Paper SM16 - PHUSE EU 2025

On the Right Track: Managing Platform Modular Studies With a Central Metadata Switchboard

Dubravka Georgijevic, Fortrea, Leiden, the Netherlands
Carla Santillan, Fortrea, Brussels, Belgium

ABSTRACT

A platform modular study is a type of adaptive trial that uses a master protocol to evaluate multiple, targeted therapies within a single disease setting. These trials may not have a planned end date, and treatment arms (and/or modules) can be added, or removed, based on predefined rules. For statistical programmers, managing these studies often feels like handling multiple studies within one database. A key challenge is the ongoing series of module-specific analyses, each requiring unique derivations and output modifications.

This paper introduces a metadata-driven solution using a central file that controls which derivation, or output, adjustments should be applied per study module, and delivery. This file is read into SAS to generate macro variables used in SDTM, ADaM, and TLF programs for conditional programming. With this approach, new deliveries require only an update to the central file - no widespread code changes - allowing for efficient and streamlined programming in modular study environments.

INTRODUCTION

At Fortrea, in recent years, we have had the opportunity to provide programming services to support platform modular studies with increasing frequency. These trials implement an adaptive design approach which use a master protocol to evaluate multiple targeted therapies, within a single disease setting, typically in Oncology. These studies are characterized by a long duration and are evolving and incremental in nature. The treatment arms (and/or modules) may be added or discontinued and may differ greatly in design. For example, an oncology study might begin with non-randomized, open-label modules using a dose-escalation design. Over time, dose-expansion components may be added, including specific cohorts within a module that are double-blind, or others that follow a cross-over design to assess food effects.

As an example of a study evolving over time with incremental data collection, one of the trials under review had its First Subject In (FSI) in Q4 2018, while the Last Subject Last Visit (LSLV) occurred in Q3 2024. During this time, the Clinical Study Protocol (CSP) underwent 13 amendments. What began as four modules involving two Investigational Products (IPs) expanded to fourteen modules, encompassing seven different IPs and twenty distinct study designs. These changes resulted in eighteen versions of the Case Report Form (CRF) and six versions of the Statistical Analysis Plan (SAP), along with corresponding updates to the Tables, Figures, and Listings (TFL) mock shells.

