PharmaSUG 2020 – Paper SS-197

Preparing a Successful BIMO Data Package

Elizabeth Li, Carl Chesbrough, and Inka Leprince, PharmaStat, LLC

ABSTRACT

In order to shorten the time for regulatory review of a new drug application (NDA) or biologic license application (BLA), more and more biotech and pharmaceutical companies prepare their Bioresearch Monitoring (BIMO) program packages as part of their initial submissions. In this paper, we walk the reader through a process of producing BIMO information, particularly the subject-level data line listings by clinical site (by-site listings) and the summary-level clinical site (CLINSITE) dataset. This paper concludes with methods of preparing electronic Common Technical Document (eCTD) documentation, such as data definition (define.xml) and reviewer's guide, to support the CLINSITE dataset. In addition, we discuss challenges as we share our experience in planning, producing, and quality control (QC) for a successful BIMO package.

INTRODUCTION

ABOUT BIORESEARCH MONITORING PROGRAM (BIMO)

According to the FDA draft guidance, *Standardized Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions*[1] and *Bioresearch Monitoring Technical Conformance Guide*[2], BIMO information is used for FDA planning of BIMO inspections in electronic form for submission of NDAs, BLAs, and supplemental applications. The draft guidance[1] states:

> "(Center for Drug Evaluation and Research) CDER's Office of Scientific Investigations (OSI) … has specific responsibility for verifying the integrity of data submitted to CDER in support of applications and supplements, and for determining whether clinical trials are conducted in compliance with applicable FDA regulations and statutory requirements..." (Page 4, Lines 138 to 142)

To facilitate the FDA's evaluation, the BIMO information consists of three parts:

1) Clinical Study-Level Information
2) Subject-Level Data Line Listings by Clinical Site (By-Site Listings)
3) Summary-Level Clinical Site (CLINSITE) Dataset

Of the aforementioned three parts of BIMO information, Part 1, clinical study-level information (see details in the draft guidance[1]), is usually collected by a sponsor's clinical operations (CO) team and prepared by its regulatory affairs (RA) team. Not all the clinical study-level information is stored in a clinical database. Hence, not all the information is in the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) or the Analysis Data Model (ADaM) datasets. Consequently, clinical study-level information cannot be easily generated from SDTM or ADaM datasets.

Therefore, this paper primarily focuses on the preparation of the second and third parts of the BIMO data package: by-site listings and the CLINSITE dataset, using SDTM or ADaM datasets.

GENERAL APPROACH FOR PREPARING BIMO DATA PACKAGE

Because of the importance of clinical data that is used for timely planning and conducting inspections during the FDA's review of NDAs/BLAs, every effort should be made to produce an accurate and clear BIMO data package. To this end, our approach was as follows:

- Draft a plan for the BIMO data package,
- Share the draft plan with FDA at a pre-NDA meeting or a similar form of communication,
- Update and finalize the BIMO data plan with feedback from the FDA reviewers,
- Execute the BIMO data plan,
- Create eCTD documentation for the CLINSITE dataset,
- QC the BIMO data package.

This paper recounts our challenges and successes in executing our finalized data plan.

PREPARING FOR SUBJECT-LEVEL DATA LINE LISTINGS BY CLINICAL SITE

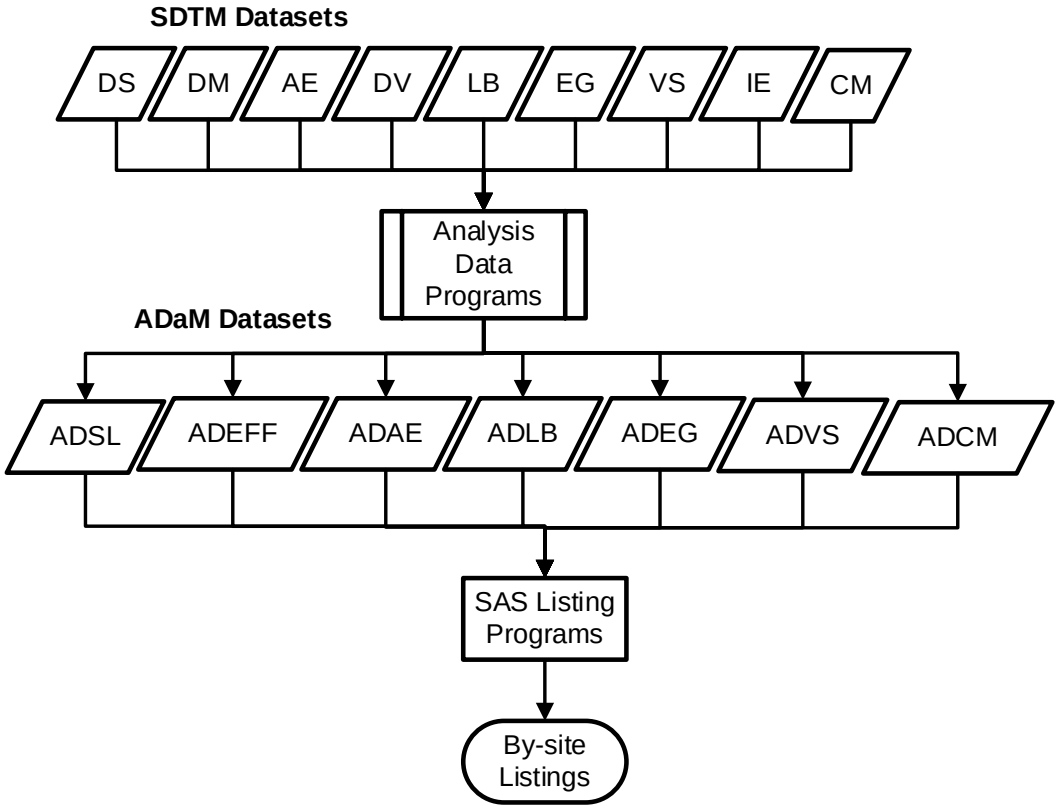
LEVERAGING EXISTING SDTM OR ADAM DATASETS

Based on the draft guidance[1], the subject-level data line listings by-site are used to verify key study data (e.g. safety and efficacy) during the BIMO inspections. By-site listings are comprised of the following key study data from "major (i.e., pivotal) studies":

1) Consented Subjects
2) Treatment Assignment
3) Discontinuation
4) Study Population
5) Inclusion and Exclusion Criteria
6) Adverse Events
7) Important Protocol Deviations
8) Efficacy Endpoints
9) Concomitant Medication
10) Safety Monitoring

The data flow diagram in Figure 1 shows an example of the data sources needed for generating the by-site listings.

Figure 1 Sample Data Source for By-site Listings



LEVERAGING EXISTING LISTING PROGRAMS

During the preparation of a regulatory submission, time is of the essence. In order to save time and resources, we used SAS® programs that produced listings for the clinical study reports (CSRs) to generate the by-site listings with minor modifications to the SAS® code. Table 1 below shows an example of CSR listings matched with required contents of BIMO by-site listings.

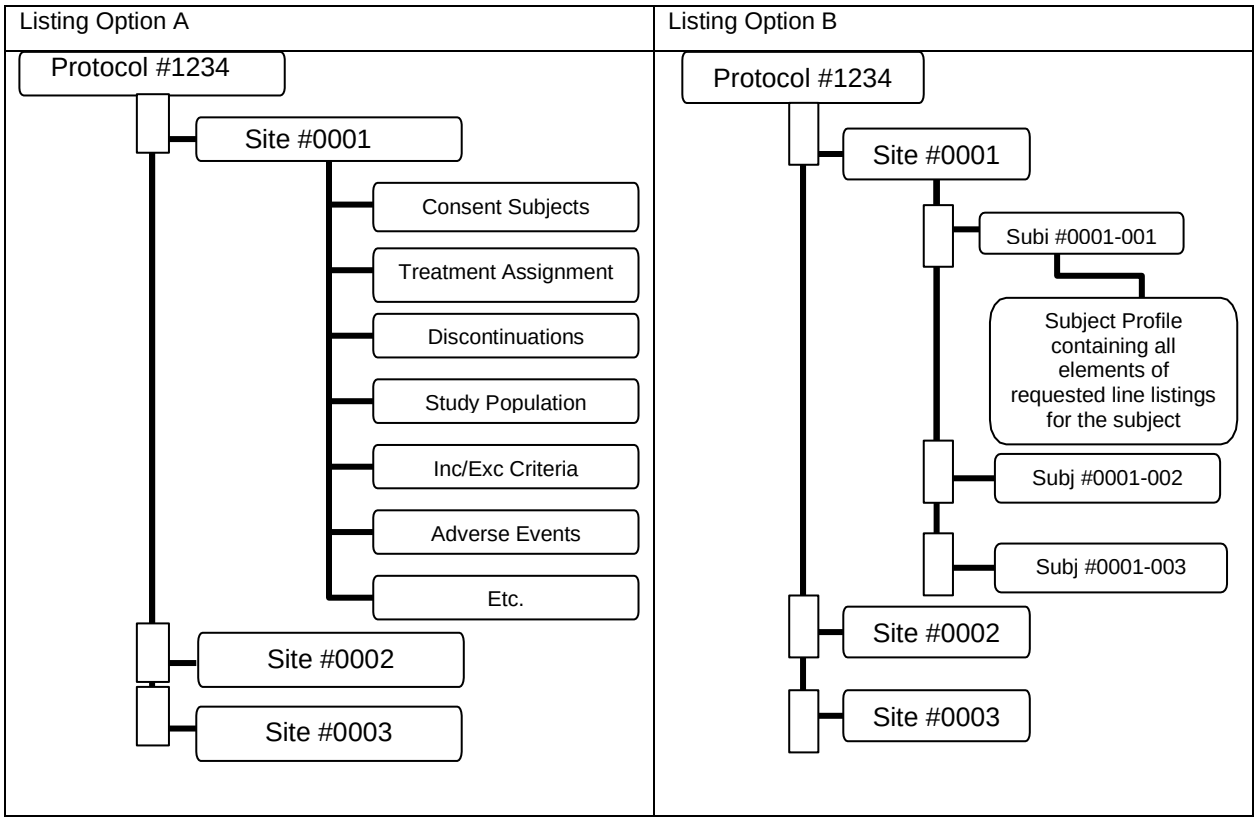
Table 1 Contents of Subject-Level Data Line Listings, by Clinical Site

