Thanks CDISC – Free Utilities

Robert Agostinelli, Sepracor Inc., Marlborough, MA

ABSTRACT

With the emergence of CDISC data standards (thanks CDISC!), there are opportunities to produce tools that check for the conformance to the standards, and more importantly, tools for data presentation/reporting. These tools could be developed in an open way, such that the entire industry can create and take advantage of work done by individuals or groups. Regarding tools to check for data conformance to the standards, OpenCDISC (www.opencdisc.org) has a free tool that will check for conformance to the SDTM standard. This paper will focus on data presentation/reporting.

INTRODUCTION

Due to the CDISC data standards, it is now easier to present subject level information from your SDTM data without lifting many fingers. The following outputs are created by using SAS® utilities (i.e. macros) that provide a good deal of subject level information from your SDTM datasets. Some of these outputs may be beneficial to Data Management staff who may be interested in subject, visit and visit date information.

Since pictures are worth a thousand words, screen shots of the output will be presented first, with a description of each output thereafter. There are three utilities presented here, where each utility may produce more than one output. The outputs are mostly XML files that can be viewed within Excel.

THE UTILITIES

UTILITY #1: SUBJECT VISIT DATES (PRODUCES 3 OUTPUTS)

Subject Visit Dates Report
Sheet name: Multiple Dates in a Visit
Output: subject_visit_dates_SDTM_2009mmdd.xls   Created on 29OCT09 11:46

Subject	Domain	Category	V1_SCREENING (1)	V2 (2)	V3 (3)	V4 (4)	V5_EOS (5)
0001-S014	IE	INCLUSION		2008-08-25			
0001-S014	LB	CHEM	2008-08-14	2008-08-25			
0001-S014	LB	HEM	2008-08-14	2008-08-25			
0001-S014	LB	TOX	2008-08-14	2008-08-25			
0001-S014	LB	UA	2008-08-14	2008-08-25			
0001-S014	MH	GENERAL	2008-08-14				
0001-S014	MH	PSYCHIATRIC	2008-08-18*				
0001-S014	PE	GENERAL	2008-08-18*				
0001-S014	PE	NEUROLOGICAL	2008-08-18*				
0001-S014	VS		2008-08-14	2008-08-25			

Figure 1a

Figure 1a (see above) of the Subject Visit Dates Report reveals subjects who have multiple visit dates within a visit identifier. In the output above, Visit 1 has 2008-08-14 for most datasets, except for the MH and PE datasets. These visit dates may be acceptable given that subjects may be allowed to leave the clinic and return a few days later during the screening period if the subject was unable to complete all screening activities on the same day.

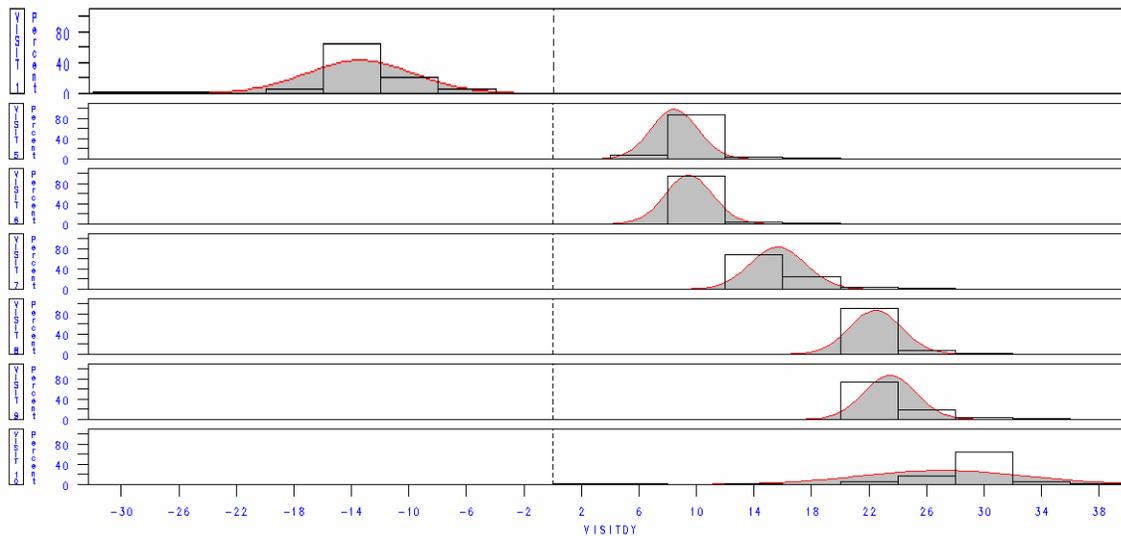*Subject Visit Dates Report*
Sheet name: Dates not Chronological
Output: subject_visit_dates_SDTM_2009mmdd.xls   Created on 29OCT09 11:46