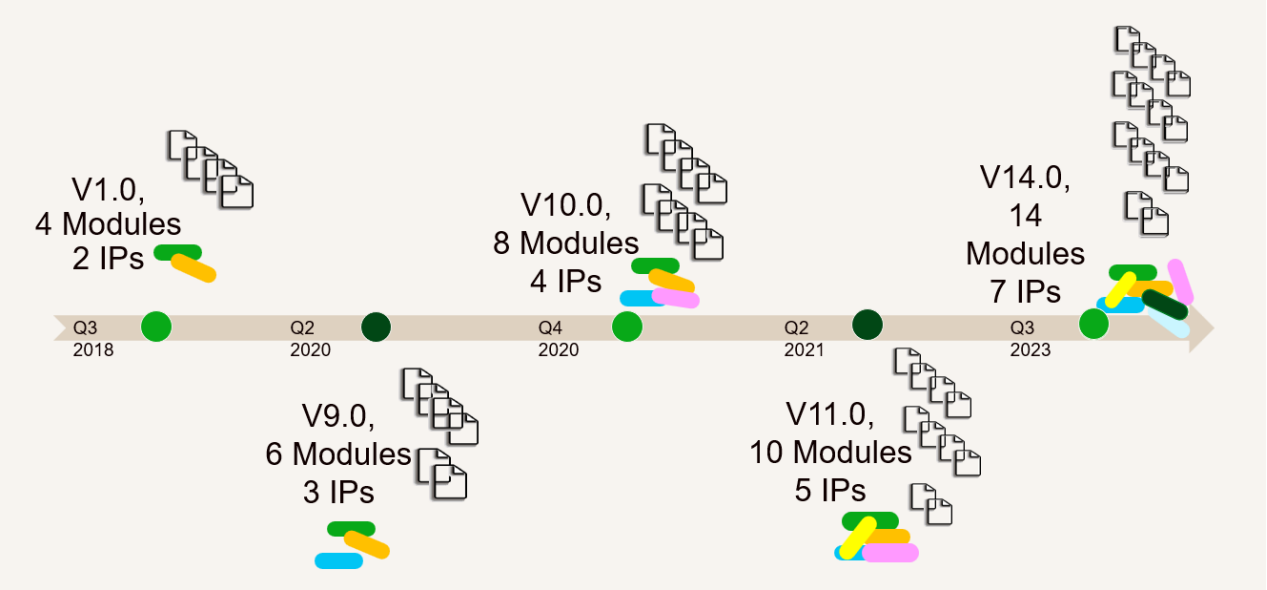

Fig. 1 – Study Protocol History – a real study example

The challenges encountered with such clinical trials are at multiple levels. We have previously described some of the impacts on Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets [1] and on Tables, Listings and Figures (TLFs) [2]

- The amount of raw data is large and ever increasing during the trial, and SDTM datasets need to support module-specific variations in trial design, collected data types, lock status of the modules (modules may lock as they complete and so the programmer is handling locked and unlocked data), differing versions of coding dictionaries.
- For ADaM datasets, baselines, analysis visit windows, exposure duration and imputations all can change from module to module depending on the design and schedules of assessments.
- Regarding TFLs, the same output, depending on the module(s) to report, can contain differences in column headers, footnotes but also in the content.

To manage this complexity, we have no one-size-fits-all solution, but instead a range of strategies that we have put in place. Previously [2] we described some of the approaches we use across these studies:

- Centralized programming (in each program, the logic is present for all modules of the study, as they are added to the protocol),
- Centralized data selection (data selection at the raw data level, in one central program, prior to creating SDTM datasets, so that SDTM datasets, ADaM datasets and transport files will contain the same version of data),
- Extensive use of macros and macro variables that grant flexibility in customizing programming code to module specificities.
- Data-driven and metadata-driven solutions to select the study modules/treatment groups applicable for the delivery in question and applicable to the analysis set used in each output.

CONTINUOUS MODULE-SPECIFIC DELIVERIES

In this paper we will focus on the situation in which no new study modules are being added, however, each module follows its own development program, with ongoing safety reviews, interim analysis, clinical study report (CSR) dry-run and final CSR. This results in the ongoing series of module-specific analyses, each requiring unique derivations and datasets- and outputs-modifications.