Required Content	*CSR Listing Number and Title* Proposed By-Site Listing Number and Title
1. Consented Subjects 2. Treatment Assignment 3. Discontinuation	*CSR Listing 16.2.1 Subject Disposition* Listing A Subject Enrollment, Treatment, and Disposition (All Screened Subjects at Site ##, Investigator = Smith)
4. Study Population 5. Inclusion and Exclusion Criteria	*CSR Listing 16.2.3.1 Subjects Excluded from Analysis Sets* Listing B Study Population and Exclusion Reasons (All Randomized Subjects at Site ##, Investigator = Smith)
6. Adverse Events	*CSR Listing 16.2.7.1 Adverse Events* Listing C Adverse Events (All Treated Subjects at Site ##, Investigator = Smith)
7. Important Protocol Deviations	*CSR Listing 16.2.2 Important Protocol Deviations* Listing D Important Protocol Deviations (All Randomized Subjects at Site ##, Investigator = Smith)
8. Efficacy Endpoints	*CSR Listing 16.2.6.1 Efficacy Data* Listing E Efficacy Endpoints (All Randomized Subjects at Site ##, Investigator = Smith)
9. Concomitant Medication	*CSR Listing 16.2.9.3 Concomitant Medications* Listing F Concomitant Medications (All Treated Subjects at Site ##, Investigator = Smith)
10. Safety Monitoring	*CSR Listing 16.2.8.1 Hematology: Complete Blood Count* Listing G1 Central Laboratory Test Results - Hematology: Complete Blood Count (All Treated Subjects at Site ##, Investigator = Smith) *CSR Listing 16.2.8.2 Chemistry: Electrolytes* Listing G2 Central Laboratory Test Results - Serum Chemistry: Electrolytes (All Treated Subjects at Site ##, Investigator = Smith) *CSR Listing 16.2.8.2 Chemistry: Renal Function* Listing G3 Central Laboratory Test Results at Serum Chemistry: Renal Function (All Treated Subjects at Site ##, Investigator = Smith) *CSR Listing 16.2.9.1 Vital Signs* Listing G4 Vital Signs and Body Weight (All Treated Subjects at Site ##, Investigator = Smith) *CSR Listing 16.2.9.2 Findings from Electrocardiogram* Listing G5 12-Lead Electrocardiogram (All Treated Subjects at Site ##, Investigator = Smith)

HAVING A DETAILED PLAN

A draft plan or key parts of the plan should be submitted for input from FDA reviewers. A similar table to the above sample Table 1 may be included in the plan to help the reviewers to confirm the proposed BIMO by-site listings containing the required study data contents. A detailed plan should also include mock-shells (layouts) that serve as a visual rendering with specifications for programmer analysts to generate the BIMO by-site listings. Equipped with the data sources and adapted code to create the by-site listings, a decision regarding format needs to be made. There are two possible formats for the listings to be provided to the Agency, see Figure 2 (duplicated from Appendix 2 of the Technical Conformance Guide[2]) below for option details. One of these options should be specified in the plan for the BIMO by-site listings.

Figure 2 Options for Subject-Level Data Line Listings, by Clinical Site



CHALLENGES

Option A or B

Since there are two options for the BIMO by-site listings, a sponsor should evaluate pros and cons for each then decide which options to use. If CSR listing programs can be easily and quickly modified, then Option A is a good choice, since the layouts of the listings are similar. If patient profiles are already generated and contain the required contents, then Option B may be a good choice, since the data are grouped by subject within each site. In our recent NDA experience, we chose Option A, since the CSR listing programs could be modified to generate the BIMO by-site listings. Within each site, the listings were ordered alphanumerically as shown in Table 1.

Multiple Studies

There is usually more than one "major (i.e., pivotal) study" in a submission. In our recent NDA experience, the data from two studies (i.e., the pivotal study and its extension study) were required by the FDA in the BIMO by-site listings. As in this case, where the majority of the subjects participated in both studies, special care should be given for the following:

- **Subject identifier**: The unique subject ID should be the same in both studies, although two clinical databases may have been used.
- **Treatment assignment**: Indicate any treatment group change from one study to the other, if applicable.
- **Study identifier**: Align study IDs with the data records that were collected during corresponding studies.
- **Sorting order**: For a given listing at each site, the listing may be ordered by subject ID, treatment group, study ID, visit date (or time point).

Screen Failure Information

Not all SDTM DM domains contain the screen failure subjects. In our recent NDA experience, only randomized subjects were included in the SDTM DM domain. In order to include all screened subjects in the Listing A Subject Enrollment, Treatment, and Disposition, we used the source data from a clinical database, along with existing ADaM datasets.

GENERATING THE BIMO BY-SITE LISTINGS

Here are the steps for generating BIMO by-site listings

1) Create base SAS® macros

Table 2 Subject-Level Data Line Listings Macros

Proposed By-Site Listing Number and Title	*CSR Listing Number and Title*	Base SAS® macro
Listing A Subject Enrollment, Treatment, and Disposition (All Screened Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.1 Subject Disposition*	l_enroll.sas
Listing B Study Population and Exclusion Reasons (All Randomized Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.3.1 Subjects Excluded from Analysis Sets*	l_excl.sas
Listing C Adverse Events (All Treated Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.7.1 Adverse Events*	l_ae.sas
Listing D Important Protocol Deviations (All Randomized Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.2 Important Protocol Deviations*	l_pd.sas
Listing E Efficacy Endpoints (All Randomized Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.6.1 Efficacy Data*	l_eff.sas
Listing F Concomitant Medications (All Treated Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.9.3 Concomitant Medications*	l_cm.sas
Listing G1 Central Laboratory Test Results - Hematology: Complete Blood Count (All Treated Subjects at Site ##, Investigator = Smith) Listing G2 Central Laboratory Test Results - Serum Chemistry: Electrolytes (All Treated Subjects at Site ##, Investigator = Smith) Listing G3 Central Laboratory Test Results at Serum Chemistry: Renal Function (All Treated Subjects at Site ##, Investigator = Smith) Listing G4 Vital Signs and Body Weight (All Treated Subjects at Site ##, Investigator = Smith) Listing G5 12-Lead Electrocardiogram (All Treated Subjects at Site ##, Investigator = Smith)	*CSR Listing 16.2.8.1 Hematology: Complete Blood Count* *CSR Listing 16.2.8.2 Chemistry: Electrolytes* *CSR Listing 16.2.8.2 Chemistry: Renal Function* *CSR Listing 16.2.9.1 Vital Signs* *CSR Listing 16.2.9.2 Findings from Electrocardiogram*	L_lab.sas, l_vs.sas, l_eg.sas

2) Generate individual PDF files of listings for each site using the above 9 SAS® macros in order of Site ID and listing number. Here is a sample SAS® code for Listing A.

```
%macro dolisting(sitenum=01);
ods tagsets.rtf file = "..\outputs\Site &sitenum Listing A Enroll.rtf"
options(continue_tag="no" order_repeat="yes") style=tempstyle;
ods tagsets.rtf anchor = "Listing_A" ;

proc report data = tlgdata.l_enroll&order  missing split = '~' spacing=1 headline spanrows ;
    column (subjid_c trt01p study_c infcons_c scrnfl_c sfrsn_c tcomp_c dtreas_c);
    by siteid_c ;
    where siteid = "&sitenum" ;
    define subjid_c     / order order = internal 'Subject~ID*'  center style=[width=0.55 in] ;
```

```
    define trt01p       / order order = internal 'Treatment'    center style=[width=0.65 in] ;
    define study_c      / display 'Study~ID'                    center style=[width=0.35 in] ;
    define infcons_c    / display 'Date of~Informed~Consent'    center style=[width=0.70 in] ;
    define scrnfl_c     / display 'Screen~Failure'              center style=[width=0.50 in] ;
    define sfrsn_c      / display 'Reason for~Screen Failure'   left   style=[width=1.30 in] ;
    define tcomp_c      / display 'Completed Study~ Date [Day]' center style=[width=1.65 in] ;
    define dtreas_c     / display 'Reason for~Discontinuation'  left   style=[width=1.00 in] ;
run ;

ods tagsets.rtf close;
%mend ;

%local ix next_name;
%do ix=1 %to %sysfunc(countw(&site_list));
   %let next_name = %scan(&site_list, &ix);
   %dolisting (sitenum = &next_name) ;
%end;
```

