

Programming Support for BIMO Deliverable

Rong Zhang, Pfizer (China) Research & Development Co. Ltd, Shanghai, China

ABSTRACT

The Bioresearch Monitoring Program (BIMO) is one of the FDA's inspection programs designed to ensure the protection of human subjects and the quality and integrity of data submitted to the agency. As more and more pivotal studies (phase 2 or phase 3) step towards the submission, it plays an increasingly important role as the cross-center guidelines for the inspections of clinical investigators, sponsors, and IRBs. Among the three parts that OSI requests sponsor to submit, programming function supports mainly on part II (Subject Level Data Listings by Site) and part III (Summary Level Clinical Site Data - SLCS). Possibly we will also be required to work with the regulatory lead or submission manager to draft the BIMO reviewer guide. This paper will focus on sharing the experience from the programmer perspective 1) the procedures of by-site listings and SLCS's generation based on OSI requirement 2) demo on how to package and sort the enormous by-site listings as well as bookmark optimization 3) typical challenges and issues encountered 4) Validation on the by-site data listings with suggested key check points.

INTRODUCTION

The BIMO program is a comprehensive program of on-site inspections and data audits designed to monitor all aspects of the conduct and reporting of FDA regulated research in support of

- New drug product approvals
- New medical device approvals
- Food and color additives
- New veterinary products

FDA uses the onsite inspections to ensure that clinical investigators, sponsors and Institutional Review Boards (IRB) comply with FDA regulations while developing investigational drugs or biologics. However, the current submission format for study data in sponsors' NDA (New Drug Applications (NDA)/subsequent New Drug Applications (sNDA) and Biologic Licensing Agreements (BLA)/subsequent Biologic Licensing Agreements (sBLA) packages does not facilitate efficient site selection for FDA because the data are prepared by subject level not site level in general. So FDA's Office of Scientific Investigation (OSI) manages and requests this BIMO program, which makes medical reviewers who are responsible for approving and disapproving a product, consult with BIMO reviewers to choose which clinical trial sites to inspect.

For each pivotal trial (PHII or PH III), OSI requests sponsor to submit the following three fields:

Part I: General study related information and comprehensive clinical investigator information (if items are provided elsewhere in submission, described location or provide link to requested information)

Part II: Subject Level Data Listings by Site.

Part III: Site Level Dataset.

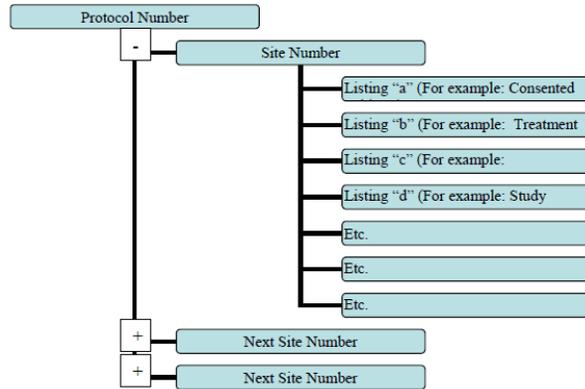
The programming group need to be familiar with the process and request within a limited time, and worked with other departments to compile the necessary BIMO requirements when the task came. We programming and statistics team were requested to support part II and part III and concurrently the clinical team was working on part I. In the following sections the author will introduce the detailed procedures of part II and part III step by step, from programming perspective.

REQUEST FOR SUBJECT LEVEL DATA LISTING BY SITE - PART II

Display 1 shows two types of formats of the subject-level data line listings by clinical site refer to appendix 2 of BIMO Technical Conformance Guide[1]:

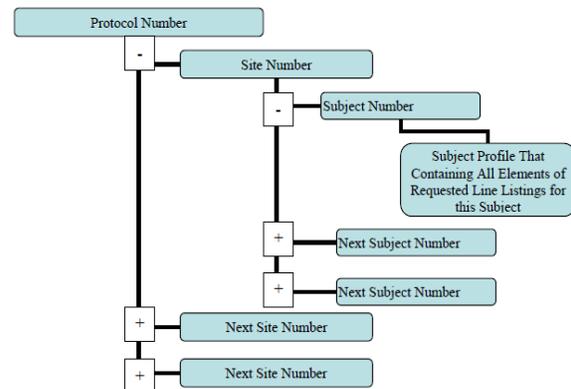
Option A:

By Site, By Listing Option A:



Option B:

By Site, By Subject Profile Option B:



Display 1. Formatting Subject-Level Data Line Listings by Clinical Site

The following procedures will be introduced based on Option A.

CREATE THE LIST OF TABLES (LOT)

Before we start to generate the by-site listings, the first thing we need to do is the preparation of the List of Tables. Unlike the clinical study report (CSR) programming, there will not be the specific paragraphs that mention BIMO in any statistical analysis document. The required by-site listings and its structure should be as follow from FDA OSI Webinar[3], see Display 2:

For each pivotal trial: Site-specific individual subject data listings (hereafter referred to as “line listings”). For each site, provide line listings for:

- a. Listing for each subject consented/enrolled; for subjects who were not randomized to treatment and/or treated with study therapy, include reason not randomized and/or treated
- b. Subject listing for treatment assignment (randomization)
- c. Listing of subjects that discontinued from study treatment and subjects that discontinued from the study completely (i.e., withdrew consent) with date and reason discontinued
- d. Listing of per protocol subjects/ non-per protocol subjects and reason not per protocol
- e. By subject listing of eligibility determination (i.e., inclusion and exclusion criteria)
- f. By subject listing, of AEs, SAEs, deaths and dates
- g. By subject listing of protocol violations and/or deviations reported in the NDA, including a description of the deviation/violation
- h. By subject listing of the primary and secondary endpoint efficacy parameters or events. For derived or calculated endpoints, provide the raw data listings used to generate the derived/calculated endpoint.
- i. By subject listing of concomitant medications (as appropriate to the pivotal clinical trials)
- j. By subject listing, of testing (e.g., laboratory, ECG) performed for safety monitoring

Display 2. Data listing structure requested by OSI

There are three things worth mentioning,

- The index letter (a,b,...j) and the data listing module are one-to-one matching, which means the letter i cannot be occupied even though the study did not collect the concomitant medications.

