Paper DS08

Overcoming Challenges of your Integrated Summary ISS/ISE Submissions

Jenet CP, Genpro Research, Trivandrum, India
Akhil Vijayan, Genpro Research, Trivandrum, India

ABSTRACT

The U.S. Food and Drug Administration (FDA) has advised that Integrated Summaries of Effectiveness and Safety (ISE, ISS) should be included in the Common Technical Document (CTD) while filing a New Drug Application (NDA). Both ISS and ISE are intended to provide a pool of databases that contain the results of all clinical trials. Integrating multiple studies from different phases of clinical trials can be challenging and it demands precise observation and planning. The most difficult and time-consuming part of the ISS/ISE process is creating the integrated datasets. Incorrect data integration leads to inaccurate analysis results that do not reflect the study drug's actual behavior. This paper explores some challenges that we face during dataset Integration (pooling of data), data harmonization (standardizing variables, dictionary up-versioning etc.), and validation in the ISS/ISE process and how to handle them with ease, alongside providing the best practices for ISS/ISE development process.

INTRODUCTION

The Integrated Summary of Safety (ISS) and the Integrated Summary of Effectiveness (ISE) are significant components in the New Drug Application (NDA). As the name implies, these are the documents comprised of integrated analyses used to analyze the efficacy and safety of an investigational product. Rather than just summaries, these include integrated analyses and clinical trial results from different therapeutic areas which are combined, and analyses are done to get combined statistical results.

The Integrated Summary of Safety (ISS) and the Integrated Summary of Effectiveness (ISE) is useful for:

- Comparison of results across multiple studies
- Identification of common related adverse events (AEs)
- Identification of common serious adverse events (SAEs)
- Assessment of efficacy in subgroups
- Identifying rare, or unexpected trends in different subgroups
- Assessment of risks and benefits of the drug

The process of combining datasets is time-consuming and challenging as several steps are involved. Before integrating datasets, it is important to identify the following points but, not limited to:

- Number of individual studies involved in the process
- Whether individual studies follow CDISC (Clinical Data Interchange Standards Consortium)/SDTM (Study Data Tabulation Model) format or non SDTM format
- Version of SDTM IG (Study Data Tabulation Model Implementation Guide)
- Version of Coding dictionary
- Version of Controlled Terminology

Once these are identified next step is the pooling of data followed by standardization and then validation. There are multiple methods used for integrating datasets. In this paper, some of the most common and efficient approaches are explained. Data harmonization happens through standardizing coding dictionaries, Implementation Guides (IG), controlled terminologies, etc. Also, it is important to check variable attributes and programming logic used in individual studies and standardize them in order to avoid inconsistency and programming errors. Validation of integrated datasets involves conformance checks with CDISC standards, checking the mapping and evaluate programming logic for important variables, etc.

The objective of this paper is to detail the challenges that are faced during dataset integration, data harmonization, and validation in ISS/ISE process along with some best practices.

DATASET INTEGRATION

Pooling of data is an important step in any integrated analysis as this helps in producing combined statistical results. As integrated summaries include more than one study, the first step is to identify the studies which are going to be part of the integrated analysis. Usually, this will be confirmed by the Sponsor and Contract Research Organization (CRO). Each study that is to be pooled for integrated analysis of safety or efficacy, should then be detailed by the study statistician in the integrated statistical analysis plan (SAP). A list of integrated analysis Tables, Listings, and Figure (TLF) outputs should also be included in SAP whose mock-ups should be provided as well. Once the integrated analysis is clear and supporting documents (electronic case report forms [CRFs], datasets specifications, etc) are available for each study, programmers can start to plan and design integration datasets. There are different approaches for integrating datasets depending on the individual studies and the most common and efficient approaches are explained below.

METHOD -1: INTEGRATING DATASET USING INDIVIDUAL ADaM DATASETS

In this method, individual ADaM (Analysis Data Model) datasets are created from individual SDTM datasets, and thus created ADaM datasets will be used to create integrated analysis datasets (Figure 1). Once programming is completed for individual SDTM and ADaM datasets, an integrated analysis dataset can be created by setting individual ADaM datasets together. When using this method programmers need to wait until the individual SDTM and ADaM datasets are validated. It would be better to fix the inconsistencies between individual studies while integrating. This might require standardizing programming logic, up-versioning of coding dictionaries, controlled terminologies, etc. If there are inconsistencies between study designs of individual studies or structural differences this also needs to be handled while integrating.

