PHUSE EU Connect 2021

Paper DS07

Integrated Summary of Safety and Efficacy Production Strategies

Navneet Agnihotri, Syneos Health®, Gurgaon, India

Sumit Pradhan, Syneos Health®, Gurgaon, India

Abstract

The integrated summary of safety (ISS) and integrated summary of efficacy (ISE) are vital components of a successful submission for regulatory approval in the pharmaceutical industry. ISS and ISE allow reviewers to easily compare individual outcomes, tracking subject's results across the entire clinical development of the investigational product, facilitating broad views of the investigational product's overall efficacy and safety profiles. With ISS and ISE, a single database is formed by pooling the results of all the clinical trials.

This paper discusses step-by-step guidance for the efficient and accurate creation of ISS and ISE datasets.

1) Approaches to integration methods for safety and efficacy analysis data pool.
2) Best practices to ensure the consistency of integrated datasets by up-versioning all data with the same coding dictionaries version, controlled terminologies version, harmonizing values of critical variables and all variable attributes.
3) Finally, additional potential compliance strategies to maintain CDISC compliance across individual studies and integrated data for safety and efficacy analysis to maintain traceability.

Introduction

Integrated Summary of Safety (ISS) and Integrated Summary of Efficacy (ISE) are Regulatory submission documents which are required to be submitted to the Food and Drugs Administration (FDA) while filing a New Drug Application (NDA). The purpose of these documents are to report the outcomes of one or more clinical trials. With ISS and ISE, a single database is formed by pooling the results of all the clinical trials. This pool of database is much larger than those of individual studies, hence, it is easier to detect significant statistical differences in the treatment groups.

Integration of data from a number of clinical trials for an Integrated Summary of Safety (ISS) and Efficacy (ISE) requires careful planning and includes the following planning steps :

- Assess the analysis and reporting requirement for the ISS/ISE.
- Consider these requirements against pre-existing study level analysis and reporting.
- Determine what data types need integrating across studies and at what level (SDTM/ADaM) the integration should occur at.

However, building integrated datasets is a challenging task as it requires the programmer to achieve consistent structures and formats while also ensuring that each dataset is CDISC-compliant.

Elements of ISS and ISE

- Assessment of summaries and statistical analysis of safety and efficacy data collected from various clinical studies.
- Evaluation of adverse event effects in various subgroups of the variable patient base.
- Impact of concomitant medications' safety and efficacy.
- Assurance of the appropriate dosage of the drugs.
- Assurance of long-term effect of the product, in case of chronic conditions.
- Effectiveness of drug in case of chronic condition and assessment of its long-term effect.
- Identification of the effects of test products when used with other medications.
- Assurance that the results of the data support the benefits of the drug and overweighs the risks.

Integrated Analysis Planning

Before any programming activities, the sponsor should evaluate and determine which studies will be part of the submission. The study statistician should then detail each study to be pooled for the ISS/ISE in an integrated statistical analysis plan (SAP) for Safety or Efficacy. The SAP should also include a list of integrated analysis tables, listings, and figure (TLF) outputs, whose mock-ups should be provided as well. Once the scope of the integrated analysis is clear and supporting documents (electronic case report forms [CRFs], datasets specifications, etc.) are available for each individual study, programmers can start to plan and design integration datasets.

Since most of the regulatory agencies expect the sponsor to submit trial data using CDISC standard format for any application that involves an ISS and ISE, this paper assumes all individual studies have already been converted to CDISC format.

Integrated datasets for ISS/ISE can be built in several different ways. The three most common approaches are described below.

Approach 1 – Integration Analysis Data Pool Using ADaM Data of Individual Studies

