

Interpreting CDISC ADaM IG through Users Interpretation

Angelo Tinazzi, Cytel Inc., Geneva, Switzerland

ABSTRACT

It is now almost three years CDISC ADaM is an official standard the pharmaceutical industries should consider in their processes and “eventually” for submission.

However the learning curve has not yet reached the peak (like for SDTM?) and therefore, also because of the nature of ADaM, companies may have applied different interpretation.

The aim of this paper is to have a systematic review of what has been published so far (CDISC, PharmaSUG, PhUSE, Forums, etc) and to evaluate how the different organizations have implemented and interpreted ADaM. Papers/slides have been identified and selected through a search on the web (e.g. Lex Jansen website indexing most of SAS proceedings). The key ‘learning’ from the most interesting “interpretation” will be discussed. The paper will also identify existing gray area; examples are but not limited to:

- How to map clinical domains not yet covered
- Complexity of data structure of developed ADaM datasets. E.g rules for creating derived records
- Use of ADaM principles toward the ‘one-proc-away’ principle.

INTRODUCTION

The document “Statistical Analysis Dataset Model: General Considerations” [1] released in December 2004, represents the first CDISC attempt for defining standards for Analysis Datasets; in this first document some general considerations for the creation, content, and related documentation for statistical analysis were described together with some key aspects in analysis dataset derivation, such as “Change from Baseline Analyses”, “Categorical Data Analyses” and “Subject-Level Analyses”. However it is only in June 2008 with the Draft for Public Comment, then with the final version in December 2009, that a stable version of the Analysis Data Model v2.1 (ADaM v2.1) [2] and the first version of the ADaM Implementation Guidance (ADaM IG v1.0) [3] were released by the CDISC Analysis Data Model Team.

With the later addition of models for Adverse Events (ADAE) [4] and Time-to-Event Analyses (ADTTE) [5], CDISC made available two additional standards for the analysis models not fitting the general BDS structure available in the initial Analysis Data Model.

Whether ADaM is required or not by the FDA, since its launch ADaM capitalized the interest of pharma/CROs; this is proven by the number of experiences shared using different channels (conferences, blogs, etc) and of course by the number of CDISC-type submission FDA-CDER has for example received (SDTM: 23% in 2010 increased to 39% in 2011; ADaM: 32% in 2011).

The ADaM Implementation Guidance (IG) is a well written document and real examples are provided so that concepts are re-enforced. Early in 2013 the updated SDTM/ADaM pilot brings additional useful information on the way ADaM should be implemented. However there are still some gray areas and space for interpretation [6]. Doubts are just around the corner and professionals working with such a model are usually trying to make their own interpretation and eventually verify if the interpretation fits with the reviewer preferences.

All of these public available materials (e.g. articles and slides) could be the right approach/place to find answers and eventually the right inspiration for the correct implementation. Also beginners reading the IG for the first time could take advantage in evaluating what other have already experienced (and understood).

Document / Guidance
Analysis Data Model (ADaM), Version 2.1
Analysis Data Model (ADaM), Implementation Guidance, Version 1.0
CDISC ADaM Validation Checks, Version 1.2
Analysis Data Model (ADaM), Examples in commonly used Statistical Analysis Methods
Analysis Data Model (ADaM), Data Structure for Adverse Event Analysis (ADAE), Version 1.0
The ADaM Basic Data Structure for Time-to-Event Analyses (ADTTE), Version 1.0
IG compliance update of CDISC-FDA Pilot 1

Table 1: Current available CDISC ADaM Documentation

PhUSE 2013

MATERIAL AND METHODS

CDISC papers presented at SAS conferences were selected through Lex Jansen website (lexjansen.com see figure 1):

- SAS Global Forum
- PharmaSUG
- PhUSE including single day events
- Other local area SAS user groups

Whenever available additional information were retrieved from the CDISC interchange conferences (Worldwide, Europe, Local User Groups), discussions/blogs such as CDSIC forum (<http://www.cdisc.org/public-discussion-forum> see figure 1) and LinkedIn CDISC Groups (CDISC, CDISC ADaM, CDISC Advocates, etc. see figure 2).



Figure 1: CDISC Presentations indexed in Lex Jansen website and CDISC Public Forum



Figure 2: Some CDISC related LinkedIn Groups

RESULTS (Quantitative)

CDISC conferences materials (e.g. power-point presentations), except for some local CDISC user groups, were not easy to retrieve, therefore the review performed was mainly based on the Pharmaceutical / SAS User Groups.

Overall 482 papers discussing CDISC topics were found; among them 102 were focusing on the implementation of ADaM principles.

PharmaSUG and PhUSE were the conferences where ADaM was more debated* with respectively 50% and 28% of the presentation made (SAS Global Forum: 12%, NESUG: 4%, Other: 6%); authors were prevalent from CROs with 58% of the presentations (Pharma: 32%, CRO+Pharma: 7%, SAS Institute: 3%).

* PhUSE 2013 data / papers are not yet available

PhUSE 2013

Figure 3 shows also how ADaM become an important and prevalent topic among all other CDISC topics since the delivery of the first Implementation Guidance first become open for public review (2008) then released (2009). However SDTM remains still the most covered topic.

These quantitative results show that, although a considerable increase was seen in the recent main conferences (PhUSE and PharmaSUG), most of the presentations focused on technical rather than methodological aspects. Several ad-hoc solutions, either end-user applications or SAS macros facilitating tasks such as managing metadata, setting automatically variable attributes, validating ADaM datasets, creating define.xml and metadata using OpenCDISC or ad-hoc solution, were presented / discussed.