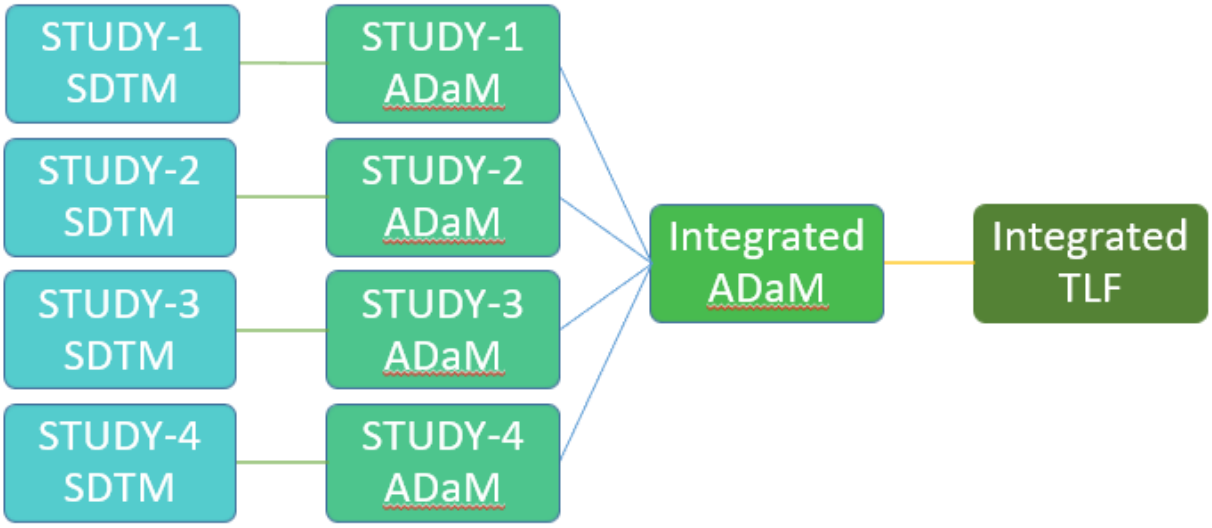

Figure 1- Integrating dataset using individual ADaM datasets

METHOD -2: INTEGRATING DATASET USING STUDY LEVEL TABULATION DATASETS

This approach is suggested when there are studies that are part of integration which are not CDISC compliant or if both standard and non-standard studies are included as study level tabulation datasets (SDTM datasets or RAW datasets), and are directly used to create integrated ADaM datasets (Figure 2). This method requires extensive programming and documentation to explain the transformations and standardization of datasets.

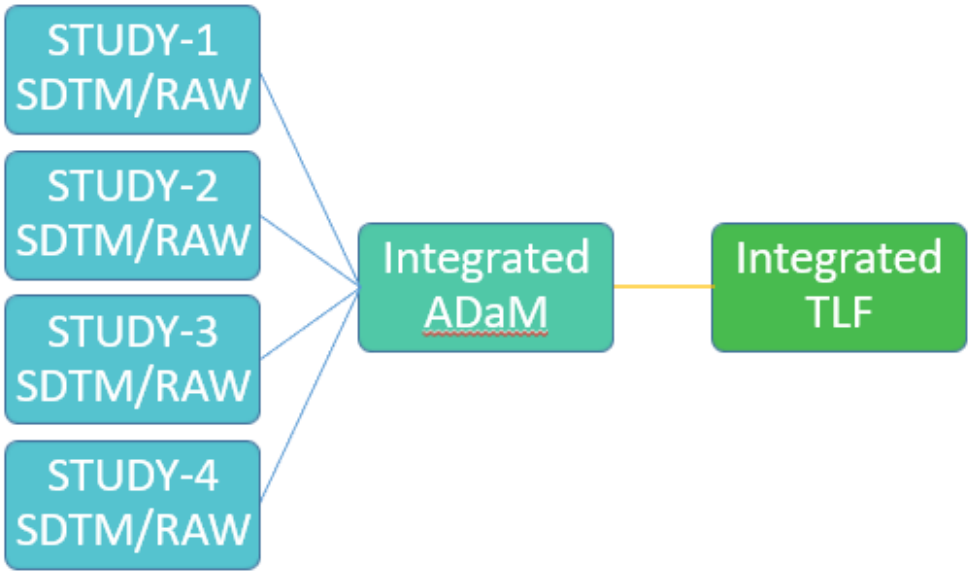

Figure 2- Integrating dataset using individual SDTM datasets

METHOD -3: INTEGRATING DATASET USING INTEGRATED SDTM DATASETS

This approach is most recommended and most commonly used, as this provides more traceability to integrated ADaM datasets and SDTM datasets. Once individual SDTM datasets are ready, supplemental datasets are merged back with corresponding parent domains. Integrated SDTM datasets can be created by stacking these processed SDTM datasets. Using these integrated SDTM datasets, integrated ADaM is created followed by TFLs as shown in Figure 3. This option can be used when all individual studies follow the CDISC/SDTM format. It enables the programmer to standardize and harmonize data at SDTM level rather than doing it at ADaM level which reduces programming efforts at ADaM level.