Here is a sample output of Listing A

BIMO for Studies 0001 and 0002 Page x of y

Listing A
Subject Enrollment, Treatment, and Disposition
(All Screened Subjects at Site 101, Investigator = Smith)

Subject ID*	Treatment	Study ID	Date of Informed Consent	Screen Failure	Reason for Screen Failure	Completed Study/ Date [Day]	Reason for Discontinuation
101-001	Active	0001	2018-11-22	No		Yes / 2019-03-01 [86]	
		0002	2019-03-02	No		Yes / 2019-12-06 [366]	
101-002	Placebo	0001	2018-11-29	No		Yes / 2019-03-06 [84]	
		0002	2019-03-07	No		Yes / 2019-12-12 [365]	
101-003*			2018-11-30	Yes	INC 3 not met		
101-004*			2018-12-12	Yes	INC 4 not met		
101-005*			2019-01-10	Yes	EXC 5 not met		
101-006	Placebo	0001	2019-01-15	No		Yes / 2019-04-23 [84]	
		0002	2019-04-24	No		Yes / 2020-02-04 [371]	

EXC = exclusion criterion; INC = inclusion criterion. Day = date - first dose date + 1, if on or after the first study drug dosing in Study 0001; Day = date - first dose date, otherwise.
* Subjects who are screen failures or did not enroll in Study 0001 are indicated. Their dates of completed or discontinued study occurred in Study 0001.

3) All individual PDF files are compiled into a single PDF of BIMO by-site listings using Adobe Acrobat®.
4) Bookmarks are added to the single PDF of BIMO by-site listings.

QC THE BIMO BY-SITE LISTING

In order to ensure quality and accuracy of the listings, companion SAS® datasets were generated prior to producing the individual PDF files of listings for each site. These companion SAS® datasets or analysis results datasets were used for QC against corresponding CSR listings. In addition, each by-site listing was compared against its corresponding CSR listing. Furthermore, selected subjects who had special events, such as met any exclusion criteria, died, or had SAEs, were cross checked against the corresponding study report to validate that the subjects and number of events were the same. These methods checked and cross-checked the BIMO listing records to ensure their accuracy.

PREPARING FOR SUMMARY-LEVEL CLINICAL SITE DATASET

HAVING A DETAILED PLAN

Based on the draft guidance[1] and Technical Conformance Guide[2], a single summary-level clinical site dataset (clinsite.xpt) should contain supporting safety and efficacy information for all major (i.e. pivotal) studies. Furthermore, the information should be summarized by study, site, and treatment arm (where applicable). When a pivotal study and its extension study are the major studies in a submission, it is a good idea to treat the extension study as a separate study. This permits reviewers to have a clearer picture of site summary level data during different phases of the study. In Appendix 3 of the Technical Conformance Guide[2], a total of 39 variables were specified. They can be classified in categories as shown in Table 3 below.

Table 3 Variable Categories in Summary-Level Clinical Site Dataset

Category	Variable Name	Data Source
a. Study Specific Information	STUDYTL, SPONCNT, SPONNAME, IND, UNDERIND, NDA, BLA, SUPPNUM	Protocols / RA
b. Site, Treatment and Analysis Population	SITEID, ARM, SAFPOP	ADaM ADSL
c. Screened Subjects	SCREEN	DM/DS or Source Clinical Database
d. Disposition	DISCSTUD, DISCTRT	ADaM ADSL
e. Endpoint Description	ENDPOINT, ENDPTYPE	SAP
f. Efficacy Variables	TRTEFFR, TRTEFFS, SITEEFFE, SITEEFFS, CENSOR	ADaM ADEFF
g. Safety Variables	NSAE, SAE, DEATH	ADaM ADAE
h. Protocol Violation	PROTVIOL	ADaM ADDV
i. Site Specific Information	FINLDISC, LASTNAME, FRSTNAME, MINITIAL, PHONE, FAX, EMAIL, COUNTRY, STATE, CITY, POSTAL, STREET, STREET1	Sites / CO

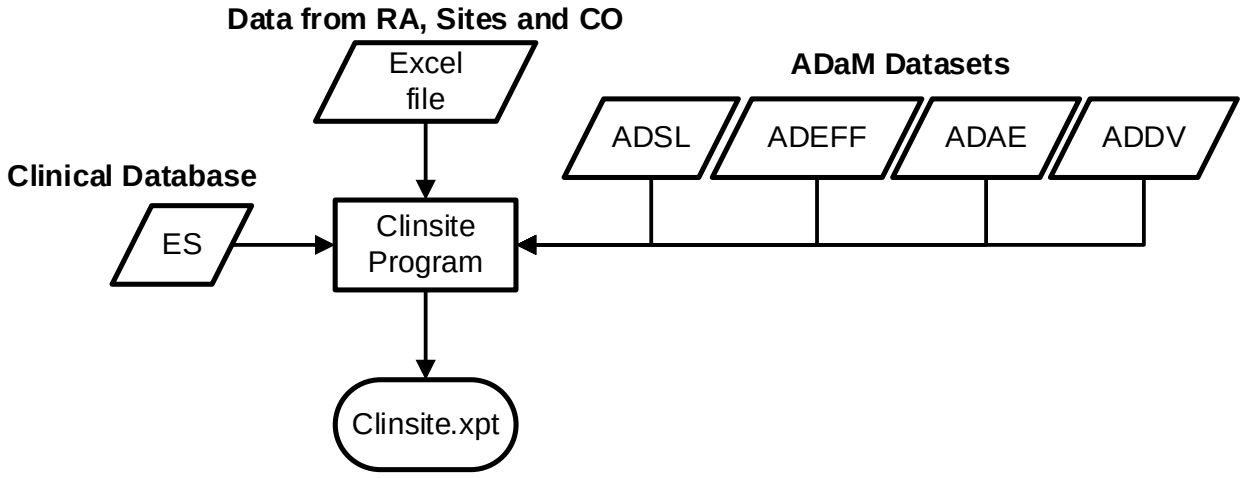
CO = clinical operations; RA = regulatory affairs; SAP = statistical analysis plan

If an extension study is primarily for safety, the efficacy related variables (variable index numbers 18 to 22 in Appendix 3 of the Technical Conformance Guide[2]) will be set to missing for that study in the CLINSITE dataset. FDA reviewer's feedback (e.g., separate pivotal study and its extension, fill values in efficacy variables for the pivotal study, etc.) was valuable for us to finalize the plan and generate the CLINSITE dataset.

LEVERAGING EXISTING ADAM DATASETS

We analyzed the 39 variables that are specified in the Technical Conformance Guide[2], to determine each variable's source data (see Table 3), which are from study protocols, regulatory affairs (RA), study site contact information, clinical operations, SAP, ADaM datasets, and clinical database source data. The data flow diagram in Figure 3 shows an example of the data sources for creating the CLINSITE dataset.

Figure 3 Sample Data Source for Summary-Level Clinical Site Dataset



CO = clinical operations; RA = regulatory affairs.

CHALLENGES

Analysis Population

In large multi-center clinical trials, some sites only screened subjects, but did not enroll or randomize any subjects. In the CLINSITE dataset, we included sites that enrolled or randomized at least one subject. In order to properly derive the efficacy variables by study, site, and treatment arm, sometimes an efficacy (or evaluable) population should be added to the CLINSITE dataset, since it is different from the safety population (SAFPOP). Table 4 below shows an example of adding an analysis population variable to the CLINSITE dataset.

Table 4 Adding a Population Variable to CLINSITE Dataset

Variable	Label	Reason
EVALPOP	Subjects with Data at Month x	This variable represents the number of subjects with data for the primary efficacy assessment at Month x in each treatment arm in Study 0001. Not all subjects in the safety population (SAFPOP) had efficacy data at Month x.

Inconsistency in the FDA Technical Conformance Guide[2]

As we followed the Technical Conformance Guide[2], we found some discrepancies in the document. Here is an excerpt from Appendix 3 of the Technical Conformance Guide[2].