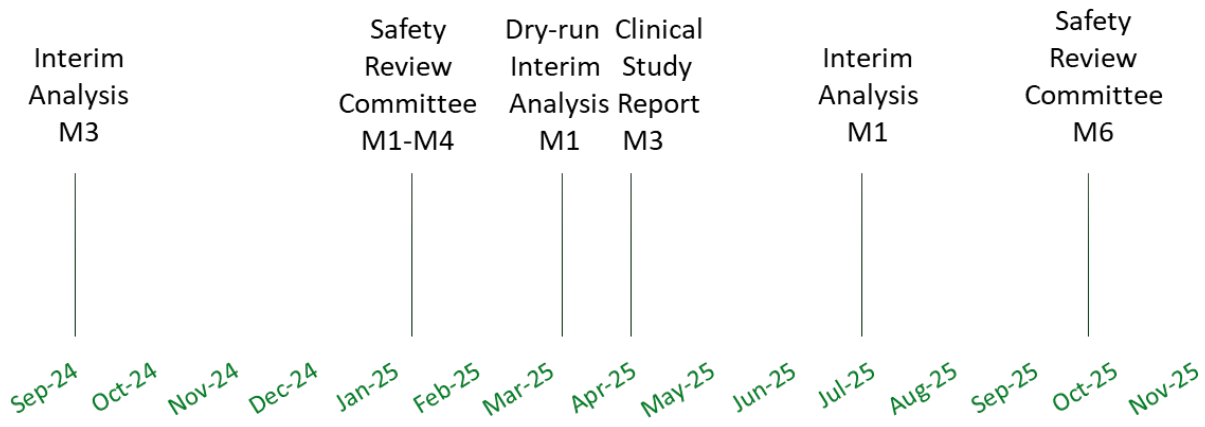

Fig. 2 – Continuous Module-specific Deliveries

These module-specific analysis releases are scheduled in rapid succession, sometimes even overlapping. Study teams often propose short timelines between releases, believing them to be feasible since the programming for SDTM, ADaM, and TLFs is already in place. However, the programs used for one release must be quickly adapted to accommodate the changes required for the next, which can be challenging under tight timelines. For this purpose, we focus on metadata-driven solutions to switch parts of code off and on.

THE CENTRAL DELIVERIES METADATA FILE

We use a central study file that stores metadata related to different study modules and analyses releases. This file stores, among other metadata, program and output file names, output titles and footnotes, and is read into SAS to apply the metadata when generating the outputs. In addition, the file contains a list of macro variables that get created in the SAS session when the file is read into SAS and are used for multiple purposes while programming.
Some of the macro variables are very general and not expected to be module-specific:

macro_var_name	macro_var_value	macro_var_purpose
meta_sponsor	Sponsor A	Text to be displayed in all outputs
meta_study	ABCD0003	Text to be displayed in all outputs
meddra_version	28.1	Dictionary version used for programs and to display in the output
whodde_version	March 2025	Dictionary version used for programs and to display in the output

We will describe some further examples of macro variable values for specific purposes of switching parts of code off or on.

SWITCHES FOR SELECTION OF SUBJECTS AND DATA CUTOFF (DCO) DATE OF THE RELEASE

For ongoing study deliveries (e.g. dry-run, interim analysis, safety review), it is common practice to analyze snapshots of data according to a particular date, i.e., the data cutoff (DCO) date for a formal analysis. In the modular studies, typically, analysis is performed on one study module only, or sometimes a combination of more than one module. Therefore, there is a need for each delivery to select data up to the DCO date, containing the desired selection of subjects based on the study modules they belong to.
We perform the data selection at the raw data level, in one central program, prior to creating SDTM datasets – this program uses macro variables to select the subjects and to data up to DCO date if needed.

macro_var_name	macro_var_value	macro_var_purpose
apply_dco	Y	Apply or not DCO rules on raw data (if not, then only subject selection based on the module will be done)
raw_dco	2025-09-02	Used as the DCO date to apply DCO rules on raw data
dcutdt	2025-09-02	Used in ADaM derivations as analysis cutoff date
module	1	Used in multiple programs to select subjects or perform module-specific derivations
meta_dco	Module 1 IA1: Data cut-off = 2025-09-02	Title to be displayed in the top left corner in all outputs

SWITCHES FOR SELECTION OF OUTPUT PROGRAMS AND TREATMENT GROUPS

Each analysis release has its own requirements, among others, in terms of subject selection (usually to include only subjects belonging to a certain study module), in terms of scope of TLF work (a list of TLF outputs to produce), and in terms of treatment groups needed (the individual groups subjects belong to, but also different ways of pooling multiple individual groups – can be module- and even release-specific).