- The OSI provides only 10 general titles for subject data, but actually the total number of listings is not limited to 10. Each index letter can be spread out the different endpoints with the suffix number. For example, if the study collected both Laboratory data and Vital Signs data, separate them as
 - j.1 By Subject Listing of Laboratory Data - Site #####
 - j.2 By Subject Listing of Vital Signs Data - Site #####
- Population is not clearly defined in the guideline, so please make sure team agrees with the gap between what we need to produce and what the study actually has.

Table 1 is an example list of BIMO tables for pivotal case study:

Purpose of Reporting Event: OSI request: Subject Level Data Listings by Site		
Table #	Table Title	Population Set
a.1	Enrollment, Randomization, Treated Status - Site #####	enrolled
b.1	Subject Listing for Treatment Assignment (Randomization) - Site #####	Intent to treat
c.1	By Subject Listing of Reasons for Discontinuations from Treatment - Site #####	Intent to treat
c.2	By Subject Listing of Reasons for Discontinuations from the Study - Site #####	Intent to treat
e.1	By Subject Listing of Eligibility - Site #####	Intent to treat
f.1	By Subject Listing of Adverse Events - Site #####	As Treated
f.2	By Subject Listing of Serious Adverse Events - Site #####	As Treated
f.3	By Subject Listing of Deaths - Site #####	As Treated
g.1	By Subject Listing of Protocol Deviations - Site #####	Intent to treat
h.1	By Subject Listing of Joint Tender and Swollen Assessment - Site #####	Intent-to-treat
h.2	By Subject Listing of Patient's Assessment of Arthritis Pain - Site #####	Intent-to-treat
h.3	By Subject Listing of Patient's Global Assessment of Arthritis - Site #####	Intent-to-treat
h.4	By Subject Listing of Physician's Global Assessment of Arthritis - Site #####	Intent-to-treat
h.5	By Subject Listing of HS-CRP - Site #####	Intent-to-treat
h.6	By Subject Listing of HAQ-DI - Site #####	Intent-to-treat
i.1	By Subject Listing of Concomitant Drug Treatments - Pre-Medication - Site #####	As Treated
i.2	By Subject Listing of Concomitant Drug Treatments - Medication for Primary Diagnosis - Site #####	As Treated
i.3	By Subject Listing of Concomitant Drug Treatments - Joint Injections - Site #####	As Treated
i.4	By Subject Listing of Concomitant Drug Treatments - Medication for Other than Primary Diagnosis - Site #####	As Treated
i.5	By Subject Listing of Concomitant Drug Treatments - Rescue Medication - Site #####	As Treated
i.6	By Subject Listing of Concomitant Nondrug Treatments - Site #####	As Treated
j.1	By Subject Listing of Laboratory Data - Site #####	As Treated
j.2	By Subject Listing of Vital Signs Data - Site #####	As Treated

Table 1. The Example of the List of tables for case study

CREATE THE MOCK SHELLS (TS)

The OSI materials do not contain mock shells for these listings. Therefore the layout is comparatively flexible and unspecific as long as the requirements from BIMO Guideline are fulfilled. From previous experience, the whole inheritance from the CSR Mock Shell with adding a subtitle or the trailing words in the main title "Site: #####" could be acceptable, after we discussed with team.

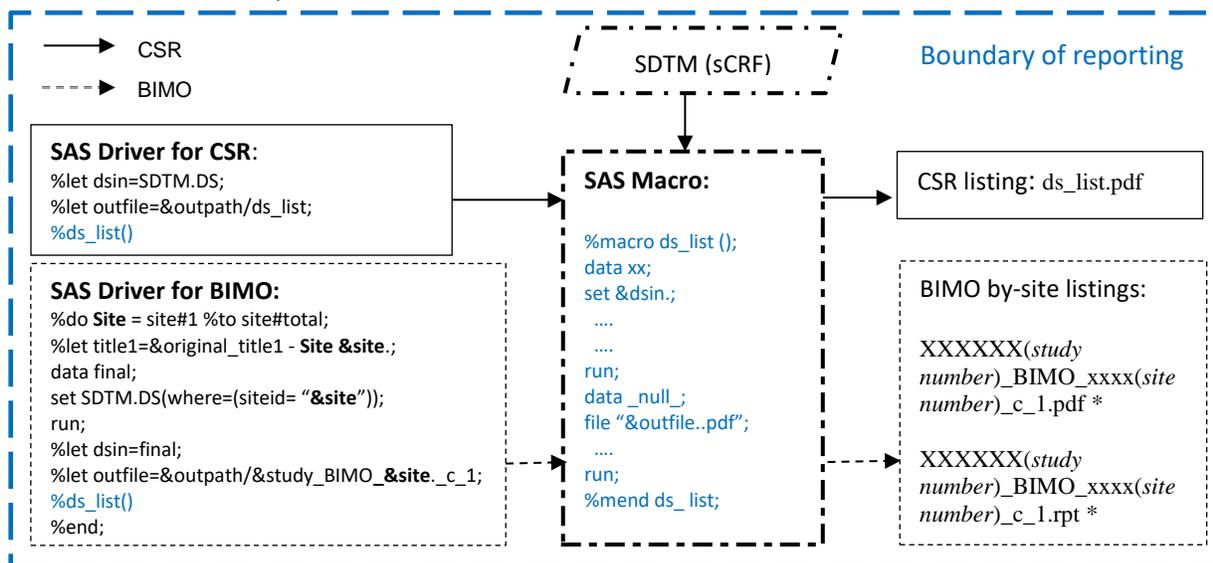
It needs to be emphasized that to ensure there is no any major discrepancies between what you have and the requirements from Bioresearch Monitoring Technical Conformance Guide[1], please read the part II (Subject-Level Data Line Listings By Clinical Site) section A (Organization of the Subject-Level Data Line Listings) very carefully for more details.