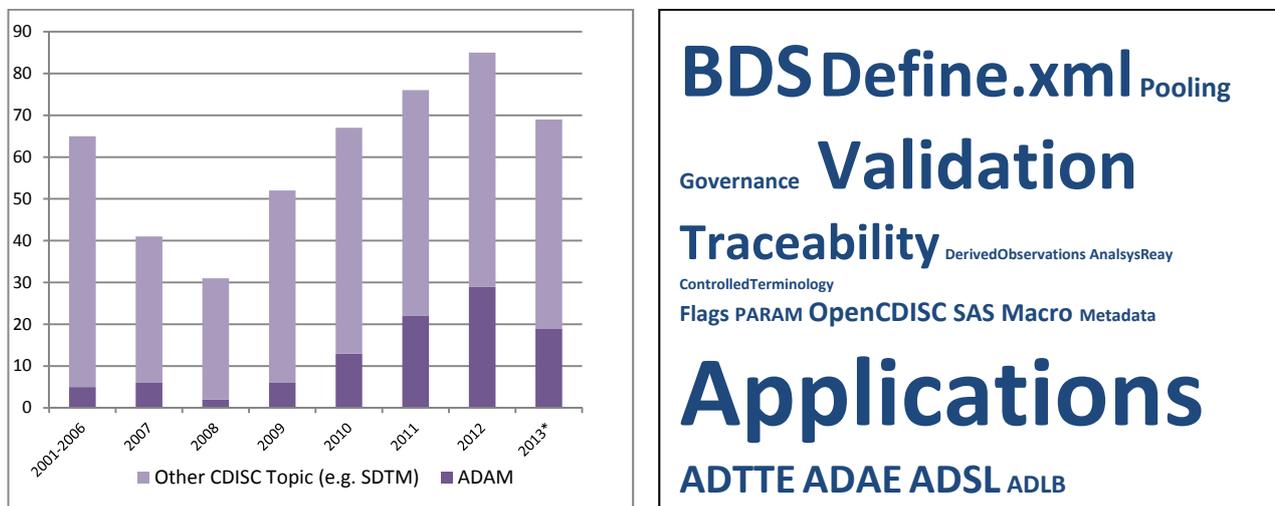


Figure 3: Nr. of papers presented at Pharma / SAS conferences by year, All CDISC related vs ADaM specific (left)
Main covered topics, font size proportional to the nr. of occurrences / nr. of articles covering the topic (right)

RESULTS (Qualitative – Questions and Answers Q&A)

As already discussed in the previous section the key topic of the presentation covering ADaM was often too generic; however a considerable number of presentations provide an in-depth review of methodological issues around the ADaM model. The key learning from the available presentations / papers is summarised here in table 2 and then discussed in more details in the following pages.

Section	Topic
Generic	The obvious rules
	Traceability
	How to map domains not fitting ADSL, BDS, ADAE and ADTTE?
	How much analysis-ready we should be?
	ADAM does not support listings
	CDER SDS Common issues
	Validation
ADSL	One or several subject-level datasets
	TRTxSdt/TRTxEdt vs TRT01P: The Oncology fight for mapping cycles date information
BDS	PARCATy cannot spit PARAM
	Misuse of Indicator variables and Criteria variables
	Deriving Rows or Adding Columns?
	How to populate TRTP and TRTA in BDS
	ADLB and how to represent / classify AVAL
	ADLB and Outputs production
Other	ADAE
	ADTTE
	POOLING: ISS/ISE Strategy
	Implementing ADaM in your organization, Governance and CRO Surveillance
	When validation rules fail and the IG requires a revision

Table 2: Main topics discussed in the selected ADaM papers/presentations

GENERIC

The obvious rules

Some rules do not need to be mentioned and mentioned again, as usually these are the type of rules that can be easily detected with OpenCDISC Validator or any ad-hoc checking tool. However let's mention the basic rules once again

- All analysis datasets name should follow the rule ADXXXXXX (rule valid for any type of analysis dataset)

PhUSE 2013

- II. Analysis datasets and variables name should not exceed the 8 characters length
- III. In addition Analysis Datasets must be SAS export files (.XPT), with the following constraints:
 - a. variable label length maximum 40 length
 - b. character variable length maximum 200 characters
- IV. Consider Analysis Dataset split to reduce the size. For example ADLB split by category, such ADLBH for Hematology, ADLBC for Chemistry, etc.

In addition it is also recommended:

- I. Analysis datasets having a label detailing the main contents
- II. When adding new variables, if ADaM variable names can be not used, use SDTM conventions (e.g. SDTM variables segments)
- III. Add "N" as a suffix to numeric variables created from SDTM character variables

Traceability

[Traceability in the ADaM Standard] [7]

[Derivations and traceability in ADaM: examples] [8]

[Examples of Building Traceability in CDISC ADaM Datasets for FDA Submission] [9]

Traceability is one of the fundamental points in the entire CDISC 'technology' and it is actually synonymous of quality as it makes possible to understand the full data flow from data collection to reporting and this is of course an important milestone in data submission.

Some of the available papers made it possible to understand further the concept of traceability and in particular:

- When it make sense using the triplets SRCDOM, SRCVAR and SRCSEQ (e.g. in ADTTE when especially when dealing with composite endpoints) and when using XXSEQ from SDTM is enough (e.g. when the source of ADaM is a single SDTM dataset)
- Keeping SDTM variables. e.g. LBSTRESC and/or LBORRESC from LB domain without any modification (i.e.same metadata)
- Use of PARAMTYP and DTYPE in a BDS structure when parameter and time-points are derived
- Use of other flags such as ANLxxFL and CRITx, or the occurrence flag in ADAE to select records used in the statistical analysis
- The importance of effective use of metadata (e.g. comment field in the define.xml) to describe the flow when full traceability is not achievable in the ADaM dataset
- The use of intermediate analysis dataset requires complex derivations and/or variable coming from multiple sources (e.g ADSL)

How to map domains not fitting ADSL, BDS, ADAE and ADTTE?