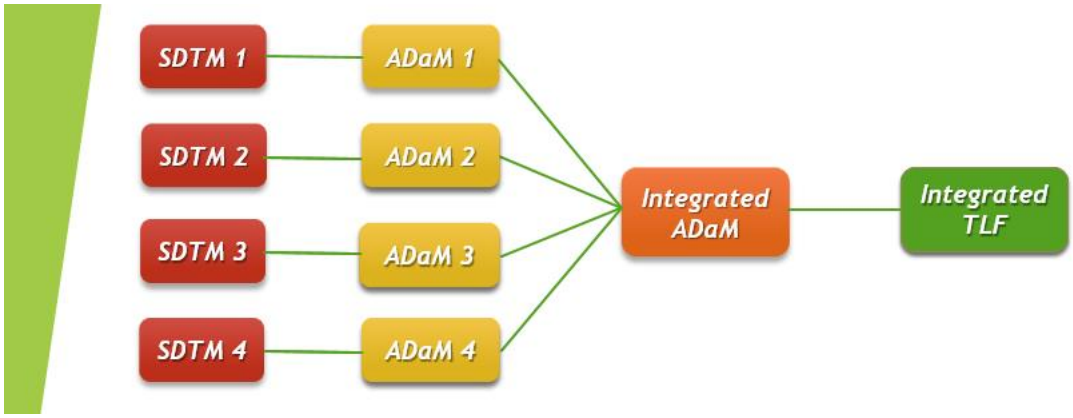

Figure 1

In this approach, the standardization effort can become extensive and get muddled due to the involvement of substantial programming effort. The programmers need to wait until the individual study's SDTM and ADaM datasets are programmed and validated. It can require extensive documentation to explain the data transformation and standardization to prevent any confusion on the reviewer's end.

Approach 2 – Integration Analysis Data Pool Directly from SDTM Data of Individual Studies

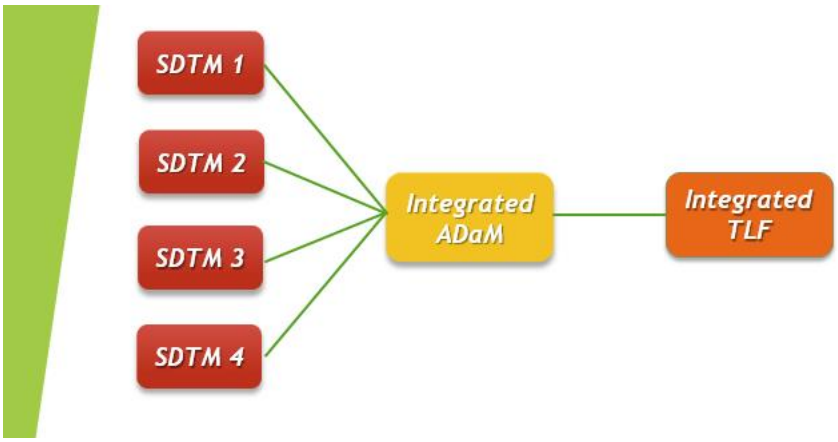

Figure 2

A possible drawback for this approach is that any common and general update in SDTM data would require updates in all the individual SDTM studies. However, this approach offers flexibility to handle inconsistencies between studies in the stacked analysis datasets.

Approach 3 – Integration Analysis Data Pool from Integrated SDTM Pool

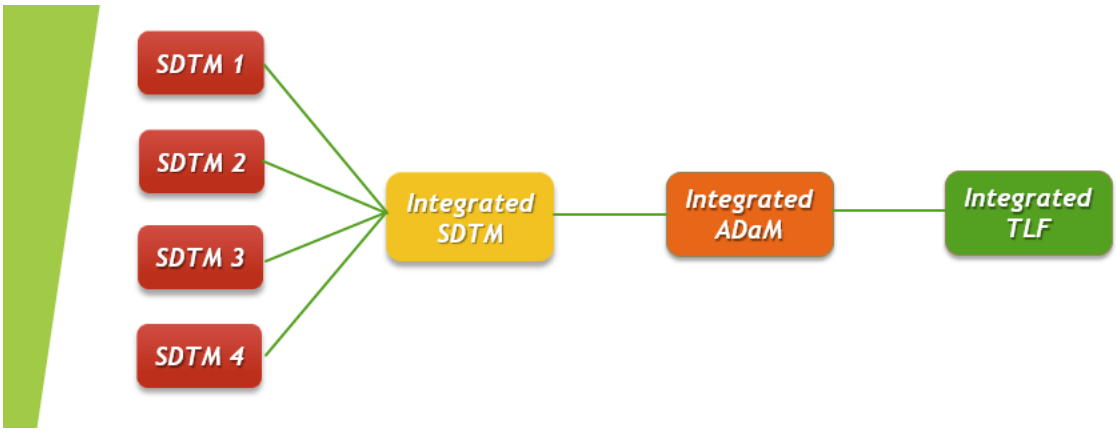

Figure 3

This option provides the sponsor higher flexibility to attain standardized and harmonized data. It also helps in defining complete traceability to integrated ADaM datasets and TLFs as the integrated SDTM pool data can be a single source of input downstream. This approach may make it easier to spot inconsistencies between studies, but it can be very time-consuming.

