PharmaSUG 2023 - Paper DS-133

Data Evaluation and Harmonization in Integrated Summary of Safety (ISS)

Ben Gao, MSD

Alex Zheng, MSD

ABSTRACT

Integrated Summary of Safety (ISS) is critical in establishing integrated safety profiles across all relevant studies within an NDA application and used to support the preparation of eCTD module 2.7.4 and 5.3.5.3. But the large number of studies pooled together requires programmers to read through and understand all the related documentation for each study pooled. Different dataset structures, underlying logic and attributes also bring new challenges to the dataset integration and harmonization steps. This paper discusses the workflow to perform a productive ISS document review and the strategies to explore various data sources and establish data integration methodology. It also provides recommendations on programming practices to resolve ISS programming issues and achieve flexibility. Examples for ISS pooling strategy, dataset structures and SAS utility macros are included and explained in this paper.

INTRODUCTION

Integrated Summary of Safety (ISS) is to analyze safety data collected from various clinical studies, which can be used to support multiple purposes, e.g., eCTD module 5.3.5.3 Reports of analyses of data from more than one study (ISS), 1.16 (Risk Management Plan) or addressing agency requests. FDA considers ISS a crucial component of the clinical safety portions of a marketing or licensing application - it is required in NDA submissions and is encouraged for BLA submissions. The integrated analyses is not strictly required for NDA submission to the MHLW (Japan) and the EMA (EU), however MHLW and EMA submission must follow the CTD format, which contains sections in line with ISS. With ISS, it is easier to detect safety signals in the treatment groups since the pooled data from multiple studies is much larger than those of individual studies.

There are multiple strategies to create integrated safety datasets: 1) from individual study SDTM to integrated SDTM then to integrated ADaM, 2) from individual study SDTM to integrated ADaM, 3) from individual study ADaM to integrated ADaM. The third approach represents the most efficient one since it bypasses the extensive integration efforts at the SDTM level and offers flexibility to handle inconsistencies between studies in the stacked analysis datasets.

In the oncology therapeutic area, it is not uncommon to see compounds with multiple ongoing trials aiming for submission. To support each submission, it is suggested to build and maintain a reusable library of reference safety datasets (RSD) for the compound's prior submitted studies. The RSD library will help avoid the integration re-work at each individual submission (Jangid and Marrer, 2021).

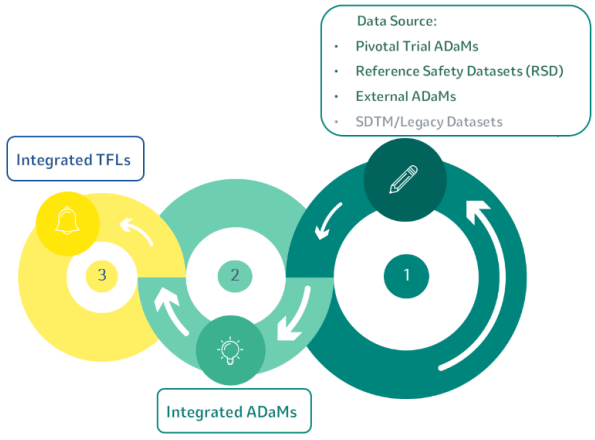

Figure 1. ISS Workflow

But even with standardized CDISC data structure and RSD library, the dataset integration of combination therapy studies can still be troublesome with the data transferred from external collaboration companies - those data usually have different structures and/or variable and parameter level derivations as compared to pivotal trial in submission. Apart from heterogeneous source data, the ISS process has its unique difficulties and subtleties than a typical CSR analysis, e.g., pooling strategy review, applicability issues, dataset attributes harmonization.

In this paper, we will go through the practices we have adopted in creating ISS deliverables for late phase oncology studies, assuming integration is conducted at ADaM level and RSD library have been established. We will introduce some key points in reviewing pooling strategy and mock tables and shed some light on how to conduct effective data evaluation. We will give a few programming tips on how to increase program reusability and flexibility during ISS programming set-up, dataset integration and table package generation.

