PharmaSUG 2017 - Paper SS09

Leveraging Study Data Reviewer's Guide (SDRG) in Building FDA's Confidence in Sponsor's Submitted Datasets

Xiangchen (Bob) Cui, Min Chen, and Letan (Cleo) Lin, Alkermes Inc., Waltham, MA

ABSTRACT

FDA issued **Study Data Technical Conformance Guide** [1] in October 2016, which stipulates "The SDRG should describe any special considerations or directions that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data." Hence SDRG not only supports regulatory review and analysis, but also establishes the **traceability** of tabulation datasets (SDTM), and source data (raw data). FDA reviewers consider traceability as an important component of a regulatory review [1]. Confidence in submitted datasets (SDTM and ADaM) can be established through traceability from the datasets, their define files and reviewer's guides (SDRG and ADRG). PhUSE released **SDRG Package v1.2** [2] on January 26, 2015, which provides a step-by-step template that helps sponsors to prepare SDRG.

This paper presents the readers how to build the traceability in SDRG, further build the FDA's confidence in sponsor's submitted datasets. The examples in this paper are from working experiences from FDA request, and NDA submission preparations of more than twenty recently-developed SDRG's from Phase I-III clinical study data.

INTRODUCTION

Per section 2.2 in **Study Data Technical Conformance Guide** [1], "The preparation of a Study Data Reviewer's Guide (SDRG) is recommended as an integral part of a standards-compliant study data submission. The SDRG should describe any special considerations or directions that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data."

Per section 8.3 Study Data Traceability in **Study Data Technical Conformance Guide** [1], "Based upon reviewer experience, establishing traceability is one of the most problematic issues associated with legacy study data converted to standardized data. If the reviewer is unable to trace study data from the data collection of subjects participating in a study to the analysis of the overall study data, then the regulatory review of a submission may be compromised. Traceability can be enhanced when studies are prospectively designed to collect data using a standardized CRF, e.g., CDASH. Traceability can be further enhanced when a flow diagram is submitted showing how data move from collection through preparation and submission to the Agency."

Sponsor can use the template developed by PhUSE released **SDRG Package v1.2** [2] on January 26, 2015 to prepare SDRG, which helps FDA reviewers to better understand SDTM datasets for meaningful analysis. The FDA's expectation of SDRG is that it should be established traceability permitting an understanding of the relationships between tabulation datasets (SDTM), and source data. Hence SDRG should be prepared in such way in order to underline the traceability, which can be achieved by aCRF, define.xml, in addition to SDRG.

This paper presents seven examples from more than twenty recently-developed SDRGs from Phase I-III clinical study data for FDA submission preparations, including the lessons learned from post submission FDA's request.

1. **An Example of Section 3.2 for Annotated CRFs in SDRG**
2. **An Example of Documenting Empty Variables in Submitted SDTM Datasets**
3. **An Example of Documenting Removing Duplicates of Study Baseline Records**
4. **An Example of Data Collected from Clinical Trial, But Not Collected Per eCRF**
5. **An Example of Handling Variables in EDC Raw Datasets, But Not Collected Per eCRF**

6. **An Example of Addressing Reviewer's Questions in Section 3.1 Overview: Additional Content of Interest**
7. **An Example of Handling Comments More Than 1000 Characters in Appendix II Computational Method**

These seven examples underline the achievement of **traceability** between tabulation datasets (SDTM), and source data. **This traceability** further helps to build **FDA's confidence** in sponsor's submitted datasets.

AN EXAMPLE OF SECTION 3.2 FOR ANNOTATED CRFS IN SDRG

Section 3.2 in SDRG template from **PhUSE released SDRG Package v1.2** [2] is for **Annotated CRFs**. **Study Data Technical Conformance Guide** [1] stipulates "When data are recorded on the CRF but are not submitted, the CRF should be annotated with the text '**NOT SUBMITTED**'. There should be an explanation in the SDRG stating why data have not been submitted."

