

## OpenCDISC Plus

Annie Guo, ICON Clinical Research, San Francisco, California

### ABSTRACT

OpenCDISC Validator is a collection of tools widely used in the industry to validate clinical trial data in compliance with the CDISC standards, including SDTM, ADaM, SEND, Define.xml, and others. The output from OpenCDISC is an Excel or CSV file, containing warnings, errors and other messages, one row per message per observation and variable(s) violating each rule. That often produces a lengthy report, and it can be laborious to cross-check SAS® data sets to resolve the messages. This paper is focused on SDTM and proposes an enhancement, OpenCDISC Plus. It merges the OpenCDISC report with SDTM data in SAS, and creates Excel files - OpenCDISC Plus SDTM - that are essentially the same as the original report, except for the addition of SDTM data alongside each message. It eliminates the need to access the physical SAS data sets or ad-hoc programming. Instead, simply check the SDTM data listed in the Plus reports for an immediate view of the entire observations that have violated the rules, and pinpoint the source of the messages. Further, we can compare two versions of Plus SDTM reports, and flag the differences between them. That is OpenCDISC Plus Comparison.

### INTRODUCTION

OpenCDISC Validator is a collection of tools widely used in the industry to validate clinical trial data in compliance with the CDISC standards, including SDTM, ADaM, SEND, Define.xml and others. Take SDTM as an example. The validator executes more than two hundreds of rules and generates a report in CSV or Excel format. OpenCDISC report (Table 1) contains messages, one message per rule violated per value(s) per variable(s) per record in each SDTM data set. For any given study there can be thousands or even tens of thousands of messages. Manually reviewing them is time consuming, not to mention that some of the messages can be difficult to interpret or to identify the underlying data issues.

This paper focuses on SDTM. It proposes a programmatic approach to the reviewing process of OpenCDISC report. The goal is to save time and provide direct interpretation by incorporating SDTM data into the report.

1	Name	Record	Variables	Values	Rule ID	Message	Category	Type
4198	CM	934	CMDOSU, CMDOSE	null, 5200	SD0035	Missing units on value	Consistency	Error
4271	CM	1183	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
4372	CM	1638	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
4373	CM	1639	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error

**Table 1. OpenCDISC report for SDTM: one message per Rule violated per Value(s) per Variable(s) per Record per SDTM data set**

### CHALLENGES AND SOLUTIONS

Two challenges with regards to OpenCDISC report are addressed in this section. Each is followed by an example and its proposed solution.

#### CHALLENGE 1 – INSUFFICIENT INFORMATION TO PINPOINT DATA ISSUES

OpenCDISC report shows key variable(s) and value(s), but in some instances they are not sufficient to resolve the matter. For example, in Table 2.1, when there are multiple rows with the same Rule ID “SD0035” and the same message “Missing units on value” pointing to the variables “CMDOSU, CMDOSE” and the values “null, null” in the CM data set, it seems to suggest a systemic problem in the data, but the values “null, null” alone do not give much clue. A workaround is to translate the rule from OpenCDISC to SAS programming, that is, “Dose Units (--DOSU) must be populated, when Dose per Administration (--DOSE) or Dose Description (--DOSTXT) is provided”. Filter CM data set and review the entire records. Alternatively, we can use the Record numbers to merge with the observation numbers in CM data set to get the same result.

The question now is, do we develop a series of SAS programs according to the rules in OpenCDISC, or just do ad-programming as needed when reviewing the report? Either way, what if there is a large number of records with that message “Missing units on value” as in Table 2.1, say, 87, but somehow the output SAS data from programming has 84 records, 3 observations short?

Name	Record	Variables	Values	Rule ID	Message	Category	Type
CM	268	CMDOSU, CMDOSE	null, 1	SD0035	Missing units on value	Consistency	Error
CM	296	CMDOSU, CMDOSE	null, 4500	SD0035	Missing units on value	Consistency	Error
CM	297	CMDOSU, CMDOSE	null, 2000	SD0035	Missing units on value	Consistency	Error
CM	1333	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
CM	1689	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
CM	1691	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
CM	1692	CMDOSU, CMDOSE	null, null	SD0035	Missing units on value	Consistency	Error
CM	1932	CMDOSU, CMDOSE	null, 400	SD0035	Missing units on value	Consistency	Error

**Table 2.1. Challenge 1: Insufficient information to pinpoint data issues**

**SOLUTION TO CHALLENGE 1 – OPENCDISC PLUS SDTM**

Import OpenCDISC report into SAS and merge with SDTM SAS data by Record number from OpenCDISC report and observation number from the SDTM data. The result is one SAS data set per domain merged, containing the same information as the OpenCDISC report, plus corresponding SDTM data alongside each message. Optionally export the SAS data to Excel for review purposes.