POOLING STRATEGY AND MOCK TABLES

ISS mock-up is one of the most important documents for us to understand the integrated analysis strategy and scope of work. It usually contains 4 sections:

1. General notes: describing background, general requirements and programming notes.
2. Pooling strategy: documenting all the studies to be pooled together and their data cutoff dates.
3. List of Tables: listing all the tables requested for ISS.
4. Mock tables: demonstrating layout and structure for each unique table.

The mock-up review should be initiated as soon as the pooling strategy is finalized. A productive review will raise critical questions at an early stage to identify potential issues, e.g., incorrect pooled data listed, strategies for external data integration. The following considerations may help during the review process.

CHECK POOLING STRATEGY

A typical pooling strategy table contains the column header to be displayed in the integrated tables and gives a full breakdown of studies and cutoff dates associated with each column. There are mainly 2 categories of columns: treatment groups for pivotal trial in a submission (e.g., column A and B) and other pooled studies (e.g., column C and D) with which to compare the safety profile. The pooled treatment columns can be further expanded depending on specific submission. Examples include:

- Column for Reference Safety Dataset (RSD) for pivotal trial's compound.
- Column for all approved trials of another compound (e.g., Drug Y) in the combination therapy.
- Column for selected studies in the same cancer indication as pivotal trials.

Column	Header	Study and cut-off
A	Pivotal Trial X + Y	PVT-001: 10FEB2021
B	Pivotal Trial X + PBO	PVT-001: 10FEB2021
C	Pooled Treatment – Drug X (RSD)	RSD-001: 01Jan2013 RSD-002: 02Feb2014 Cohort A RSD-003: 03Mar2015 RSD-004: 04Apr2016 Part 2
D	Pooled Treatment – Drug Y	EXT-001: 05May2017 EXT-002: 06Jun2018 EXT-003: 07Jul2019 EXT-004: 08Aug2020 Stratum C

Table 1. Pooling Strategy Table Example

Every effort should be made to guarantee the pooling strategy is accurate. For in-house trials in the RSD, regulatory liaisons may provide a tracking spreadsheet listing all the study numbers and corresponding cut-off dates and population number. For external collaboration trials, ADRG section 2.2 Protocol Design in Relation to ADaM Concepts is the place to look up as it usually contains a table documenting the study information of the trial, e.g., protocol design, population, indication, dosing period.

Pivotal Trial X + Y	Pivotal Trial X + PBO	Pooled Treatment – Drug X (RSD)	Pooled Treatment – Drug Y

Table 2. Column Headers

Further confirmation with team (statistician, regulatory) is needed when:

- There exist studies not listed in aforementioned documents.
- There are studies listed for multiple pooled columns (belongs to different cohorts or stratums).
- Only a fraction of RSD studies is needed for indication-specific comparison.

CHECK APPLICABILITY

At the mock-up review stage, applicability comes in two aspects: data and relevance.

Data

Sometimes not all studies have the necessary data to support certain integrated analysis of a pooled treatment group. Although a comprehensive data review may not be feasible at this stage, browsing through all available data sources and accompanying define.xml will help identify possible missing data.

For example, an ISS analysis may be interested in comparing concomitant medication usage for AE of interest between pivotal trial's combination therapy (compound X + Y) and pooled monotherapy studies of compound Y. Of the four monotherapy Y studies in column D, two studies (EXT-001 and EXT-003) may not have ADCM created. Valid questions should be raised to determine if there is scientific value in comparing concomitant medication usage against this pooled treatment group when only a subset of studies have the data. If yes, more instructions should be added to the mock-up to clarify the background and decision. Footnotes or column headers should be updated to convey clearer messages to reviewers.

Relevance

The inherent logic of some tables may not apply to all the pooled columns. Examples are:

- Column B (drug X + placebo) may not be displayed for tables summarizing significant AE resulting in drug Y discontinuation or interruption.
- If AE of interest (AESI) is mostly induced by drug X, AESI should not be applied to monotherapy Y.
- Summary of Exposure for Drug X should not include monotherapy Y columns.

If needed, clinical, safety and statistician team should be consulted as to whether columns mentioned above can be left out.

