PhUSE 2012

Paper CD04

Interpreting ADaM standards with OpenCDISC

Trupti Bal, Cytel Inc, Pune, India
Madhura Paranjape, Cytel Inc, Pune, India

ABSTRACT

Creating CDISC ADaM standard analysis datasets is a challenge. How do we make sure of the compliance for ADaM datasets before submitting it to client or regulatory agencies? Usually a lot of time is spent in performing QC of analysis datasets. Is there an efficient way to handle QC?

OpenCDISC is an open source community which provides tools to check compliance of CDISC standards such as SDTM, ADaM and define.xml. The recent addition is ADaM compliance rules. The rules included in OpenCDISC are very helpful in understanding ADaM standards.

This paper will focus on ADaM validation using OpenCDISC. Understanding OpenCDISC messages is not straight forward at times. The overview of error/warning messages will be included in the paper. Examples from real life studies will be used to explain possible resolutions for the validation messages.

INTRODUCTION

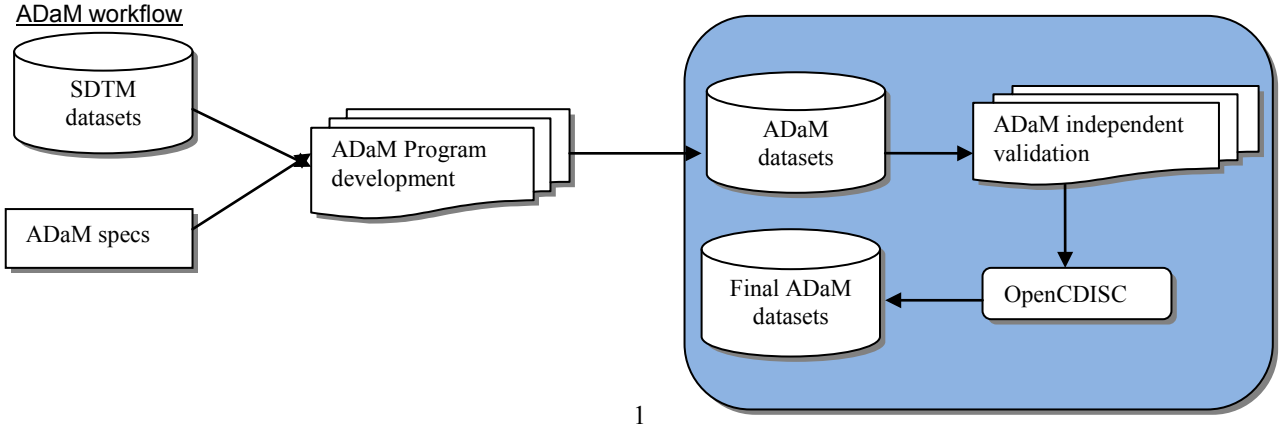
CDISC ADaM standards have evolved over the years and are being adapted by the clinical research industry rapidly. Validation checks for ADaM have been updated very recently. Adapting ADaM is challenging inspite of the guidelines. Ensuring the compliance to guidelines is difficult in situations for programmers where they are pressurized for time before submission or they are new to the industry and standards. Additionally, novelties to standard and stringent timelines impede QC activities to ensure compliance to standard.

OpenCDISC Validator is a popular tool with people in the workflow since it is readily available, easy to use and educational at the same time. The tool was initially developed for SDTM implementation checks. ADaM checks were added in version v1.2.1 in 2011. The latest version OpenCDISC Validator v1.3 has been released in March 2012.

The Validator has 130 rules to check the compliance of datasets with ADaM standards. These are stratified further by category (Consistency, Format, Limit, Metadata, Presence, System and Terminology) and severity (Error, Warning and Notice). The results of the checks performed are compiled in an excel file which has four excel sheets namely Dataset Summary, Issue Summary, Details and Rules

Although the results of OpenCDISC Validator are very useful most of times, in some cases these may not be applicable to the study you are working on. Since the OpenCDISC Validator current version was released before latest validation checks, currently the tool is not in complete sync with the latest checks. So it is better to be careful and decide which errors are applicable to any particular study. In this paper we will discuss some cases that can be seen during regulatory data submissions that are focused on the highlighted portion of the ADaM workflow.