There are certain fields in aCRF, which are annotated as "Not Submitted". Some of them are collected and built into EDC data, and some of them are not collected or built into EDC data. Table 1 shows an example of fields from AE CRF. The highlighted ones at the bottom are built to facilitate certain operational processes including data cleaning and dynamically creating additional forms in the electronic data capture system. The first field "Verbatim Term" was for site use only, and was not built in EDC dataset. These data were collected, but not built into EDC.

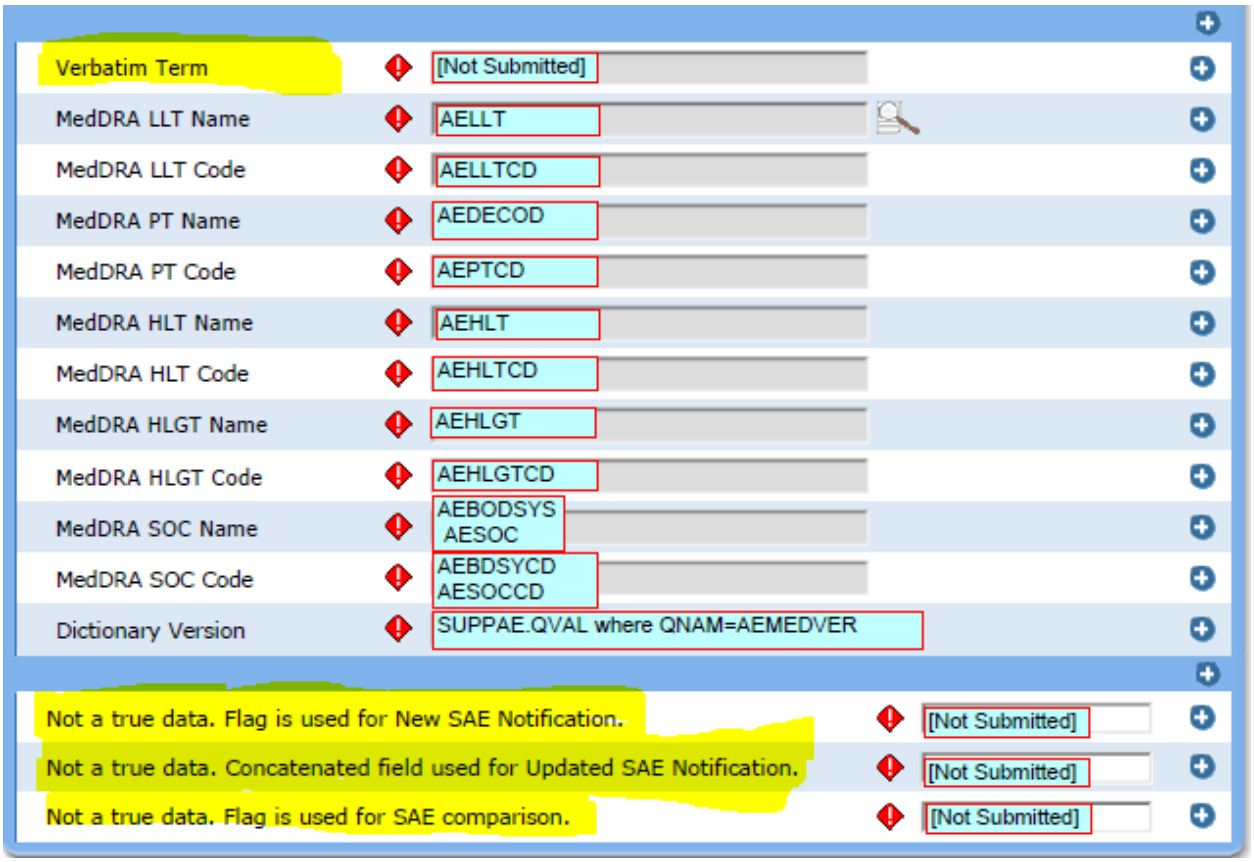


Table 1 Part of CRF for AE Domain

Table 2 shows an example of fields from DA CRF. The highlighted items at the bottom are redundant as their information was already captured by other variables. For example, field 'Number of Tablets Missed' was not submitted as the information was also collected by the field 'Total Number of Tablets Missed'.