LOOK INTO MOCK TABLES

Each mock table should be reviewed to ensure it is adjusted for the variances among studies and conforming with submission requirements.

Column Headers and Table Scope

Regulatory team may prepare a document detailing the eCTD components, within which the sections for eCTD 2.7.4 (Summary of Clinical Safety) and 5.3.5.3 (Integrated analysis of safety) should be carefully reviewed. This document stipulates the accurate language to describe pooled treatment headers and lists all the required tables to support both sections in eCTD. It is important to check against this document in the early stage of the process to avoid late-stage scope change.

Footnotes

Footnotes in ISS tables should account for the differences brought forth by different treatment groups. For example:

- AE/SAE follow-up period may vary across studies.
- Time conversion factor may be slightly different, e.g., from days to months/years.
- Lab toxicity may be graded upon different CTCAE versions.

Reviewing the footnotes may help dig out some deeper problems for the whole team to discuss and decide. For instance, team can decide if duration of exposure in months (converted with a different factor) should be reconverted for external data; or if different CTCAE grading criteria impact the lab toxicity grade comparison.

Alignment with CSR and Standards

ISS tables should be consistent with pivotal trial tables in terms of displayed contents, e.g., subgroups in AE summary tables, duration categories in exposure tables, title/subtitle wordings. Mock tables should conform with the latest company/TA standards.

PROGRAMMING ENVIRONMENT SET-UP

Multiple ISS packages may be requested to support submissions at different regions (e.g., US, EU, China), therefore it is convenient to use a utility macro to create subfolders for storing the outputs and documents of each package. The basic folder structure is illustrated in Figure 2. This macro can:

- Create separate subfolders to store programs and outputs for designated ISS packages (e.g., US, EU, China etc.).
- Create subfolders to store documents (mock-up, programming tracking sheet) and ADaM datasets.
- Copy ADaM input files (e.g., AE of interest, SMQ etc.) to study metadata folder.
- Copy company and TA standard macros to study macro library.

In consideration of the different data sources, a startup program can be created to set up libnames linking to all source data. It defines macro variables holding macros search path, SAS options, ADaM spec, input file names etc. Programmers should verify that the linked datasets reflect the correct data cut and MedDRA version.

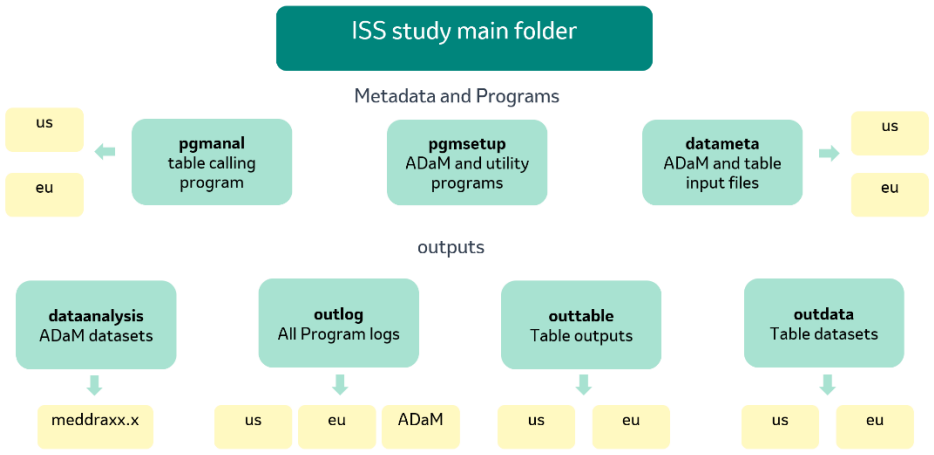

Figure 2. Folder Structure

DATA ASSESSMENT AND HARMONIZATION

With the environment for all the source data and supporting documents correctly set up, the next step is to evaluate different data sources and harmonize them into the same concept. Despite being the most time-consuming step, a thorough data evaluation will sort out the commonality and distinction among all contributing data sources and equip programmers with all the information at disposal when leveraging different normalization methods.