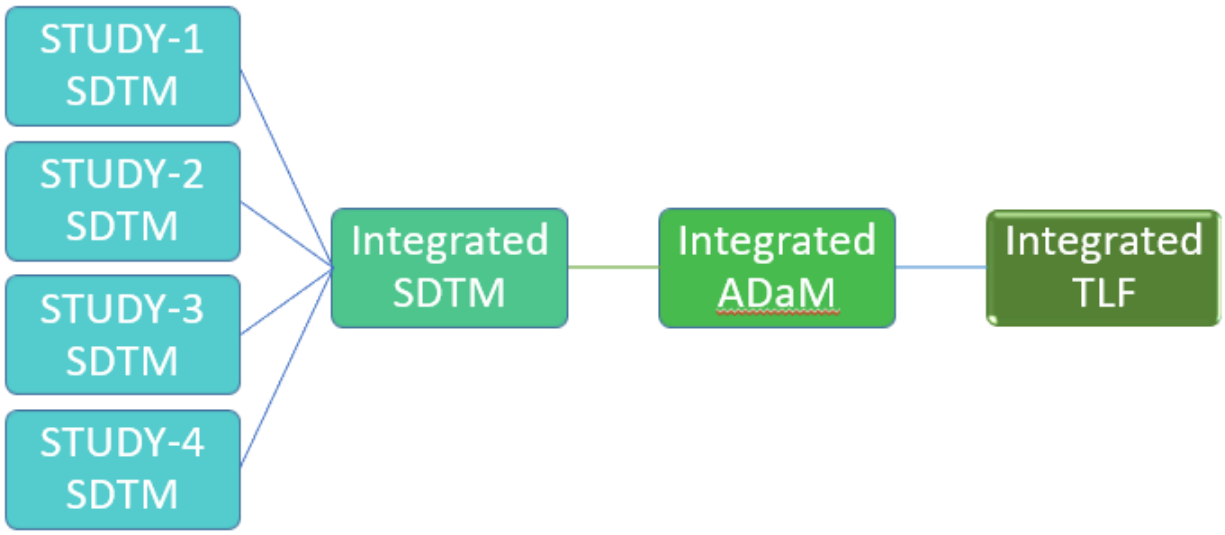

Figure 3- Integrating dataset using integrated SDTM datasets

DATA HARMONIZATION

Figure 4 shows different steps involved in harmonization of data.

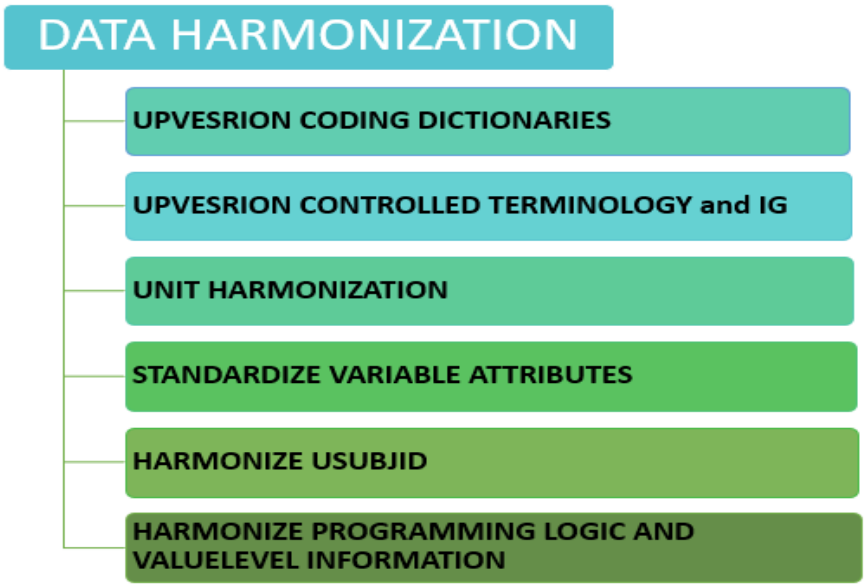

Figure 4-Different steps involved in dataset harmonization

UPVESRION CODING DICTIONARIES

It is important to ensure that all individual studies use a single version of coding dictionaries for the submission, as individual studies may follow different versions of the Medical Dictionary for Regulatory Activities (MedDRA) and WHODrug Dictionary. New versions of these dictionaries are released every few years and it is recommended that all studies follow a single version which is the latest. However, this can be decided by the Sponsor. Method 3 allows the programmer to easily implement the same standards for all studies. This helps to avoid inconsistency and version mismatches between studies.