ADaM workflow



The cases have been based on Phase II/ III studies of different study designs. For these studies, SDTM datasets were created and validated as per Implementation Guide. These SDTM datasets were used for creating Analysis Datasets as per ADaM standards. The OpenCDISC Validator version used in each case may be different, since these cases have occurred over a period of time. Hence, some of the errors or warnings may not be seen with the current OpenCDISC Validator version. However, the objective here is to detail the lessons learnt about implementation of ADaM standards, along with understanding of OpenCDISC messages and interpretation of these messages to ensure compliance.

CASE 1

BACKGROUND: In a randomized, parallel trial, planned treatment was given from Visit 1. ECG assessments were performed as shown in the schedule below. The OpenCDISC validator used during this study was v1.2.1.

ECG Assessment and treatment Schedule:

	Screening	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Visit7	Visit8	Visit9
ECG	X	X				X				X
TREATMENT1		X	X	X	X	X	X	X	X	X
TREATMENT2		X	X	X	X	X	X	X	X	X

ANAYSIS ECG DATASET

Missing TRTP

USUBJID	TRTP	PARAMCD	AVAL	VISIT
ABC1201		HR	70	Screening
ABC1201	TRT A	HR	70	Visit 1
ABC1201	TRT A	HR	72	Visit 5
ABC1201	TRT A	HR	71	Visit 9

On running OpenCDISC Validator checks, the following error was seen.

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADEG	AD1008	Null value in variable marked as Required	Error	476

Rule description:

Rule ID	Description	Category	Severity
AD1008	Required variables (where Core attribute is 'Req') cannot be null for any record.	Presence	Error

This is an issue under the category “Presence”. As the error name suggests, OpenCDISC validator checks for the presence of “REQUIRED” variables. This error was issued for variable ‘Planned Treatment’ i.e. TRTP/TRTPN.

In any trial, planned treatment is given after “SCREENING” visit. Ideally there cannot be any valid value that can be entered for “Screening” visit, since there is no planned treatment at this visit.

TRTP is one of the variables that have core attribute as “Required” as per ADaM standard. However records with null values may be allowed since ADaM allows null values for variables with core “Required”. OpenCDISC Validator treated a ‘Required’ variable with the definition similar to SDTM where a “Required” variable cannot have null values.

The problem at hand appears to be with identifying variables that can be treated as “Required” as per SDTM standards. In other words we need to have a clear classification of variables that can or cannot have null values. Variables such as USUBJID, STUDYID, ARM, and PARAM cannot have null values, while TRTP can be left as blank in the right context. We decided to ignore this error. However, what we really need is to classify variables, since this error can be seen for any BDS dataset, including ADEG. This error has been fixed in the current version of OpenCDISC Validator. However, this error was an eye-opener to realize the distinction between core attributes of SDTM and ADaM. It also helped realize that there may be set of ADaM required variables which shouldn’t follow SDTM rule of required and this is acceptable as per ADaM IG.

CASE 2

BACKGROUND: In a study similar to Case 1, ECG measurements were taken as displayed in the schedule. The baseline was defined as the record at “Screening” visit. If the results were missing, the baseline was to be kept as

missing. The assessments were performed only once during "Screening" and hence there was only one record for this visit.

ECG Assessment and treatment Schedule:

	Screening	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Visit7	Visit8	Visit9
ECG	X	X	X	X	X	X	X	X	X	X
TREATMENT1		X	X	X	X	X	X	X	X	X
TREATMENT2		X	X	X	X	X	X	X	X	X

ANAYSIS ECG DATASET

Single Baseline per Subject

USUBJID	PARAMCD	AVISIT	ABLFL	AVAL	BASE	BASETYPE
ABC1201	HR	Screening	Y	69	69	
ABC1201	HR	Visit 1		70	69	
ABC1202	HR	Screening	Y	.	.	
ABC1202	HR	Visit 1		71	.	

Not included

Following error was seen on running OpenCDISC.

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADEG	AD1002	Expected BASETYPE is not present within dataset	Warning	1

Rule description:

Rule ID	Description	Category	Severity
AD1002	Variables described in ADaM as Expected should be included in the dataset.	Presence	Warning

This warning is related to the presence of "EXPECTED" variables. Since the variable "BASETYPE" was not added to the dataset, OpenCDISC issued this warning.

