

Generating Readable and User-friendly Patient Profile Output in xlsx Format

Yunxiao Zhao, Haihe Biopharma Co., Ltd.

ABSTRACT

Checking data points of a specified subject is high-frequency and important work for clinical research physicians (CRPs), statisticians, and SAS programmers. Usually, the user could review data in sas7b dataset or rich text format (rtf) report. However, it is hard to review data across different domains and re-analyze the data in the rtf report. It takes lots of steps to locate a specified subject's data points when using the SAS review tool, which requests basic knowledge of software operation. In addition, there are no links between datasets, especially the supplemental qualifiers (SUPP) datasets and main datasets. The user must review datasets one by one in the SAS review tool.

This macro is used to generate patient profiles in the xlsx format. The user could just use excel settings and functions to filter, calculate and summarize data. In addition, the user could review data across different datasets quickly by clicking on hyperlinks. The xlsx patient profiles are easy to locate, review and re-analyze for all users.

INTRODUCTION

It is time-consuming work to review subjects' data in clinical trials. CRPs need to review data to grasp information of subjects. Statisticians and SAS programmers need to check data to confirm potential data issues, detail programming logic, and actual value of the specified variable. Usually, the user needs to review data across multiple datasets, such as from subject-level listing to record-level listing, from record-level listing to subject-level listing, or even from one record-level listing to another record-level listing. It is hard to switch from one dataset to another dataset in the rtf report and the SAS review tool.

The patient profile in xlsx format is easier for users to select and cross-review data. This macro is based on the SAS code and SDTM datasets. The displayed output is standardized. Therefore, this macro could be applied to new studies without any modification.

PATIENT PROFILE OUTPUT

This macro could generate one profile for all subjects. Usually, the number of subjects in clinical trials is large, especially in phase II or phase III studies. The coder could decide to generate multiple profiles, one file for one subject. Besides subject ID, any demographic variable could be used to separate profiles into different xlsx files, such as site ID, sex, race, and arm.

There are three parts to this patient profile output.

INDEX OF ALL SHEETS

The first part is the index of all sheets, as in Figure 1. The sheet names displayed are sourced from the SDTM dataset label. All data related to the subjects would be included. If the user wants to go to the sheet of a specified listing, e.g., Figure 2 and Figure 4, they could just click on the sheet name.

	A	B	C
1	Demographics		
2	Comments		
3	Subject Elements		
4	Subject Disease Milestones		
5	Subject Visits		
6	Concomitant or Prior Medicat		
7	Exposure		
8	Procedures		

Figure 1. The output of the “index” sheet

SUBJECT-LEVEL LISTING

The second part is the subject-level listing, including the demographic variables and hyperlinks to detailed listings of interest for each subject. The excel output is filtered for all columns. The first row and first three columns are frozen. So, they could not be scrolled with the scroll bar. This output format makes it easy to select and read data. There is only one record (one row) for one subject.

For example, Figure 2 is the patient profile, which is grouped by sex, and only subjects whose sex='M' are included in this file. The subjects with Sex='M' are highlighted with yellow. If the user wants to open the xlsx file for SEX='F', they could just click on the value “F” in the column “SEX”.

	A	B	C	P	Q	R	S	T
1	Unique Subject Identifier	Study Identifier	Domain Abbreviation	Age Units	Sex	Race	Planned Arm Code	Description of Planned Arm
2	TEST-01	TEST	DM	YEARS	M	ASIAN		
3	TEST-02	TEST	DM	YEARS	M	ASIAN	TEST1	TEST1
4	TEST-04	TEST	DM	YEARS	F	ASIAN	TEST1	TEST1
5	TEST-05	TEST	DM	YEARS	M	ASIAN		
6	TEST-06	TEST	DM	YEARS	M	ASIAN		
7	TEST-07	TEST	DM	YEARS	M	ASIAN	TEST2	TEST2
8	TEST-08	TEST	DM	YEARS	F	ASIAN		

Demographics | Comments | Subject Elements | Subject Disease Milestones | Subject Visits

Figure 2. The output of demographic variables in the subject-level listing

The default listings of interest are exposure, medical history, vital sign, ECG test, lab tests, disease responses, adverse events, concomitant or prior medication, and disposition. The coder could update the keyword parameter in the %macro statement easily to add or remove any domains. If the specified subject had records in the listing of interest, the value “Link to Tab” is displayed and the hyperlink is active, as in Figure 3.