The user may opt for any of the above approaches to develop integrated analysis data pool of ISS and ISE based on the analysis requirements, efforts in terms of time and data consistency and traceability.

Data Harmonization Strategies

➢ Standardizing Coding Dictionaries

Since new versions of coding dictionaries are released every few years, it is very likely that not all individual studies used in the integration are coded with the same version of dictionaries. For submission, it is required that all studies use the same version (preferably latest) of the Medical Dictionary for Regulatory Activities (MedDRA), WHODrug and Common Terminology Criteria for Adverse Events (CTCAE) on applicable domains data to avoid data differences arising from version mismatches. The sponsor need to confirm a single version for each above mentioned coding dictionary to implement the same standards across all individual studies without changing the individual study specific requirements.

For instance, if the latest MedDRA version 23.1 has been used in one of the contributing studies and lower MedDRA versions in other studies for coding verbatim terms, all the contributing studies should be upgraded (if applicable) to use the same drug dictionary.

Generally, reconciling coding dictionary versions is completed by coding specialists in the data management department before data is pooled to create integrated ISS/ISE datasets. However, it can be handled at SDTM/ADaM programming end as well, if required.

➢ Standardizing Controlled Terminology Usage

The CDISC Controlled Terminology for SDTM and ADaM are released in each quarter of the year and there are possibilities that different trials may have implemented different versions of CDISC Controlled terminologies at a given point based on study requirements and the Controlled terminology available during the study development. Similar to coding dictionaries, a single version of CDISC Controlled terminology should be used to validate the individual contributing study before creating a pool data to avoid discrepancies owing to version mismatches. All these discrepancies should be rectified before the data is considered final for the integration analysis use. Also, the same Controlled terminology version will be used to validate ISS/ISE pooled data.

For instance, if a CDISC Controlled terminology of 2021-06-25 has been used to validate codelist values for one of the contributing studies and lower versions to validate other studies, all the studies should be validated via same CDISC Controlled terminology version of 2021-06-25 to avoid data discrepancies and standards issues during and post data pooling.

Note – In cases where the same version of drug dictionaries or the CDISC Controlled terminologies could not be applied to all contributing studies, the specifics should be documented in Reviewer's Guide and Define.xml.

➢ Harmonize STUDYID

As per CDISC, STUDYID is a clinical trial identifier and is a required variable which has a unique value across all the domains in a study, be it SDTM or ADaM. Since, for the integration of ISE and ISS, multiple individual studies are used to create a pool of data, the STUDYID for the pooled data has to be a unique value across standards, domains and records. The STUDYID in a pooled data cannot be distinct values of each contributing study but rather a single unique value. This value of STUDYID for

ISS/ISE pooled data could be either value of STUDYID of one of the contributing studies or a customized value based on sponsor decision. This harmonization of STUDYID will be managed at programming level of ISS/ISE, be it at SDTM or ADaM level based on the integration method opted by the programmer.

