**PhUSE 2013**

**Paper CD08**

# ADaM datasets

# - Standard and submission ready

Karin Fleischer Steffensen, Ferring Pharmaceuticals, Copenhagen, Denmark

## 1.0 ABSTRACT

The CDISC ADaM Team has defined standards for ADaM dataset structures as ADSL, BDS, TTE and Adverse Event - in theory! Unfortunately it's not that easy in practice. There are some gaps, unclear definitions and other challenges for programmers trying to create Vanilla ADaM datasets to support trial reporting (and as internally company standard).

Should we try to create analysis datasets, that follow the ADaM Model completely or as standardised as possible? We experience that definitions are lacking, examples of more complex endpoints and trial designs (e.g. pooled studies, extension- or crossover studies) are needed. In our company we have met these challenges and created our own ADaM Model interpretation, which refers to the most updated ADaM guideline and apply to ADaM fundamental principles, but with some "tailor-made" variables and designs. But with minor changes it is submission ready.

This paper will raise questions to the ADaMIG and show examples of how we have solved some of the issues.

## 2.0 INTRODUCTION

In 2009 Ferring implemented an analysis dataset standard based on at that time current ADaM Model. This was done with some struggles due to the early stage of the ADaM Model. Today the ADaM Model documentation is much more mature and will continue to develop in the future. This means that it is now easier to build standard upon the ADaM Model without violating it. Standardization and horizontal data structure often do not go hand in hand. Here the BDS (Basic Data Structure) analysis datasets with vertical structure are perfect for that purpose. Also the vertical structure in SDTM is suitable for standardization.

At Ferring we prioritize to build standardized analysis datasets to support regular trial reporting and repositories, and then secondly creating analysis data as close to submission standards as possible. When upgrading our company ADaM Model version from 2009, we realized that definition and/or explanations of some important issues were either lacking or unclear, and we were challenge to solve it while trying to adhere to the principles discussed in the CDISC ADaM Model Version 2.1[1] and CDISC ADaMIG v.1.0[2] documents.

### 2.1 ANALYSIS DATASETS BUILD FOR OUTPUT

A large part of the analysis datasets we create within a trial can be standardized with respect to structure and a large core of variables. When specifying the ADaM datasets we have the final output or displays in mind. This means that we actually start at the end by specifying standard output. Having standardized output makes it much easier to build standard analysis datasets to support the standard code.

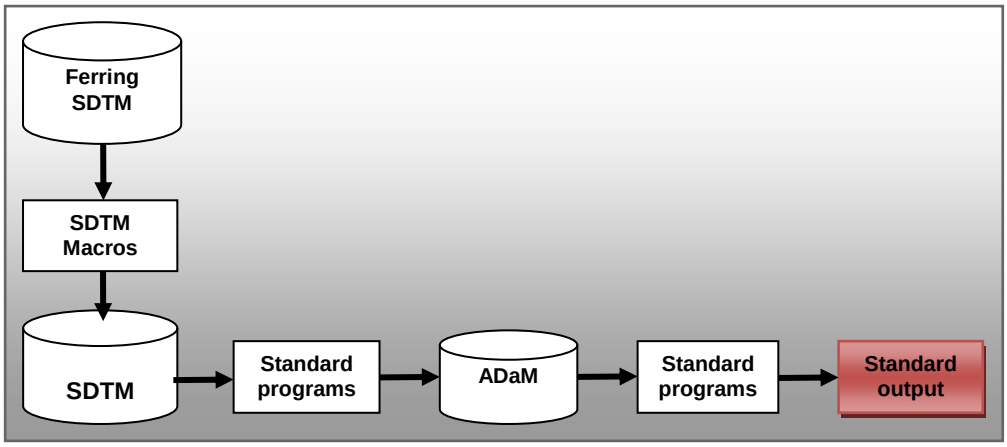



Fig. 1

One of the primary reasons for upgrading the ADaM version was due to implementation of new procedures at Ferring. We have built submission ready SDTM data from eCRF, where some derivations, earlier done in ADaM, are moved back into SDTM, e.g. baseline flags, --STRESN, and TRTP. This would lead to a violation of some ADaM rules if we did not update our current ADaM version.