Establishing a compound-level RSD library will simplify the whole evaluation process as the data structures are fully integrated. But the co-development of combination regimens with external companies requires the pooling of external analysis datasets (e.g., compound Y in Table 2. Column Headers) for complete safety profile review. The recommended practices in this process are described below.

POOL SUFFICIENT DATA

Sufficient means pooling adequate but not excessive data to fulfill all the required analysis.

The first step toward creating compliant integrated analysis datasets is to check whether each dataset contains the ADaM-IG required variables, as well as TA/cancer indication specific variables. Then go through the mock-up and ensure the source datasets covered the data needed for every table.

On the other hand, programmers should review SAP/mock-up and confirm with ISS statistician about the specific data (e.g., population, period, phase, subgroup, parameters) to be summarized on each table. Sometimes, only a subset of data should be used in the analysis:

1. ISS AE summary might not be interested in toxicity grade change (often mapped to SDTM FA domain) summary, so only AE records with worst grade need to be included.
2. ISS population is normally set to All Subjects as Treated (ASaT), so screen failure, non-randomized and randomized but not treated subjects can be left out from integrated ADSL.
3. A multi-phase (e.g., neoadjuvant and adjuvant) pivotal study may create 3 separate sets of tables for neoadjuvant, adjuvant and combine phases, but it is likely that ISS tables for lab tests only focus on the combined phase. Only combined phase records need to be kept in the integrated ADLB.

DISSECT EXTERNAL DATA

External data can pose real challenges to data normalization efforts due to different ADaM implementations across companies. Start the data assessment by exploiting all possible documents, e.g., ADRG, define.xml and ADaM spec to familiarize with trial designs, data structures and algorithms. If necessary, some checking programs will help corroborate the understanding.

Analysis Treatment Period

For multi-period studies, understanding the trial design is critical to identify the ASaT population and to ensure the correct treatment dates are tabulated which would be used for tables like duration of exposure or exposure-adjusted AE. ADRG section 2.2 should document the ADaM structure in relation to analysis period, treatment variables (TRTxxP/A) and dates (TRxxSDT/EDT), especially for studies with complex design.

In the following tables, Drug Y 100 mg OD continuous dosing period is the target treatment analysis period, which may span several periods defined by TRTxxP/A. ADRG section 2.2 documented the definition of a set of newly created treatment variables (e.g., ANTRTP, ANTRTA, ANTRTSDT, ANTRTEDT) to identify this period easily (ANTRTA='Drug Y 100 mg OD'). We referenced this design and introduced those 4 variables to represent the main treatment period in the integrated ADSL. See Table 5. ADSL Spec Example for the normalization methods.

ANTRTSDT ANTRTEDT

TRT01P/A	TRT02P/A	TRT03P/A	TRT04P/A
Drug Y 50 mg BD	Chemo 1	Drug Y 100 mg OD + Chemo 1	Drug Y 100 mg OD
Drug Y 50 mg BD	Chemo 2	Drug Y 100 mg OD + Chemo 2	Drug Y 100 mg OD

Table 3. Study 1 Treatment Period

ANTRTSDT ANTRTEDT

TRT01P/A	TRT02P/A	TRT03P/A	TRT04P/A
Drug Y 100 mg BD	Chemo 1	Drug Y 100 mg OD	Drug Y 100 mg OD
Drug Y 100 mg BD	Chemo 2	Drug Y 100 mg OD	Drug Y 100 mg OD

Table 4. Study 2 Treatment Period

Variable Name	Variable Label	Define Derivation
ANTRTP	Planned Trt. for Pooled Anal. Period	Pivotal Trial: TRT02P; Pooled Treatment X: TRT01P; Pooled Treatment Y: ANTRTP.
ANTRTA	Actual Trt. for Pooled Anal. Period	Pivotal Trial: TRT02A; Pooled Treatment X: TRT01A; Pooled Treatment Y: ANTRTA.
ANTRTSDT	Date of First Exp. in Pooled Period	Pivotal Trial: TR02SDT; Pooled Treatment X: TR01SDT; Pooled Treatment Y: ANTRTSDT.
ANTRTEDT	Date of Last Exp. in Pooled Period	Pivotal Trial: TR02EDT; Pooled Treatment X: TR01EDT; Pooled Treatment Y: ANTRTEDT.