Variable Index	Variable Name	Variable Label	Type	Controlled Terms or Format	Notes or Description
18	TRTEFFR	Treatment Efficacy Result	Num	Floating Point	Summary statistic for each primary efficacy endpoint by treatment arm at a given site.
19	TRTEFFS	Treatment Efficacy Result Standard Deviation	Num	Floating Point	Standard deviation of the efficacy result (TRTEFFR) for each primary efficacy endpoint by treatment arm at a given site. If N=1, set to “0”.
20	SITEEFFE	Site-Specific Treatment Effect	Num	Floating Point	Site-specific treatment effect reported using the same representation as reported for the primary efficacy
21	SITEEFFS	Site-Specific Treatment Effect Standard Deviation	Num	Floating Point	Standard deviation of the site-specific treatment effect (SITEEFFE). If N=1, set to “0”.

Here is an excerpt from Appendix 4 of the Technical Conformance Guide[2].

STUDYID	SITEID	ARM	SAFPOP	DISCSTUD	ENDPOINT	ENDTYPE	TRTEFFR	TRTEFFS	SITEEFFE	SITEEFFS
ABC-123	001	Active	26	3	Percent Responders	Binary	0.48	0.0980	0.34	0.1405
ABC-123	001	Placebo	25	4	Percent Responders	Binary	0.14	0.0694	0.34	0.1405
ABC-123	002	Active	23	2	Percent Responders	Binary	0.48	0.1042	0.33	0.1427
ABC-123	002	Placebo	25	4	Percent Responders	Binary	0.14	0.0694	0.33	0.1427
ABC-123	003	Active	27	3	Percent Responders	Binary	0.54	0.0959	0.35	0.1448
ABC-123	003	Placebo	26	5	Percent Responders	Binary	0.19	0.0769	0.35	0.1448

By comparing the variable label and description, Variables TRTEFFS and SITEEFFS were specified as standard deviation in the Appendix 3. However, the values in the variables, shown in Appendix 4, appear to be standard errors. We interpreted that the variable TRTEFFS is "Treatment Efficacy Asymptotic Standard Error" and variable SITEEFFS is "Site-Specific Treatment Effect Asymptotic Standard Error".

Efficacy Variables

Due to small sample sizes at some clinical sites, the variable SITEEFFS (Site-Specific Treatment Effect Asymptotic Standard Error) may not be reasonably estimated. Exact confidence limits rely on exact distributions and do not rely on an asymptotic standard error. Providing these confidence limits for the proportion of responders may add value. These variables can provide information for traceability of efficacy variable derivations and/or supplemental efficacy information. Table 5 below shows examples of efficacy variables that can be added to the CLINSITE dataset.

Table 5 Adding Efficacy Variables to CLINSITE Dataset

Variable	Label	Reason to Add to CLINSITE
NRESP	Responders at Month x	This variable represents the number of responders (defined as XXXXXXXX) in each treatment arm in Study 0001. Using both EVALPOP and NRESP, the variable Treatment Efficacy Result (TRTEFFR) can be properly derived.
SITEELCL	Site-Specific Treat Effect 95% Exact LCL	The exact 95% limits are provided for additional information for the site-specific treatment effect.
SITEEUCL	Site-Specific Treat Effect 95% Exact UCL	

LCL = lower confidence limit; Treat = treatment; UCL = upper confidence limit.

Consistency Between the By-Site Listing and CLINSITE Dataset

In the Technical Conformance Guide[2] about the by-site listings, the important protocol deviations, as reported in the NDA or BLA, are to be listed. However, in Appendix 3 of the Technical Conformance Guide[2] regarding the CLINSITE dataset, the description (see excerpt below) of the protocol violations calls for all types of violations (i.e. not limited to only significant deviations).

Variable Index	Variable Name	Variable Label	Type	Controlled Terms or Format	Notes or Description
26	PROTVIOL	Number of Protocol Violations	Num	Integer	Total number of protocol violations at a given site by treatment arm as defined in the clinical study report. A protocol violation is an unplanned excursion from the protocol that is not implemented or intended as a systematic change. This value should include multiple violations per subject and all violation types (i.e., not limited to only significant deviations).

Our SDTM data only included important protocol deviations. These protocol deviations are included in the CSR listings and the by-site listings. To be consistent with the CSRs and by-site listings, the CLINSITE dataset only includes these same important protocol deviations. For the purpose of full disclosure, this summary/reporting is documented in the CLINSITE dataset reviewer's guide.

Variable Labelling

In the Appendix 3 of the Technical Conformance Guide[2] about the CLINSITE dataset, variable attributes, such as variable name and variable label are provided. Three variables have labels more than 40 characters in length (listed in Table 6 below). Due to the limitation of 40-character label length in SAS® transport file (*.xpt), these variable labels have been modified.

Table 6 Adding Efficacy Variables to CLINSITE Dataset

Variable	Label in Appendix 3 of BIMO Technical Conformance Guide	Label Used in CLINSITE Dataset
DISCTRT	Number of Subject Discons from Study Treatment	No of Subjects Disc from Study Treatment
TRTEFFS	Treatment Efficacy Result Standard Deviation	Treatment Efficacy Standard Error
SITEEFFS	Site-Specific Treatment Effect Standard Deviation	Site-Specific Treat Eff Standard Error

Disc = discontinued; Discons = discontinued; Eff = effect; No = number; Treat = treatment.

GENERATING THE CLINSITE DATASET AND ECTD DOCUMENTATION

Here are the steps for generating the BIMO CLINSITE dataset

1) Draft specifications for deriving variables for the CLINSITE dataset (note that red font variables are not in the Technical Conformance Guide[2], but added for traceability of efficacy variable derivations and/or supplemental efficacy information)

Table 7 Sample CLINSITE Dataset Specifications

Variable Name	Variable Label	Type	Length	Origin	Specification
SITEID	Study Site Identifier	Char	2	ADSL.SITEID	A site will have 1 to 4 records (2 treatment arms and 2 studies), depending on the number of subjects per site and if the site participated in one or both studies. **Do not include data for the sites at which subjects were screened but no subjects enrolled in Study 0001.**

Table 7 Sample CLINSITE Dataset Specifications

Variable Name	Variable Label	Type	Length	Origin	Specification
EVALPOP	Subjects with Data at Month x	Num	8	Derived	For records where STUDYID = '0001', EVALPOP = subject counts in ADEFF.PARAMCD = 'PRIMARY' by ADEFF.SITEID and ADEFF.TRT01P. Set to 0, if no subject count by ADEFF.SITEID and ADEFF.TRT01P. Set to _Blank_ for STUDYID='0002' records.
NRESP	Responders at Month x	Num	8	Derived	For records where STUDYID = '0001', NRESP = subject counts in ADEFF.PARAMCD = 'PRIMARY' and ADEFF.AVALC = 'Y' by ADEFF.SITEID and ADISTAT.TRT01P. Set to 0, if no subject count by siteid and arm. Set to _Blank_ for STUDYID='0002' records.
TRTEFFR	Treatment Efficacy Result	Num	8	Derived	For STUDYID = '0001': TRTEFFR = NRESP/EVALPOP by siteid and arm. Keep 3 decimal places. If EVALPOP> 0 and NRESP = 0 then TRTEFFR = 0. If EVALPOP = 0 then TRTEFFR = _Blank_. For STUDYID='0002', Set to NRESP to _Blank_.
TRTEFFS	Treatment Efficacy Standard Error	Num	8	Derived	TRTEFFS = sqrt (TRTEFFR * (1-TRTEFFR)/EVALPOP) [If EVALPOP = 1 or TRTEFFR = 0 set TRTEFFS to 0; if EVALPOP = 0, set TRTEFFS = _Blank_] for STUDYID = '0001' by siteid and arm. Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.
SITEEFFE	Site-Specific Treatment Effect	Num	8	Derived	For STUDYID = '0001' for a given site: SITEEFFE = TRTEFFR (when ARM = Active) minus TRTEFFR (when ARM = Placebo). Populate to both ARMs within a site. Keep 3 decimal places. For sites that have only one arm, set SITEEFFE to missing (_blank_). Set to _Blank_ for STUDYID='0002' records.

SITEELCL	Site-Specific Treat Effect 95% Exact LCL	Num	8	Derived	For STUDYID = '0001', obtain 95% exact lower limit for SITEEFFE by siteid using ADEFF.AVAL variable, ADEFF.SITEID and ADEFF.TRT01PN where ADEFF.PARAMCD = 'PRIMARY':
					
					ods output RiskDiffCol2=riskdiff;
					proc freq data=adeff;
					tables trt01pn*aval/chisq riskdiff(CL=EXACT);
					exact riskdiff;
					by siteid;
					run;
					ods output close;
					SITEELCL = round(- riskdiff.ExactUpperCL, 0.0001); where riskdiff.Row = 'Difference'. Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.

SITEEUCL	Site-Specific Treat Effect 95% Exact UCL	Num	8	Derived	For STUDYID = '0001', obtain 95% exact lower limit for SITEEFFE by siteid using ADEFF.AVAL variable, ADEFF.SITEID and ADEFF.TRT01PN where ADEFF.PARAMCD = 'PRIMARY':
					
					ods output RiskDiffCol2=riskdiff;
					proc freq data=adeff;
					tables trt01pn*aval/chisq riskdiff(CL=EXACT);
					exact riskdiff;
					by siteid;
					run;
					ods output close;
					SITEELCL = round(- riskdiff.ExactLowerCL, 0.0001); where riskdiff.Row = 'Difference'. Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.