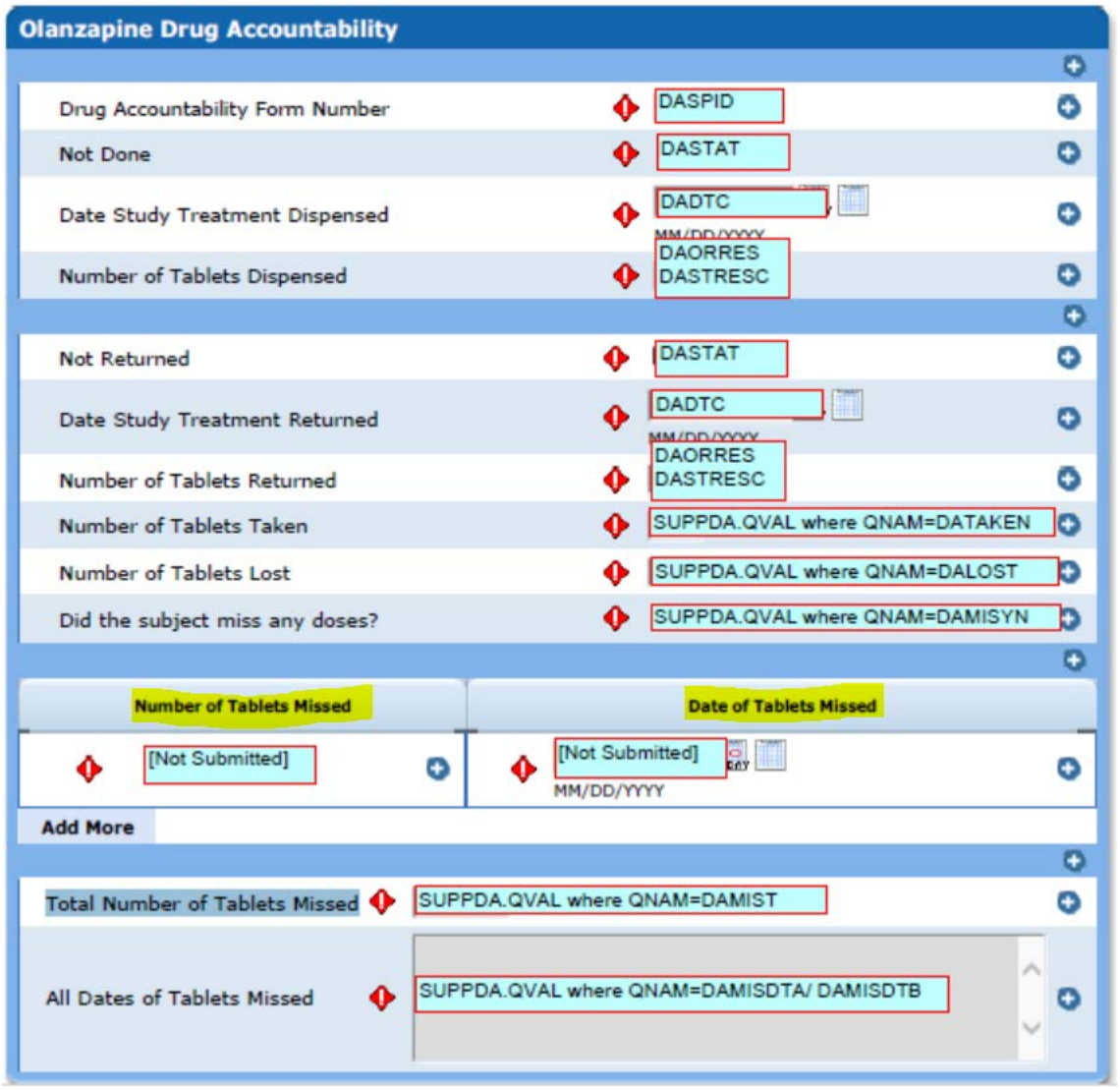


Table 2 DA Domain CRF

Seven categories were summarized for an aCRF field annotated as “Not Submitted” in Section 3.2 in our SDRG. We listed them in Table 3 to explain the reasons for “Not Submitted” for each field in aCRF pages. The example below is copied from one of SDRGs. This practice follows the FDA’s guideline “Description of all sponsor decisions related to data standard implementations”, and creates traceability by showing how data move from collection through preparation and submission to the Agency.