This proposed solution is called OpenCDISC Plus SDTM. Table 2.2, Table 2.1 counterpart, is an example. It displays associated SDTM data alongside each message, including CMCAT, CMTRT, and CMDOSTXT. In this example, there are two issues identified. First of all, the records in green mark suggest programming changes to split CMDOSTXT into CMDOSE and CMDOSU to resolve the messages. Secondly, the records in yellow mark, depending on the nature of CMCAT and CMTRT, may need to query the missing dose unit. USUBJID, CMSTDTCT and CMENDTC (not shown in Table 2.2) are available in the OpenCDISC Plus SDTM file to facilitate querying.

Name	Record	Variables	Values	Message	USUBJID	CMTRT	CMCAT	CMDOSE	CMDOSTXT	CMDOSU
CM	268	CMDOSU, CMDOSE	null, 1	Missing units on value	252-01002014	Z-PACK	PRIOR AND CONCOMITANT MEDICATIONS	1		
CM	296	CMDOSU, CMDOSE	null, 4500	Missing units on value	252-01002017	THERAPY	PRIOR RADIATION THERAPY	4500		
CM	297	CMDOSU, CMDOSE	null, 2000	Missing units on value	252-01002017	THERAPY	PRIOR RADIATION THERAPY	2000		
CM	1333	CMDOSU, CMDOSE	null, null	Missing units on value	252-01242001	THERAPY	PRIOR RADIATION THERAPY		.4500CGY	
CM	1689	CMDOSU, CMDOSE	null, null	Missing units on value	252-05032003	CAPECITABIN	PRIOR CANCER THERAPY		(4000 MG)	
CM	1691	CMDOSU, CMDOSE	null, null	Missing units on value	252-05032003	OXALIPLAIN	PRIOR CANCER THERAPY		(260 MG)	
CM	1692	CMDOSU, CMDOSE	null, null	Missing units on value	252-05032003	XELODA	PRIOR CANCER THERAPY		(4000 MG)	
CM	1932	CMDOSU, CMDOSE	null, 400	Missing units on value	252-05042006	SOL. RINGERI	PRIOR AND CONCOMITANT MEDICATIONS	400		

**Table 2.2. OpenCDISC Plus SDTM displaying SDTM data alongside each OpenCDISC message**

**CHALLENGE 2 – LACK OF MECHANISM TO IDENTIFY CHANGES FROM ONE REPORT TO NEXT**

When there are two OpenCDISC reports corresponding to two versions of SDTM data and we have reviewed the old report, theoretically only the new messages in the new report need review. How do we identify those new ones? This can be difficult in particular when there are large numbers of same messages with same variables and values.

For example, in Table 3.1, assume this is a new OpenCDISC report indicating, say, 379 LB records that violate Rule ID SD0047, "--ORRES is null but --STAT or --DRVFL are not specified". There is an old OpenCDISC report indicating 345 records in the old LB data set violating the same rule. Just because we reviewed the old report and disregarded the messages due to ongoing data collection on CRF and central lab does not mean we can assume it holds true for all of the 379 records in the new LB data set. Even though the Record numbers in the reports are uniquely linked to the observation numbers in the two LB data sets, in this instance there is no way to flag changes between the two reports.

Name	Record	Variables	Values	Rule ID	Message	Category	Type
LB	145	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	146	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	321	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	322	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	323	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	465	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	739	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	861	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	1045	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	1430	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	1561	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	1886	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning
LB	1887	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning

**Table 3.1. Large number of LB records, 379, violating Rule ID SD0047**

## SOLUTION TO CHALLENGE 2 – OPENCDISC PLUS COMPARISON

Create OpenCDISC Plus SDTM files for the two OpenCDISC reports. Compare the two files and spit out new records that are in the new file but absent from the old file. Save the result in SAS, and optionally export the data to Excel.

This proposed solution is called OpenCDISC Plus Comparison. Table 3.2 is an example, which is Table 3.1 counterpart. In this example, the OpenCDISC Plus Comparison file identifies 75 out of the total 379 LB records that violates the rule are new ones. Further, while most of the new records indicate lagging issues between CRF data and central lab, others may need queries. For instance, the record in yellow mark in Table 3.2 is missing pregnancy test result. Knowing that pregnancy test comes from CRF only, site needs to either enter the pregnancy test result or mark Not Done on CRF.