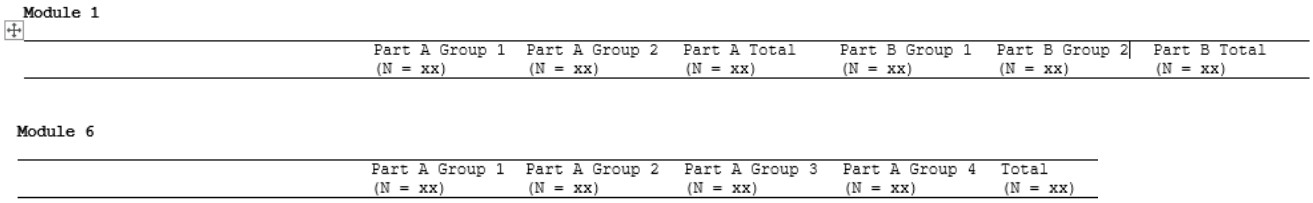

Fig. 3 – Treatment groups per module

Our central file flags outputs/programs to be selected when producing a release in the 'program_select' column:

output_number	program_name	output_name	program_select
14.x.y.1	table_xy	table_xy_1	CSRM3/M1IA1/SRCM6
14.x.y.2	table_xy	table_xy_2	CSRM3/M1IA1
14.x.y.3	table_xy	table_xy_3	CSRM3/M1IA1
14.x.y.4	table_xy	table_xy_4	CSRM3
14.x.z.1	table_xz	table_xz_1	CSRM3/M1IA1/SRCM6
14.x.z.2	table_xz	table_xz_2	
14.x.z.3	table_xz	table_xz_3	M1IA1
14.x.z.4	table_xz	table_xz_4	SRCM6

In addition, from the central file we define macro variables similar to this example:

macro_var_name	macro_var_value	macro_var_purpose
current_release	M1IA1	Used to select tlf programs to run
m1ia1	Y	Used in tlf macros to select trt columns and in multiple programs to perform release-specific derivations
srcm6	N	Used in tlf macros to select trt columns and in multiple programs to perform release-specific derivations

When read into SAS, the combination of macro variable current_release and the data from the 'program_select' column enables a run-all type of program to execute the appropriate selection of TLF programs and outputs.

SWITCHES FOR ANALYSIS SETS DEFINITION

Ongoing study deliveries (e.g. dry-run, interim analysis, safety review) often classify subjects into interim analysis sets. These follow the Statistical Analysis Plan (SAP) definitions of analysis sets, with the additional requirement that subjects should have received first dose of study drug at least xx months (or xx weeks) prior to the data cutoff date, where xx depends on study module and the nature of the delivery – in our example, for an SRC delivery the required time from DCO (xx months or weeks) is shorter than for an interim analysis. To avoid having to update the ADaM ADSL dataset program for each delivery, the program uses the macro variables from the central metadata file:

macro_var_name	macro_var_value	macro_var_purpose
iafl	Y	Apply IA definitions of interim analysis sets in ADSL
srcfl	N	Apply SRC definitions of interim analysis sets in ADSL
iasaftxt	3 months	Footnote text for outputs using IA definition of interim safety analysis set
iaefftxt	17 weeks	Footnote text for outputs using IA definition of interim evaluable for efficacy analysis set
iasafdays	90	Number of days from DCO date to use in the IA definition of interim safety analysis set
iaeffdays	118	Number of days from DCO date to use in the IA definition of interim evaluable for efficacy analysis set
srcsaftxt	3 months	Footnote text for outputs using SRC definition of interim safety analysis set
srcefftxt	13 weeks	Footnote text for outputs using SRC definition of interim evaluable for efficacy analysis set
srcsafdays	90	Number of days from DCO date to use in the SRC definition of interim safety analysis set
srceffdays	91	Number of days from DCO date to use in the SRC definition of interim evaluable for efficacy analysis set

SWITCHES FOR OTHER SPECIFIC MODULE DIFFERENCES

Some more examples of module- and release-specific requirements are shown below.

Table 14.x.y.z.w Progression-free survival (Evaluable for efficacy analysis set)

Module x

	Part A Total (N = xx)	Part B Group 1 (N = xx)	Part B Group 2 (N = xx)	Part B Total (N = xx)
Total events, n (%)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)

Individual Part A cohorts are not displayed.