UPVESRION CONTROLLED TERMINOLOGY and IG

Similar to coding dictionaries, implementation guidelines and controlled terminologies are released every few years and every quarter respectively. Hence it needs to be up-versioned to a single version to avoid mismatches and future programming errors.

Once all studies are up-versioned to a single version (preferably the latest) the same version will be used to validate ISS/ISE datasets. Table 1 shows an example of up-versioning different coding dictionaries and controlled terminologies for four different studies.

Study	MedDRA v	WHO	SDTM IG v	SDTM CT v
STUDY-1	20	B3 SEP2017	3.2	2018-12-21
STUDY-2	17	B2 SEP2014	3.1.3	2017-12-22
STUDY-3	17	B2 SEP2014	3.1.3	2014-06-27
STUDY-4	21	GB3 MAR2018	3.2	2018-12-21
UP-VERSION	21	GB3 MAR2018	3.2	2018-12-21

Table 1- Up-Versioning of coding dictionaries, controlled terminology and IG

UNIT HARMONIZATION

It is possible that ongoing studies are part of the submission, and is required to update units each time when eDC (Electronic Data Capture) is updated. In such cases, it will be helpful, if a spreadsheet is created which contains collected units (eg: lab), standard units and conversion factors, etc. This spreadsheet can be updated when new data comes and can be used in the program.

STANDARDIZE VARIABLE ATTRIBUTES

As multiple studies are involved in integrated submission there can be cases where the same variables in different studies have different attributes. Integration of datasets would be easier if these inconsistencies in variable attributes on type, label, length, etc. can be identified and fixed. This can be done using simple PROC CONTENTS and PROC COMPARE procedures in SAS. If the maximum real length from individual studies is identified and used for the integrated dataset, this issue can be solved.

HARMONIZE USUBJID

A subject should have unique identification for each trial. As per CDISC SDTM standards, SUBJID is the identification within a study, and USUBJID is used as a unique subject identifier across studies in the submission. If the same subject participated in multiple studies, this subject will have a different SUBJID in each trial but USUBJID must be unique across studies (Figure 5).

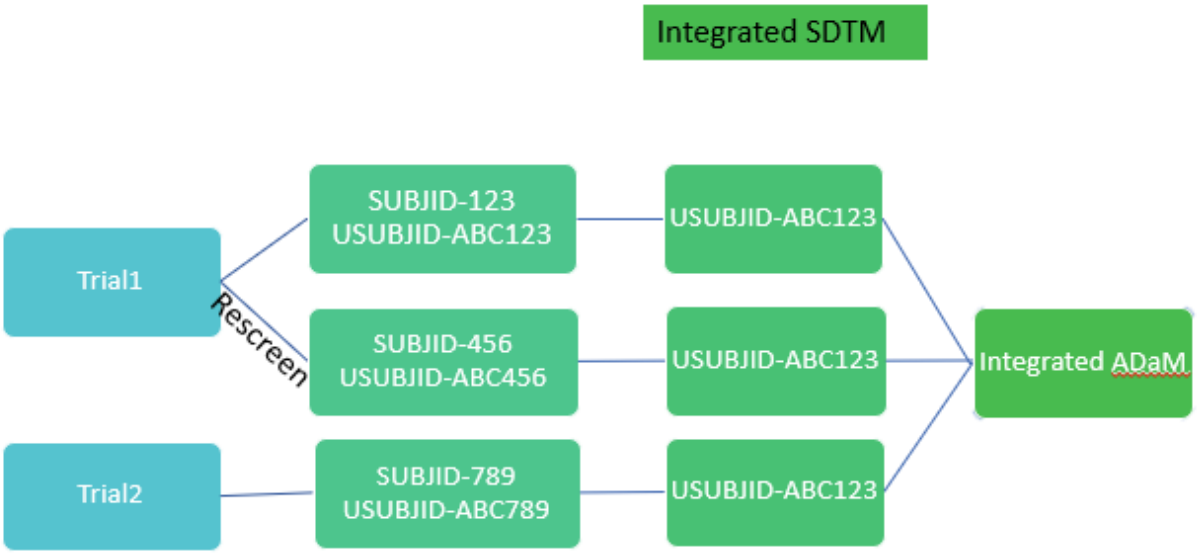

Figure 5- USUBJID Harmonization

HARMONIZE PROGRAMMING LOGIC AND VALUELEVEL INFORMATION

Once all inconsistencies which are mentioned above are identified and rectified, now is required to harmonize the derivation logic of variables. There is a possibility that derivation logic used for the same variables in individual studies (different datasets) can be different (eg: individual studies from multiple CRO/Sponsor). Different derivation logic can cause programming errors. This mostly happens in ADaM datasets. The below tables show inconsistent programming logic used for the Analysis flag for different studies. This is another challenge that is faced during data harmonization. This can be tackled by identifying the changes and standardizing them prior to dataset integration.