Table 7 Sample CLINSITE Dataset Specifications

Variable Name	Variable Label	Type	Length	Origin	Specification

SITEEFFS	Site-Specific Treat Eff Standard Error	Num	8	Derived	For STUDYID = '0001', obtain standard error for SITEEFFE by siteid using ADEFF.AVAL variable, ADEFF.SITEID and ADEFF.TRT01PN where ADEFF.PARAMCD = 'PRIMARY':
					
					ods output RiskDiffCol2=riskdiff;
					proc freq data=adeff;
					tables trt01pn*aval/chisq riskdiff(CL=EXACT);
					exact riskdiff;
					by siteid;
					run;
					ods output close;
					SITEEFFS = round(ase,0.0001); where riskdiff.Row = 'Difference'.
					
					Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.

CENSOR	Number of Censored Observations	Num	8	Assigned	CENSOR = _blank_

Sample SAS® code

```
%*------------------------------------------------------------------------ ** ;
%* Derive variables TRTEFFR TRTEFFS SITEEFFE SITEEFFS                     ** ;
%* from ADEFF                                                             ** ;
%* SITEEFS is the asymptotic standard error from PROC FREQ                ** ;
%*------------------------------------------------------------------------ ** ;
proc sort data=SRCDATA.adeff out=efficacy (keep = siteid trt01pn aval avalc);
   where PARAMCD = 'PRIMARY' and AVAL>. ;
   by siteid;
run;
ods output RiskDiffCol2=riskdiff;
proc freq data= efficacy;
   tables trt01pn*aval/chisq riskdiff(CL=EXACT);
   exact riskdiff;
   by siteid;
run;
ods output close;

data riskdiff4 (keep=siteid SITEEFFE SITEEFFS SITEELCL SITEEUCL);
   set riskdiff;
   where Row = 'Difference';
   if risk >. then SITEEFFE = round(-risk,0.001);
   if ase > . then SITEEFFS = round(ase, 0.0001);
   if ExactUpperCL>. then SITEELCL = round(- ExactUpperCL,0.0001);
   if ExactLowerCL>. then SITEEUCL = round(- ExactLowerCL,0.0001);
run;
```

2) Develop primary (production) program
3) Develop secondary (validation) program independently
4) QC the dataset
    a. Compare the primary and secondary CLINSITE datasets
    b. Verify frequency counts against ADaM datasets

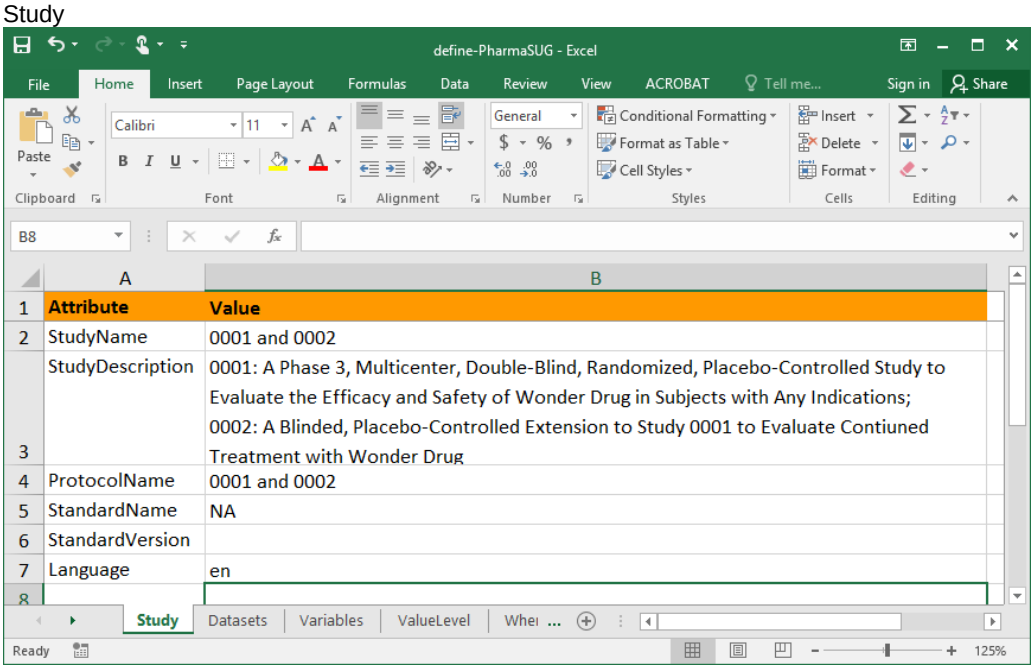
CREATING REVIEWER'S GUIDE

Currently, there are no official CLINSITE Reviewer's Guide. However, we believe a reviewer's guide will be very helpful for FDA reviewers when using the CLINSITE dataset. We adapted the template for the analysis data reviewer's guide (ADRG) to create the CLINSITE data Reviewer's Guide, which contains the following sections:

CREATING DATA DEFINITION

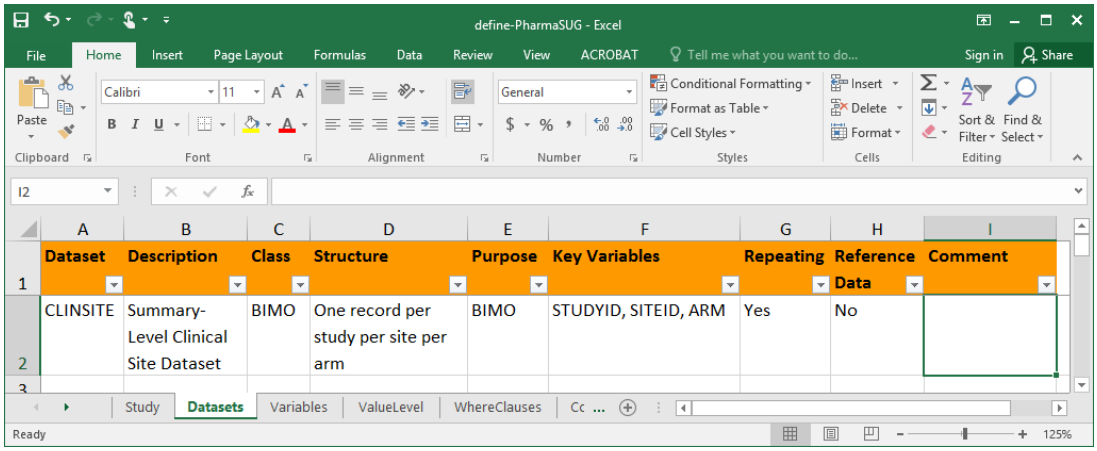
Data definition (define.xml) for CLINSITE was created using Pinnacle 21 Community® version 3.0. In order to use the Pinnacle software, an Excel file can be prepared. The Excel file contains the following spreadsheets:

1) Study



	A	B
1	**Attribute**	**Value**
2	StudyName	0001 and 0002
3	StudyDescription	0001: A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Wonder Drug in Subjects with Any Indications; 0002: A Blinded, Placebo-Controlled Extension to Study 0001 to Evaluate Contiuned Treatment with Wonder Drug
4	ProtocolName	0001 and 0002
5	StandardName	NA
6	StandardVersion	
7	Language	en
8		