The fields defined in aCRF that are not tabulated in the SDTM datasets have been annotated as “Not Submitted”.

> ***For fields that have been excluded in EDC database but in CRFs, certain data elements are collected to facilitate certain operational processes including data cleaning and dynamically creating additional forms in the electronic data capture system.***
>
> ***All fields that have been annotated as “Not Submitted” meet the following criteria:***
>
> > ***1) Only used to trigger an event or record reported for a subject.***
> >
> > ***2) Not needed for SDTM or Analysis.***
> >
> > ***3) Not available in raw data, for site use only.***
> >
> > ***4) Redundancy. The information was already captured by other variables.***

5) Not mapped to SDTM because there were no results.

6) To facilitate certain operational processes including data cleaning and dynamically creating additional forms in the electronic data capture system.

7) Only used to differentiate admission & discharge elements, for deriving start date & end date.

All variables annotated as "Not Submitted" in Annotated CRFs are listed below, along with their reasons categorized above.

CRF Name	*Variable*	*Label*	*1)*	*2)*	*3)*	*4)*	*5)*	*6)*	*7)*
AE		*Verbatim Term*			*X*				
AE		*Not a true data. Flag is used for New SAE Notification*						*X*	
AE		*Not a true data. Concatenated field used for Updated SAE Notification*						*X*	
AE		*Not a true data. Flag is used for SAE comparison*						*X*	
AEYN	*AEYN*	*Were any Adverse Events experienced?*	*X*						
DA	*DAMISS*	*Number of Tablets Missed*				*X*			
DA	*DAMISDT*	*Date of Tablets Missed*				*X*			
PAT	*PATIEN*	*Subject Caption*		*X*					
PAT		*Subject Initials*			*X*				
SER	*SERDAT*	*Date of Collection*					*X*		
SER	*SERPERF*	*Was the serology sample collected?*	*X*						
SITE	*STATE*	*State*		*X*					
SITE	*STCFN*	*Investigator First Name*		*X*					
SITE	*STCITY*	*City*		*X*					
SITE	*STCLN*	*Investigator Last Name*		*X*					
SITE	*STNAME*	*Site Name*		*X*					

Table 3 An Example of Summary Table of Listing Fields on aCRF annotated as "Not Submitted".

AN EXAMPLE OF DOCUMENTING EMPTY VARIABLES IN SUBMITTED SDTM DATASETS

When submitting clinical study data in electronic format to the FDA, it is preferable to submit as few unnecessary variables as possible which have all missing values [3]. This kind of variable is called empty variable. If all empty variables by their domains within submitted SDTM were detected by our validation rules, they will be summarized along with the explanation in Section **4.3 Additional Conformance Details** in SDRG. It presents FDA reviewers both more clarity of and better understanding of these empty variables, and further a clearer understanding of the submitted datasets. The example below is from one of SDRGs and it serves as an example of how to document these empty variables.

The table below shows all variables with all values missing in the submitted SDTM datasets. They were either not collected or assigned with missing values following CDISC SDTMIG rules. They were reported during the study for decision making whether to be kept in SDTM submission.