Table 5. ADSL Spec Example

Data Restructuring to Company/TA Standards

Lab toxicity grade tables are frequently generated in oncology studies for safety reporting, e.g., a summary table for worst post-baseline grade with increased and/or decreased lab values. Starting from ADaM-IG v1.2, new bi-directional toxicity grade variables (ATOXGRL, ATOXGRH, BTOXGRL, BTOXGRH) were added to address bi-directional toxicity reporting needs. But this design was not implemented in some external data (See Table 6. ADLB Comparison between Pivotal Trial and External Data) we received. After reviewing ADRG and data, a preliminary comparison was summarized below.

	Pivotal Trial ADLB	External Data ADLB
Parameter	Keep the original PARAMCD	One SDTM lab record is split into 2 records: <PARAMCD>DE and <PARAMCD>IN
Toxicity Grade	Low direction: ATOXGRL High direction: ATOXGRH	Toxicity grade was recorded in AVAL
Worst Post-baseline Grade flag	Low direction: ANL01FL High direction: ANL02FL	AOCC01FL: flags the first occurrence of the maximum severity (worst post baseline grade) of each parameter for each subject.

Table 6. ADLB Comparison between Pivotal Trial and External Data

Before normalizing the 2 distinct structures, the following points should be considered:

- Adhere to the pivotal trial's data structure if applicable, i.e., combining --DE an --IN parameters onto the same record. This structure is compatible with standard macros for creating lab tables.
- The number of split records in the low and high directions should be identical.
- The restructuring should not compromise the traceability to original lab record, i.e., carried-over variables (e.g., AVISIT, ADT, ATOXGR, LBSEQ) should stay unchanged.
- Check if the same parameter is represented differently (different PARAMCD or PARAM) between two data sources and if the directionality of each parameter is aligned with pivotal trial.
- There should be only one worst post-baseline toxicity grade (denoted by AOCC01FL) in either direction. Evaluate if this flag is period based.

After confirming answers to above questions, it was assured that the external ADLB was programmed as per our understanding and integrating it would still conform to internal standards and ADaM-IG rules.

USUBJID	PARAMCD	AVISIT	ADT	ATOXGR	ANRIND	AVAL	AOCC01FL	LBSEQ
EXT-002-02	MG	VISIT 1	01Jan2021	4	L	110		1
EXT-002-02	MG	VISIT 2	01Feb2021	2	H	78		2
EXT-002-02	MG	VISIT 3	01Mar2021	0	N	53		3
EXT-002-02	MG	VISIT 4	01Apr2021	3	H	96		4
EXT-002-02	MGDE	VISIT 1	01Jan2021			4	Y	1
EXT-002-02	MGDE	VISIT 2	01Feb2021			0		2
EXT-002-02	MGDE	VISIT 3	01Mar2021			0		3
EXT-002-02	MGDE	VISIT 4	01Apr2021			0		4
EXT-002-02	MGIN	VISIT 1	01Jan2021			0		1
EXT-002-02	MGIN	VISIT 2	01Feb2021			2		2
EXT-002-02	MGIN	VISIT 3	01Mar2021			0		3
EXT-002-02	MGIN	VISIT 4	01Apr2021			3	Y	4

Table 7. ADLB Structure in External Data

USUBJID	PARAMCD	AVISIT	ATOXGRL	ATOXGRH	ANRIND	ATOXGR	ANL01FL	ANL02FL	LBSEQ
EXT-002-02	MG	VISIT 1	4	0	L	4	Y		1
EXT-002-02	MG	VISIT 2	0	2	H	2			2
EXT-002-02	MG	VISIT 3	0	0	N	0			3
EXT-002-02	MG	VISIT 4	0	3	H	3		Y	4

Table 8. Final Integrated ADLB structure

Variable Normalization

The variable normalization step demands great attention to details. With external data increasing the difficulties, we have a few suggestions below:

- Be cautious when there are multiple variables holding similar information. For instance, there were two ADAE variables (AEREL and RELGR1) both showing causalities, but their relationship was not explained in the define.xml. Programmers should align with team on the correct logic and create a new variable (AREL) to normalize AE relationship from multiple source data.