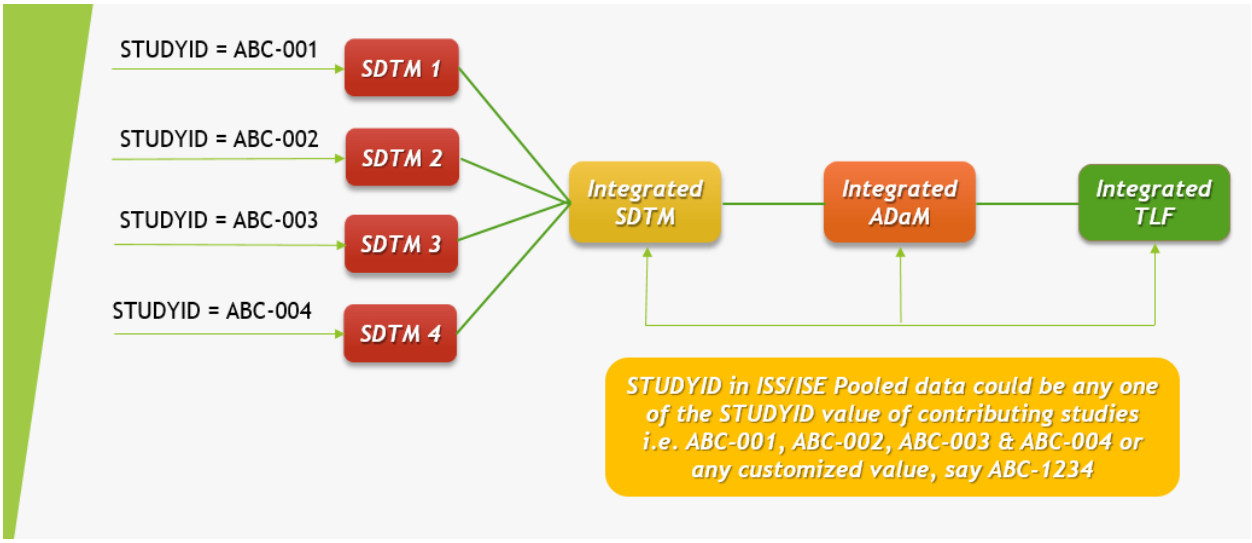

Figure 4

➢ Harmonize SUBJID and USUBJID

In CDISC SDTM world, the SUBJID is the individualized ID of the subject participating in a study. The subject must have a unique identification for each trial. Similar to STUDYID, to identify a subject uniquely across all studies for all applications or submissions involving the product, a unique subject identifier (USUBJID) should be assigned and included in all datasets. The USUBJID is required in all datasets containing subject-level data. USUBJID values must be unique for each trial participant (subject) across all trials in the submission portfolio.

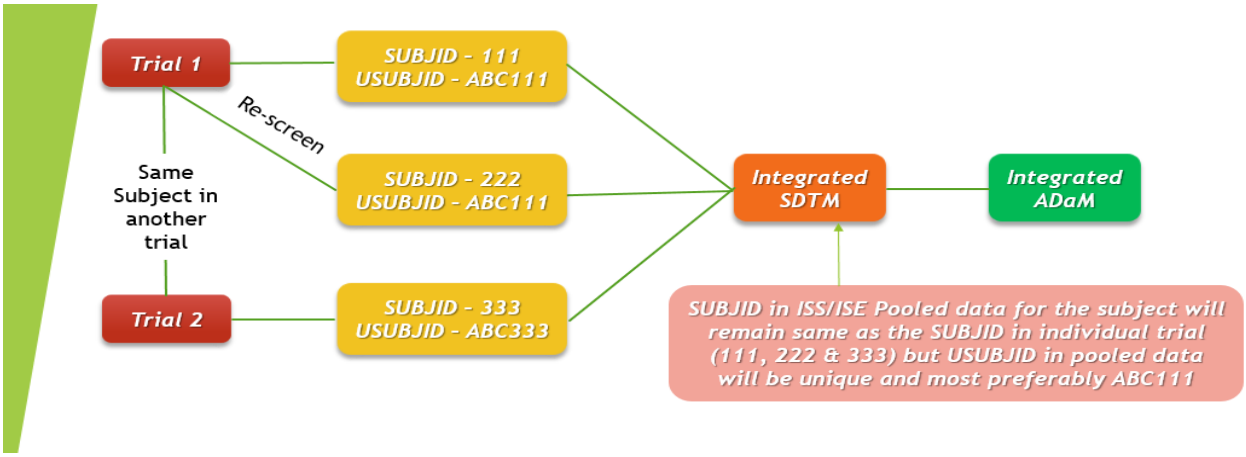

Figure 5

As suggested in above figure, the SUBJID of a subject in pooled data would be same as that in individual trials. However, the USUBJID of a subject which participated in more than one trial would be same in pooled data. The sponsor should control the list of the unique trial

subjects across portfolio which can support harmonizing and verify the USUBJID data information at the integration level as exemplified in Figure 5.

Note - The same person who participates in multiple clinical trials (when this is known) must be assigned the same USUBJID value in all individual trials.

➢ Harmonize Variable Attributes

Assuming individual study datasets have the same dataset structure and all coding dictionaries and controlled terminologies across all studies are up-versioned with a single version, the next step is to ensure that the variables with the same name across individual studies have the same attributes before stacking them together. Reconciling attributes is essential for preventing future programming errors and warnings. Attribute checks on variable types, formats, labels, length, etc. should be performed and in case of any discrepancy, the same should be rectified in applicable studies before data pooling is done.