Name	Variables	Values	Rule ID	Message	Category	Type	usubjid	Ibtest	Ibcat	visit
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1092003	Laboratory Data	CHEMISTRY	End of Treatment
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1092003	Laboratory Data	COAGULATION	End of Treatment
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1092003	Laboratory Data	HEMATOLOGY	End of Treatment
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1092003	Laboratory Data	URINALYSIS	End of Treatment
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102003	Chorionadotropin Beta	PREGNANCY TEST	Unscheduled
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102003	Laboratory Data	CHEMISTRY	Unscheduled 25.01
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102009	Laboratory Data	HEMATOLOGY	Cycle 4 Day 15
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102014	Laboratory Data	CHEMISTRY	Cycle 2 Day 1
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102014	Laboratory Data	HEMATOLOGY	Cycle 1 Day 15
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102014	Laboratory Data	HEMATOLOGY	Cycle 2 Day 1
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102014	Laboratory Data	HEMATOLOGY	Cycle 2 Day 15
LB	LBSTAT, LBORRES	null, null	SD0047	--ORRES is null but --STAT or --DRVFL are not specified	Consistency	Warning	1102015	Laboratory Data	HEMATOLOGY	Cycle 2 Day 15

Table 3.2. OpenCDISC Plus Comparison identifying 75 new records in LB file violating Rule ID SD0047

## BONUS – OPENCDISC PLUS COMMENTS

If we document review comments as additional column(s) in OpenCDISC report, the comments can be incorporated into OpenCDISC Plus Comparison, and carried over for those matched records between the two OpenCDISC reports. This way we can keep OpenCDISC report up to date with new review comments while automatically accumulating past, but valid, review comments from previous reports.

This is called OpenCDISC Plus Comments, as in Table 4. It rebuilds the entire new OpenCDISC report, plus one or more additional columns as a placeholder to store comments coming from the old OpenCDISC report. In this example, the first few rows are matched records between the new and the old OpenCDISC reports, and therefore the comments from the old OpenCDISC report are appended to the rebuilt new file. The last few rows do not have any comments. That means either they are new records per OpenCDISC Plus Comparison, or if they are matched records they did not have comment in the old OpenCDISC report.

Name	Record	Variables	Values	Rule ID	Message	Category	Severity	Comment
DM	2	ARMCD, RFENDTC	TRT180, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	Okay due to no dosing
DM	7	ARMCD, RFENDTC	TRT120, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	Okay due to no dosing
DM	8	USUBJID, ARMCD	09-02-0021006, PLA	SD0006	No baseline result in VS for subject	Presence	Warning	Okay due to no raw data
DM	8	USUBJID, ARMCD	09-02-0021006, PLA	SD0006	No baseline result in QS for subject	Presence	Warning	Okay due to no raw data
DM	8	USUBJID, ARMCD	09-02-0021006, PLA	SD0006	No baseline result in LB for subject	Presence	Warning	Okay due to no raw data
DM	8	USUBJID, ARMCD	09-02-0021006, PLA	SD0006	No baseline result in EG for subject	Presence	Warning	Okay due to no raw data
DM	8	USUBJID, ARMCD	09-02-0021006, PLA	SD0070	No Exposure record found for subject	Presence	Warning	Okay due to no raw data
DM	8	RFSTDTC, ARMCD	null, PLA	SD0087	RFSTDTC is not provided for a randomized subject	Consistency	Warning	Okay due to no raw data
DM	8	ARMCD, RFENDTC	PLA, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	Okay due to no raw data
DM	12	ARMCD, RFENDTC	TRT180, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	
DM	13	ARMCD, RFENDTC	TRT180, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	
DM	15	ARMCD, RFENDTC	TRT120, null	SD0088	RFENDTC is not provided for a randomized subject	Consistency	Warning	

Table 4. OpenCDISC Plus Comments carrying over past comments from old OpenCDISC report and rebuilding new OpenCDISC report

## IMPLEMENTATION

The OpenCDISC Plus tool is developed based on the above Challenges and Solutions as the user requirements of the tool.

## PROGRAMMING TECHNIQUES

There are three basic programming techniques.

- Base SAS is the programming environment and provides user interface.
- SAS macros are modules of various functionalities to process OpenCDISC reports and SDTM data sets.
- ODS Tagsets.ExcelXP and the program ExcelXP.sas provided by SAS convert SAS data to XML files.