Domain	Variable Order	Variable	Variable Label	Total Number of Observations	Core
DA	*14*	*VISITNUM*	*Visit Number*	*Xxxx*	*Exp*
	13	*DASTAT*	*Completion Status*	*Xxx*	*Perm*
DM	*11*	*DTHDTC*	*Date/Time of Death*	*Xx*	*Exp*
	12	*DTHFL*	*Subject Death Flag*	*Xx*	*Exp*
LB	*22*	*LBSPCCND*	*Specimen Condition*	*Xxxx*	*Perm*
	30	*LBTPT*	*Planned Time Point Name*	*Xxxx*	*Perm*
	31	*LBTPTNUM*	*Planned Time Point Number*	*Xxxx*	*Perm*
SUPPAE	*10*	*QEVAL*	*Evaluator*	*Xxx*	*Exp*
SUPPDM	*10*	*QEVAL*	*Evaluator*	*Xxx*	*Exp*
	4	*IDVAR*	*Identifying Variable*	*Xxx*	*Exp*
	5	*IDVARVAL*	*Identifying Variable Value*	*Xxx*	*Exp*
TA	*9*	*TATRANS*	*Transition Rule*	*Xx*	*Exp*
TS	*10*	*TSVCDVER*	*Version of the Reference Terminology*	*Xx*	*Exp*
TV	*6*	*ARMCD*	*Planned Arm Code*	*Xx*	*Exp*
	7	*ARM*	*Description of Planned Arm*	*Xx*	*Perm*

Table 4 An example of Documenting Empty Variables in Submitted SDTM Datasets

AN EXAMPLE OF DOCUMENTING REMOVING DUPLICATES OF STUDY BASELINE RECORDS

The duplicates of study baseline records are usually removed from raw data by the study data management before data base lock. One of our studies is an open-label long-term safety study. Some subjects who enter this study have completed the treatment period of antecedent studies. Baseline assessments will be conducted during the 4-week follow-up period of the antecedent study. Due to an EDC database build "error", there are some duplicates of baseline records. The study data management couldn't remove these duplicates due to the budget and resource. However these duplicates must be deleted in ADaM programming to support TFL programming, in addition to FDA submission [4]. The best solution is to remove these duplicates in SDTM programming. Documentation of the removals of duplicates is very critical for the traceability. The SDTM define.xml is a good place for this solution. We will provide more detailed description in SDRG under section 4.3 **Additional Conformance Details**.

The following is from SDRG for this study as an example.

For Rollover Subjects, Visit 2 will be the first visit of this study and there will not be a separate screening visit. EDC database was built to collect Visit 2 of this study only if Visit 2 occurred more than 10 days of the subject's end of treatment visit in the prior study. However, when the sites collected the baseline records, duplicate records were collected at Visit 2 as baselines for findings domains LB, EG, VS, and QS. We identified xxxx Cases of Duplicates for xx Subjects across these domains. Due to the budget and resources needed in cleaning the duplicate records from raw data, we removed the duplicate records in SDTM programming with the following process.

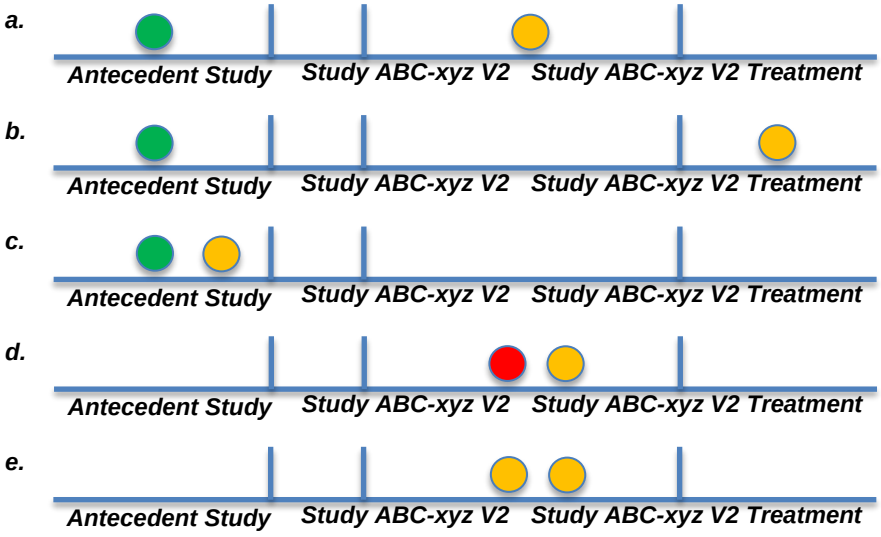
(1) Identification of the types of the duplicate baselines