On clicking on the “Link to Tab”, it would lead to the detailed listing for the specified subject, and the records for that subject would be highlighted. For example, when clicking “Link to Tab” of the subject “TEST-02” in the “Exposure” column, it would go to a sheet as in Figure 4, and rows 2-3 for the subject “TEST-02” are highlighted in the exposure domain.

	A	B	C	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	Unique Subject Identifier	Study Identifier	Domain Abbreviation	Exposure	Medical History	Vital Sign	ECG Test Result	Laboratory Test Result	Disease Response and Clinical	Adverse Events	Concomitant or Prior Medication	Disposition
2	TEST-01	TEST	DM			Link to Tab						Link to Tab
3	TEST-02	TEST	DM	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab
4	TEST-04	TEST	DM									
5	TEST-05	TEST	DM			Link to Tab						Link to Tab
6	TEST-06	TEST	DM			Link to Tab	Link to Tab	Link to Tab				Link to Tab
7	TEST-07	TEST	DM	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab	Link to Tab
8	TEST-08	TEST	DM									

Demographics | Comments | Subject Elements | Subject Disease Milestones | Subject Visits

Figure 3. The output of hyperlinks to detail listings in the subject-level listing

RECORD-LEVEL LISTINGS

The third part is record-level listings, including all interventions, events, findings, and other information for each subject. Data in different domains are displayed in different sheets. Figure 4 is an example of exposure listing. From this listing, if the user wants to go to the subject-level information for a specified subject, they could just click on the subject's ID which serves as a hyperlink.

For example, it would go to a sheet as in Figure 3 and row 3 for the subject "TEST-02" is highlighted in the demographics domain when clicking on the value "TEST-02" in the "Unique Subject Identifier" column.

	A	B	C	D	E	F	G	H	I	J	K	L
	Unique Subject Identifier	Study Identifier	Domain Abbreviation	Sequence Number	Sponsor-Defined Identifier	Name of Treatment	Dos	Dose Unit	Dose Form	Dosing Frequency per Interval	Route of Administration	Reason for Dose Adjustment
1	TEST-02	TEST	EX	1.1		TEST		10 mg	TABLET	QD	ORAL	
2	TEST-02	TEST	EX	2.2		TEST		20 mg	TABLET	QD	ORAL	
3	TEST-03	TEST	EX	1.1		TEST		20 mg	TABLET	QD	ORAL	
4	TEST-04	TEST	EX	2.4		TEST		20 mg	TABLET	NA	ORAL	Adverse Event
5	TEST-05	TEST	EX	3.5		TEST		20 mg	TABLET	QD	ORAL	
6	TEST-06	TEST	EX	4.6		TEST		20 mg	TABLET	QD	ORAL	
7	TEST-07	TEST	EX	5.7		TEST		20 mg	TABLET	NA	ORAL	Adverse Event
8	TEST-08	TEST	EX	6.8		TEST		20 mg	TABLET	QD	ORAL	
9	TEST-09	TEST	EX	7.9		TEST		20 mg	TABLET	QD	ORAL	

Figure 4. The output of record-level listings

The SUPP datasets are merged and combined with the main datasets. Therefore, the user could review all information of one domain in one sheet.

In this patient profile, the user could cross-review data from subject-level to record-level, from record-level to subject-level with just one click. The user also could cross-review data from one record-level to another record-level with just two clicks, e.g., click on subject ID to go to the subject-level listing and click on "Link to Tab" to go to the targeted record-level listing.

KEYWORD PARAMETERS OF MACRO

This macro had only seven keyword parameters. Five of them have a default setting. The coder could just define the folders of source data and output to generate the profiles with the default settings.

DATAFOLD: the folder location of the source SDTM datasets in sas7b format.

OUTFOLD: the folder location used for storing the outputs.

SUB2REC: the datasets which need a subject-level to record-level hyperlink. The character "|" is used to connect the names of datasets. The default value is "EX|MH|VS|EG|LB|RS|AE|CM|DS". The coder and reviewer could decide to add or remove any datasets. If the dataset in the list does not exist or there are no records in the dataset, the dataset would be ignored. If there is at least one record in the dataset, the macro will generate a column in the subject-level listing. The value "Link to Tab" is displayed and a hyperlink is added for the subjects who had records in the dataset.

DISPLABEL: displaying the label or name of datasets and variables. This parameter could be assigned as "Y" or "N" and the default value is "Y". The label will be displayed when "Y" is assigned, and the name will be displayed when "N" is assigned. If the user is familiar with the CDISC variable-naming abbreviations, displaying the name is clear enough for the reviewer.