Source Data		Integrated ADAE
AEREL	**RELGR1**	**AREL**
	Related	
NOT RELATED	Not Related	Not Related
POSSIBLY RELATED	Related	Related
PROBABLY RELATED	Related	Related

Table 9. AEREL vs RELGR1

- Examine the define.xml as there could be variables holding the desired information. A very common case in ISS is to normalize action taken with study medications in ADAE, which are often used to summarize AE resulting in the entire treatment or individual drug discontinuation. In pivotal trial, we have SDTM variables AEACN for overall action and ACNx (where x is medication number) for action taken with individual drug. But for external Drug Y data, no such variables exist. We eventually found a flag variable (AEN001FL), which is set to Y if there were any action taken equal to 'DRUG WITHDRAWN' in toxicity change records (SDTM FA domain).
- Begin with the end in mind. Variable normalization should not be a simple stacking of variables but should facilitate the creation of downstream tables. For example, AE occurrence summary was requested in our ISS package to compare the total number of AE events (episodes) across treatment groups. But this episode definition did not exist or differentiated greatly among data sources, so we created a new AEPISODE variable to harmonize values (see Table 10. Episode Info Breakdown). With this new AEPISODE, the table reporting macro logic is simplified as there is no need to write post-processing steps separately for each data source.

Source	Background	Original Format	Normalized AEPISODE
Pivotal Trial	No episode related variable. One worst AE record is one episode (ACAT1='WORST AE').	N/A	Pivotal Trial: derive episode number per USUBJID, AEDECOD. Concatenate this number with AEDECOD with '-'; Pooled Treatment X: concatenate AEDECOD and EPISODE; Pooled Treatment Y: retain the original value.
Pooled Treatment - Drug X (RSD)	Variable EPISODE exists: derived per USUBJID, AEDECOD.	Numeric: 1, 2, 3	
Pooled Treatment - Drug Y	Variable AEPISODE exists: derived per USUBJID, AELLT.	Character: Anemia-1, Anemia-2, Anemia-3	

Table 10. Episode Info Breakdown

PROGRAMMING TIPS

ISS programming should be initiated after performing mock-up and data review, but it can be tricky because:

- There is no ADaM-IG guidance on how to represent integrated treatment groups.
- Different variable lengths will cause annoying log issues.
- Codelist need to be re-created to accommodate all pooled databases.
- Different packages require its own set of programming set-up.

We presented below some programming tips to increase efficiencies and reusability.

CREATE POOLED TREATMENT GROUPS

ADaM-IG introduced a set of standard treatment variables (e.g., TRTxxP, TRTxxA, TRxxPGy, TRxxAGy) to support treatment-based analysis. But none of their definitions fit perfectly into ISS needs. For example, programmers cannot use TRTxxP(A) directly because each study was designed differently and TRTxxP(A) cannot be modified due to traceability concerns. TRxxPGy or TRxxAGy is still period-based so it cannot represent continuous dosing that spans multiple periods.

A workaround at our end is to create one population flag (APTwFL) for each pooled column. A utility macro *%iss0trtgrp* was created to select population and assign pooled treatment groups at TLF call programs. The macro variables (pop_where1, iss_column_ord1, iss_column_where1, iss_column_label1) fed into this macro was centralized global macro variables defined in startup program. Multiple sets of them can be defined to accommodate changing column number in different ISS tables.