Please also notice that the shells we created were just for internal use only.

GENERATE BY-SITE LISTINGS

Our reporting model is creating a macro (.sas file) and a corresponding driver (.sasdrv file) to call that macro. The advantage of this method is we do not need to have another new macro for BIMO by-site listing considering it shares the same layout with the CSR one. We just prepared a new driver to do the loop through all the sites, and what's more, this demo driver can be used as a template and routinely applied to all the other listings. Obviously this action can reduce the operation time to a certain extent.

Figure 1 is the flow map for the generation of CSR listings and BIMO listings per site. Both input and output are different but intermediate process shares the same.



*It's recommended but not required to have file name that contains STUDYID, SITEID then alphabet as suffix

Figure 1. The flow map for CSR listings and BIMO listings

There is another irregularity is that we reported a null table even if there's no data qualified for some site.

PACKAGE BUNDLE BY-SITE LISTINGS

Now all the listings per site have been ready in the folder. Imagine that the study involved 127 sites and there were 23 listings altogether (see Display 2), then we have 127*23=2921 number of listings, more than 4000 pages. So, how to pack such huge amounts of listings and organized by sites? Now I will share what to do.

SAS Programs Preparation (Attached in [Appendix A](#))

- Bundle.sas: This is a program to read the files in table folder and looks for any pdf that begins with "XXXXXX_BIMO". It also builds HTML file (study_sites.htm).
- Bookmarks_fix.sas: This is a program to read the .csv of bookmarks that is exported from Adobe 11. It inserts bookmarks for each site and nests the report bookmarks under each site. At the end, it saves a new .csv file that can be used to create the correct bookmark and imported back through Adobe 11.

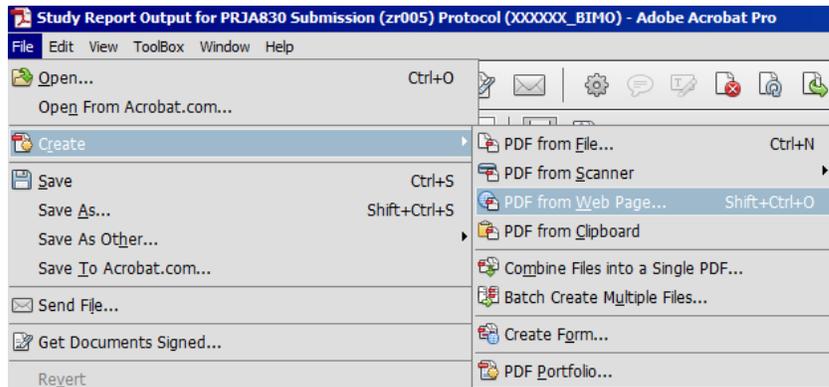
Process of Running SAS Code

- Run the Bundle.sas under Linux prompt. The code will create study_sites.htm file.
- Using study_sites.htm file to pack pdf files via Adobe.
 1. Open Output 1. study_sites.htm created by Bundle.sas by using the Adobe with plug-in program ISIToolbox.

Table Number	User Description	The PDF File
site: 2017		
a_1	Enrollment, Randomization, Treated Status - Enrolled Population - Site 2017	XXXXXX_BIMO_2017_a_1.pdf
b_1	By Subject Listing for Treatment Assignment (Randomization) - Intent to Treat - Site 2017	XXXXXX_BIMO_2017_b_1.pdf
c_1	By Subject Listing of Reasons for Discontinuations from Treatment - As Treated - Site 2017	XXXXXX_BIMO_2017_c_1.pdf
c_2	By Subject Listing of Reasons for Discontinuations from the Study - Intent to Treat - Site 2017	XXXXXX_BIMO_2017_c_2.pdf
d_1	By Subject Listing of Not Per Protocol Reasons - Intent to Treat - Site 2017	XXXXXX_BIMO_2017_d_1.pdf
e_1	By Subject Listing of Eligibility - Intent to Treat - Site 2017	XXXXXX_BIMO_2017_e_1.pdf
f_1	By Subject Listing of Adverse Events - As Treated - Site 2017	XXXXXX_BIMO_2017_f_1.pdf
f_2	By Subject Listing of Serious Adverse Events - As Treated - Site 2017	XXXXXX_BIMO_2017_f_2.pdf

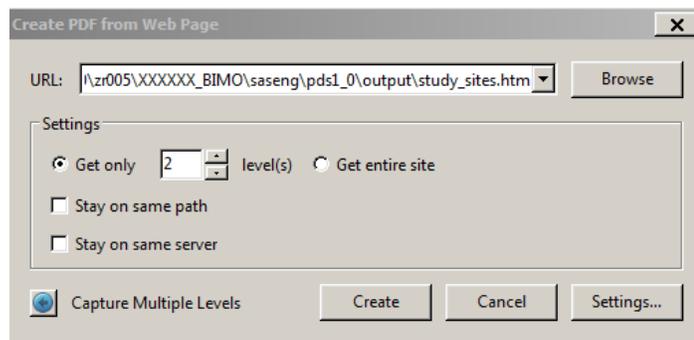
Output 1. study_sites.htm

2. Choose 'PDF from Web Page' under the 'File' menu.



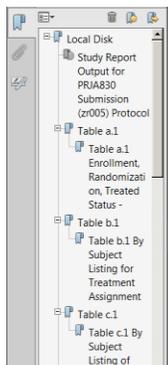
Output 2. Snapshot of 'PDF from Web Page'

3. Click 'Capture Multiple Levels' to expand the details, use the 'Browse' button to navigate to the HTML, and MUST set levels to 2 to get the bookmarks, then click 'Create'. This process takes several minutes to load the pdf files, depending on how many tables need to be downloaded.



Output 3. Snapshot of PDF Creation Step Forward

4. Finally, save the concatenated pdf as Output 4. study_sites.pdf. Now you can see the contents page has been created, however the bookmark does not fit to BIMO requirement. P.S. the first content page is just the specific but not required from OSI.