According to ADaM, BASETYPE variable has Core attribute as "CONDITIONAL" and is generated when there is more than one definition of baseline. Since baseline had a single definition and only one record with ABLFL = "Y" was observed, addition of the variable "BASETYPE" was not necessary. The question was why did OpenCDISC warn us against not including this variable? The answer lies in the interpretation of core attributes. As per ADaM a "CONDITIONAL" variable is one that is generated based on the needs of the data. However, OpenCDISC Validator treats all Conditional variables as "EXPECTED" variables as per SDTM standard, where the variable has to be included in the dataset whether or not it has non-null values.

We decided to ignore this warning. BASETYPE is one of the few variables for which we have seen this warning. This situation is study based, and unless we can customize the settings to suit our study we cannot have an entirely clean report.

CASE 3

BACKGROUND: For an Integrated Safety Summary assignment, data from 3 studies with different designs was pooled. The subjects enrolled in each of these studies were distinct. The analysis defined per SAP was study specific, so it was not necessary to map the visit and visit number across all studies. Instead we set AVISIT=VISIT and AVISITN=VISITNUM. The laboratory assessments were performed on visits as shown in the schedule below.

Laboratory Schedule:

Study		Visit1	Visit2	Visit3	Visit4	Visit5	Visit6
111	Screening	Day 1	Week 2	Week 4	Month 2	Month 4	Month 6
112		Day 1	Day 10	Month 1	Month 6	Month 12	Follow UP
113		Day 1	Month 1	Month 2	Month 3	Month 4	Month 5

ISS ANAYSIS LABORATORY DATASET

STUDYID	PARAMCD	VISIT	AVISIT	AVISITN
ABC111	ALT	VISIT 1 (DAY 1)	VISIT 1 (DAY 1)	1
ABC111	ALT	VISIT 2 (WEEK 2)	VISIT 2 (WEEK 2)	2
ABC112	ALT	VISIT 1 (DAY 1)	VISIT 1 (DAY 1)	1
ABC112	ALT	VISIT 2 (DAY 10)	VISIT 2 (DAY 10)	2
ABC113	ALT	VISIT 1 (DAY 1)	VISIT 1 (DAY 1)	1
ABC113	ALT	VISIT 1 (MONTH 1)	VISIT 1 (MONTH 1)	2

AVISIT not consistent across studies

We ran a check on ADLB dataset and the result was as follows:

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADLB	AD0110	Inconsistent value for AVISIT	Error	861

Rule description:

Rule ID	Description	Category	Severity
AD0110	For a given value of Analysis Visit numeric (AVISITN), all values of Analysis Visit (AVISIT) should be the same	Consistency	Error

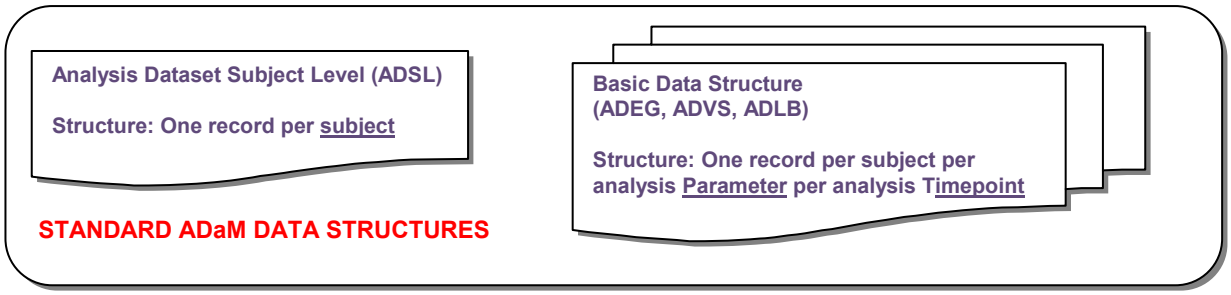
This is the issue under the category “Consistency”. According to consistency principle of ADaM standards, one should have AVISIT (Visit names) and AVISITN (Numeric version of visit names) consistent for a given PARAMCD.