There are 5 scenarios of duplicate baselines shown as below:

a. Duplicate baselines were collected pre-dose at the same visit and different dates, 1 rollover and 1 current study (ABC-xyz) V2 records

b. ***Duplicate baselines were collected at the same visit and different dates, 1 rollover pre-dose and 1 current study (ABC-xyz) V2 post-dose records.***
c. ***Duplicate baselines were collected at the same visit and different dates; 2 rollover pre-dose records.***
d. ***Duplicate baselines were collected at the same visit and same date with different results; 2 current study (ABC-xyz) V2 pre-dose records***
e. ***Duplicate baselines were collected at the same visit and same date with same results; 2 current study (ABC-xyz) V2 pre-dose records***

These scenarios can be shown in the figure below, in which two yellow circles indicate same results, red and yellow circles indicate different results, and green and yellow circles either same results or different results.



(2) The rules used to remove the duplicate baselines

Any duplicate baselines with missing values will be deleted from the database. If there are still non-missing duplicate baselines, the following rules were agreed across the clinical team to remove the duplicate records for each scenario from a-e:

a. ***If study ABC-xyz V2 visit pre-dose record is non-missing, delete the records from antecedent study.***
b. ***Use antecedent study pre-dose record as baseline.***
c. ***Choose the latest non-missing record***
d. ***Choose the latest non-missing record***
e. ***This scenario contains the true duplicate records. Keep one record only.***

AN EXAMPLE OF DATA COLLECTED FROM CLINICAL TRIAL, BUT NOT COLLECTED PER ECRF

Protocol Deviation Log is an external dataset for each study per our SOP. It documents major and/or minor occurrences of protocol deviations during a clinical trial. Each deviation will be reviewed by the clinical study team. Before each study database lock, certain items will be reconciled with clinical database. Certain categories will be reported in CSR. This log is the source data for creating DV (Protocol Deviation) domain, along with clinical data derived from the rules specified in Statistical Analysis Plan (SAP). For example, “Lack of Adherence with Study Medication” is based on the treatment adherence. This deviation is listed under Section 3.3 SDTM Subject Domains below. The following is from one of our SDRGs as an example.

DV – Protocol Deviations

A protocol deviation log was provided by Clinical Study Team and converted to an external datasets, to provide protocol deviation information in category "Did not Meet Inclusion/Exclusion". DV dataset was generated for Major deviations ONLY.

AN EXAMPLE OF HANDLING VARIABLES IN EDC RAW DATASETS, BUT NOT COLLECTED PER ECRF

Certain variables in each EDC raw dataset are not collected per eCRF from the clinical trial. They are created per EDC database programming. They are part of EDC raw datasets. We had a FDA's request to SDTM mapping information from one of our post FDA submissions. FDA reviewers would like to know which variables in EDC raw datasets were not submitted to the agency. Table 5 shows an example of these raw variables from raw datasets: AE and CGIS. These variables were not collected in the eCRF, but were created by EDC database programming.

Raw Dataset Name	*Raw Variable Name*	*Raw Variable Label*
AE	*VISITDT*	*Visit Date*
AE	*VISITN*	*Visit Name*
AE	*VISITNO*	*Visit No*
AE	*CRFINSNO*	*Form Instance*
AE	*CRFNAME*	*Form Name*
AE	*CRFSTAT*	*Form Status*
AE	*GRPINSNO*	*Group Instance*
AE	*SITEC*	*Site Caption*
AE	*STNAME*	*Site Name*
AE	*STNO*	*Site No*
AE	*TRCAP*	*Trial Caption*
AE	*TRNAME*	*Trial Name*
CGIS	*VISITDT*	*Visit Date*
CGIS	*CRFINSNO*	*Form Instance*
CGIS	*CRFNAME*	*Form Name*
CGIS	*CRFSTAT*	*Form Status*
CGIS	*GRPINSNO*	*Group Instance*
CGIS	*SITEC*	*Site Caption*
CGIS	*STNAME*	*Site Name*
CGIS	*STNO*	*Site No*
CGIS	*TRCAP*	*Trial Caption*
……	*……..*	*……*