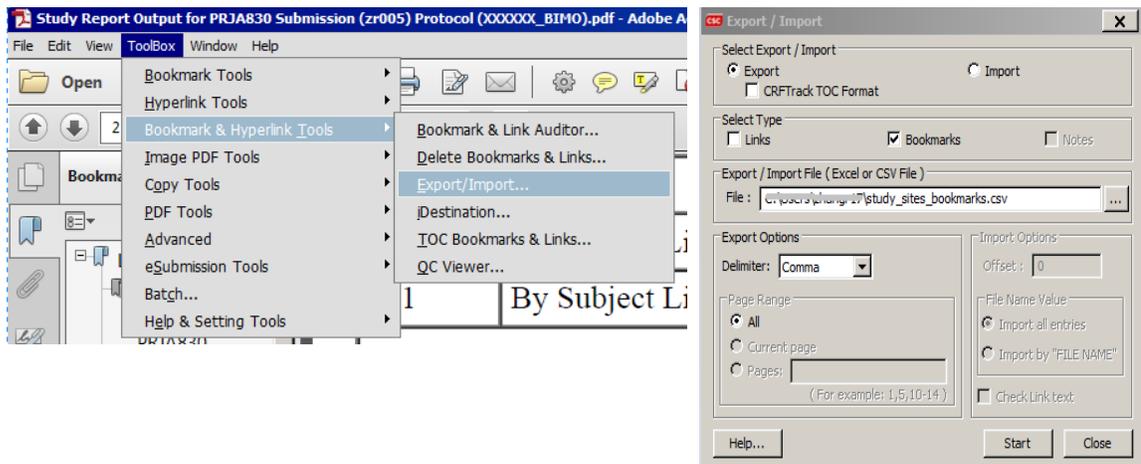


Study Report Output for PRJA830 Submission (zr005) Protocol (XXXXXX_BIMO)		
Table Number	User Description	The PDF File
site: 2017		
a_1	Enrollment, Randomization, Treated Status - Enrolled Population - Site 2017	XXXXXX_BIMO_2017_a_1.pdf
b_1	By Subject Listing for Treatment Assignment (Randomization) - Intent to Treat - Site 2017	XXXXXX_BIMO_2017_b_1.pdf
c_1	By Subject Listing of Reasons for Discontinuations from Treatment - As Treated - Site 2017	XXXXXX_BIMO_2017_c_1.pdf

**Output 4. study_sites.pdf
(with the Initial Bookmarks)**

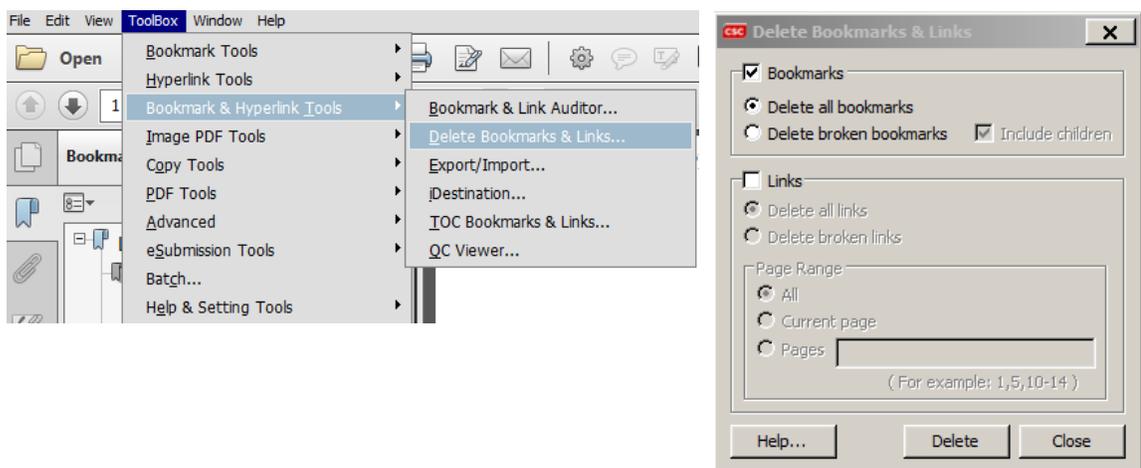
Bookmark Optimization

- Use Bookmarks & Hyperlink choice under the ISIToolbox menu to “Export” old bookmarks. Select only the bookmarks and save the file to study_sites_bookmarks.csv.



Output 5. Snapshot of Bookmark Export

- Run program Bookmarks_fix.sas. This program writes out a new .csv with the same name as the input, with “_fix” appended to the name - study_sites_bookmarks_fix.csv in the same folder.
- Back to Adobe, use the bookmarks and hyperlinks tools under the ISIToolbox menu to delete the old bookmarks and import the new ones:
 - Uncheck ‘Links’, just delete the bookmarks from the concatenated pdf.



Output 6. Snapshot of Bookmark Deletion

- Similar to ‘Export’ procedure, import the new bookmarks through the file study_sites_bookmarks_fix.csv, finally we can get the correct bookmarks and save as Output 7. study_sites_new.pdf.



Output 7. study_sites_new.pdf
(with the fixed Bookmarks)

KEY CHECK POINTS FOR LISTING VALIDATION

We performed the validation by manual considering it's hard to generate another pdf bundle to compare the differences. Here are some key check points from previous experience,

- At least 50% subjects data need to be reviewed
- Pay more attention to those discontinued subjects
- Footnote relevant to the specific data handling should be cared in the corresponding site listings.
- Confirm that the serious adverse event data (Death data) comply with safety database or Argus database if there is any difference between two sources.