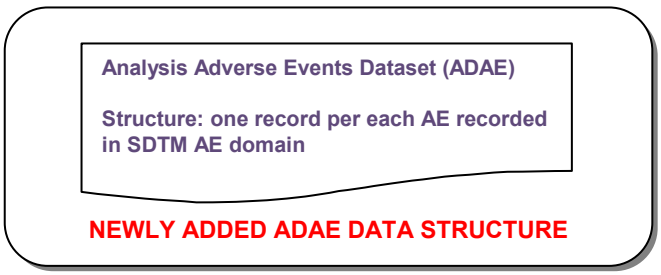
Since the VISIT/VISITNUM was consistent only within a study and not across the 3 studies, OpenCDISC issued an error. Initially, we decided not to add AVISIT /AVISTN in ADLB. Effectively we got an error message from OpenCDISC. As ADaM IG may not strictly cover for ISS/ISE studies, we had thought of having this solution as an easy work around. However, we mapped VISIT/VISITNUM to AVISIT/AVISITN that was consistent across studies and managed to get a cleaner report.

While working on this study we realized a couple of things that need contemplation. Firstly, since ADaM guidelines do not strictly cover ISS/ISE submissions, how much ADaM compliant should datasets be for such submissions? And secondly, is it possible to have a different set of rules for these submissions?

CASE 4

BACKGROUND: For almost all studies safety data is collected and thus it makes Analysis Adverse Events Dataset (ADAE) a must. Since adverse events are crucial for safety data analysis, it is essential to check the compliance of this dataset. The datasets for this study were ADSL, ADEG, ADVS, ADLB, and ADAE. The data structures were as shown in the diagram.





When we passed the datasets through OpenCDISC validator we got the following message for ADAE dataset

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADAE	MISSING_CONFIG	Configuration Unavailable	Warning	1

As the message says this warning is about configuration. We understood that OpenCDISC validator can identify datasets with either BDS or ADSL structure only. Adverse Events dataset does not follow either of these structures and hence is not identified by Validator.

ADSL follows the one record per subject structure. OpenCDISC Validator identifies ADSL dataset from the name and label of the dataset. BDS datasets have one or more records per subject, per analysis parameter; per analysis time-point. There are central variables such as PARAM, and AVAL for a dataset with BDS structure. The BDS structure is identified by OpenCDISC Validator if any one of the central variables i.e. PARAM, PARAMCD, AVAL, AVALC, or ASTDT is present.

The ADAE dataset follows one record per subject per adverse event recorded in SDTM AE domain. The recently released guidelines describe ADAE as a structure similar to SDTM, but with added variables. The current OpenCDISC version does not support this structure yet, but we can hope that it does in the future versions.

CASE 5

BACKGROUND: Analysis Dataset Medical History is generally a part of submissions.

ANAYSIS MEDICAL HISTORY DATASET

USUBJID	MHBODSYS	MHDECOD	MHSTDTC	ASTDT
ABC1201	Psychiatric Disorders	Anxiety	2009-11-11	2009-11-11
ABC1201	Vascular Disorder	Hypertension	2009-11-12	2009-11-12
ABC1201	Psychiatric Disorders	Insomnia	2008-06-10	2008-06-10
ABC1201	Cardiac Disorders	Palpitations	2008-10-09	2008-10-09

Core Variable for BDS structure included in non-BDS dataset

The results of OpenCDISC validator were as follows.

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADMH	AD1005	Neither AVAL nor AVALC are present in dataset	Error	133

Rule description:

Rule ID	Description	Category	Severity
AD1005	At least one analysis value (numeric or character) is not present in BDS	Presence	Error

This is again an error from the category “Presence”. The error is related to the presence of AVAL/AVALC variables. AVAL/AVALC is a required variable for a dataset that follows the BDS structure.

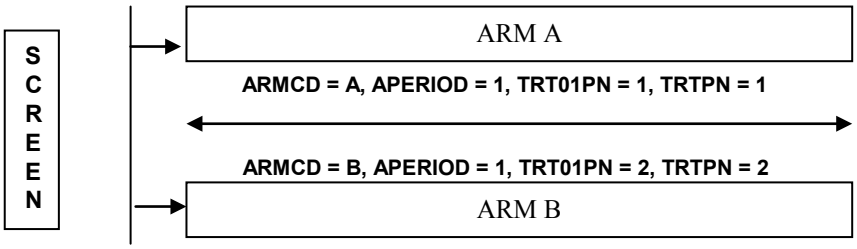
As per SDTM, Medical History, Concomitant Medications and Adverse Events belong to the Events Class and follow similar structure. Ideally, if ADAE does not follow BDS, ADMH or ADCM should not follow BDS. The real question here is why ADMH got categorized as BDS dataset. If OpenCDISC Validator could identify this dataset as BDS, it meant that one of the core variables for BDS was present in the dataset. ASTDT had been generated in the dataset that forced the dataset being treated as BDS.