Programming Guidance:
In case of Modules for which no subject included in Part B and individual Part A cohorts should be displayed, not present footnote "Individual Part A cohorts are not displayed.".
Add the footnote: "BRCAm plus corresponds to BRCA1/BRCA2/PALB2/RAD51C/RAD51D mutations." in all outputs where BRCAm plus grouping is displayed

Fig. 4 – Treatment groups selection example needed only for a specific output

macro_var_name	macro_var_value	macro_var_purpose
daysafteot	35	Number of days (xx) to use in ADaM derivations and in footnotes that need ADSL.EOTDT+xx
brcafn	Y	Display or not the module 1 specific footnote with the definition of BRCAm plus
partB	Y	Used for efficacy outputs to NOT display individual part A cohorts when partB=Y

CONCLUSION

Our aim is to avoid widespread code changes when having to switch between module-specific releases with short timelines. Through the extensive use of macros and macro variables that grant flexibility in customizing programming code to module specificities, combined with a central metadata file that determines the macro variable values when read into SAS, a new release becomes a matter of switching the appropriate items on or off in the central file, and rerunning the programs, as shown in the following example of switching from an IA analysis for module 1 (M1IA1) to an SRC analysis for module 6 (SRCM6):

macro_var_name	macro_var_value
meta_sponsor	Sponsor A
meta_study	ABCD0003
meddra_version	28.1
whodde_version	March 2025
current_release	M1IA1
module	1
m1ia1	Y
srcm6	N
raw_dco	2025-09-02
dcutdt	2025-08-02
meta_dco	Module 1 IA1: Data cut-off = 2025-09-02
apply_dco	Y
iafl	Y
srcfl	N
iasaftxt	3 months
iaefftxt	17 weeks
iasafdays	90
iaeffdays	118
srcsaftxt	3 months
srcefftxt	13 weeks
srcsafdays	90
srceffdays	91
daysafteot	35
brcafn	Y
partB	Y

macro_var_name	macro_var_value
meta_sponsor	Sponsor A
meta_study	ABCD0003
meddra_version	28.1
whodde_version	March 2025
current_release	SRCM6
module	6
m1ia1	N
srcm6	Y
raw_dco	2025-10-02
dcutdt	2025-10-02
meta_dco	Module 6 SRC: Data cut-off = 2025-10-02
apply_dco	N
iafl	N
srcfl	Y
iasaftxt	3 months
iaefftxt	17 weeks
iasafdays	90
iaeffdays	118
srcsaftxt	7 weeks
srcefftxt	13 weeks
srcsafdays	49
srceffdays	91
daysafteot	42
brcafn	N
partB	N

In terms of validation of the metadata-driven switches, most have easily noticeable effects on the TLF outputs and therefore require only a visual review of the TLFs. However, some switches influence analysis derivations and necessitate a statistician's review of both the derivation and the corresponding TLF output. Both the visual inspection and the derivation check are part of the standard process of preparing any analysis release.

REFERENCES

[1] Carla Santillan, Rakesh Reddy
Challenges implementing SDTM and ADAM in oncology platform trials
CDISC+TMF EUROPE INTERCHANGE 2024

[2] Francesca Dall' Acqua, Dubravka Georgijevic
Challenges in Oncology Modular studies: Fortrea approach
PHUSE EU CONNECT 2023 Birmingham

ACKNOWLEDGMENTS

We would like to express our warm gratitude to Vivien Mitchell and Nicky Lawrence for taking the time to review this paper and providing detailed and very helpful feedback.

CONTACT INFORMATION

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

Dubravka Georgijevic
Fortrea Netherlands BV
Schipholweg, 111 2316 XC
Leiden, The Netherlands
dubravka.georgijevic@fortrea.com

Carla Santillan
Fortrea Belgium SRL
Avenue du Bourget 50
Brussels 1130
Belgium
carla.santillan@fortrea.com