REQUEST FOR SITE LEVEL DATASET - PART III

This part is optional. The process of the site level dataset is requested by the agency. The programmer needs to check with the team at the beginning to ensure if this part is really needed. We can explain to the agency that the data includes variable "SITEID", so they can filter the dataset by site. If it still requires, just like what we do for a SDTM, a dataset specification needs to be reviewed by and agreed within the team. Thus, a corresponding define file is necessary to prepare after we create the dataset. There are totally 39 variables requested from the Bioresearch Monitoring Technical Conformance Guide[1]. Table 2 is an example from pivotal case study, and some derivations and instructions have been provided in column 'Derivation':

Seq	Do-main	Variable Name	Variable Label	Type	Leng-th	Codelist/ Controlled Terms or Formats	Derivation
1	DE	STUDYID	Study Number	Char	10	String	ADSL.STUDYID
2	DE	STUDYTL	Study Title	Char	200	String	Informal study title from the protocol. Please streamline the words if the total length is greater than 200 characters
3	DE	SPONNO	Sponsor Number	Num	8	Integer	Team provide
4	DE	SPONNAME	Sponsor Name	Char	60	String	Set to "Pfizer Inc." (Just the example)
5	DE	IND	IND Number	Num	8	6 digit identifier	Team provides.
6	DE	UNDERIND	Under IND	Char	1	(NY)	Team provides.
7	DE	NDA	NDA Number	Num	8	6 digit identifier	Team provides. Set to blank if not applicable
8	DE	BLA	BLA Number	Num	8	6 digit identifier	Team provides. Set to blank if not applicable

Seq	Do-main	Variable Name	Variable Label	Type	Length	Codelist/ Controlled Terms or Formats	Derivation
9	DE	SUPPNUM	Supplement Number	Num	8	Integer	Team provides. Set to blank if not applicable
10	DE	SITEID	Site ID	Char	4	String	ADSL.SITEID. Only sites with at least one randomized subject are included.
11	DE	ARM	Treatment Arm	Char	200	(ARM)	ADSL.ARM, including Screening Failure
12	DE	SAFPOP	Number of Subjects in Safety Population	Num	8	Integer	Total number of subjects randomized at a given site per treatment group (ADSL.SAFFL = "Y")
13	DE	SCREEN	Number of Subjects Screened	Num	8	Integer	Total number of subjects screened at a given site (ADSL.ARMCD ne "SCRNFAIL")
14	DE	DISCSTUD	Number of Subject Discons from Study	Num	8	Integer	Number of subjects discontinued from study (ADSL.EOSSTT="DISCONTINUED")
15	DE	DISCRT	Number of Subject Discons from Study Treatment	Num	8	Integer	Number of subjects discontinued from study treatment (ADSL.EOTSTT="DISCONTINUED")
16	DE	ENDPOINT	Endpoint	Char	120	String	Primary endpoints or key secondary endpoints
17	DE	ENDPTYPE	Endpoint Type	Char	18	(ENDPTYPE)	Value list: "Discrete" (binary, categorical...), "Continuous", "Time-to-Event", "Other"
18	DE	TRTEFFR	Treatment Efficacy Result	Num	8	Floating point	Statistics for the primary endpoints. Need to confirm with Statistician
19	DE	TRTEFFS	Treatment Efficacy Result Standard Deviation	Num	8	Floating point	Standard deviation results for the endpoints. The STD calculation method varies with the endpoint type. Need to confirm with Statistician e.g.: $TRTEFFS = \sqrt{TRTEFFR * (1 - TRTEFFR)}$
20	DE	SITEEFFE	Site-Specific Treatment Effect	Num	8	Floating point	Same representation as reported for the primary efficacy result. Need to confirm with Statistician. e.g.: $(TRTEFFR \text{ for active treatment arm}) - (TRTEFFR \text{ for placebo group})$ if TRTEFFR not missing for either; Set to missing for Placebo treatment arm

Seq	Do-main	Variable Name	Variable Label	Type	Length	Codelist/ Controlled Terms or Formats	Derivation
21	DE	SITEFFS	Site-Specific Treatment Effect Standard Deviation	Num	8	Floating point	Same representation as reported for the primary efficacy result. The STD calculation method varies with the endpoint type. Need to confirm with Statistician. e.g.: $STD = \sqrt{((n1-1)*(STD1^{**2}) + (n2-1)*(STD2^{**2}))/((n1+n2-2))}$ (where n1=arm 1 sample size, n2=arm 2 sample size)
22	DE	CENSOR	Number of Censored Observations	Num	8	Integer (0,1)	Set to blank if not time-to-event endpoint
23	DE	NSAE	Number of Non-Serious Adverse Events	Num	8	Integer	Total number of non-serious AEs at a given site within safety population. Need to confirm with Statistician and Clinician. e.g.: We count the events only once per USUBJID, AEDECOD, AESEV/AETOXGR, AESTDTC, AEENDTC per given site and treatment arms.
24	DE	SAE	Number of Serious Adverse Events	Num	8	Integer	Total number of serious AEs at a given site within safety population. Need to confirm with Statistician and Clinician. e.g.: We count the events only once per USUBJID, AEDECOD, AESEV/AETOXGR, AESTDTC, AEENDTC per given site and treatment arms.
25	DE	DEATH	Number of Deaths	Num	8	Integer	Total number subjects of deaths at a given site by treatment within safety population. Need to confirm with Statistician and Clinician.
26	DE	PROTVIOL	Number of Protocol Violations	Num	8	Integer	Per Protocol Violation list provided by Clinician
27	DE	FINLDISC	Financial Disclosure Amount	Num	8	Floating Point	Team provides. Set to blank if not applicable
28	DE	LASTNAME	Investigator Last Name	Char	80	String	keep last name only for the primary investigator
29	DE	FRSTNAME	Investigator First Name	Char	80	String	keep last name only for the primary investigator
30	DE	INITIAL	Investigator Middle Initial	Char	20	String	Keep it if the middle name initials exist for the primary investigator