[Common Misunderstanding about ADaM Implementation] [10]

If an analysis dataset does not fit into any existing CDISC models, 'Class of dataset' should be defined as 'Other' and ADaM rules do not need to be necessarily followed. However it would be a good approach to at least follow some basic ADaM principles so that the development of 'Other' derived datasets within your organization follows a standard approach.

An example of possible rules/steps to be followed for 'Other' class dataset could be as follows:

- Keep structure from SDTM (e.g. exposure)
- Copy every key variables from ADSL
- Use SDTM/ADaM principles to create new variables
- Some of the BDS variables can be used e.g. TRTx, CRITx
- Use ADaM principles for creating new observations
- Apply existing ADaM models with data domains of the same "family". E.g. Use ADAE model for standard medical history

How much analysis-ready we should be?

["Analysis ready" - Considerations, Implementations, and Real World Applications] [12]

[Laboratory Analysis Dataset (ADLB): a real-life experience] [13]

[Linkedin ADaM Group Discussion http://www.linkedin.com/groupItem?view=&gid=3092582&type=member&item=245409684&qid=379c7b1b-df19-4772-acf1-3d279ef5b245&trk=group_most_popular-0-b-ttl&goback=%2Egmp_3092582&_mSplash=1]

While traceability is one of the key points in the entire CDISC process, the concept of "analysis-ready" is a key principle in ADaM. As in the examples used in the IG, the most ideal "analysis-ready" dataset is the one "one-proc away", that means an analysis dataset is ready for the use in the output programs, where minimal data processing is performed such as selecting the population to use.

It has also to be clarified that the "Analysis ready" concept does not include the data manipulation steps required to produce optimal output layout (e.g. when for example SAS procedure output is not satisfactory).

The concept of "analysis-ready" therefore, if correctly applied, could reduce the number of steps (and places) where complex data manipulations are performed in the output programs. This implies that in a validated environment mistakes in data derivation if mainly performed in the production of analysis dataset, can be easily detected and with a better accuracy (e.g. the QC programmer pointing out at specific subject/time-point where the two derivation algorithms did not match).

As an example let's consider exposure. Although this is not a data-domain supported by any ADaM models, one may create ADEX by retrieving data from SDTM EX, transposing additional variables stored in SUPPEX (e.g. further details concerning treatment modifications), adding derived variables as a results of simple derivations (e.g. calculation dose as mg/m2 if in the eCRF only total dose by time-point is collected) and getting core variables from ADSL. However this may not be enough to support output production when summary statistics such number of administrations received per subject,

PhUSE 2013

number of administrations not compliant with protocol per subject (e.g. delayed or reduced), etc., need to be described (e.g. number of subjects receiving at least 6 cycles/administration in an oncology study). Deriving this information in the output programs, for example for number of administrations received counting the number of records in exposure with dose not missing, could be avoided and a good approach would be to derive this information in a separate analysis dataset. For example we may create an additional exposure dataset (e.g. ADEXSUM) using a BDS model, having ADEX as input dataset, where each parameter contains a summary statistics for the single subjects. For example see figure 4 from which we can deduct subject 0102-0004 has received overall three administrations (PARAMCD=NADAM), none of them overdosed (PARAMCD=OVERD) over a period of almost 1 months (PARAMCD=TDUR) resulting in a 'Relative Dose Intensity' of more than 120% (PARAMCD=RDI). All the parameters, including the relative dose intensity in categories as specified in the SAP (variable AVALCAT1), where then used in an output program to create the summary table shown in figure 5.

SITEID	SUBJID	PARAMCD	PARAM	AVAL	AVALCAT1
0102	0004	NADM	Nr. of Administrations	3	
0102	0004	OVERD	Nr. of Overdoses	0	
0102	0004	RDI	Relative Dose Intensity (%)	123.52941176	>110%
0102	0004	TDUR	Treatment Duration (months)	1.6755646817	

Figure 4: An example of BDS-like dataset to store summary information by subject (exposure)

Characteristics Statistics	Placebo (N=60)	75 mg q3w + SoC (N=58)	150 mg q3w + SoC (N=60)
Last initiated administration (Number of adm.)			
Min	1	1	1
Q1	4.0	4.0	4.0
Median	6.0	6.0	6.0
Q3	9.0	8.0	10.0
Max	24	25	24
Relative Dose Intensity (%)			
N	0	58	60
Missing	60	0	0
Mean	-	99.01	97.45
SD	-	4.561	7.958
Min	-	85.7	66.7
Median	-	100.00	100.00
Max	-	123.5	123.5
<=75%	N (%) 0	0	3 (5.0)
>75-<=90%	N (%) 0	2 (3.4)	2 (3.3)
>90-<=110%	N (%) 0	55 (94.8)	54 (90.0)
>110%	N (%) 0	1 (1.7)	1 (1.7)

Figure 5: A summary tables using an analysis dataset with the structure shown in figure 4

Some users also introduced the concept of 'lookup analysis dataset' a dataset that contains additional information non-subjects related; for example information about laboratory parameters, such as additional classification driving outputs production. Although this can be a good idea to improve / automatize the outputs production process, I think this is outside the scope of ADaM.

ADaM does not support listings

[An Evaluation of the ADaM Implementation Guide v1.0 and the Analysis Data Model v2.1] [14]

[Considerations for CSR Output Production from ADaM Datasets] [15]

A common criticism is that ADaM is not suitable for creating listings containing derived values. This is true but we should note that ADaM models are mainly built with the submission in mind and as of today analysis data submission should mainly cover key efficacy and safety analysis, thus not all analysis data domains and not all items need to be submitted. However organizations are adopting ADaM not just at the time of submission, but as an official working method, therefore a practical approach has to be defined so that analysis datasets can also support production of regular reports, such as Clinical Study Report or any other internal analysis task.