2) Datasets

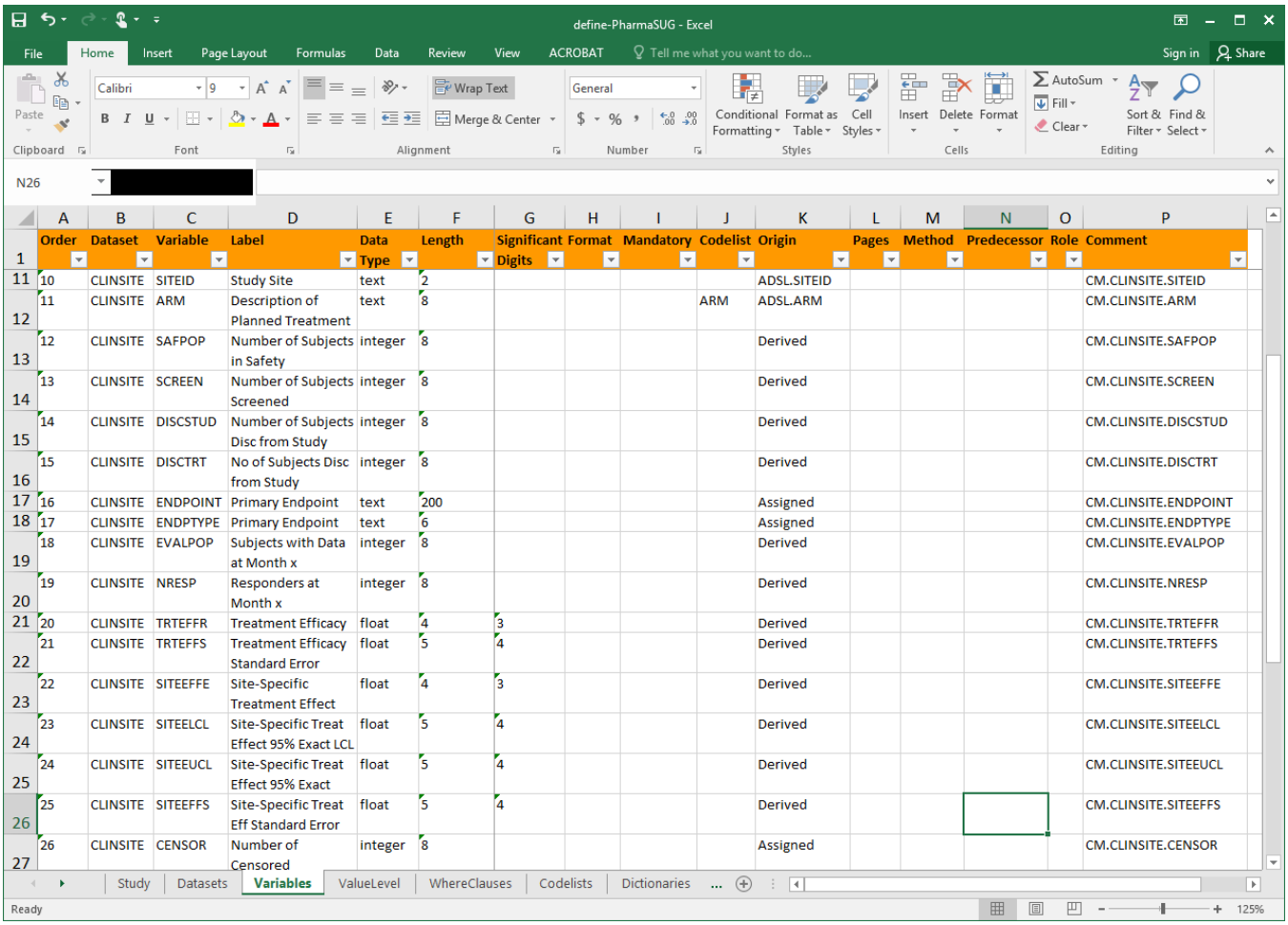


	A	B	C	D	E	F	G	H	I
1	**Dataset**	**Description**	**Class**	**Structure**	**Purpose**	**Key Variables**	**Repeating**	**Reference Data**	**Comment**
2	CLINSITE	Summary-Level Clinical Site Dataset	BIMO	One record per study per site per arm	BIMO	STUDYID, SITEID, ARM	Yes	No	
3									

Note: FDA may request for source datasets to the CLINSITE dataset. Examples of source datasets are:

a. ADSL (subject level analysis dataset)
b. ADAE (information related to SAEs)
c. ADDV (information related to protocol deviations)
d. ADEFF (information related to efficacy)
e. If information about screen failure subjects is not available in SDTM DS dataset, a raw (source) data that contains such information may be provided.

3) Variables



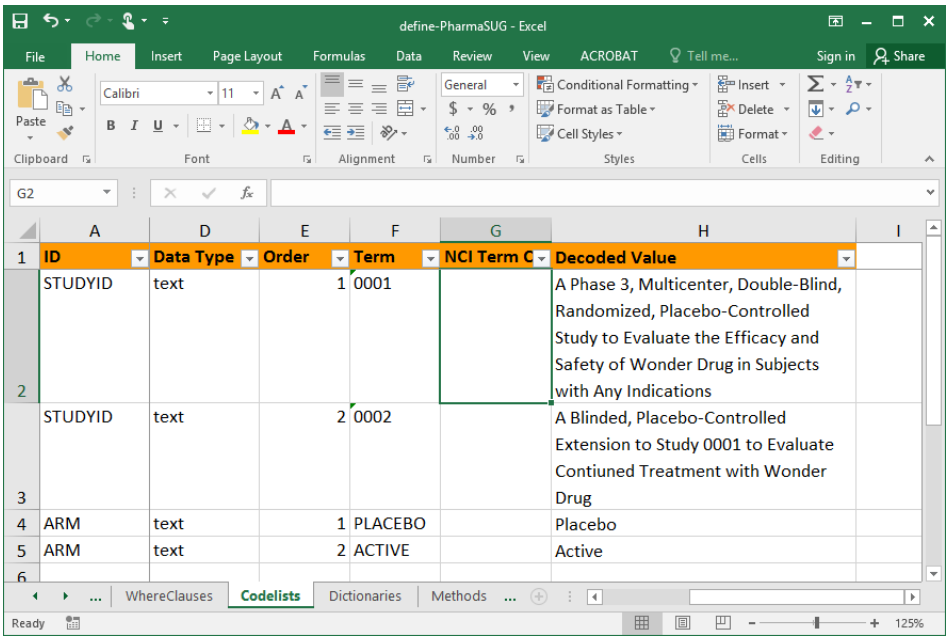
	Order	Dataset	Variable	Label	Data Type	Length	Significant Digits	Format	Mandatory	Codelist	Origin	Pages	Method	Predecessor	Role	Comment
11	10	CLINSITE	SITEID	Study Site	text	2					ADSL.SITEID					CM.CLINSITE.SITEID
12	11	CLINSITE	ARM	Description of Planned Treatment	text	8				ARM	ADSL.ARM					CM.CLINSITE.ARM
13	12	CLINSITE	SAFPOP	Number of Subjects in Safety	integer	8					Derived					CM.CLINSITE.SAFPOP
14	13	CLINSITE	SCREEN	Number of Subjects Screened	integer	8					Derived					CM.CLINSITE.SCREEN
15	14	CLINSITE	DISCSTUD	Number of Subjects Disc from Study	integer	8					Derived					CM.CLINSITE.DISCSTUD
16	15	CLINSITE	DISCTRT	No of Subjects Disc from Study	integer	8					Derived					CM.CLINSITE.DISCTRT
17	16	CLINSITE	ENDPOINT	Primary Endpoint	text	200					Assigned					CM.CLINSITE.ENDPOINT
18	17	CLINSITE	ENDPTYPE	Primary Endpoint	text	6					Assigned					CM.CLINSITE.ENDPTYPE
19	18	CLINSITE	EVALPOP	Subjects with Data at Month x	integer	8					Derived					CM.CLINSITE.EVALPOP
20	19	CLINSITE	NRESP	Responders at Month x	integer	8					Derived					CM.CLINSITE.NRESP
21	20	CLINSITE	TRTEFFR	Treatment Efficacy	float	4	3				Derived					CM.CLINSITE.TRTEFFR
22	21	CLINSITE	TRTEFFS	Treatment Efficacy Standard Error	float	5	4				Derived					CM.CLINSITE.TRTEFFS
23	22	CLINSITE	SITEEFFE	Site-Specific Treatment Effect	float	4	3				Derived					CM.CLINSITE.SITEEFFE
24	23	CLINSITE	SITEELCL	Site-Specific Treat Effect 95% Exact LCL	float	5	4				Derived					CM.CLINSITE.SITEELCL
25	24	CLINSITE	SITEEUCL	Site-Specific Treat Effect 95% Exact	float	5	4				Derived					CM.CLINSITE.SITEEUCL
26	25	CLINSITE	SITEEFFS	Site-Specific Treat Eff Standard Error	float	5	4				Derived					CM.CLINSITE.SITEEFFS
27	26	CLINSITE	CENSOR	Number of Censored	integer	8					Assigned					CM.CLINSITE.CENSOR

Since the structure of our CLINSITE dataset was relatively simple, we did not use the following spreadsheets.

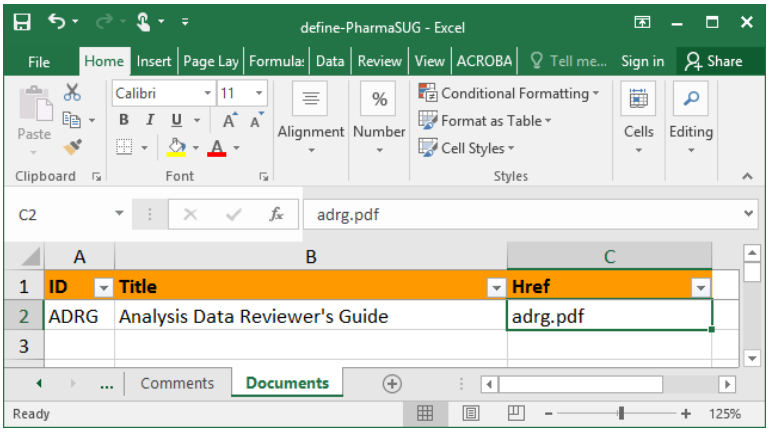
4) ValueLevel (not used)
5) WhereClauses (not used)
6) Dictionary (not used)
7) Method (could have been used, but was not used)

Any of the above should be used, if a CLINSITE dataset contains relevant information.