## 3.0 PROBLEMS AND SOLUTIONS

In the following some examples of issues in the ADaM Model/ADaMIG and how to solve them while still adhering to the ADaM principles will be discussed. Some easy solutions are mentioned, but also more difficult ones where the solution is not straight forward.

### 3.1 AREL, ASEV, ASER

According to the CDISC ADaM Model for Adverse Events[3] some of the variables carried forward from the AE SDTM domain have corresponding analysis variables, which are not required, but permissible. Since we are using these variables directly in output/displays, they are always present in our ADAE. In the ADaM guideline they are present if an imputation is done, but we use them not only for imputation, but also for changing the text from uppercase to sentence case or correct the text to match what is specified for our standard output.

### 3.2 AETHNIC, ARACE, ASEX

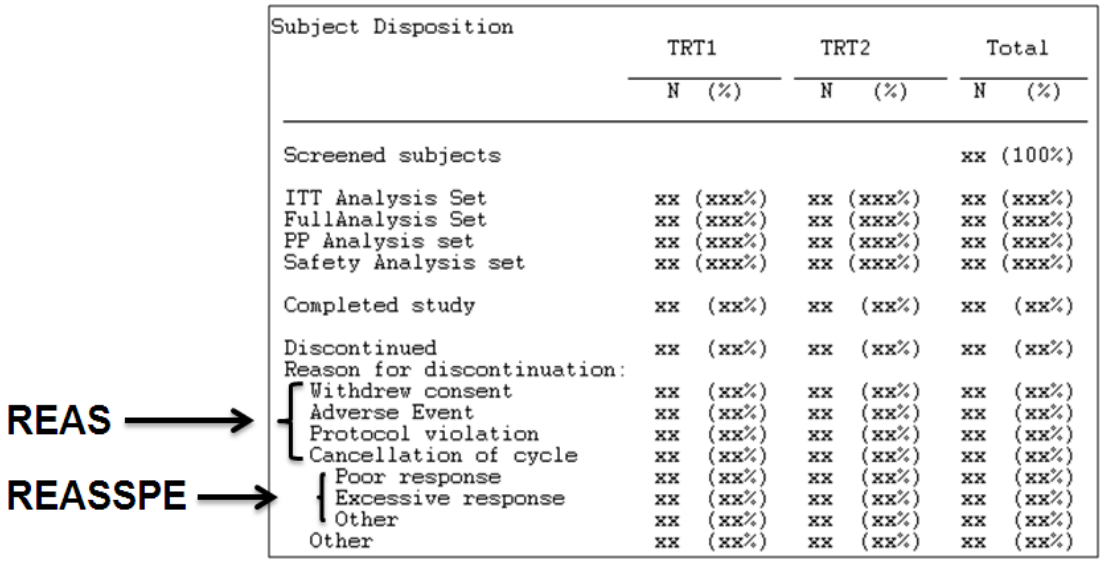
We have chosen to define some additional demographic variables as permanent in our ADSL. It is not a violation of the model and guideline to add these variables, which are sponsor-defined and follow the standard variable names (the 'A' in front indicates an analysis variable) and naming conventions defined in the ADaMIG. The reason for including these variables is, as with the AE variables, to be able to use the content directly in our standard demographics table.

| Demographics | TRT 1 |
|---|---|
| Age (years), N (%) | |
| Female | 20 (48%) |
| Male | 22 (52%) |
| Total | 42 (100%) |
| Ethnicity, N (%) | |
| Hispanic or latino | 2 (5%) |
| Not hispanic or latino | 40 (95%) |
| Total | 42 (100%) |
| Race, N (%) | |
| American indian/alaska native | |
| Asian | |
| Black or african american | 1 (2%) |
| White | 41 (98%) |
| Total | 42 (100%) |