A practical approach would be to name the additional variables supporting listings production with some naming convention so that they can be easily (programmatically) identified and removed at the time of submission. An example is to start the variable name with a character like '_', and use a SAS statement such as `DROP _;` to drop these variables at the time of analysis dataset generation at the time of submission.

PhUSE 2013

CDER SDS Common issues

[ADaM Implications from the “CDER Data Standards Common Issues” and SDTM Amendment 1 Documents] [16]

“The ADaM document says that we should create ADaM data to support all of our analyses. The CDER document tells sponsors to ‘submit analysis datasets with their application to support key efficacy and safety analyses’, implying that we may not need to submit some of the analysis data we create and use for generating analysis results.

If we’re going to call a submission CDISC-compliant, it should then include SDTM, ADaM ADSL, and the ADaM data that supports our key analyses. Although we might create additional ADaM datasets for non-key analyses, these may not need to be included in a submission to CDER”.

The paper is a good a review of “CDER Common Data Standards Issue Document” [17] where requirements for data submission with either SDTM or ADaM are discussed. Although as a consequence of such a document CDISC had to amend SDTM with the SDTM IG 3.1.2, “nothing stated in the CDER document contradicts what is in the current ADaM IG”. However in the document we can see some good suggestions and re-enforcement of concepts based also on the current experience FDA had we data submission packages received so far:

- USUBJID should be consistent within and between SDTM and ADaM, this mean the variable should be derived the same way everywhere
- Among all possible CDISC implementation approach, the ‘linear’ one is the one that is suggested, which is also the one most suitable for data integration / pooling
- We should not rely on SDTM derived variables and simply cut & paste from it (e.g. study day) and we therefore need to verify if these variables ‘respect’ the SAP requirements
- We need to make sure important no ‘important’ variables are stored in any SUPPQUAL dataset, and if so we should make sure in the mapping process we don’t forget them

Validation

[SDTM, ADaM and define.xml with OpenCDISC] [18]

[Interpreting ADaM standards with OpenCDISC] [19]

OpenCDISC is commonly accepted as the tool to be used for the validation of both SDTM and ADaM datasets and the report provided by the tool is required as part of the submission or used by the sponsors to check the CRO deliverables. In addition to the user manual, several good tutorials/workshops have been run at different conferences proposing real cases of studies/data. The tool is definitively accepted as the gold standard for validating SDTM and ADaM using the “Validation Rules”. However the outputs and the clarity of the error and warning message are still the object of some criticism and that’s why some companies have also developed their own internal tool.

ADSL

One or several subject-level datasets

[CDISC ADaM Application: Does All One-Record-per-Subject Data Belong in ADSL?] [20]

[Designing and Tuning ADaM Datasets] [21]

[ADaM on a Diet: Preventing Wide and Heavy Analysis Datasets] [22]

As clearly stated in the IG section 3.1 “it is not the intent or the purpose of ADSL to take every single data in a study and include them as variable in ADSL”. However both the IG and the FDA with the “CDER Common Data Standard Issue Document”, they state that ADSL should contain various important baseline characteristics, such as demographics, treatment variables, population indicators, trial dates, stratifications and subgrouping variables. So how do we know which variables should be included in ADSL and which not? Because the building of the ADaM is driven by the SAP, by looking at it and at the tables to be produced, you can store in ADSL only variables supporting Demography, Baseline Characteristics and Disposition Tables and create additional Analysis SL dataset for supporting other types of outputs such as in oncology the prior cancer history (e.g. stage, histotype, etc.). Nevertheless information that usually goes to ADSL, such as disposition, may be stored in an additional SL dataset if the analysis requires more details (e.g. when handling reason for discontinuation for multiple treatments and/or multiple trial periods).

In some papers it is also suggested to use BDS structure to store additional SL variables introducing in a way the concept of SUPPQUAL in ADaM; such a model would also improve the traceability which is often an issue in ADSL especially when information are derived from multiple SDTM datasets or as a result of computation from several SDTM records (e.g. the first trial treatment date).

Another interesting proposed approach is also to create a baseline datasets (ADBL or ADBASE) where to store all baseline values from other ADaM (e.g. ADVS, ADLB, etc.).

Whatever approaches you would like to take with ADSL development, “prevent ADSL from being a dumping ground of nice to have variables” and keep ADSL slim.

TRTxxSDT/TRTxxEDT vs TRT01P: The Oncology fight for mapping cycles date information

[Linkedin ADaM Group Discussion http://www.linkedin.com/groupItem?view=&gid=3092582&type=member&item=228552283&qid=2742e47e-d7cd-44a5-9fc6-083901aa2f76&trk=group_items_see_more-0-b-ttl&_mSplash=1]

A repeated questions I was asked where I was part of a sponsor ADaM governance team, was about how to handle oncology cycle information in ADSL. This was also the topic of a Linkedin Group:

Linkedin Discussion Group

(Question)

In the IG, the description for TRxxSDT says "Also useful in designs where multiple periods exist for the same treatment (i.e., multiple cycles of the same study treatment)." We take that to mean that TRxxSDT can store the start date for each cycle. If we have one TRxxSDT variable for each cycle (TR01SDT to TRxxSDT), do we need to have corresponding TRTxxP (TRT01P – TRTxxP) variables in ADSL? That seems redundant because the value of TRTxxP will be the same for all of them. Bear in mind, the statisticians want this to be modeled as a 'one period' study, but want to use TRTxxSDT for the various cycle start dates without all of the corresponding TRTxxP variables. We, the programmers had set up two "periods", one for the loading doses (where only one treatment is given), followed by treatment where both treatments are given, so our TRT01P and TRT02P values are different.