Table 5 is the flowchart of OpenCDISC Plus. There are four up-arrow and down-arrows, showing four arms: Plus SDTM, Comparison Template (“Comp. Templt”), Plus Comparison (“Plus Comp.”), and Plus Comments (“Plus Cmnts”). Each of the four arms starts from the rectangle at the center top. It reads in OpenCDISC report(s) and SDTM data sets as source data, and collects user input in the forms of macro variables and macro calls via the SAS program “user.sas”. The left-right arrow in the middle of the flowchart, with the text “Run user.sas”, separates user interface (above the left-right arrow) from program execution (below the left-right arrow). When the program “user.sas” is submitted, the data coming from OpenCDISC report(s) and SDTM data sets are processed by various macros according to user-defined macro variables and macro calls, and XML or text files are created at the bottom of each arm.

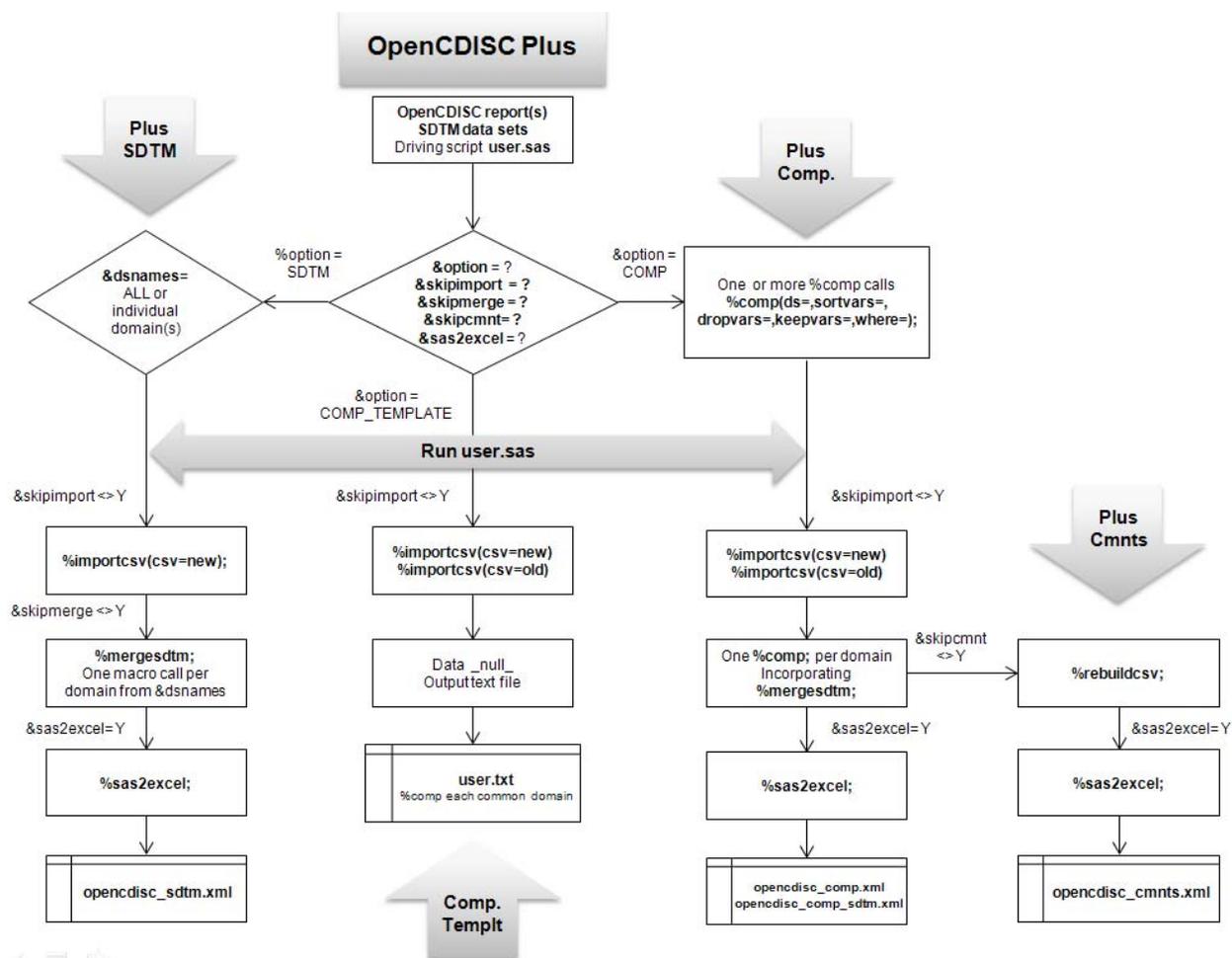


Table 5. OpenCDISC Plus flowchart

## MODULAR DESIGN

The functionalities of OpenCDISC Plus are divided into self-contained SAS macros.