Subject	Domain	Category	Reference Start Date	V1_SCREENING (1)	V2 (2)	V3 (3)	V4 (4)	V5_EOS (5)
0031-S015	LB	CHEM	2008-07-10	2008-07-02	2008-07-09		2008-08-07	
0031-S015	LB	HEM	2008-07-10	2008-07-02	2008-07-09		2008-08-07	
0031-S015	LB	TOX	2008-07-10	2008-07-02	2008-07-09		2008-08-07	
0031-S015	LB	UA	2008-07-10	2008-07-02	2008-07-09		2008-08-07	
0031-S015	MH	GENERAL	2008-07-10	2008-07-02				
0031-S015	MH	PSYCHIATRIC	2008-07-10	2008-07-02				
0031-S015	PC	ANALYTE	2008-07-10		2008-07-31	2008-07-23*	2008-08-07	
0031-S015	PE	GENERAL	2008-07-10	2008-07-02	2008-07-09			
0031-S015	PE	NEUROLOGICAL	2008-07-10	2008-07-02				
0031-S015	QS	Suicidality Tracking Scale (STS)	2008-07-10	2008-07-02	2008-07-09	2008-07-23	2008-08-07	2008-09-17
0031-S015	VS		2008-07-10	2008-07-02	2008-07-09	2008-07-23	2008-08-07	2008-09-17

Figure 1b

Figure 1b (see above) of the Subject Visit Dates Report reveals subjects who have visit dates that are not chronologically ascending within a dataset. In the example above, the PC dataset has a Visit 3 date (2008-07-23) that is not larger than the Visit 2 date (2008-07-31). Upon further examination, the issue is with the Visit 2 date, which most likely should be 2008-07-09 to be consistent with all of the other datasets. In essence, this report is a visit dates 'data scrubber'.

Figure 1c

Figure 1c (see above) of the Subject Visit Dates Report reports histograms of the actual relative days from dosing (i.e. VISITDY) by visit using a particular dataset of interest. Here you can see the histogram at the top of the graph revealing most screening visits were performed two weeks before the start of treatment. You can also see how dispersed the last histogram is, and this makes sense since the last histogram is for the End of Study visit, which can occur any time after dosing.

UTILITY #2: SUBJECT VISIT COUNTS

Subject/Visit Count Report
Subject ID Variable: USUBJID   Visit Number Variable: VISITNUM   Visit Variable: VISIT
Output: subject_visit_counts.xls   Created on 03JUL09 09:29

Visit Number	Visit Text	TV	DS	IE	LB	MH	PE	SV	VS
1	SCREENING	X	69	7	43	36	36	69	36
2	DAY 1 DOSING	X	69		36		36	36	36
3	DAY 2 POST DOSE	X		1	35		36	35	35
9	UNSCHEDULED	X			8		2	8	2
99	TERMINATION/END OF STUDY	X	36		36		36	36	36
	TOTAL		69	8	43	36	36	69	36

Figure 2

The Subject Visit Counts output (see Figure 2) reports the number of subjects found in each dataset, for each visit. Notice that the Trial Visit (TV) dataset has 'X', since the Trial Visit dataset does not include the subject identifier in the dataset. An 'X' denotes that a particular visit is identified within the Trial Visit dataset.