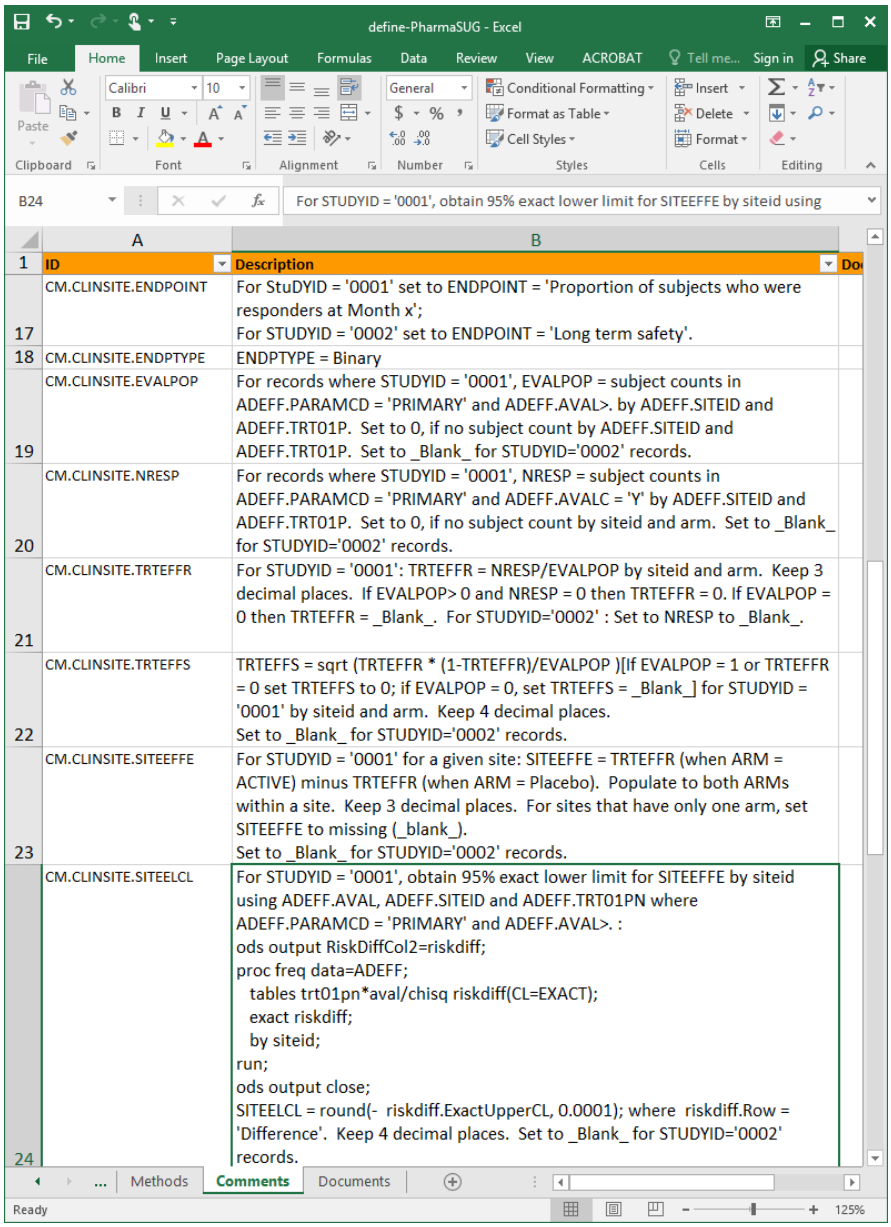
8) Codelists



9) Documents



10) Comments

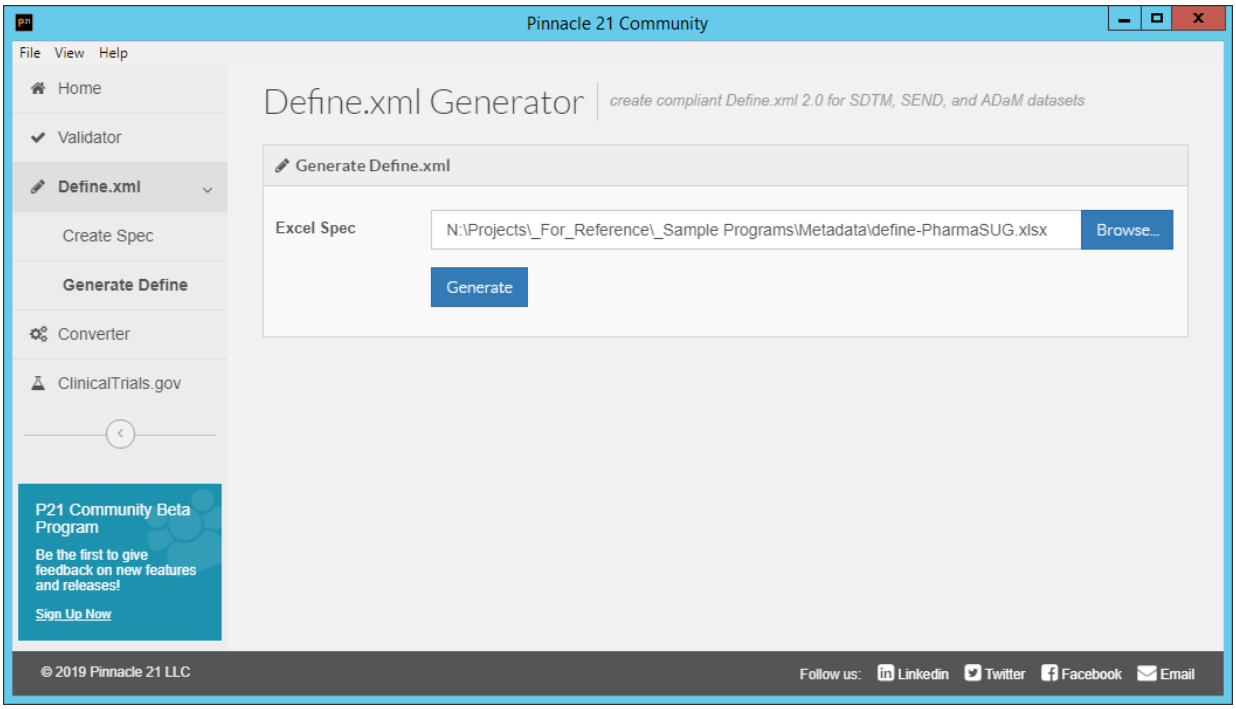


	A	B
1	ID	Description
17	CM.CLINSITE.ENDPOINT	For StuDYID = '0001' set to ENDPOINT = 'Proportion of subjects who were responders at Month x'; For STUDYID = '0002' set to ENDPOINT = 'Long term safety'.
18	CM.CLINSITE.ENDPTYPE	ENDPTYPE = Binary
19	CM.CLINSITE.EVALPOP	For records where STUDYID = '0001', EVALPOP = subject counts in ADEFF.PARAMCD = 'PRIMARY' and ADEFF.AVAL>. by ADEFF.SITEID and ADEFF.TRT01P. Set to 0, if no subject count by ADEFF.SITEID and ADEFF.TRT01P. Set to _Blank_ for STUDYID='0002' records.
20	CM.CLINSITE.NRESP	For records where STUDYID = '0001', NRESP = subject counts in ADEFF.PARAMCD = 'PRIMARY' and ADEFF.AVALC = 'Y' by ADEFF.SITEID and ADEFF.TRT01P. Set to 0, if no subject count by siteid and arm. Set to _Blank_ for STUDYID='0002' records.
21	CM.CLINSITE.TRTEFFR	For STUDYID = '0001': TRTEFFR = NRESP/EVALPOP by siteid and arm. Keep 3 decimal places. If EVALPOP> 0 and NRESP = 0 then TRTEFFR = 0. If EVALPOP = 0 then TRTEFFR = _Blank_. For STUDYID='0002' : Set to NRESP to _Blank_.
22	CM.CLINSITE.TRTEFFS	TRTEFFS = sqrt (TRTEFFR * (1-TRTEFFR)/EVALPOP)[If EVALPOP = 1 or TRTEFFR = 0 set TRTEFFS to 0; if EVALPOP = 0, set TRTEFFS = _Blank_] for STUDYID = '0001' by siteid and arm. Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.
23	CM.CLINSITE.SITEEFFE	For STUDYID = '0001' for a given site: SITEEFFE = TRTEFFR (when ARM = ACTIVE) minus TRTEFFR (when ARM = Placebo). Populate to both ARMs within a site. Keep 3 decimal places. For sites that have only one arm, set SITEEFFE to missing (_blank_). Set to _Blank_ for STUDYID='0002' records.