- %importcsv(csv=) is to import OpenCDISC report in CSV file into SAS, where &csv = “New” or “Old” referring to the current New OpenCDISC report, and in the case of OpenCDISC Plus Comparison, the previous Old OpenCDISC report.
- %mergesdtm is to merge OpenCDISC report with SDTM data. This macro is run once per domain.
- %comp(ds=,sortvars=, dropvars=,keepvars=,where=) is to run OpenCDISC Plus Comparison. The local macro variables are to be defined by user prior to submitting user.sas. The first one, &ds, is required and specifies single domain name for the comparison. The other four local macro variable are optional; &where is to subset the

data being compared, &dropvars and &keepvars to keep or drop select SDTM variable(s) from the comparison, and &sortvars to sort output data in this order.

It is advised to intelligently drop unneeded SDTM variables to increase the odds of matched records and thus identify true changes between New and Old OpenCDISC reports.

- %rebuildcsv refers to OpenCDISC Plus Comments. It rebuilds the New OpenCDISC report, with the addition of user-entered past review comments carried over from the Old OpenCDISC report.
- %sas2excel converts SAS data to XML file. It utilizes ODS Tagsets.ExcelXP and the program ExcelXP.sas provided by SAS.

## DATA FLOW

Data flow is controlled by user-define macro variables.

- &option = "SDTM", "COMP\_TEMPLATE", or "COMP" specifies one of the OpenCDISC Plus solutions to run.
- &dsnames = "ALL" or space-delimited domain names, e.g., "AE CM DS", is applicable only when &option = "SDTM". User chooses to create OpenCDISC Plus SDTM file for all applicable domains or select ones.
- &skipimport = "Y" or null is associated with %importcsv, the import process of OpenCDISC report. When user repeats OpenCDISC Plus, and the OpenCDISC report has been saved in SAS the first the tool runs, set this macro variable to "Y" to skip %importcsv and save time.
- &skipmerge = "Y" or null is associated with %mergesdtm, the data merging process of OpenCDISC report and SDTM SAS data sets. This macro variable is applicable only when &option = "SDTM". When user runs OpenCDISC Plus SDTM repeatedly and the merged data sets have been saved in SAS the first time, set &skipmerge = Y to skip %mergesdtm and save time.
- &sas2excel = "Y" or null is associated with %sas2excel, the conversion process from SAS data to XML file. This is made optional to save time because user may not always need the data converted to XML, and thus skip %sas2excel.

## APPLICATIONS

This section presents practical uses of the OpenCDISC Plus tool.

### OPENCDISC PLUS SDTM

OpenCDISC Plus SDTM is the first step when utilizing the tool.

Sample user.sas

```


** Specify prefix to be used in output XML file names;
let studyno=ABCPHARM;

** Define filename for current New OpenCDISC report in CSV file;
filename new "/pub/ABCPHARM/SDTM/reports/new.csv";

** Define libname for SDTM data corresponding to the New OpenCDISC report;
libname newsdtm "/pub/ABCPHARM/SDTM/data/2012-03-15";

** User selects to run OpenCDISC Plus SDTM;
let option = SDTM;

** Include all domains that are present in the New OpenCDISC report;
let dsnames= ALL;

** Run %importcsv and do not skip because this is the first time running
OpenCDISCPlus SDTM for this OpenCDISC report;
let skipimport=;

** Run %mergesdtm and do not skip because this is the first time running
OpenCDISC Plus SDTM for this OpenCDISC report.
Subsequent runs for the same report can skip %mergesdtm by setting &skipmerge=Y;
let skipmerge=;

** &skipcmnt is not applicable when &option = SDTM;
let skipcmnt=;


```

```

*** Macro %comp is not applicable when &option = SDTM.
    User-defined %comp macro calls are collected with %usercomp to be passed all at once;
%macro USERCOMP;
*** %comp(ds=,sortvars=, dropvars=,keepvars=,where=);
%mend;

*** Define libname to store output SAS data sets from OpenCDISC Plus;
libname temp "pub/ABCPHARM/SDTM/reports";

*** User selects to create OpenCDISC Plus SDTM file in XML format;
%let sas2excel=Y;

** Include macros;
%inc "/pub/templates/OpenCDISCPlus_macros.sas";

** Execute macros to run OpenCDISC Plus SDTM according to user input;
%all;

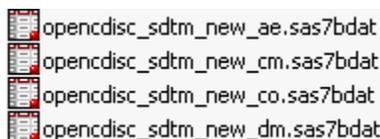
```

After the program user.sas is submitted and macros are run according to user input, output files are created in the libname "outlib". The following are sample output files.