USUBJID	TRT01P	TRT01A	APT1FL	APT2FL	APT3FL	APT4FL
PVT-001-01	Drug X + Y	Drug X + Y	Y			
PVT-001-02	Drug X + PBO	Drug X + PBO		Y		
RSD-001-01	Drug X 150 mg	Drug X 150 mg			Y	
RSD-001-02	Drug X 200 mg	Drug X 200 mg			Y	
EXT-001-01	Drug Y 100 mg OD	Drug Y 100 mg OD				Y
EXT-001-02	Drug Y 150 mg OD	Drug Y 150 mg OD				Y

Table 11. Population Flags in Integrated ADSL

```
%macro iss0trtgrp (in=, out=, trt=, trtc=, isw=, iso=, isl=);
 %let count_pipes = %sysfunc(countw(&iso.,|));
 %if %eval(&count_pipes>0) %then %do;
 data &out.;
    length &trtc.  $200;
    set &in.;

    %do i = 1 %to &count_pipes;
       %scan(&isw.,&i,|) then do;
          &trt. = %scan(&iso.,&i,|);
          &trtc. = "%scan(&isl.,&i,|)";
          &trtc. = strip(&trtc.);
          output;
       end;
  %end;
 run;
%end;
%mend iss0trtgrp;

%let pop_where1=%str(APT1FL='Y' or APT2FL='Y' or APT3FL='Y' or APT4FL='Y');
%let iss_column_ord1=% nrbquote (1|2|3|4);
%let iss_column_where1=%nrbquote(if APT1FL='Y')|
                       %nrbquote(if APT2FL='Y')|
                       %nrbquote(if APT3FL='Y')|
                       %nrbquote(if APT4FL='Y');
%let iss_column_label1 =%nrbquote(Pivotal Trial X + Y
                                 |Pivotal Trial X + PBO
                                 |Pooled Treatment Drug X (RSD)
                                 |Pooled Treatment Drug Y);
```

HARMONIZE VARIABLE ATTRIBUTES

Stacking datasets from different sources may cause log issues and truncation issues leading to potential data loss. To address this, a utility macro *%iss0len0chk* was created to assign variable length

automatically. It will loop through each data source and search for variables in the same name. If there is a hit, it will compare their types and only retain the character ones. Finally, it calculates the largest length for same-named variables and creates a macro variable holding a SAS length statement (e.g., length var1 $20 var2 $30 var 3 $25;). Programmers can simply copy and paste it when setting together different datasets.

```
%***  loop through all data and obtain variable length  ***;
%let libnum=%sysfunc(countw(&libs., |));
%do k=1 %to &libnum.;
  %let lib&k.=%qscan(&libs., &k., |);
  %let dat&k.=%qscan(&dats., &k., |);

  proc sql noprint;
   create table attr&k._%left(%trim(&&lib&k..)) as
   select distinct libname as lib_%left(%trim(&&lib&k..)), memname, name,
                   length as len_&k._%left(%trim(&&lib&k..)),
                   type as type_&k._%left(%trim(&&lib&k..))
   from dictionary.columns
   where upcase(libname)="%qupcase(&&lib&k..)" and upcase(memtype)="DATA"
         and upcase(memname)="%qupcase(&&dat&k..)"
   order by name;
  quit;
%end;


%***  combine variables across libraries to check length   ***;
data attr_&dataIn._com(drop=lib_: memname);
  merge %do k=1 %to &libnum.;
        attr&k._%left(%trim(&&lib&k..))(in=in&k.)  %end;;
  by name;
  if %do k=1 %to %eval(&libnum. - 1); in&k. or %end; in&libnum.;
run;
```

CREATE CODELIST FORMATS

In ISS study, some codelist should be re-defined to allow the unification of variable codelists from different studies. But mapping codelist values and adapting to updated codes in programming can be time-consuming and annoying, especially when there are external datasets. We can save those manual efforts by creating a utility macro to import the codelist file and generate a catalog of formats.

Codelist look up table should be set up in advance through ADaM spec or a separate document. After importing the look up table, format datasets will be created to meet the requirements of 'PROC FORMAT'. For both numeric and character formats, the dataset must contain the variables FMTNAME, START, and LABEL. TYPE can be set as 'C' or 'I' to denote character formats or numeric informats respectively. **Error! Reference source not found.** provides an example of the dataset structure. After the dataset is ready, format catalog can be created in the library via PROC FORMAT and the CNTLIN option. Figure 4 shows one informat generated from the codelist look up table and numeric variable ARELN can be created by *input(AREL, areln.)*.