(Some answers provided)

It is true the ADaM IG make reference to cycles, but my suggestion is to keep one period if there is no cross over or if in any case there is not a multiple period treatment where different therapies are used. I also suggest to check with your stats what it is the purpose of having all cycles date in ADSL especially if there is no a fixed nr cycles by protocol. E.g. The subject continue the treatment until for example progression (you may end up with subjects re-cycling 20 or more times).

My suggestion is still to not create such a variable for cycles in ADSL and if you really need cycle date create a separate analysis dataset and use AVISIT in a BDS-like structure if you have to derive the cycle name/number from VISIT or from other variables (eg date boundaries).

Each "xx" used in ADSL refers to a specific period in the study. When using variables with "xx", we must be consistent. In other words, TR03SDT must correspond to TRT03P. Also note that if there is more than just one period, you must also then include the TRTSEQP variable to show how the treatment changed over time.

For ADSL you first need to decide what your analysis period(s) is/are. Often it's a direct match to the ARM information from SDTM DM, but there's no requirement that it must be.

How are the data being analyzed? And by that I mean how are your tables broken down? If the majority of the time you're dealing with a single treatment period, then you might want to use just 01. You can always create an exposure analysis dataset to handle a more detailed analysis of that type of data.

BDS

PARCATy cannot spit PARAM

[Designing and Tuning ADaM Datasets] [21]

[ADaM Implementation Guide Status Update] [6]

The variable PARCATy can only be used to group different parameters together, not to further describe and/or split a parameter into different categories. For example in oncology the tumor response is often evaluated by independent reviewers in addition to the evaluation made by the investigator; in this case, to fulfill ADaM IG and the validation checks, you may need to either create two parameters, one containing the investigator assessment and one containing the independent reviewer assessment, or to store the two assessments into two separate datasets.

Correct use of the PARCATy variable is for example the case where primary and secondary endpoints in ADTTE are grouped together by using the PARCAT1 variable. Another such example would be to group the ECG parameters PR, RR, HR, QT, QTcF, QTcB, and QRS as Centrally collected parameters by using PARCAT1 = 'Central' and differentiate this from the INTp (Interpretation) parameter which has PARCAT1 = 'Local'.

In the end PARCATy can be not used as a qualifier of PARAM and therefore PARAM has not qualifiers.

Misuse of Indicator variables and Criteria variables

[Common Misunderstanding about ADaM Implementation] [10]

[Flags for Facilitating Statistical Analysis Using CDISC Analysis Data Model] [11]

Quite often there is a misuse of variables with clear purpose stated In the IG. This is the case of Indicator Variables (XXFL) when used to classify observation (e.g. classifying a significant abnormal lab values should be done using AVALCATx variable and not an XXFL variable) or criteria variables when used to flag observations (e.g. identifying pre-treatment observation while you should use PREFL variable).

Deriving Rows or Adding Columns?

[Designing and Tuning ADaM Datasets] [21]

[Common Misunderstanding about ADaM Implementation] [10]

[Adding new Rows in the ADaM Basic Data Structure. When and How] [23]

[Derived observations and associated variables in ADaM datasets] [24]

We derive observations in ADaM for two main purposes:

- To impute for missing time-point by applying different imputation methods (e.g. LOCF, MEAN, etc.)
- To create summary observation within time-point and parameter, for example when you have 'triplicate' observations in ECG measurements where you can either derive the MEAN or the MEDIAN

In a derived datasets in addition to records copied from source SDTM, additional records may be also added when you create new parameters as a function of two other (or more) parameters. This process and ADaM requirements are pretty clear and we "only" have to make sure, while satisfying rules for creating new columns vs creating new records as described in IG section 4.2, ADaM dataset does not become a monster of MB or GB.

PhUSE 2013

Where I think instead the IG and its models are a bit vague is for example with DTYPE=ENDPOINT and how to handle situation where “heavy derivations” are required. A classic example is exposure as discussed in the section for “analysis-ready”, where often you need derivations as a function of one or more variables, then as a function of such a new variables from multiple records (and parameters). As previously discussed in this case an approach would be to create a separate analysis dataset.

There are other situations where different approaches can be followed. For example when you have to derive analysis datasets supporting shift-tables where you want to present last value prior to 1st drug administration (baseline) vs worst value on treatment. In this case you can use flags (standard and/or ad-hoc) to identify the baseline observation and the worst on treatment observation for each parameter you want to analyze.

How to populate TRTP and TRTA in BDS

The way the purpose of TRTP and TRTA in BDS is described in the IG may make the ADaM user thinking the variables should be only populated in periods / days where a subject received a treatment. In a recent ADaM mapping for a study requiring the approval of a Sponsor ADaM team, we identified an issue related to follow-up safety data occurring in between two phases of a cross-over study. Because the programmer did not assign any ‘Actual Treatment’ (TRTA) to the observations occurred in between the treatment period, some key safety information were not reported. By reading at ADaM IG section 3.2.2:

TRTA is a record-level identifier that represents the actual treatment attributed to a record for analysis purposes. TRTA indicates how treatment varies by record within a subject and enables analysis of crossover and other multi-period designs. TRTxxA (copied from ADSL) may also be needed for some analysis purposes, and may be useful for traceability and to provide context. TRTA is required when there is an analysis of data as treated and at least one subject has any data associated with a treatment other than the planned treatment.

The words “actual treatment attributed” in the sentence above, make me thinking that we are not obliged to assign the treatment that was received the day when a particular assessment was made (like in the case of the assessments occurred in between the two periods). Instead we have to assign a treatment, planned for TRTP or actual for TRTA, driven by the analysis (as a consequence observation occurred prior to randomization and prior to study treatment start, can have both TRTP and TRTA assigned).

In addition in the latest CDISC pilot, in the analysis of adverse events (ADAE) they assigned TRTA to all AEs regardless if the AE started before the first treatment (non-emergent) or after the last treatment (of course then in the analysis you don’t select AE that are not treatment emergent).