Table 5 An Example of Variables in EDC Raw Dataset, But Not Collected Per eCRF from the Clinical Trial

Since SDRG template does not provide the instruction about this kind of information, we add a paragraph, shown below, to describe it after **Section 4.3 Additional Conformance Details** of SDRG.

Certain variables in each EDC raw dataset are not collected per eCRF from the clinical trial. They are created per EDC database programming. They are part of EDC raw datasets, and dropped from SDTM programming. Hence they are not submitted in the submission. However the information is FDA submission-ready.

AN EXAMPLE OF ADDRESSING REVIEWER'S QUESTIONS IN SECTION 3.1 OVERVIEW: ADDITIONAL CONTENT OF INTEREST

Study Data Reviewer's Guide Completion Guidelines [2] recommends that additional contents of interest can be summarized under Section 3.1 and provides the suggestion of additional content of interest, listed below.

Additional content may include, but is not limited to the following:

- ***Description of any study history or timing relevant to the submitted data (e.g. interim data cutoff, data differences due to protocol amendments, etc.).***
- ***Location of key safety, efficacy, or other data of special interest.***
- ***Explanation of the mapping of death information in the subject level datasets. Explain any differences in the occurrences (frequencies) of death across the datasets.***
- ***Document the location of adjudication data and the method used to differentiate these data from data collected at the investigational site.***
- ***Document any notable subjects of interest within the context of the study.***
- ***Description of the reference start date including any differences in the definition across subjects and description of the calculation of study days. These should align with the definitions in define.xml.***
- ***If you are documenting an extension study, include description(s) of any data that have been copied from or are located in another study in the submission.***

A review tool at FDA, JumpStart, is run by data experts in the Office of Computational Science of FDA to run initial analyses and flag areas for follow-up by the review teams.

FDA "JumpStart" uses SDTM tabulation data as the basis for exploratory analysis [4]:

The Data Fitness session analyzes all datasets submitted for a trial

The Safety Analysis session focuses primarily on DM, DS, EX, AE, LB, and VS

JumpStarting the Regulatory Review Process: The Review Perspective [4] summarizes FDA reviewer's questions for safety review by JumpStart. One of the questions under "Safety Population" is that "Are there any subjects who were randomized but not treated?" To address this question, we should list all these subject numbers under Section 3.1 Overview: Additional Content of Interest. Definitely we can provide other information to address reviewer's common questions summarized in [4]. We provide an example here to create awareness for providing more information in SDRG to address FDA reviewer's questions in addition to ones suggested in **SDRG Package v1.2** [2].

An example of this sort of listing is shown below in one of the SDRGs:

The subjects who were randomized, but not received study drugs were listed below.

123456-001, ….., etc.

EXAMPLE OF HANDLING COMMENTS MORE THAN 1000 CHARACTERS IN APPENDIX II COMPUTATIONAL METHOD

One of the rules in Pinnacle 21 points out that "FDA's Clinical Trial Repository (CTR) software currently has a maximum length of 1000 characters for data attributes in Define.xml". If the computational methods exceed 1000 characters, we will provide the detailed computational method in SDRG at Appendix II with a link to SDRG added in Derivation/Comment Column of define.xml.