Fig. 2

### 3.3 REASON FOR DISCONTINUATION

A disposition table is always present in the clinical trial report, containing information of screened, randomized, discontinued etc. Here we find that a definition of reason for discontinuation and defined variables are lacking in the ADaM guideline/model. We have therefore defined our own variables (REAS and REASSPE) which are used in the standard code for the standard output for the disposition table (see figure 3). The information is found in the DS domain in SDTM and the variables are placed in our ADSL. The content of the variable, e.g. the reason for discontinuation, is not standard, but the names of the variables are standardized.




| Subject Disposition | TRT1 | TRT2 | Total |
|---|---|---|---|
| | N (%) | N (%) | N (%) |
| Screened subjects | | | xx (100%) |
| ITT Analysis Set | xx (xxx%) | xx (xxx%) | xx (xxx%) |
| FullAnalysis Set | xx (xxx%) | xx (xxx%) | xx (xxx%) |
| PP Analysis set | xx (xxx%) | xx (xxx%) | xx (xxx%) |
| Safety Analysis set | xx (xxx%) | xx (xxx%) | xx (xxx%) |
| Completed study | xx (xx%) | xx (xx%) | xx (xx%) |
| Discontinued | xx (xx%) | xx (xx%) | xx (xx%) |
| Reason for discontinuation: | | | |
| Withdrew consent | xx (xx%) | xx (xx%) | xx (xx%) |
| Adverse Event | xx (xx%) | xx (xx%) | xx (xx%) |
| Protocol violation | xx (xx%) | xx (xx%) | xx (xx%) |
| Cancellation of cycle | xx (xx%) | xx (xx%) | xx (xx%) |
| Poor response | xx (xx%) | xx (xx%) | xx (xx%) |
| Excessive response | xx (xx%) | xx (xx%) | xx (xx%) |
| Other | xx (xx%) | xx (xx%) | xx (xx%) |
| Other | xx (xx%) | xx (xx%) | xx (xx%) |

Fig. 3

### 3.4 EVALUATION AND CLINICAL SIGNIFICANCE

When we want to create standard shift tables for laboratory parameters e.g. with change from normal reference ranges, we use ANRIND, BNRIND and SHIFT1 (See figure 4) as recommended in the ADaMIG. And SHIFT1 contains exactly what we want to show in the output, e.g. Low -> High, or Low -> Normal etc. See figure 5 below. So here it is possible to create a standard which is ADaM compliant and ready for a submission.

| SUBJID | PARAMCD | AVAL | ANRLO | ANRHI | BNRIND | ANRIND | SHIFT1 |
|---|---|---|---|---|---|---|---|
| XXX1 | ALP | 28 | 35 | 105 | Normal | Low | Normal -> Low |

Fig. 4

| Test / shift | TRT 1 N | n (%) |
|---|---|---|
| Alanine Aminotransferase (IU/L) | | |
| Normal -> High | 45 | 1 (2%) |
| Normal -> Low | 45 | |
| Low -> High | 45 | |
| Low -> Normal | 45 | 1 (2%) |
| High -> Low | 44 | |
| High -> Normal | 44 | 1 (2%) |

Fig. 5

But when we want to create a standard shift table for markedly abnormal changes and clinical significance, we do not find any information described in either the ADaM Model or ADaMIG. Is it a criterion, a flag, a categorization of AVAL or something entirely different?
The only place where this is mentioned is in a paper from PharmaSUG 2012[4]. Here it is not recommended to define the clinical significance assessment as a flag or an indicator variable, but instead as a categorization of AVAL, e.g. AVALCATy.
According to the ADaMIG Table 3.2.4.1 Analysis Parameter Variables for BDS Datasets AVALCATy is a categorization of AVAL, not necessary a one-to-one mapping of AVAL. But what if you have a different range for men and women, which means that AVAL = 9.5 for women is in the category AVALCAT1 = "Normal", but for men AVALCAT1 = "Abnormal, NCS". This means that you have a different AVALCAT1 for the same value of AVAL. In the ADaM Validation Checks[5] check number 221 states that: "AVALCATy must have the same value for all records within a parameter for a given value of AVAL", which indicates that AVALCATy is not suited for this purpose. So in our Ferring standard we choose to create our own variable AEVL. (See figure 6)