ADLB and how to represent/classify AVAL

[Laboratory Analysis Dataset (ADLB): a real-life experience] [13]

Although it would be good as suggested by several users if ADaM team could revise the rule, PARCATx can be not used as a PARAM classifier. Some users have for example proposed to use PARCATx to classify the PARAM based on the actual contents of AVAL. For laboratory for example we could have AVAL representing the original collected value, the value converted to the SI unit and the normalized value where normalization is applied. While I do not have any solution for the normalized value where also traceability would be difficult to document, my opinion is that AVAL for laboratory should always contain standardized value, that is SDTM LBSTRES, while if really needed because of traceability or because we want to support listing programs through ADaM datasets, additional variables coming from source SDTM should be kept with their original name and attributes, that is for example for original value SDTM.LBORRES.

ADLB and Outputs Production

[Producing Clinical Laboratory Shift Tables From ADaM Data] [25]

[Using the ADaM ADAE Structure for Non-AE Data] [26]

Some users have discussed laboratory analysis datasets from the outputs production point of view. The topic under discussion here is often how to represent the information for the production of shift tables where baseline value is presented against worst on treatment value. While baseline observation can be clearly identified with existing BDS flags, one may decide to select the worst observation directly in the output program, for example by calculating the worst CTACE grade by subject per MedDRA SOC and PT. An alternative can be to use additional “instruments” provided by the IG such as the CRITxx or the ANLxx variables to flag/classify the records to be used in the shift table, or we can derive either new parameters or new records (e.g. creating derived records containing worst observations having DTYPE=MINIMUM and/or DTYPE=MAXIMUM depending on the type of parameter).

An interesting solution for handling for example the presentation of worst derived CTCAE grade [27] is to make use of an ADAE structure having as adverse event any laboratory parameter occurrence of grade > 0 and with term matching the classification in the NCI-CTCAE criteria and make the use of “occurrence” flags to identify worst observations (e.g. AOCCIFL “1st Max Sev./Int. Occurrence Flag”, see ADAE topic for an introduction to “Occurrence” flags).

PhUSE 2013

[LinkedIn ADaM Group Discussion http://www.linkedin.com/groupItem?view=&gid=3092582&type=member&item=113403311&qid=14cc651c-92d7-4aa3-b60b-05afe8f9ebc5&trk=group_items_see_more-0-b-ttl&_mSplash=1]

The topic was also discussed in a linkedin group:

LinkedIn Discussion Group

(Question)

ADaM IG is does not provide much details about Treatment Emergent Lab Abnormalities in BDS. I could think of following 3 options:

- I. Adding a record: It is a parameter function, should an additional record be created? If so, can we leave AVAL missing? If there are multiple post baseline records with same tox grade, which record should one duplicate?
- II. Adding a variable: What variable should capture this? What about "Max PB TXEmerg Tox Grade by subject"?
- III. Create a separate dataset

(Best or Possible Answer)

I would choose option II in your original message i.e. create a variable. You can use ANLzzFL flag variable to flag records in your ADLB dataset that meet the max post baseline TOXGR criteria. Page 34/35 of IG. Regarding "on treatment" piece, you can use ONTRTFL/ONTRTFN variables to identify records that represent information when the subject was on treatment

ADAE

[Hands-On ADaM ADAE Development] [28]

[Using the ADaM ADAE Structure for Non-AE Data] [26]

ADAE was released in May 2012 to answer to one of the gap in the initial set of models and IG, that was the lack of supporting models to 'multiple occurrences'. The model can be applied straightforward by following some 'standard' steps:

- Take all records of AE, and all variables needed for analysis or traceability
- Transpose SUPPAE and merge useful variables
- Merge appropriate variables from ADSL
- Derive any additional variables needed for ADAE

Like the 'parent' domain in SDTM, ADAE will contain the same number of records as in SDTM that is only data about occurred AEs (e.g. subjects with no AE will be not present in ADAE); however an AE may be duplicated in ADAE because of multiple coding path.

Although the ADaM team is working on a more generic model supporting 'multiple occurrences', ADAE is a model that could be already used for other similar data domains such as MH (Medical History) and CM (Concomitant Medications). See also previous discussion ADLB for CTCAE criteria.

ADTTE

[Multiple applications of ADaM Time-to-Event Datasets] [29]

[Creating the Time to Event ADaM Dataset. The Nuts and Bolts] [30]

Together with ADAE, ADTTE was one of the last major ADaM "improvement", although some variables and concepts in the BDS structure already allowed for some ADTTE-like mapping.

The two papers provide a good overview of purpose and main characteristics of the ADTTE model.

Because ADTTE is based on BDS, you may have several parameters representing different 'time-to-event' endpoints in the same Analysis Dataset. Like other type of data you may use PARCATy variable to classify type of TTE endpoints. However a better solution would be to organize time-to-event endpoints with respect to their purpose. For example you may decide to have all the efficacy time-to-event endpoints in an analysis dataset called ADTTEEFF (for example Overall Survival) and all safety time-to-event endpoints in an analysis dataset called ADTTESAF (for example time to first Pain); also, because you may want to clearly identify primary efficacy endpoint, you could decide to store in one dataset only the primary efficacy endpoint all other efficacy endpoints in a separate analysis dataset.

POOLING: ISS/ISE Strategy

[Approaches to Creating ADaM Subject-Level Analysis Datasets (ADSL) for Integrated Analyses] [31]

[ADaM or SDTM? A Comparison of Pooling Strategies for Integrated Analyses in the Age of CDISC] [32]

[ADaM in a Pool! A Concept on how to Create Integrated ADaM Datasets] [33]

[ADaM Implications from the "CDER Data Standards Common Issues" and SDTM Amendment 1 Documents] [16]

[Strategies for Implementing SDTM and ADaM Standards] [34]

This is one of the hot topics and unfortunately the current IG doesn't give enough guidance. Some considerations for pooling strategies have been also introduced in the "CDER SDS Common Issues" previously discussed.