SUBDAT: the name of the subject-level dataset. The default value is "DM". This dataset is used to display the subject-level dataset, e.g., Figure 2.

SUBID: the name of the subject ID variable. The default value is "USUBJID". This variable is used to identify the subjects uniquely in all the datasets.

SEPVAR: the name of a variable to separate patient profiles in different files. The default value is "SUBID". Any variables in the subject-level dataset could be used to separate the patient profiles, such as site ID, sex, race, and arm. If there is a missing value, "UK" would be assigned as the name of the patient profile.

MACRO CODE

This macro does not require complex programming skills. But programming knowledge in different areas is used, such as dictionary tables, macro variables, macro function, %DO loop, and excel output. The coder needs a good and wide understanding of SAS code and SDTM data structure.

1. Getting the dataset names that related to subjects. We could use dictionary tables to get the metadata of the datasets. The dataset "dat_att" includes all datasets which had at least one record. The dataset "var_att" includes all datasets related to subjects.

```
proc sql;
  create table dat_att as
  select libname,memname,memlabel,nobs
  from sashelp.vtable
  where libname in("SOUDATA") and memtype='DATA' and nobs>0;
quit;
proc sql;
  create table var_att as
  select memname,name
  from sashelp.vcolumn
  where libname in("SOUDATA") and memtype="DATA" and name="&subid.";
quit;
```

2. Combining all the SUPP datasets to the main datasets if SUPP datasets exist.
3. Defining a series of macro variables, including dataset name, dataset label, and values to separate the profiles. There are many ways to define the macro variable. "call symput" is used in this macro.

```
data _null_;set dat_att3 end=last;
  call symput('dat' || strip(put(_n_,best.)),strip(memname));
  call symput('datdis' || strip(memname),strip(memlabel));
  if last then call symput('datnum',strip(put(_n_,best.)));
run;
```

4. Getting the location of subject-level listing to record-level listing, e.g., from row No. X to row No. Y. We could use the data step to get the location of each record-level listing for each subject.
5. Defining the excel output format. "ods excel" is used to specify the excel format. "autofilter" turns on the filtering for the columns, "gridlines" prints the grid lines, "flow" enables wrap text, "sheet_name" specifies the name of sheet, "frozen_headers" specifies the first row to not scroll, "frozen_rowheaders" specifies the first three columns to not scroll.

```
ods excel file="&outfold.\xxx.xlsx" options(autofilter='all'
gridlines='yes' flow='tables' sheet_name="xxx" frozen_headers='yes'
frozen_rowheaders='3' );
```

6. Defining the output columns and hyperlinks. "proc report" is used to specify the columns in the output. "call define" is used to define the style and hyperlinks in the output.

```
proc report data=xxx split='^';
  column (_all_ );
  define subn/noprint;
  compute usubjid;
    if usubjid ne "" then do;
      length urlstring $200.;
```

```
urlstring="#'&sheetname.'!'"||strip(subn)||':'||strip(subn);  
call define(_col_,'URL',urlstring);  
call define(_col_,"STYLE","style=[color=blue]");  
end;  
endcomp;  
run;
```

7. Use macro variables and %DO loop to generate productions, file by file, sheet by sheet.

CONCLUSION

This macro could generate patient profiles in the xlsx format. It is easy to select, read, cross-review, and re-analyze data in the xlsx outputs. This output would be very useful in helping CRPs, statisticians, and SAS programmers to review data.

REFERENCES

Shun Dai. "Macro-Ods Excel Output". 2017. Available at https://mp.weixin.qq.com/s?search_click_id=1813017789691375092-1657865474077-813785&_biz=MzlyNTQ5NDI5MA==&mid=2247483982&idx=1&sn=86ae0ea96c1ddde32a812c37a1a99a27&chksm=e87f9633df081f253234744172dbc57a0563ee55eb42db8da980c5e805dcd1cf52dcd58a41f2&scene=0&subscene=10000&clicktime=1657865474&enterid=1657865474&sessionid=0&ascene=65&fasttmpl_type=0&fasttmpl_fullversion=6241523-zh_CN-zip&fasttmpl_flag=0&realreporttime=1657865474103#rd

CONTACT INFORMATION

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

Yunxiao Zhao
Haihe Biopharma Co., Ltd.
Yunxiao.Zhao@haihepharma.com

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