An example can be shown as below in one of the SDRGs:

Appendix II: Computational Method

1. Derivation Description for AE.EPOCH:
Only Populated when AESTDTC is a complete date. For Group 1 Subjects (Visit 1a date is missing):if both DM.RFSTDTC and DM.RFENDTC are missing, EPOCH='SCREENING'; Else if DM.RFSTDTC is missing and DM.RFENDTC is not missing, and AE Start Date > date part of RFENDTC, then EPOCH='FOLLOW-UP'; Else if DM.RFSTDTC is missing and DM.RFENDTC is not missing then EPOCH=''; Else if DM.RFSTDTC is not missing and AE Start Date < date part of RFSTDTC, then EPOCH='SCREENING'; Else if DM.RFSTDTC is not missing and AE Start Date >= date part of RFSTDTC and (AE Start Date <= date part of RFENDTC or RFENDTC is missing), then EPOCH='BLINDED TREATMENT'; Else if DM.RFSTDTC is not missing and AE Start Date>date part of RFENDTC, then EPOCH='FOLLOW-UP'; For Group 2 Subjects (Visit 1a date is not missing):If AE Start Date < Visit 1a date then EPOCH = 'SCREENING'; Else if RFSTDTC is not missing and AE Start Date < date part of RFSTDTC then EPOCH='RUN-IN'; Else if RFSTDTC is not missing and date part of RFSTDTC<=AE Start Date and (AE Start Date<=date part of RFENDTC or RFENDTC is missing) then EPOCH='BLINDED TREATMENT'; Else if RFSTDTC is not missing and AE Start Date>date part of RFENDTC then EPOCH='FOLLOW-UP'; Else if both RFSTDTC and RFENDTC are missing then EPOCH='RUN-IN'; Else if RFSTDTC is missing and RFENDTC is not missing, and AE Start Date > date part of RFENDTC, then EPOCH='FOLLOW-UP'; Else if DM.RFSTDTC is missing and DM.RFENDTC is not missing then EPOCH=''

CONCLUSION

This paper presents seven examples from more than twenty recently-developed SDRGs from Phase I-III clinical study data for FDA submission preparations including the lessons learned from post submission FDA's request. To share our experience in preparing SDRGs is to present the readers the idea: how to achieve the **traceability** of between tabulation datasets (SDTM), and source data and the **traceability** is further helping to build **FDA's confidence** in sponsor's submitted datasets.

REFERENCES

[1] U.S. Department of Health and Human Services, Food and Drug Administration, Study Data Technical Conformance Guide: Technical Specifications Document. November 2016. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf.

[2] Pharmaceutical Users Software Exchange, Study Data Reviewer's Guide Final Work Packages: SDRG Package v1.2 2015-01-26. Available at http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide.

[3] Min Chen; Xiangchen (Bob) Cui. Automatic Detection and Identification of Variables with All Missing Values in SDTM/ADaM Datasets for FDA Submission. Proceedings of the Pharmaceutical SAS® Users Group Conference, PharmaSUG 2012. Available at https://pdfs.semanticscholar.org/fa48/2715ffb8f21d081b8e225de89f172c513751.pdf.

[4] Rosario, 2015, Lilliam, Ph.D. Director, Office of Computational Science, Office of Translational Science, CDER, "JumpStarting the Regulatory Review Process: The Review Perspective", Keynote Address at PharmaSUG 2015 Conference, Orlando, FL, Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf

ACKNOWLEDGMENTS

Appreciation goes to Rennie Castelino for his valuable review and comments.

CONTACT INFORMATION

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

Name: Xiangchen (Bob) Cui, Ph.D.
Enterprise: Alkermes, Inc.

Address: 852 Winter Street
City, State ZIP: Waltham, MA 02451

Work Phone: 781-609-6038
Fax: 781-609-5855
E-mail:xiangchen.cui@alkermes.com

Name: Min Chen, Ph.D.
Enterprise: Alkermes, Inc.
Address: 852 Winter Street
City, State ZIP: Waltham, MA 02451
Work Phone: 781-609-6047
Fax: 781-609-5855
E-mail: min.chen@alkermes.com

Name: Letan (Cleo) Lin, M.S.
Enterprise: Alkermes, Inc.
Address: 852 Winter Street
City, State ZIP: Waltham, MA 02451
Work Phone: 781-609-6380
Fax: 781-609-5855
E-mail: letan.lin@alkermes.com