Seq	Do-main	Variable Name	Variable Label	Type	Length	Codelist/ Controlled Terms or Formats	Derivation
31	DE	PHONE	Investigator Phone Number	Char	200	String	Team provides. Phone number of the primary investigator. Include country code for non-US numbers.
32	DE	FAX	Investigator Fax Number	Char	200	String	Team provides. Fax number of the primary investigator. Include country code for non-US numbers.
33	DE	EMAIL	Investigator Email Address	Char	200	String	Team provides. Email address of the primary investigator.
34	DE	COUNTRY	Country	Char	3	ISO 3166-1-alpha-3	DM.COUNTRY 3 letters ISO 3166 country code
35	DE	STATE	State	Char	40	String	Team provides. Unabbreviated state or province in which the site is located. If not applicable, keep blank.
36	DE	CITY	City	Char	50	String	Team provides. Unabbreviated city, county, or village in which the site is located.
37	DE	POSTAL	Postal Code	Char	30	String	Team provides. Postal code in which site is located. If not applicable, keep blank.
38	DE	STREET	Street Address	Char	200	String	Team provides. Street address and office number at which the site is located.
39	DE	STREET1	Street Address Continued	Char	200	String	Continued street address if insufficient space for variable STREET

Table 2. Site Level Dataset DE

TYPICAL CHALLENGES AND ISSUES ENCOUNTERED

CHALLENGES AND ISSUES FOR PART II

- Previously we mentioned there were more than 2000 number of listings and 4000 pages that need to be packed. It's too miserable to find the issues after we package the bundles. So our solution was we chose 5 sites by-site listings to make a test, and then expanded to all sites once we confirmed the small bundles had reflected the correct layout and information.
- Sometimes a footnote indicating the special data handling for some subject is required to add to the CSR listings. Considering of this, we recommend keeping the same footnote in BIMO listing for this certain site.

CHALLENGES AND ISSUES FOR PART III

- Except for those efficacy endpoints' calculation, the investigator and address information are difficult to handle. For example, the full name of the principal investigator who comes from Non-US country is 'Halpern, Ari Stiel Radu'. Apparently we cannot identify 'Ari' is the middle name or belongs to the first name. By the same token, we will encounter the same issue on address segmentation. Hence, please request the site to provide the 'clean' information for each cell separately (First name, Last name, Middle name, State, City, Postal...) if possible.

CONCLUSION

Pfizer has its own reporting system and server, so we have to take these into the consideration when coping with the BIMO requests, especially for by-site listing package. Although what I introduced is based on our Pfizer experience, I still expect the method and the whole procedure can bring you some inspiration and enlightenment. Meanwhile I hope the reader can learn from our effort, also plan a good process ahead of your study submission.

REFERENCES

- [1]. "Bioresearch Monitoring Technical Conformance Guide - FDA, Containing Technical Specifications", February 2018. Available at <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formsubmissionrequirements/ucm332468.pdf>
- [2]. "FDA Bioresearch Monitoring (BIMO) Checklist", April 2013, Available at <https://oprs.usc.edu/files/2017/05/FDA-BIMO-Checklist-Copyrighted-pdf.pdf>
- [3]. Meeker-O'Connell, Ann and Kassim, Sean, "Overview of Information Requested by CDER OSI for NDA and BLA Submissions", Oct 2012. Available at <https://collaboration.fda.gov/p44198603/>

ACKNOWLEDGMENTS

Thank you to Pfizer colleagues Oharu, Norihiko and Pasquino, Alan for their creating the whole process of bundle BIMO site level listing for Pfizer submission.

RECOMMENDED READING

I suggest viewing the FDA's OSI webinar and reading the reference materials Bioresearch Monitoring Technical Conformance Guide[1] and FDA Bioresearch Monitoring (BIMO) Checklist[2] listed in the Reference section above.

CONTACT INFORMATION

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

Rong Zhang
Pfizer (China) Research & Development Co. Ltd
Ascendas Lotus Business Park, Building 3, 60 Naxian Road, Pudong Zhangjiang Hi-Tech Park
Shanghai, China 201210
021-28935955
Rong.Zhang@pfizer.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX A: SAS CODE

Bundle.sas:

```

**** Automatic macro variables Setup ****;
%let TABSPROT=;
%let projcode=;/*project number*/
%let prot=;/*protocol number*/
%let htmlwhr= ; /*where clause*/
%let htmlout=study_sites.htm;
%let ordervars=site rpt;
%let byvar=site;
%let output=;

%macro bundle;
*****
**** Specification 1. Set the locations of the files ****;
*****
**** Set a fileref to the folder containing all of the PDFs ****;
filename pdfpath "&tabsprot";
libname SAVED "&output";

** Build the path to HTML file to produce. Written directly to the OUTPUT folder **;
%let htmlfile=&output/&htmlout;
%let tocfile=%sysfunc(tranwrd(&htmlfile, .htm, _bookmarks.csv));

*****
**** Specification 2. Read the PDFS from the table folder ****;
*****
data pdf_list;
  length filename $80;
  **** Open the folder to read the filenames ****;
  DirID=dopen("PDFPath");
  **** Check if the open was successful ****;
  if DirID ne 0 then do;
    **** Loop through all of the files in the directory ****;
    Do file=1 to dnum(DirID);
      filename=dread(DirID,file);
      **** THIS SECTION MAY NEED TO BE MODIFIED TO BE USED IN OTHER STUDIES ****;
      **** Keep only the specified files ****;
      if filename="&prot.-" and upcase(scan(filename,-1,','))="PDF" then do;
        *** pull out the parts of the filename ***;
        site=scan(filename,2,"-");
        rpt =scan(filename,3,"-");
        rpt=tranwrd(rpt,".pdf","");
        section=substr(rpt,1,1);
        output;
      end;
    else put "NOTE: Dropping " filename=;
  end;
  DirId=dclose(dirid);
end;
else put "'ERR'%str()'OR': Unable to read from &tabsprot";
run;

**** Apply any special filters the user wants ****;
data pdf_list;
  set pdf_list;
  &htmlwhr ;
run;

**** Generate a list of the sites/reports found ****;
proc tabulate f=4. missing;