| SUBJID | PARAMCD | SEX | AVAL | BASE | BEVL | AEVL | SHIFTy |
|---|---|---|---|---|---|---|---|
| **XXX1** | **PLAT** | **F** | **9.5** | **12.1** | **Abnormal NCS** | **Abnormal CS** | **Abnormal NCS –> Abnormal CS** |
| **XXX2** | **PLAT** | **M** | **9.5** | **15.9** | **Normal** | **Abnormal NCS** | **Normal –> Abnormal NCS** |
| **XXX3** | **PLAT** | **M** | **14.7** | **13.8** | **Abnormal NCS** | **Normal** | **Abnormal NCS -> Normal** |
| **…** | | | | | | | |

Fig. 6

Since we create a (standard) shift table from the ADaM dataset, we need both a variable containing the baseline value, and a variable containing the values which is used when looking at the change from baseline. So we choose also to create a variable BEVL containing the evaluation at baseline. (See fig.6)

#### 3.5 AVALCATyN

Another question is raised when investigating AVALCATy. For many character variables a related numeric variable is defined (one-to-one) in the ADaMIG. But for AVALCATy and also BASECATy this is not the case. There are no numeric variables connected to these two in ADaMIG. Maybe it is because the variable names are already the permitted 8 characters long. But that is exactly why it is much more important to define the numeric variables in the ADaMIG, instead of we all invent a name for this numeric variable which probably will NOT be the same across companies/standards. To apply to the "rule" that if a variable converted into numeric, is already an 8-character variable name, the last character may be removed prior to appending the "N", then it must be AVALCAyN or? The solution we found for this is to allow a length of 9, e.g. AVALCATyN, and remove the variable from submission data. The variables must then only be used for sorting and NOT sub setting!

#### 3.6 RELATIVE TIME VARIABLES AND END OF TRIAL DATE

In our analysis datasets we need duration variables. *DY is a definition of durations but can't contain 0, which we want for use in tabulations! So we need to define other variables that can be used for this

purpose. From the ADaMIG we choose ARELTM, the time relative to an anchor time. But we also need an analysis relative START and END time. This is not specified precisely in the ADaMIG, so we have decided to create our own based on rules specified in the ADaMIG Section 3 Standard ADaM Variables, "General Timing Variable Conventions". It is rule no. 9: "Names of timing start variables end with an S followed by the two characters indicating the type of timing (e.g. SDT, STM)", and rule no. 10 equal to no. 9 but for timing end variables, e.g. ARELSTM (Analysis Relative Start Time) and ARELSTM (Analysis Relative End Time).
Another time variable, not regarding duration, but time/day of "event" is the "End of trial" date. There are always confusions regarding when the end of trial (EOT) occur. Is it the last day of exposure (+ a window?), the last visit in the study or the date of the last measurement available? It would make life so much easier for the programmers, if an EOT date was defined in ADSL just as TRTSDT and TRTEDT are.

#### 3.7 OpenCDISC ADaM Validator

The OpenCDISC ADaM Validator[6] is a free tool that can be downloaded from www.opencdisc.org. It is a very good tool for checking, that the analysis datasets are in compliance with the guidelines and hence ready for a submission. The OpenCDISC ADaM Validator is based on the validation checks published from the CDISC ADaM team.
We tried it on our analysis datasets and here we got some expected errors, but also an unexpected error. It was regarding a rule that we have not discovered and did not find clearly described in the ADaMIG. It is the relationship between PARCATy and PARAMN (and indirectly PARAMCD/PARAM). In figure 7 there is a one-to-one mapping between PARAM, PARAMCD and PARCAT1. And in each PARCAT1 category, there is a one-to-one mapping also with PARAM, PARAMCD and PARAMN, but only within each PARCAT1. And this one-to-one mapping is known from the ADaMIG. According to the OpenCDISC ADaM Validator an error occur, since PARAMN = 1 for two different PARAMCD/PARAM even though it is for two different PARCAT1N values (1 and 2). So the right way is to define PARAMN = (3, 4, 5, 6, 7) where PARCAT1N = 2. And this means that there has to be a one-to-one mapping between all the variables PARCATy/PARCATyN, PARAMCD, PARAM and PARAMN.