Among the available user experiences one of the key question is whether the pooling should be done in SDTM, "to pool all the single study SDTMs and to derived the analysis datasets from the pooled SDTMs", or/and in ADaM, "to pool the single study analysis datasets and to make all needed derivations during the pooling process".

Pros and Cons have been discussed and it has been conceived the correct strategy depends on the status of CDISC implementation within your organization and of course having a linear process in place where all your studies are mapped to SDTM then ADaM is derived from the SDTM will facilitate your pooling strategy.

[LinkedIn ADaM Group Discussion http://www.linkedin.com/groupItem?view=&gid=3092582&type=member&item=234994040&qid=3176dc3d-5687-4c4a-85ae-1d533ed62d8c&trk=group_items_see_more-0-b-ttl&_mSplash=1]

PhUSE 2013

The topic is also hot in the forum such as the one in LinkedIn where users are asking for example “if there is any guideline in CDISC as what should be going as input datasets into ISS/ISE ADaM (individual study ADaM vs. individual study SDTM datasets)”. As previously said CDISC and the ADaM team has not tackled this topic yet. There is however an ADaM sub-team working on content, but not process; the “Computational Sciences Symposium” has a working group on Traceability that will be touching on this subject (http://www.phusewiki.org/wiki/index.php?title=Traceability_and_Data_Flow).

[CDISC Public Discussion Forum <http://bbs.cdisc.org/bbs/forums/thread-view.asp?tid=3450>]

Related to the pooling topic is the handling of follow-up studies. This was the topic of some discussion in CDISC public forum. The issue the user had to solve in this case while pooling studies of which one was the continuation of the other, was the USUBJID variable duplication in ADSL. The answer is in the CDER Data Standards Common Issues allowing for duplication in ADSL (see previous topics).

IMPLEMENTING ADAM IN YOUR ORGANIZATION, GOVERNANCE AND CRO SURVEILLANCE

[Defining the Governance and Process of Implementing ADaM across an Organization] [35]

[The 5 Biggest Challenges of ADaM] [36]

The above two papers gave some practical idea about to set-up a “Standard Governance Structure”, where processes are defined, teams are formalized, Library created and maintained.

“Beginning the development process of ADaM across your organization can be an overwhelming task. This is even more evident if you try to approach this challenge across the entire organization all at once. The key is to start small and build iteratively different levels of standards and refining the standards and process along the way. Standards development must be a continuously iterative process that is refined along the way”.

This is exactly the same process we implemented for a company I work with. We started small with a team of statisticians and statistical programmers, reviewing the models and IG, organizing training for the rest of the team, let the team starting to use the models and making their interpretation; we later formalized an ADaM team reviewing initial interpretation of study teams, providing advices. Study teams were also allowed to use ADaM principles where SDTM was not yet in place for internal task (e.g. not for submission). Because the ADaM team was also made of statistical programmers working actively on reporting tasks, with the experience gathered and the scenario generated (e.g. with questions from study team), a sponsor ADaM IG and standard models were developed and governed (e.g. peer review of any new domains and/or modifications, granting approval to major deviations to standard); for example clinical domain specific models were developed for a number of BDS-type datasets, such as laboratory, vital signs and ECG.

Las but not least having in place a sponsor-IG clarifying as much as possible any possible misunderstanding and clarifying up-from which way the sponsor like more, will results in having better and compliant ADaM datasets from the involved CRO.

WHAT ELSE?

WHEN VALIDATION RULES FAIL AND THE IG REQUIRES A REVISION

Character Tests with AVAL/AVALC and deriving rows

[Common Misunderstanding about ADaM Implementation] [10]

IG section 3.2.4 requires that either AVAL or AVALC are mapped in a BDS dataset and if both are present there should be a one-to-one relationship; for example if you to map questionnaires with categorical outcome in a BDS-type ADaM dataset, you may want to store both the code (e.g. AVAL=1,2,etc.) and the decode (e.g. AVALC=VERY BAD, BAD, etc.). Now suppose you want to create a score in the same analysis dataset summarizing multiple scores per subject and because ADaM allows it you want to do it in the same ADaM dataset by adding a derived record with the average of all numeric score per subject. The average will obviously contain decimals and only AVAL will be mapped violating the IG rule and generating an error if OpenCDISC will be executed. One may suggest the IG and therefore the check to be changed so that the rule is only valid “on the rows on which both AVAL and AVALC are populated.

SAME NAME, SAME VALUE, SAME METADATA BUT....IT SHOULD BE MEANINGFUL!!!!

“Any ADaM variable with the same name as an SDTM variable is required to be a copy of the SDTM variable, and its label, attributes, and values cannot be modified”. This simple rule from the ADaM IG is often misunderstood or it could be least ‘violated’ in some cases. As an example consider the reason for end of treatment or end of study usually stored in the DS SDTM domain (Disposition), keeping the same SDTM name for the variable containing the reason of the end of study may be not appropriate as the variable DSDECOD contains also information about other disposition event. It would be better to use meaningful name such as ESDREAS.

CONCLUSION

This systematic review of experience in ADaM datasets development shows that there is still space for further development for the CDISC ADaM team. The CDISC ADaM Team has targeted 2014 for the release of IG version 1.2; taking inspiration from what the users has experienced so far it would be a good approach for them to see what needs to be clarified or added. One idea could be to launch a survey among the CDISC users to see what they would like to see in next IG; this is for example the same approach used by SAS Institute since 1976 with the SASware Ballot for the release of new versions / new functionalities (<http://support.sas.com/community/ballot/>).