FMTNAME	START	END	LABEL	TYPE
agegr8n	< 65	< 65	1	I
agegr8n	65 - 74	65 - 74	2	I
agegr8n	75 - 84	75 - 84	3	I
agegr8n	85+	85+	4	I
agegr8n	Unknown	Unknown	9	I

Figure 3. Format Dataset Structure

INFORMAT NAME: @ARELN LENGTH: 11 NUMBER OF VALUES: 2
MIN LENGTH: 1 MAX LENGTH: 40 DEFAULT LENGTH: 11 FUZZ: 0

START	END	INVALUE(VER. V7|V8 22APR2023:00:39:40)
Not Related	Not Related	1
Related	Related	2

Figure 4. Numeric Informat

This utility macro can be implemented in every ISS study to generate formats automatically and reduce the modification of ADaM programs. It's more efficient to maintain the look up table to separate data integration from the codelist maintenance. Besides, this macro can also check erroneous codelist mapping after integrated datasets are generated.

INCREASE PROGRAM REUSABILITY

Various deliverables in ISS may share some common outputs, thus it is good practice to centralize analysis related information as global macro variables in set-up programs, so table calling programs can make full use of them to minimize potential updates needed when addressing ISS requests from different regions or even different studies. Figure 5 shows the example to update package environment: population, subtitles and footnotes, column headers, different subgroup variables and labels (e.g., age groups, regions), output prefix (e.g., us0 for US ISS tables) and suffix. Programmers can just run %define0pkg to switch between different package environments (e.g., US, EU, China) without creating multiple startup files. There is no need for any further modification for table programs, which literally simplifies the process and increases programing reusability and efficiency.

```
%macro define0pkg(pkg=);   ➡ pkg can be us/ eu/rmp /chn etc.

libname   lptod    "&protpath/outdata/&pkg.";
filename  fptotb   "&protpath/outtable/&pkg.";
%if %lowcase(&pkg.)=%str(us) %then %do;
  %let us_eu_code=&pkg.;
  %let regn_grp=%str(region1^region1n);   ➡ US Region
  %let regn_lbl=%str(Region (US, Ex-US));
  %let regn_var=%str(REGION1);
  %let popn=5;                ➡ Number of column
  %let prefix=%lowcase(%trim(&pkg.))0c;   ➡ Use for creating different output names
  %let iss_column_ord1 = %str(1|2|3|4);
%let population_where1 = %str(A1FL='Y' or A2FL='Y' or OLATFL='Y' or %trim(%qupcase(&us_eu_code.)B)FL='Y');
%end;          ➡ Use for selecting US population
```

Figure 5. Set up Specified Environment to Generate Multiple Outputs

CONCLUSION

This paper outlines the basic principles in reviewing ISS pooling strategy and mock tables from aspects of submission document crosschecking, data applicability and table layout/structure. It elaborates how to perform data normalization and harmonization with different trial designs, data structures and variable derivations. It also proposes programming enhancement tips in combining data and setting up flexible programing environment to address different ISS requests. The presented examples should give ISS programmers preliminary sense to kick off ISS preparation and improve programming efficiency.

REFERENCES

Jangid, Ghanshyam and Marrer, Jane. 2021. "Processes and Techniques for creating Integrated Summary of Safety (ISS).", *PHUSE US Connect 2021*, Paper DS03.

SAS 9.4 and SAS Viya 3.5 Programming Documentation. "Results: FORMAT Procedure".

https://documentation.sas.com/doc/en/pgmsascdc/v_038/proc/p0owa4ftikc2ekn1q0rmpulg86cx.htm

ACKNOWLEDGMENTS

We would like to thank Nicole Zhang and Amy Gillespie for their comments and review on this paper.

CONTACT INFORMATION

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

Ben Gao
MSD China
+86 (21) 22112893
peng.gao4@merck.com

Alex Zheng
MSD China
+86 (21) 22112510
ze.wei.zheng@merck.com