The most obvious help from this error was to drop the variable ASTDT from ADMH. Secondly, it helped to understand how OpenCDISC detects the dataset structure. We need to confirm the structure of the dataset and the variables that need to be generated before passing through Validator. If variables additional to SDTM are not required, it will be better to use SDTM dataset and not go for ADaM.

CASE 6

BACKGROUND: - In this case we wish to highlight upon the variable TRTP. This case is particularly true when study design is randomized parallel or open label. Figure A is a randomized, parallel trial

FIGURE A: Randomized, parallel study design



On checking datasets, we received the same error for TRTP in each of the BDS datasets. Here error for ADVS is displayed.

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADVS	AD1001	Required TRTP is not present within dataset	Error	40

Rule description:

Rule ID	Description	Category	Severity
AD1001	Variables described in ADaM as Required must be included in the dataset.	Presence	Error

Again, this error is related to the category "Presence". TRTP has been defined as one of the variables which has core attribute as 'Required' and hence needs to be included. In this case, we had decided not include the variable TRTP.

When a study has randomized, parallel design, the treatment within each period is distinct. In an open-label single period study, the planned treatment will always remain same. The reason for not including TRTP is that, we have sufficient variables such as APERIOD and TRTxxP for performing analysis. Hence TRTP is not entirely required for analysis in such scenarios.

In scenarios such as above, TRTP can be populated conditionally in the dataset. One may argue that this is a deviation from ADaM Validation checks. It certainly is. However, it can be addressed in ADaM guidelines. This can be tackled in the OpenCDISC validator too, if it allows user to add input on the ADaM as a framework for checks - e.g. study design, individual study or ISS/ISE.

CASE 7

BACKGROUND: In a placebo controlled study, data was collected for safety and efficacy. The laboratory data was collected as per visit schedule. Generally the values for a laboratory dataset are not imputed. In this study, majority of results were indefinite and thus, hampered the analysis. So the SAP defined imputation for such values as follows:

> **SAP:** For laboratory assessments if majority of results are indefinite, imputation of these values will be considered. Thus if the result has indefinite value, the upper limit, lower limit or median values need to be used in the summary. (for example, 0.1 can be used if the result is <0.1, >0.1, +0.1, or 0.1+)

ANAYSIS DATASET LABORATORY

USUBJID	PARAMCD	VISIT	AVAL	AVALC
ABC01023	GGT	SCREENING	4	>4
ABC01023	GGT	VISIT 1	3	3
ABC01023	GGT	VISIT 3	1	>1

AVAL ^= AVALC

The results of OpenCDISC validator were as follows.

OpenCDISC Result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADLB	AD0149	Inconsistent value for AVALC	Error	579

Rule Description:

Rule ID	Description	Category	Severity
AD0149	For a given value of Analysis Value Numeric (AVAL), all values of Analysis Value Character (AVALC) should be the same for a given PARAMCD	Consistency	Error

The above error is under the category "Consistency" which states that AVAL and AVALC should have one to one relation. In the current study values in AVAL were derived to suit the SAP.

There were differing opinions within the team as to how the rule could be implemented. First, was to keep AVAL as missing for such records. The data at hand had many results in such format, which meant a large amount of data loss. Hence we could not follow this preferred approach. The other option was to drop AVALC altogether. Instead standard character result variable (LBSTRESC) from SDTM could be added. However this meant that another dataset was needed for tests with character result. It was not good for the study to implement the first option and we were too pressurized for time to implement the second option so we submitted the data as is with notes in the define document.

CASE 8

BACKGROUND: In a randomized controlled study, the subjects went through Screening and Run-In Period. Subjects were on control treatment during "Screening" visit and during Run-In period. On successful completion of Run-In period the subjects were assigned to treatment arms. Assessments made during the Run-In period were to be considered for analysis according to SAP. The planned treatments defined in ADSL are as shown in the data below.