24	CM.CLINSITE.SITEELCL	For STUDYID = '0001', obtain 95% exact lower limit for SITEEFFE by siteid using ADEFF.AVAL, ADEFF.SITEID and ADEFF.TRT01PN where ADEFF.PARAMCD = 'PRIMARY' and ADEFF.AVAL>. :
		ods output RiskDiffCol2=riskdiff;
		proc freq data=ADEFF;
		tables trt01pn*aval/chisq riskdiff(CL=EXACT);
		exact riskdiff;
		by siteid;
		run;
		ods output close;
		SITEELCL = round(- riskdiff.ExactUpperCL, 0.0001); where riskdiff.Row = 'Difference'. Keep 4 decimal places. Set to _Blank_ for STUDYID='0002' records.

To create the define.xml,

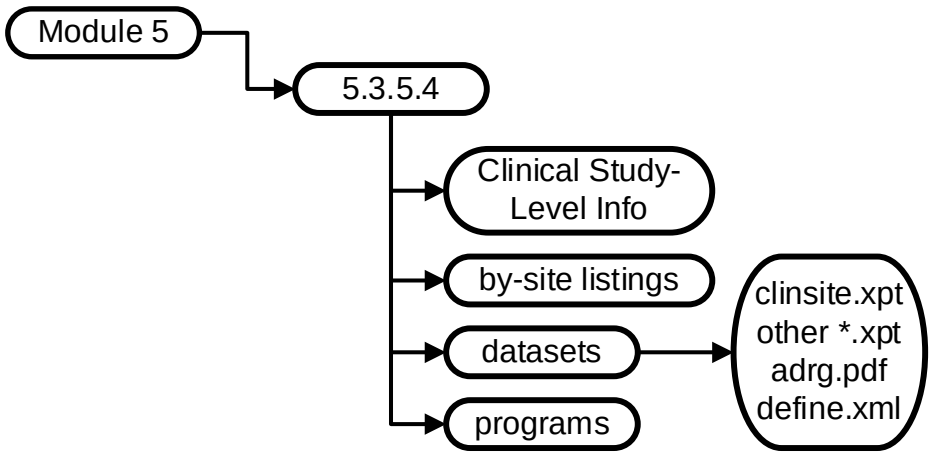
1) Open Pinnacle 21 software
2) Select “Define.xml” on the left panel
3) Press “Browse” to selected the input Excel file as prepared above
4) Press “Generate” (see next section for partial image of the define.xml)



PUTTING IT ALL IN THE ECTD BACKBONE

Structure of BIMO eCTD data documentation is shown in the figure below.

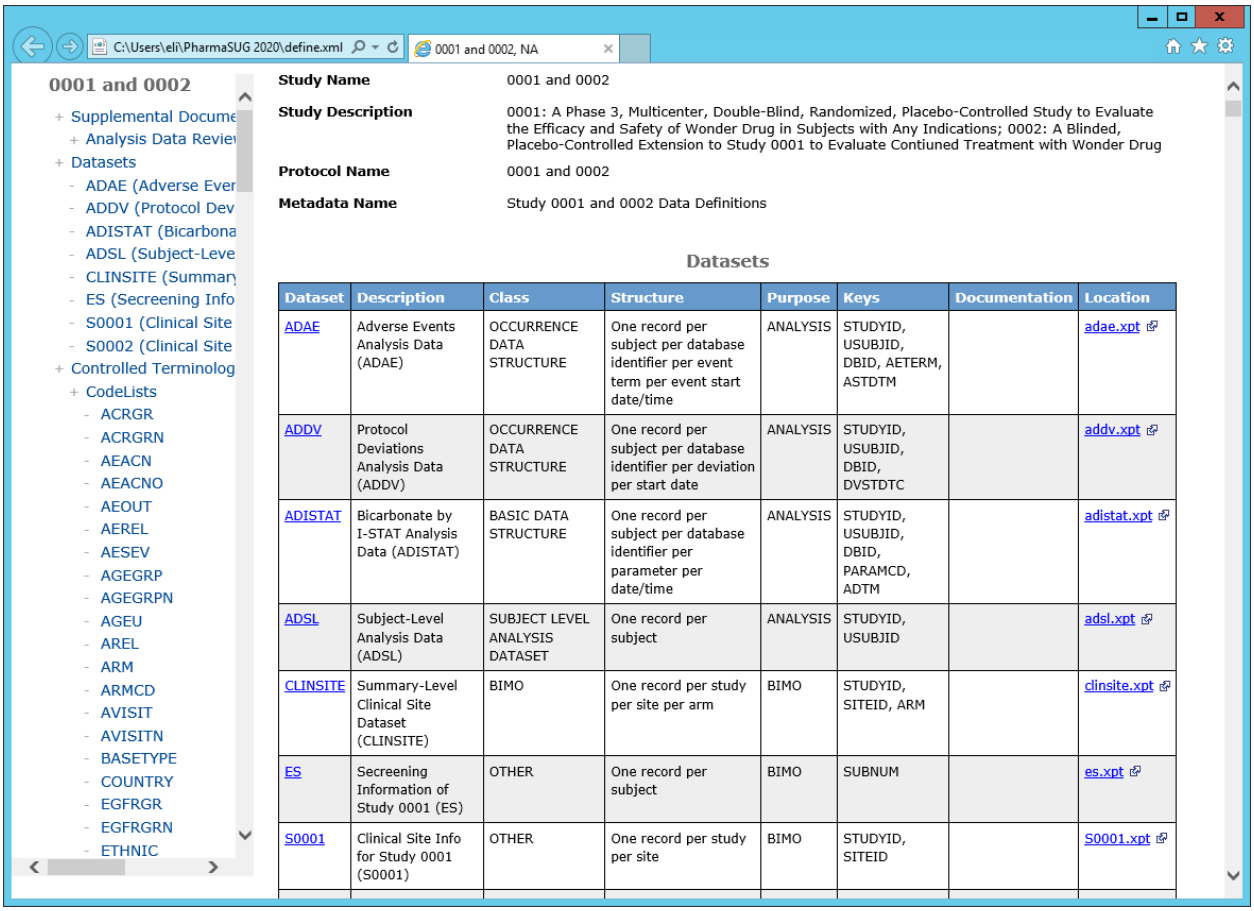
Figure 4 Sample BIMO eCTD Structure



The “datasets” folder contains

- CLINSITE dataset
- source datasets
- adrg.pdf (reviewer’s guide)

- define.xml (data definition)



The “programs” folder contains the SAS® program and macro(s) that generated the CLINSITE dataset.

CONCLUSION

At recent PharmaSUG meetings, papers covered the topic of BIMO packages on standardizing the generation of CLINSITE dataset[3], implementing BIMO for multiple studies[4], building a BIMO reviewer’s guide[5], creating listings and the CLINSITE dataset[6], as well as an overview of the OSI requests for BIMO[7]. In this paper, we show a comprehensive approach to create a BIMO data package, from subject-level data line listings by clinical site, the CLINSITE dataset, the CLINSITE dateset reviewer’s guide, define.xml, to submission of programs and source data. We share our experience of overcoming challenges during the process. Finally, we illustrate quality control of a BIMO data package to ensure the highest quality data are submitted to the FDA.

REFERENCES

1. FDA (February 2018) *Standardized Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions Guidance for Industry* (Draft Guidance) https://www.fda.gov/media/85056/download
2. FDA (February 2018) *BIORESEARCH MONITORING TECHNICAL CONFORMANCE GUIDE* https://www.fda.gov/media/85061/download
3. D. Michel and J Maynard (2019) *Clinical Development Standards for FDA Bioresearch Monitoring (BIMO) Submissions.* PharmaSUG 2019 Conference Proceedings PharmaSUG-Paper SS-030
4. R. Valluru and H. Dyavappa (2019) *Multiple Studies BIMO Submission Package – A Programmer’s*

Perspective. PharmaSUG 2019 Conference Proceedings PharmaSUG-Paper SS-162
5. K Kundarapu, J Low, and M Haloui (2019) *Sponsor Considerations for Building a Reviewer's Guide to Facilitate BIMO Review*. PharmaSUG 2019 Conference Proceedings PharmaSUG-Paper SS-240
6. C. S. Kahlon, D. Tirumalasetti, B. Busa, and K. Kooken (2018) *Programmer's Guide for OSI Deliverables – Creation of Site Level Summary Dataset and Automation of BIMO Listings Generation*. PharmaSUG 2018 Conference Proceedings PharmaSUG-Paper SS-16.
7. E Lin, W Cui, R. Li, and Y. Teng (2018) *Preparing the Office of Scientific Investigations (OSI) Requests for Submissions to FDA*. PharmaSUG 2018 Conference Proceedings PharmaSUG-Paper EP15

CONTACT INFORMATION

Your comments, suggestions, and questions are most welcome. Please contact the authors at:

Elizabeth Li,

PharmaStat, LLC (www.pharmastat.com)

elizabethli@pharmastat.com

Carl Chesbrough,

PharmaStat, LLC (www.pharmastat.com)

cchesbrough@pharmastat.com

Inka Leprince

PharmaStat, LLC (www.pharmastat.com)

ileprince@pharmastat.com