- Output from %importcsv is the New OpenCDISC report in SAS data set.



- Output files from %mergesdtm are merged OpenCDISC-SDTM data sets, one file per domain merged. The following are partial list.



- Output from %sas2excel is a XML file. It contains a TOC tab, i.e., Table of Contents (Table 6.1), followed by multiple worksheets (Table 6.2), one worksheet per SAS data set created by %mergesdtm. The TOC tab has hypertext links, linked to the domain worksheets in the same XML file.

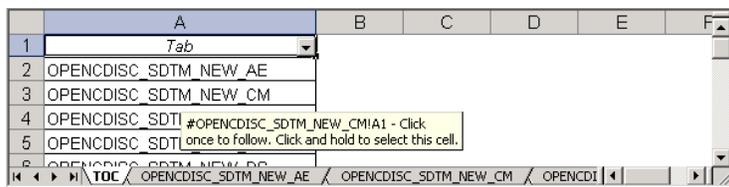


Table 6.1. OpenCDISC Plus SDTM file: hypertext links linked from TOC tab to domain worksheets

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Name	Record	Variables	Values	Rule_ID	Message	Category	Type	STUDYID	DOMAIN	USUBJID	CMSEQ	CMSPID	CMTRT	CMDECO
2	CM	7	CMDOSU	OTHER	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	7	14	VACCINE	VACCINE
3	CM	8	CMDOSU	APPLICABL	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	8	6	LIDEX E	IDE
4	CM	10	CMDOSU	OTHER	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	10	13	MONISTAT	E NITRATE
5	CM	11	CMDOSU	OTHER	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	11	8	OLUX FOAM	OL
6	CM	13	CMDOSU	APPLICABL	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	13	10	CREAM	S AND
7	CM	14	CMDOSU	UNKNOWN	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	14	5	B12	B12 NOS
8	CM	17	CMDOSFRQ	4	CT0018	CMDOSFRQ	Terminology	Warning	ABCSTUDY	CM	1001001	17	2	AVASTIN	AB
9	CM	17	CMDOSU	COLLECTE	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	17	2	AVASTIN	AB
10	CM	18	CMDOSFRQ	4	CT0018	CMDOSFRQ	Terminology	Warning	ABCSTUDY	CM	1001001	18	4	AVASTIN	AB
11	CM	18	CMDOSU	COLLECTE	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	18	4	AVASTIN	AB
12	CM	19	CMDOSU	COLLECTE	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	19	6	AVASTIN	AB
13	CM	20	CMDOSFRQ	1	CT0018	CMDOSFRQ	Terminology	Warning	ABCSTUDY	CM	1001001	20	8	AVASTIN	AB
14	CM	20	CMDOSU	COLLECTE	CT0049	CMDOSU	Terminology	Warning	ABCSTUDY	CM	1001001	20	8	AVASTIN	AB

Table 6.2. OpenCDISC Plus SDTM file: one worksheet per domain merged

## COMPARISON TEMPLATE

The purpose of Comparison Template is to facilitate user-defined %comp macro calls that are required by OpenCDISC Plus Comparison. The output template is a text file named "user.txt" containing one %comp macro call per common domain in both New and Old OpenCDISC reports. Domains that are found in the New OpenCDISC report but not in the Old OpenCDISC report do not run %comp.

Sample user.sas

```


** User selects to create a template for %comp macro calls;
%let option = COMP_TEMPLATE;

** Define filenames for New and Old OpenCDISC reports in CSV files;
filename newcsv '/pub/ABCPHARM/SDTM/reports/new.csv';
filename oldcsv '/pub/ABCPHARM/SDTM/reports/old.csv';

** Define libnames for New and Old SDTM data corresponding to the two OpenCDISC reports;
libname newsdtm '/pub/ABCPHARM/SDTM/2012-03-15';
libname oldsdtm '/pub/ABCPHARM/SDTM/2012-01-10';

** Run %importcsv and do not skip because this is the first time
    running Comparison Template for this pair of New and Old OpenCDISC reports;
%let skipimport=;

** Execute macros to run Comparison Template and create user.txt;
%all;


```

Output files are described as follows.