ANAYSIS DATASET SUBJECT LEVEL

USUBJID	TRT01PN	TRT02PN	TRT03PN
ABC12301	1	1	2

TRT00PN????

ANAYSIS LABORATORY DATASET

USUBJID	AVISIT	APERIOD
ABC12301	SCREENING	0
ABC12301	RUN-IN	1
ABC12301	PERIOD 1	2
ABC12301	PERIOD 2	3

The results of OpenCDISC validator were as follows.

OpenCDISC Result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADLB	AD0102	BDS APERIOD xx must have corresponding ADSL variable TRTxxP	Error	756

Rule Description:

Rule ID	Description	Category	Severity
AD0102	The analysis period (APERIOD) will be a number like 01. There must be a corresponding planned treatment variable for this period (TRT01P)	Consistency	Error

The above error is under the category "Consistency" which states APERIOD (Analysis Period) must be the same as one of the xx treatment periods. For "Screening" visit we had described the APERIOD as 0, as it is not a real Analysis Period. This forced OpenCDISC to throw the error.

In this study, we had described three planned treatments. According to rule the only possible values for APERIOD were 1, 2 and 3. Although this did not affect the analysis, there was a deviation from ADaM standards. We fixed this by setting APERIOD as blank for such visits.

CASE 9

BACKGROUND: Below dataset is representative of any BDS dataset where Time variable needs to be imputed.

ASTDT	ASTTM	ASTMF
2012-07-07	00:00	Y
2012-07-10	15:32	

Use Codelist: DATEF or TIMEF?

On performing validation using OpenCDISC Validator following message was issued.

OpenCDISC result:

Issue Summary				
Source	Rule ID	Message	Severity	Found
ADTT	AD0040	Variable ending in *TMF is not in TIMEF codelist	Error	3

Rule description:

Rule ID	Description	Category	Severity
AD0040	A variable with a suffix of DTF (Date Imputation Flag) has a value that is not within Controlled Terminology for DATEF (D= Day, M=Month, Y=Year)	Terminology	Error

This is an error related to the use of Controlled Terminology. According to the guidelines, we need to use certain terminologies for date/time imputation flags. We decided to check with the "Rules" sheet for the description of error. On checking we found the description had the following explanation: - "A variable with a suffix of DTF (Date Imputation Flag) has a value that is not within Controlled Terminology for DATEF (D= Day, M=Month, Y=Year)". On cross-checking with ADaM we found out that the Terminology that can be used is H=Hour, M=Minutes, S=Seconds.

So the value "Y" here needs to be changed to "H" since the hour has been imputed. Although the Validator has given the right error, the description is not so apt. This may suggest a change in the Rule description in the Validator.

CONCLUSION

ADaM guidelines are quite different than SDTM, since they allow the choice of structure and variables to be included as per analysis needs. This flexibility makes the guidelines both a boon as well as a bane. It is a boon since the datasets can be suited to analysis as per the study needs. However since most of us work on variety of study designs, implementation/interpretation of guidelines may change as per context. Thus it is it difficult to ensure compliance.

OpenCDISC Validator is an excellent tool that helps in implementing the guidelines to the best possible level. It provides us with quick reviews and helps fix majority of issues. Since the ADaM guidelines are new to the industry, the ADaM validation by OpenCDISC is not as mature as SDTM validation by OpenCDISC. Also, there is a continuous update to the validation checks. Thus, it makes validation through independent programming essential along with the use of the tool.

REFERENCES

1. CDISC ADaM Implementation Guide Version 1.0 CDISC Analysis Data Model Team
2. The Analysis Data Model, Version 2.1 (ADaM 2.1)
3. http://www.opencdisc.org/
4. ADaM validation check v1.2

ACKNOWLEDGMENTS

We would like to thank all our colleagues at Cytel who shared their experiences, patiently answered all our queries, read through our drafts and gave us constructive feedbacks.

CONTACT INFORMATION

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

Trupti Bal trupti.bal@cytel.com
Madhura Paranjape madhura.paranjape@cytel.com

Cytel Statistical Software & Services Pvt. Ltd.
6th Floor, Lohia-Jain IT Park – A Wing, Survey #150, Paud Road, Kothrud,
Pune 411 038, India