Below is an example showing differences in the length of VISIT variable of SDTM across four individual studies (Figure 6). Attribute checking can be done easily with a PROC CONTENTS and PROC COMPARE procedure in SAS.

Study	Variable Name	Variable Label	Variable Length
SDTM 1	VISIT	Visit	$20
SDTM 2	VISIT	Visit	$50
SDTM 3	VISIT	Visit	$200
SDTM 4	VISIT	Visit	$80

Figure 6 – Variable with Same Name but Different Length Attribute

➢ Harmonize Derivation Algorithm and Value Level Metadata Across Studies

Once the above-mentioned pointers are identified and harmonized, it's time for the programmer to ensure value level harmonization of variables. The variables with same name must be derived using same algorithm across all contributing studies. Additionally, variables with same name must contain consistent and similar applicable content across studies before stacking individual datasets together. Even Qualifier Value Names (QNAM), Qualifier Value Labels (QLABEL) and Qualifier Values (QVAL) of variables displayed in Supplemental Qualifiers (SUPP--) datasets in SDTM must be consistent across studies (if they represent same information in studies). Different variable content could result in severe programming errors as they present different meanings.

Variables with Same Name Must Have Similar and Consistent Values Across Studies

Figure 7 below presents an example showing inconsistency of the relationship between VISIT and VISITNUM variables at SDTM level across studies. Before stacking the individual datasets together to generate integrated pooled data, the value of VISIT when VISITNUM is equal to 40 in Study 3 needs to be adjusted to be consistent with Study 1 and Study 2, and VISITNUM in Study 4 needs to be reassigned to ensure data consistency across all studies.

Visitnum	Study 1 Visit	Study 2 Visit	Study 3 Visit	Study 4 Visit
10	Screening	Screening	Screening	
20	Visit 1	Visit 1	Visit 1	Screening
30	Visit 2	Visit 2	Visit 2	Visit 1
40	Follow-Up	Follow-Up	Visit 3 – Follow-Up	Visit 2
50				Follow-Up

Figure 7 – Inconsistent Values for Same Variable across Studies

Both at SDTM and ADaM level, performing a quality check on data values for consistency on categorical values could be an essential step to ensure if special attention is required when combining data in integrating all datasets. For example :

a) Are values of variables (--CAT, --TESTCD, PARAMCD etc.) capitalized where applicable as per CDISC standard?
b) Do the classified categories (LBSCAT, DSSCAT, FAOBJ) have the same wording, spacing, abbreviations, capitalization, etc.?
c) Does the same variable with the same value have the same meaning?

The consistency in value level data of same variables across studies is a critical step in ensuring correct and harmonized pooling of integrated data. The programmer needs to be 100% compliant in this regard to avoid inconsistencies or inefficient analysis results from pooled data.

Variables With Same Name Must Have Consistent Derivation Algorithm Across Studies

To ensure effective analysis outcomes from integrated pooled data, it is important that the similar variables across studies are derived using same constant approach and ease in expected outputs. If the programming algorithm is inconsistent within and across studies, the same has to be updated and rectified before creating a pooled data.

Figure 8.1 below presents an example showing inconsistency in derivation logic of DTHFL variable in DM domain of SDTM. The derivation of DTHFL in Study 2, 3 and 4 has to be updated with respect to Study 1 so that all the possible scenarios are covered while deriving DTHDL and that no discrepancy is available in individual study data for DTHDL variable either due to programming inconsistencies or due to data issues.

Study	Variable Name	Variable Label	Programming Notes
SDTM 1	DTHFL	Subject Death Flag	Set to Y if AE outcome is "Fatal" or Serious Event criteria "Results in Death" is checked or Reason for Subject Discontinuation from Study is "Death"
SDTM 2	DTHFL	Subject Death Flag	Set to Y if AE outcome is "Fatal" or Reason for Subject Discontinuation from Study is "Death"
SDTM 3	DTHFL	Subject Death Flag	Set to Y if Reason for Subject Discontinuation from Study is "Death"
SDTM 4	DTHFL	Subject Death Flag	Set to Y if AE outcome is "Fatal" or AE Serious Event "Results in Death" is checked

Figure 8.1 – Same Variable with Different Derivation Logic

Figure 8.2 provides an example of inconsistent programming logic used when deriving the same variable, ANL01FL, in the efficacy dataset, ADEFF, across all individual studies. Though all four studies have the same variable named ANL01FL in dataset named ADEFF, the algorithm used is different in all the studies. Before stacking ADEFF from these four individual studies to create an integrated ADEFF dataset, the programmer and study statistician should agree to apply one consistent programming logic to derive ANL01FL and discuss how to update any other needed derivations.