- Two SAS data sets are created by %importcsv for New and the Old OpenCDISC reports, respectively.

 opencdisc\_new\_csv.sas7bdat  
 opencdisc\_old\_csv.sas7bdat

- The template user.txt consists of one or more partially completed %comp macro calls.

 user.txt

```


** Domains in both new and old OpenCDISC reports **;
%comp(domain=AE, sortvars=name variables values message, dropvars=, keepvars=, where=);
%comp(domain=CM, sortvars=name variables values message, dropvars=, keepvars=, where=);
%comp(domain=CO, sortvars=name variables values message, dropvars=, keepvars=, where=);
%comp(domain=DM, sortvars=name variables values message, dropvars=, keepvars=, where=);
%comp(domain=DS, sortvars=name variables values message, dropvars=, keepvars=, where=);
** BI in new report but not in old report;
** TM in new report but not in old report;


```

## OPENCDISC PLUS COMPARISON

The comparison of New and Old OpenCDISC Plus SDTM files is performed via %comp, one macro call per domain. The local macro variables in each %comp macro call are defined by user. They are inserted into the driving script user.sas, within the macro %usercomp that works as a placeholder.

Sample user.sas

```


** User selects to run Plus Comparison;
%let option = COMP;

** Skip %importcsv because both New and Old OpenCDISC reports are already in SAS
    after having run COMP_TEMPLATE;
%let skipimport=Y;

** One or more %comp macro calls are included within &usercomp;
%macro USERCOMP;
  %comp(domain=LB,
        sortvars=name variables values message usubjid lbtest lbcats visit ,
        dropvars=,
        keepvars=usubjid lbtest lbcats visit lborres lbstat,
        where= %nrstr(rule_id = 'SD0047'));


```

```

%comp(domain_=DM,
      sortvars_=name variables message usubjid,
      dropvars_=invid age dmdtc dmdy,
      keepvars=,
      where= );
%mend;

*** User chooses to skip rebuilding New OpenCDISC report this time;
%let skipcmnt=Y;

*** User selects to create Plus Comparison file in XML format;
%let sas2excel=Y;

** Exceute macros to run Plus Comparison according to user input;
%all;

```

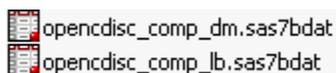
The first %comp macro call is for LB domain where &ds = LB. The local macro variable &where narrows down the comparison by keeping only those records violating Rule ID SD0047, that is "--ORRES is null but --STAT or --DRVFL are not specified". The SDTM variables associated with this Rule are LBTEST, LBCAT, LBORRES and LBTEST that along with USUBJID and VISIT are kept in the data comparison as defined by the local macro variable &keepvars. The output data from the comparison are sorted according to the local macro variable &sortvars. Default sort order is Name, Variables, Values, and Messages from OpenCDISC report. In this example, user adds USUBJID, LBTEST, LBCAT and VISIT to the sorting.

The other local variable &dropvars is the opposite of &keepvars. If user wishes to keep most of the SDTM variables except for a few in the data comparison, &dropvars may be used instead. For example, in the second %comp macro call, for DM domain, the SDTM variables INVID, AGE, DMDTC and DMDY are dropped and the rest variables from DM domain are kept implicitly.

Every time %comp is executed, first of all it runs %mergesdtm to merge OpenCDISC reports and SDTM data. Then it runs Proc Sql with Except and Union options to compare the two merged data sets and output the changes between the two.

Output files from OpenCDISC Comparison include the following.

- Each %comp macro call creates one SAS data set. In the sample user.sas, there are two %comp macro calls, and thus two output SAS data sets, for LB and DM, respectively.



- The above SAS data sets are optionally converted to XML.



Table 7.1 is OpenCDISC Plus Comparison file, one tab per domain compared. Each tab contains the different records from either New or Old OpenCDISC reports. The extra column Source indicates which one of the two OpenCDISC reports the records come from, either New or Old. User can review the data between New and Old and identify changes in the New OpenCDISC report.

	B	C	D	E	F	G	H	I	J	K	N
1	Variables	Values	Rule ID	Message	Category	Type	usubjid	lbtst	lbcat	visit	source
2	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002002	Laboratory Data	HEMATOLOGY	Cycle 2 Day 15	New
3	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002003	Laboratory Data	CHEMISTRY	Unscheduled 5.01	Old
4	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002003	Laboratory Data	HEMATOLOGY	Unscheduled 5.01	Old
5	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	CHEMISTRY	Screening	New
6	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	CHEMISTRY	Unscheduled 27.01	New
7	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	CHEMISTRY	Unscheduled 29.01	New
8	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	CHEMISTRY	Unscheduled 31.01	New
9	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Screening	New
10	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 27.01	New
11	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 28.01	New
12	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 29.01	New
13	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 30.01	New
14	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 31.01	New
15	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	HEMATOLOGY	Unscheduled 31.02	New
16	LBORRES	null, null	SD0047	null but --	Consistency	Warning	ARQ197-A-U252-01002004	Laboratory Data	URINALYSIS	Screening	New

Table 7.1. OpenCDISC Plus Comparison: showing changes between New and Old OpenCDISC reports

- The macro %mergesdtm creates a pair of OpenCDISC Plus SDTM files in SAS for each domain compared. These files take into account the local variables passed from each %comp.