```

```

        table site,rpt/rts=10;
        class site rpt;
run;
proc tabulate f=4. missing;
    table rpt,n/rts=10;
    class rpt;
run;

*****;
**** Specification 3. Get the titles that correspond to each report ****;
*****;
proc sort data=pdf_list out=rpt_list;
    by site rpt;
run;

**** For each distinct report, look for the .RPT file and try to read the title ****;
data title_list;
    set rpt_list(rename=(rpt=start));
    length label $200;
    label="";
    keep filename fid rc fmtname start label fnrc fnrc2 clrc;
    **** Assign a fileref to the .rpt file and attempt to open for input ****;
    fnrc=filename("FILEREF",&tabsprot/"||tranwrd(filename,".pdf", ".rpt"));
    fid=fopen("FILEREF");
    **** Only read if we were able to open the file ****;
    if fid then do;
        **** Read the first three lines (continue only if each read worked) ****;
        if fread(fid)=0 then if fread(fid)=0 then if fread(fid)=0 then
rc=fget(fid,label,200)=0;
        **** Close the .rpt file ****;
        clrc=fclose(fid);
    end;
    **** Build a dataset that can create a format ****;
    fmtname="TITLES";
    fnrc2=filename("FILEREF");
run;

proc sort data=pdf_list;
    by filename;
run;
proc sort data=title_list;
    by filename;
run;

data pdf_list;
    merge pdf_list title_list(keep=filename label);
    by filename;
run;

*****;
**** Specification 4. Generate an HTML file containing a table that lists the PDF****;
**** files. This table is created with a header row for the site****;
**** followed by a row for each table for that site. ****;
*****;
**** Get the list of tables in the appropriate order ****;
proc sort data = pdf_list;
    by &ordervars;
run;

data htmltable;
    file "&htmlfile" new;
    set pdf_list end=last_pdf;
    by &ordervars;

```

```

*****;
**** Create the html header ****;
*****;
if _n=1 then do;
  put @ 1 "<HTML><HEAD>";
  put @ 1 "<TITLE>Study Report Output for &projcode Submission Protocol
(&prot)</TITLE>";
  put @ 1 "</HEAD><BODY>";
  put @ 1 "<font face='Times New Roman' size='1'>";
  put @ 1 "<br><br><br><br>";
  put @ 1 "<table border='1' cellspacing=0 cellpadding=4 bordercolordark>";
  put @ 1 "<tr> <th colspan=3 align=left valign=top height=20>Study Report
Output for &projcode Submission Protocol (&prot)";
  put @ 1 "</th></tr>";
  put @ 1 "<tr align=center><h1><th>Table Number</th><th width=240>User
Description</th>";
  put @ 1 "<th>The PDF File</th></tr></h1>";
end;

*****;
**** Create the body of the table ****;
*****;

**** Create a header row for the first record per site ****;
%if &byvar ne %str() %then %do;
  if first.&byvar then do;
    put @ 1 "<tr> <th colspan=3 align=left valign=top>&byvar.: " &byvar
"</th></tr>";
  end;
%end;

**** Create a row for each report ****;
quotes = "'";
files="<A
HREF="||quotes||'../table/'||compress(filename)||quotes||">"||compress(filename)||"</A
>";
  put @ 1 "<tr>";
  put @ 1 "<td>" rpt "</td>";
  put @ 1 "<td>" label "</td>";
  put @ 1 "<td>" files "</td>";
  put @ 1 "</tr>";

*****;
**** Create the html footer ****;
*****;
if last_pdf then do;
  put @ 1 "</table></font></BODY></HTML>";
end;

run;

*****;
**** Specification 5. Save the list of reports ****;
*****;
data SAVED.%sysfunc(tranwrd(&htmlout, .htm, ));
  set htmltable;
  title=label;
  drop dirid file quotes files label;
run;

proc tabulate;
  table rpt*title,n/rts=90;

```

```

class rpt title;
run;

%mend;

```

Bookmark_fix.sas:

```

*** Path to the bookmark file and the pdf report dataset from bundle.sas ***;
%let path=&output;
**** Name of the bookmark file exported by Adobe ****;
%let bookmarks=study_sites_bookmarks;
**** Name of the pdf dataset created by bundle.sas ****;
%let report_list=study_sites;
**** Name of the variable in the pdf dataset that defines the groups ****;
%let byvar=site;
%let maxlevels=6;

options linesize=130;
*****;
**** Read the dataset produced by bundle.sas ****;
*****;
libname saved "&path";
data pdf_list;
    set saved.&report_list;
    by &byvar;
    retain &byvar._num 1 pdf_num 0;
    **** Generate the new numbers to indicate the nesting ****;
    **** BYVAR is the upper level, and the PDF becomes the second level ****;
    if first.&byvar then do; &byvar._num=&byvar._num+1; pdf_num=1; end;
    else pdf_num+1;
    **** Remember the first title so we can identify it in bookmarks ****;
    if _n_=1 then call symput("firsttable",'Table a.1');
run;

Title "List of Expected tables. N represents the number of &byvar";
proc tabulate;
    table pdf_num*rpt*title,n/rts=90;
    class pdf_num rpt title;
run;

*****;
*** Divide the bookmarks into header and bookmarks specific to each PDF ***;
***(Adobe generally creates a root bookmark and a bookmark for the HTML table)***;
*****;
data master_bookmarks pdf_bookmarks;
    infile "&path/&bookmarks..csv" delimiter="," MISSOVER DSD lrecl=32767;
    retain pdfs 0;
    length concatpdf $100 type $10 indent $20 bookmark $200 action $50
magnification $20 destpg $20 destfile $20 fullpath $50 zoom $20 destl $20 destr $20
destb $20 destt $20 left $20 right $20 top $20 bottom $20 color $20 style $20
openin $50 open $20 namdes $20 absolu $20 higlig $20 linesty $20;
    input @;
    _infile_=compress(_infile_,"");
    input concatpdf $
        type $
        indent $
        bookmark $
        action $
        magnification $
        destpg $
        destfile $