PhUSE 2013

In the end implementing ADaM at your organization means taking decisions, but it is more important that decisions are 'stable' so that within your organization ADaM mapping will be harmonized as much as possible. Therefore it is suggested to write your own Implementation Guidance where you try to identify areas not clear in the CDISC ADaM IG and then make your own decision on the approach to be followed. Of course in the end if you have the chance to go to the FDA you have to make sure the reviewer is happy with whatever approach you have followed!

REFERENCES

- [1] CDISC, "Statistical Analysis Dataset Model: General Considerations," vol. Version 1.0, 2004.
- [2] CDISC, "Analysis Data Model Version 2.1," December 2009.
- [3] CDISC, "ADaM Implementation Guide, Version 1.0," December 2009.
- [4] CDISC, "Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis," 2012.
- [5] CDISC, "The ADaM Basic Data Structure for Time-to-Event Analyses," 2012.
- [6] JK Troxell, "ADaM Implementation Guide Status Update," in *CDISC UG Atlantic*, 2013.
- [7] E, Lombardi, "Traceability in the ADaM Standard," in *PharmaSUG*, 2013.
- [8] KG Malley, "Derivations and traceability in ADaM: examples," in *CDSIC UG Washington DC*, 2012.
- [9] X. Cui, "Examples of Building Traceability in CDISC ADaM Datasets for FDA Submission," in *SAS Global Forum*, 2012.
- [10] N. Freimark et al, "Common Misunderstanding about ADaM Implementation," in *PharmaSUG*, 2012.
- [11] C, Feng X, Li N, Wang, "Flags for Facilitating Statistical Analysis Using CDISC Analysis Data Model," in *PharmaSUG*, 2013.
- [12] E. Lin, ""Analysis ready" - Considerations, Implementations, and Real World Applications," in *PharmaSUG*, 2012.
- [13] P Dostie, "Laboratory Analysis Dataset (ADLB): a real-life experience," in *CDISC Europe Interchange*, 2013.
- [14] C Price, "An Evaluation of the ADaM Implementation Guide v1.0 and the Analysis Data Model v2.1," in *PhUSE*, 2009.
- [15] A Noller, "Considerations for CSR Output Production from ADaM Datasets," in *PhUSE*, 2012.
- [16] S Minjoe, "ADaM Implications from the "CDER Data Standards Common Issues" and SDTM Amendment 1 Documents," in *PharmaSUG*, 2012.
- [17] FDA, "CDER Common Data Standards Issue Document," 2011.
- [18] A, Ringelberg T, Sherman, "SDTM, ADaM and define.xml with OpenCDISC," in *PharmaSIG*, 2013.
- [19] T Bal, M Paranjape, "Interpreting ADaM standards with OpenCDISC," in *PhUSE*, 2012.
- [20] S Minjoe, "CDISC ADaM Application: Does All One-Record-per-Subject Data Belong in ADSL?," in *PharmaSUG*, 2012.
- [21] S. Zhu, "Designing and Tuning ADaM Datasets," in *PharmaSUG*, 2013.
- [22] D. Van Krunckelsven Dirk, "ADaM on a Diet: Preventing Wide and Heavy Analysis Datasets," in *PhUSE*, 2011.
- [23] M Widel, S Minjoe, "Adding new Rows in the ADaM Basic Data Structure. When and How," in *SAS Global forum*, 2013.
- [24] AR Vidhyadharan, "Deriving observations and associated variables in ADaM datasets," in *PharmaSUG*, 2013.
- [25] R. Bingi, "Producing Clinical Laboratory Shift Tables From ADaM Data," in *PharmaSUG*, 2011.
- [26] M. Widel, S. Minjoe, "Using the ADaM ADAE Structure for Non-AE Data," in *SAS Global Forum*, 2013.
- [27] NCI, "NCI-CTCAE Terminology Criteria Common for Adverse Events Version 4.03".
- [28] S, Minjoe, "Hands-on ADaM ADAE Development," in *PharmaSUG*, 2013.
- [29] HL Chen, H Wang, "Multiple applications of ADaM Time-to-Event Datasets," in *PharmaSUG*, 2012.
- [30] N Brucken, P Slagle, "Creating the Time to Event ADaM Dataset. The Nuts and Bolts," in *PharmaSUG*, 2012.
- [31] T, Cinch N, Freimark, "Approaches to Creating ADaM Subject-Level Analysis Datasets (ADSL) for Integrated Analyses," in *PhUSE*, 2012.
- [32] J Guettner, A Cuza, "ADaM or SDTM? A comparison of Pooling Strategies for Integrated Analyses in the Age of CDISC," in *PhUSE*, 2012.
- [33] A Cuza, "ADaM in a Pool! A Concept on how to Create Integrated ADaM Datasets," in *PhUSE*, 2012.
- [34] S J Kenny, MA Litzsinger, "Strategies for Implementing SDTM and ADaM Standards," in *PharmaSUG*, 2005.
- [35] C. Decker, "Defining the Governance and Process of Implementing ADaM across an Organization," in *PhUSE*, 2011.
- [36] T Peterson et al, "Th 5 Biggest Challenges of ADaM," in *NESUG*, 2010.
- [37] C Teng, M Coughlin, "Outsourced Data Integration Project with CDISC SDTM & ADaM Deliverables," in *NESUG*, 2010.

PhUSE 2013

ACKNOWLEDGMENTS

I would like to thank Emmanuel Abgrall and Dean Shults of the Cytel Geneva offices for their in-depth review and for providing valuable comments and suggestions.

CONTACT INFORMATION

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

Angelo Tinazzi

Cytel Inc.

Route de Prè-Bois 20

1215 Geneva – Switzerland

+41 765359946

Email: angelo.tinazzi@cytel.com

Web: www.cytel.com

Brand and product names are trademarks of their respective companies