- The above SAS data sets are optionally converted to XML.



Table 7.2 is an example. This file is of the same structure as OpenCDISC Plus SDTM, consisting of the domains processed by user-defined %comp macro calls. It is for look-up and serves as a supplemental to OpenCDISC Plus Comparison file in Table 7.1.

**Table 7.2. OpenCDISC Plus SDTM from new and old reports, supplemental to Plus Comparison report**

### OPENCDISC PLUS COMMENTS

OpenCDISC Plus Comments is a continuation of OpenCDISC Plus Comparison. The driving script user.sas is the same as the one for OpenCDISC Plus Comparison, except for &skipcmnt = "Y".

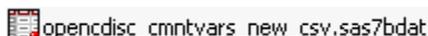
Sample user.sas

```

%** User selects to rebuild New CSV file and include comments coming from Old CSV file.
  Comments are carried over for those match records between New and Old OpenCDISC reports.;
%let skipcmnt=;
    
```

Output files created from %rebuildcsv include the following.

- Rebuilt new OpenCDISC report is saved in SAS.



- The above SAS data set is optionally converted to XML when &sas2excel = "Y".

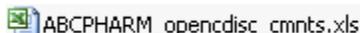


Table 8 is an example of OpenCDISC Plus Comments file. This file is the same as the New OpenCDISC report, with the additional column "comment1". The text on the column "comment1" was past review comments manually entered by user, carried over for those matched records between New and Old OpenCDISC reports.

	A	B	C	D	E	F	G	H	I
1	Var	Repor	Variables	Values	Rule ID	Message	Category	Type	comment1
2	DM	7	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	Data collection ongoing
3	DM	8	RFSTDTCT, ARMCD	null, DRUG/PLA	SD0087	RFSTDTCT cannot be null for randomized subject	Consistency	Warning	
4	DM	8	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	
5	DM	11	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	Data collection ongoing
6	DM	13	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	Data collection ongoing
7	DM	15	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	Data collection ongoing
8	DM	16	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	
9	DM	19	USUBJID, ARMCD	01002016, DRUG/PLA	SD0006	No baseline result in VS for subject	Presence	Warning	Data collection ongoing
10	DM	19	USUBJID, ARMCD	01002016, DRUG/PLA	SD0006	No baseline result in QS for subject	Presence	Warning	Data collection ongoing
11	DM	19	USUBJID, ARMCD	01002016, DRUG/PLA	SD0006	No baseline result in LB for subject	Presence	Warning	Data collection ongoing
12	DM	19	USUBJID, ARMCD	01002016, DRUG/PLA	SD0006	No baseline result in EG for subject	Presence	Warning	Data collection ongoing
13	DM	19	USUBJID, ARMCD	01002016, DRUG/PLA	SD0070	No Exposure record found for subject	Presence	Warning	Data collection ongoing
14	DM	20	RACE	CAUCASIAN	CT0029	terminology codelist	Terminology	Warning	
15	DM	20	ARMCD, RFENDTC	DRUG/PLA, null	SD0088	RFENDTC cannot be null for randomized subject	Consistency	Warning	

**Table 8. OpenCDISC Plus Comments: rebuilt New OpenCDISC report including past review comments**

## **CONCLUSION**

OpenCDISC Plus focuses on SDTM and streamlines the reviewing process of OpenCDISC report by incorporating SDTM data into the report. There are two proposed solutions: OpenCDISC Plus SDTM associates OpenCDISC report with SDTM data so that the entire SDTM records are available at a glance; OpenCDISC Plus Comparison between two OpenCDISC Plus SDTM files help identify changes, and thus avoids repeatedly reviewing the same data issues from one report to the next. Supplemental to OpenCDISC Plus Comparison, OpenCDISC Plus Comments rebuilds OpenCDISC report and keeps track of past review comments.

## **ACKNOWLEDGMENTS**

The author would like to extend a special thanks to Robert Stemplinger for his ongoing support for progress, creativity and innovation.

## **CONTACT INFORMATION**

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

Name: Annie Guo  
Enterprise: ICON Clinical Research  
City, State ZIP: San Francisco, CA, USA  
Work Phone: 215-616-6597  
Fax: 215-240-7595  
E-mail: [annie.guo@iconplc.com](mailto:annie.guo@iconplc.com)  
Web: [www.iconplc.com](http://www.iconplc.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.