system options 2) Define sub-macros 3) Build up dataset templates 4) Record the initiation status (used later during cleaning-up)
2	Checking Parameter and Dataset	1) Check existences of LIBREFs, datasets, formats, or additional requirements such as on EPOCH 2) Check user input conformance with defined rules and requirements 3) Set defaults (unless specified otherwise): for example, POPU to SAFFL and ADSL_VAR to SITEID 4) Prepare for derivations, e.g., along with checking the validity of macro parameters, parse the parameters so that they are derivation-ready 5) Write the outcomes of checks to SAS LOG window and Execution Summary Report
3	Derivation	1) Perform variable derivations: first global variables, then project and study variables per macro parameters 2) Prepare the dataset and associated documents for output generation
4	Output	Store ADaM dataset and associated documents in the designated folder
5	Cleaning-up	Reset settings to the status prior to the macro call (such as deleting intermediate datasets)

2. Macro Call Instance

The macro call instance shown below contains all of the available macro parameters.

```
%ADCM(
  SDTMLIB   =%str(SDTM),           /*provide libname SDTM as source data location   */
  ADaMLIB   =%str(ADaM),           /*provide libname ADaM as output location         */
```

```

POPU      =%str(SAFFL),          /*apply SAFFL as analysis population set          */
ADSL_VAR  =%str(sex # race),    /*request SEX/RACE copied from ADSL to ADCM      */
STDT_IMPUT=Y,                  /*apply defined imputation method to CMSTDTC missing */
ENDT_IMPUT=Y,                  /*apply defined imputation method to CMENDTC missing */

ACAT      =%str(CMDECOD/ACAT/format=$RESCUE),
/*derive ACAT var for rescue medications by applying a user-defined format $RESCUE on CMDECOD*/
AVALFL    =%str(PRIORFL/Prior Concom Med Flag/ASTDT<TRTSDT # CONCOMFL/Concom Med Flag/ASTDT>=TRTSDT),
/*derive PRIORFL and CONCOMFL analysis variables for PRIOR and CONCOMITANT medication analyses*/

SUPPQ     =%str(CMDOSF00),      /*besides default ATCy/ATCyCD, request adding additional QNAM: CMDOSF00 to ADCM*/
WHODRUG   =%str(March 2021),   /*assign WHODRUG dictionary version: March 2021 to ADCM with WHODRUG variable*/
SPECS     =%str(Y)             /*request generation of ADCM specs with Excel format and saved in ADaM location*/
);

```

In above macro call instance, a predefined SAS format \$RESCUE is required which is enforced in macro checking step. The \$RESCUE format lists the WHO-DD standardized medication names stored within the CMDECOD variable that should be categorized as a rescue medication.

Discussion

The end-to-end streamlined data processing is the best practice in clinical data processing. Under this umbrella, from programming perspective, there are multiple linear components, including CRF annotation, SDTM mapping, ADaM development, and TLF generation. Technically, it's feasible to achieve automations for any component. The only challenge presented would be automating efficacy ADaM. However, the challenge would be resolved with multiple efficacy ADaM macros by therapeutic areas or indications other than one overall macro. Automation of efficacy ADaM dataset development is currently under investigation.

Contact Information

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

Name: Chengxin Li
Enterprise: Regeneron Pharmaceutical Inc.
Address: 777 Old Saw Mill River Road
City, State ZIP: Tarrytown, NY 10591
Work Phone: (914) 847-3687
E-mail: chengxin.li@regeneron.com

Name: Tingting Tian
Enterprise: Regeneron Pharmaceutical Inc.
Address: 777 Old Saw Mill River Road
City, State ZIP: Tarrytown, NY 10591
Work Phone: (914) 847-3150
E-mail: tingting.tian@regeneron.com

Name: Toshio Kimura
Enterprise: Regeneron Pharmaceutical Inc.
Address: 777 Old Saw Mill River Road
City, State ZIP: Tarrytown, NY 10591
Work Phone: (914) 847-1463
E-mail: toshio.kimura@regeneron.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.