This output can be useful to check that each visit and visit number combination is unique, given that we know that SDTM programmers may need to remap the raw data visit identifiers to ensure uniqueness. For example, a raw dataset may have a visit number of 9 for an unscheduled visit in the Vital Signs data, but the Lab data uses 8 to represent unscheduled visits. The Subject Visit Counts output can be reviewed to identify these cases.

UTILITY #3: SUBJECT GRID

Subject ID Variable: USUBJID

Subjects by Dataset

USUBJID	AE	CM	DM	DS	EX	LB	MH	PE	VS	_Total_
0001-S001				X		X				2
0001-S002	X		X	X	X	X	X	X	X	8
0001-S003	X		X	X	X	X	X	X	X	8
0001-S004	X		X	X	X	X	X	X	X	8
0001-S005				X		X				2
0001-S006				X		X				2
0001-S007				X						1
0001-S008	X		X	X	X	X		X	X	7
0001-S009	X			X		X				3
0001-S010				X						1
0001-S011	X		X	X	X	X	X	X	X	8
0001-S012	X	X	X	X	X	X		X	X	8
0001-S013	X	X	X	X	X	X	X	X	X	9
0001-S014				X						1
0001-S015	X	X	X	X	X	X	X	X	X	9
0001-S016				X		X				2
0001-S017	X		X	X	X	X	X	X	X	8
0001-S018	X		X	X	X	X	X	X	X	8
0001-S019	X		X	X	X	X	X	X	X	8
0001-S020				X		X				2
0001-S021	X			X		X				3
0001-S022	X		X	X	X	X	X	X	X	8
0001-S023	X		X	X	X	X	X	X	X	8
0001-S024	X		X	X	X	X	X	X	X	8
0001-S025				X		X				2
0001-S026	X			X		X				3
0001-S027	X		X	X	X	X	X	X	X	8
0001-S028				X						1
0001-S029				X						1
0001-S030				X		X				2
0001-S031				X		X				2
0001-S032	X		X	X	X	X	X	X	X	8
0001-S033	X		X	X	X	X	X	X	X	8
0001-S034				X						1
0001-S035				X		X				2
0001-S036				X		X				2
0001-S037	X		X	X	X	X	X	X	X	8
0001-S038				X		X				2
0001-S039	X			X						2
Total	22	3	18	39	18	32	16	18	18	39

Figure 3

The Subject Grid output (see Figure 3) reports the existence of subjects in each dataset and the total number of subjects found in each dataset. This report can be visually examined to confirm that every subject that exists in the study is at least represented in the DM dataset. The above output violates this SDTM rule.

CONCLUSION

The point of this presentation is to show how data standards can be leveraged for a variety of purposes. In our case, utilities can be created to find raw data issues such as visit date issues. Moreover, issues with SDTM dataset construction, such as not ensuring unique combinations of visit and visit number combinations, or not including every subject in the demographics dataset.

Moreover, we as an industry, owe it to patients everywhere, to find the most effective means in answering clinical research questions. The development of data standards provides us with a basic framework to do just that. Now it is on all of us to develop and share tools to allow for more informed and faster decision-making.

REFERENCES

CDISC: www.cdisc.org

OpenCDISC: www.opencdisc.org

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

If you are interested in receiving any of these SAS® macros, feel free to contact me. Use these macros at your own risk. These macros may also work on raw datasets, be it CDASH, or your company standard. Also, they can be further enhanced to suit your needs.

Please feel free to contact the author:

Robert Agostinelli
Sepracor, Inc.
84 Waterford Drive
Marlborough, MA 01752-7010

Work Phone: (508) 357-7690
E-mail: robert.agostinelli@sepracor.com
Web: www.sepracor.com