Study	Variable Name	Variable Label	Programming Notes
ADaM 1	ANL01FL	Analysis Flag 1	Sort the USUBJID AVISITN CHG ATPTN and assign Y for the last record within each Visit (where ATPTN>0)
ADaM 2	ANL01FL	Analysis Flag 1	Sort the USUBJID AVISITN CHG ATPTN and assign Y for the last record within each Visit (where ATPTN>0 and CHG is not missing)
ADaM 3	ANL01FL	Analysis Flag 1	Sort the USUBJID AVISITN CHG ATPTN and assign Y for the last record within each Visit (where ATPTN>0 and DTYPE is missing)
ADaM 4	ANL01FL	Analysis Flag 1	Sort the USUBJID AVISITN CHG ATPTN and assign Y for the last record within each Visit (where ATPTN is not missing)

Table 2- Example for inconsistent derivation logic

Another major point to be considered here is the value level information. The programmer needs to ensure that the value level information of variables is consistent before creating the integrated dataset. For instance, AETERM, SEX, RACE, etc. variables should have values in the same letter case. Similarly, if there are related variables like AVISIT, AVISITN, the one-to-one mapping should be consistent across studies and domains.

DATA VALIDATION

Once data standardization, harmonization, and integration of datasets are completed, now it's time to validate the final ISS/ISE datasets. Each dataset should be compliant with CDISC standards. This can be checked through Pinnacle 21 validator. Each of the issues in P21 report should be checked and fixed. If there are any unaddressed errors or warnings in the pinnacle 21 report, this should be documented in the study data reviewer's guide (SDRG) or analysis data reviewer's guide (ADRG).

It is necessary to have a statistical analysis plan (SAP) that clearly explains the statistical analyses that are performed for integrated analysis. Definitions analysis flags, baseline flags, treatment group, analysis populations and handling of missing data must be included in SAP. There must be separate mapping specifications for integrated SDTM and integrated ADaM datasets created based on SAP prior to the data integration process.

As the pinnacle 21 validator might not identify all issues, it would be better to cross-check the datasets with SAP and mapping specifications of integrated analysis ISS/ISE. Programmers can ensure that all studies include the domains required for integration and that their mapping contents are consistent. Also, it needs to be verified that the supplemental qualifiers have the same QNAM and QLABEL if supplemental domains are used for integrated datasets. Individual ADaM datasets should also be subjected to the domain and variable mapping validation.

CONCLUSION

Integrated Summaries of Effectiveness and Safety (ISE, ISS) are integral part of successful submission for New Drug Application (NDA) in US. The most difficult and time-consuming part of the ISS/ISE process is creating the integrated datasets. This paper detail about challenges that are faced during the data integration process, different approaches that can be used for data pooling, data harmonization tips.

REFERENCES

Bharath Donthi & Lingjiao Qi – “Best Practices for ISS/ISE Dataset Development”, PharmaSUG 2019
Priyank H Patel – “Navigating FDA requirement: ISS/ISE build strategies”, PhUSE Connect 2019
U.S. Food & Drug Administration**,** Study Data Technical Conformance Guide v4.8 (June 2021)
Kishore Pothuri & Kesava Vajjala – “Case Study: Integrated SDTM to Integrated ADaM to TLFs” SI08

ACKNOWLEDGMENTS

The content, ideas and recommendations presented in this paper are all developed from experiences in our career. These experiences come through previous companies, various industry leaders, colleagues, mentors, conferences, and direct experience.

CONTACT INFORMATION

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

Author Name: Jenet CP
Company: Genpro Research
Address: Second Floor, Nila Building Technopark, Campus
City / Postcode: Trivandrum, Kerala, India
Work Phone: +91 8281895063
Email: jenet.cp@genproresearch.com
Web: www.genproresearch.com

Co-author Name: Akhil Vijayan
Company: Genpro Research PVT LTD
Address: Second Floor, Nila Building Technopark, Campus
City / Postcode: Trivandrum, Kerala, India
Work Phone: +91 9061208074
Email: akhil.vijayan@genproresearch.com
Web: www.genproresearch.com