```

```

        fullpath $
        zoom $
        destl $
        destr $
        destb $
        destt $
        left $
        right $
        top $
        bottom $
        color $
        style $
        openin $
        open $
        namdes $
        absolu $
        higlig $
        linesty $
    ;
**** Find the title of the 1st PDF to determine where bookmarks start ****;
if bookmark="&firsttable" then pdfs=1;
**** Pull apart the numbers from the INDENT variable (looks like x.y.z) ****;
array level[&maxlevels] 11-1&maxlevels;
do l=1 to &maxlevels;
    level[l]=input(scan(indent,l,"."),?? best.);
end;
if pdfs then output pdf_bookmarks;
else output master_bookmarks;
drop pdfs l;
run;

data pdf_bookmarks;
    set pdf_bookmarks;
    retain &byvar._num 1;
    if strip(bookmark)="&firsttable" then &byvar._num=&byvar._num+1;
run;

data pdf_bookmarks;
    set pdf_bookmarks;
    retain pdf_num 0;
    by site_num 12 13 14 15 16;
    if first.site_num then pdf_num=0;
    if first.12 then pdf_num=pdf_num+1;
run;

proc sort data=pdf_list;
    by site_num pdf_num;
run;

data pdf_bookmarks;
    merge pdf_bookmarks(in=a) pdf_list(in=b keep=site_num pdf_num title);
    by site_num pdf_num;
    if a;
    if length(bookmark)>10 then bookmark=title;
run;

data pdf_bookmarks;
    set pdf_bookmarks;
    by site_num pdf_num notsorted;
    retain destpg_pre;
    if first.pdf_num then destpg_pre=destpg;
    else destpg=destpg_pre;
    drop destpg_pre;
run;

```

```

*****;
**Delete the ROOT bookmarks and promote the HTML bookmark to the 1st level**
*****;
data master_bookmarks;
  set master_bookmarks;
  if indent="1" then delete;
  if indent="1.1" then do;
    l2=.;
    openIn="ACROBAT_DEFAULT";
    open="OPEN";
  end;
run;

*****;
**Get a list of the actual PDFs. Keep first bookmark**
*****;
data actual_pdf_list;
  set pdf_bookmarks;
  by l1-l&maxlevels;
  if last.l2;
run;

*****;
**Join list of actual PDFS with the expected PDF list that has group var **
*****;
data actual_pdf_list;
  merge actual_pdf_list pdf_list;
run;

*****;
**** Ensure the join worked properly ****
*****;
data check; set actual_pdf_list;
  if bookmark ne title;
  put "%str(ER)ROR FOUND: Difference between expected and actual PDFS";
run;
TITLE "%str(ER)ROR FOUND: Difference between expected and actual PDFS";
proc print width=min; var title bookmark site indent site_num pdf_num;run;

*** Fix the bookmark for tables that have no bookmark ***;
data mod;
  set check;
  indent=strip(indent)||'.1';
  bookmark=title;
  destt='557';
  l3=1;
  drop filename site rpt section;
run;

data pdf_bookmarks;
  set pdf_bookmarks mod;
  if length(bookmark)<10 then OPEN='CLOSE';
proc sort;
  by l1-l3;
run;

data actual_pdf_list;
  set actual_pdf_list;
  if missing(l3) then do;
    l3=1;
    indent=strip(indent)||'.1';
  end;

```

```

    if bookmark ne title then bookmark=title;
    destt='557';
proc sort;
    by l1-l3;
run;

*****;
**** Join PDF bookmarks with the actual list to get the new numbering ****;
*****;
data pdf_bookmarks;
    merge pdf_bookmarks actual_pdf_list;
    by l1-l3;
run;

*****;
* Replace the old numbers L1&L2 with the new numbering based on the BYVAR *;
*****;
data pdf_bookmarks;
    set pdf_bookmarks;
    l1=&byvar._num;
    l2=pdf_num;
run;

*****;
*Create a list of bookmarks for BYVAR (PDFS will be nested under this one)*;
*****;
data &byvar._bookmarks;
    set actual_pdf_list;
    by &byvar._num;
    if first.&byvar._num;
    bookmark="&byvar.: "||&byvar;
    l1=&byvar._num;
    l2=.;
    l3=.;
    destt='612';
    open='CLOSE';
run;

*****;
** Sort datasets **;
*****;
proc sort data = master_bookmarks;
    by l1-l&maxlevels;run;
proc sort data = pdf_bookmarks;
    by l1-l&maxlevels;run;
proc sort data = &byvar._bookmarks;
    by l1-l&maxlevels;run;

*****;
**** Assemble the header, pdf and byvar bookmarks into a single dataset ****;
*****;
data all_bookmarks;
    set master_bookmarks pdf_bookmarks &byvar._bookmarks;
    by l1-l&maxlevels;

    **** Rebuild the INDENT variable with the new numbers ****;
    array level[&maxlevels] l1-l&maxlevels;
    if l1 ne . then do;
        indent=compress(put(l1,best.));
        l=2;
        do while (level[l] ne . and l<=&maxlevels);
            indent=trim(indent)||"."||compress(put(level[l],best.));
            l=l+1;

```

```
end;
end;

**** The bookmark and the indent string need to have quotes ****;
if _n_>1 then do;
  concatpdf=quote(trim(concatpdf));
  bookmark=quote(trim(bookmark));
  indent=quote(trim(indent));
  color=quote(trim(color));
end;

**** Save this dataset in the .CSV format needed by Adobe ****;
file "&path/&bookmarks._fix.csv" delimiter="," lrecl=32767;
put  concatpdf
      type
      indent
      bookmark
      action
      magnification
      destpg
      destfile
      fullpath
      zoom
      destl
      destr
      destb
      destt
      left
      right
      top
      bottom
      color
      style
      openin
      open
      namdes
      absolu
      higlig
      linesty
      ;
run;

Title "Final list of tables";
proc report nowd;
columns indent bookmark;
define indent/display width=12;
define bookmark /display width=80 flow;
run;
```