| PARCAT1 | PARCAT1N | PARAMCD | PARAM | PARAMN |
|---|---|---|---|---|
| Infertility history | 1 | DUR | Duration of infertility (months) | 1 |
| Infertility history | 1 | PREASON | Primary reason for infertility | 2 |
| Reason for infertility | 2 | UNXINF | Unexplained infertility | 1 |
| Reason for infertility | 2 | TUBALINF | Tubal infertility | 2 |
| Reason for infertility | 2 | MIMF | Mild male factor | 3 |
| Reason for infertility | 2 | MOMF | Moderate male factor | 4 |
| Reason for infertility | 2 | SEMF | Severe male factor | 5 |

Fig. 7

Finally we also had some errors due to mismatch of label or length of label. Some we found and corrected but others we simply could not see! E.g. label for TRTSDT "Date of First Exposure to Treatment". Where is the mismatch compared to the ADaMIG? I assume that this is a bug in the OpenCDISC ADaM Validator?

The OpenCDISC ADaM Validator creates a log with all errors and warning including details about these findings. Sometimes there are hundreds of records, but it is rather easy to select only the most essential findings and information violating the ADaM model. Of course it will not replace the validation process, but it is a good tool which is very easy to use as an extra check.

## 4.0 CONCLUSION

Most important to Ferring is to have one common analysis dataset standard that is effective for trial reporting across all trials and which can easily be upgraded to be submission ready i.e. ADaM compliant. In Ferring we have built a standard with some minor deviations/additions to the formal standard. However with a few additional steps (e.g. generating transport files, truncate variables and add/remove few variables) it can be made submission ready. One could question if we are following the ADaM Model and guideline or if we have a set of analysis data sets that are ADaM-like?

Today we have built a much more robust standard especially due to the more vertical structure in ADaM e.g. BDS datasets but also because ADaM is much more detailed today. This means that is easier to enforce and maintain the standards. The ADaM Model/IG is created for submissions for a single trial, which means that of course there is a lack of definitions and examples in the guideline when trying to implement standards from it. But as mentioned here in this paper, the issues we are dealing with are not crucial. However to make sure that we all call the variables the same (with the same meaning or definition) there is a need for specifying all "necessary and relevant" variables no matter if they are complicated or not.
So did we build true ADaM datasets, meaning ready for submission, or did we build ADaM-like datasets suitable for company standard? In my opinion we have done both and hereby proved that it is possible to build a company standard, which is very close to submission ready.

## 5.0 REFERENCES

1. CDISC ADaM Model Version 2.1
2. CDISC ADaM Implementation Guide Version 1.0
3. CDISC ADaM Data Structure for Adverse Event Analysis Version 1.0
4. "Common Misunderstandings about ADaM Implementation" by Nate Freimark, Susan Kenny, Jack Shostak and John Troxell
5. CDISC ADaM Validation Checks Version 1.2
6. OpenCDISC ADaM Validator

## 6.0 ACKNOWLEDGEMENTS

It can be very difficult to define variables and analysis datasets which have to follow the ADaM Model strictly. And it does not make the task easier, when you want ***both*** to build an ADaM standard ***and*** compliant and submission ready ADaM datasets. So we would never have come that far without these discussions about interpretation due to different views on things depending on the different therapeutic areas and experiences. So I want to thank all of my colleagues in Ferring Global Biometrics, for their patience, commitment and contribution to finding useful solutions.

## 7.0 CONTACT INFORMATION

Karin Fleischer Steffensen
Senior Statistical Programmer
Global Biometrics
International PharmaScience Center
Ferring Pharmaceutical A/S
Kay Fiskers Plads 11
2300 Copenhagen S, Denmark
Phone: +45 28 78 78 39
Fax: +45 28 17 68 39
Email: kast@ferring.com