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)

Figure 8.2 - Same Variable with Different Derivation Logic

➢ Re-evaluating and Re-deriving Key Analysis Variables in Integrated Pooled Data

Developing Integrated Pooled data for ISS/ISE is not simply stacking of individual studies data but far more than this. There are certain key variables in SDTM and ADaM end which need to be re-derived based on analysis requirements.

For instance, in case of developing Integrated SDTM pool data from individual studies SDTM data, Subject Reference Start Date/Time (DM.RFSTDTC) needs to be re-evaluated and decided for each subject so that other parameters (Baseline Flags, Study day, Treatment Emergent Flag etc.) dependent on RFSTDTC could be derived from it and hence be consistent in pooled data for all subjects.

Similar would be the case with Date/Time of Informed Consent (DM.RFICDTC), Subject Reference End Date/Time (DM.RFENDTC) and Date/Time of End of Participation (DM.RFPENDTC). The programmer and the study statistician shall consult with sponsor and decide over how the above variables would be taken care of in pooled data. Generally, RFICDTC and RFSTDTC shall be the earliest date/time per subject and RFPENDTC as the last date/time per subject across all contributing studies.

Alike in SDTM, certain analysis variables may have to be re-derived at ADaM end (if developing Integrated Analysis pool data from individual studies ADaM data) based on the analysis requirements. For instance, the programmer may need to re-derive Baseline Record Flag (ABLFL), Analysis Flag (ANL01FL and similar) or re-phrase AVISIT/AVISITN to maintain consistency in data across subjects and hence obtain relevant analysis outputs.

Other Potential Compliance Strategies

- Before being pooled to create integrated analysis datasets, the SDTM and ADaM datasets in each individual study should have undergone conformance checks (e.g., Pinnacle 21) with CDISC standards. Any warnings and errors found in the conformance check report should have been properly addressed. Unaddressed warnings and errors should be documented in the study data reviewer's guide or analysis data reviewer's guide.
- Separate SDTM and ADaM specifications have to be in place for Integrated ISS/ISE for reference to check for the mapping and derivation of variables in each CDISC Standard.
- One of the CDISC rules for ADaM is the one-to-one mapping requirement for designated variable pairs. After stacking all individual ADaM datasets together to create the integrated dataset, it is worthwhile to check all CODE and DECODE variables to ensure they maintain a one-to-one mapping.
- The SAP for ISS/ISE should list details on statistical methods and statistical analysis rules for developing integrated analysis datasets and programming integrated TLFs including

definitions for baseline flag, integrated analysis treatment groups, integrated analysis populations and missing data.

Conclusion

Integrated analysis planning is a complex process by nature. The data analysis cannot be simply executed by stacking multiple tabulations or analysis data of the different trials together. This paper defines several factors thats must be considered and planned to preserve the integrity and accuracy of the data. It requires step by step planning based on analysis requirements and regulatory authorities submission expectations. Following the right approach and techniques from the beginning shall assist in efficient and accurate creation of ISS and ISE datasets.

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). https://www.fda.gov/media/147233/download
U.S. Food & Drug Administration**,** FDA Data Standards Catalog v7.2 (07-01-2021) https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

Acknowledgement

The authors would like to acknowledge the "Material Review Board" group of Syneos Health for their thoughtful review of this manuscript.

Contact Information

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

Navneet Agnihotri
Syneos Health®, Principal Statistical Programmer
Building No. 14, Tower B, DLF Cyber City, Gurgaon - 122002, Haryana, India
E-mail: navneet.agnihotri@syneoshealth.com
LinkedIn: https://www.linkedin.com/in/navneet-agnihotri-78540553/

Sumit Pratap Pradhan
Syneos Health®, Principal Statistical Programmer
Building No. 14, Tower B, DLF Cyber City, Gurgaon - 122002, Haryana, India
E-mail: sumit.pradhan@syneoshealth.com
LinkedIn: https://www.linkedin.com/in/sumit-pradhan-71133345/