Paper CD13 – PhUSE 2009

An Implementation of ADaM standards NOT driven by a submission

Karin Fleischer Steffensen, Ferring Pharmaceuticals, Copenhagen, Denmark
Gitte Dam Jepsen, Ferring Pharmaceuticals, Copenhagen, Denmark

1.0 ABSTRACT

The Biometrics department has implemented CDISC SDTM over a period of two years. The logical progression was then to implement ADaM (Analysis Data Model).
The implementation of data standards should not only be driven by a submission. The main reason for implementing ADaM at Ferring were to have standardized analysis datasets to support the regular trial reporting and secondly to be as close to submission standards as possible.

To successfully conduct this implementation it was of upmost importance to:

- Have sufficient resources allocated
- Have support from managers and colleagues
- Brainstorm with colleagues and discuss/interpret the ADaM implementation
- Conduct a pilot study in parallel and not *after* finalization of standards.

Extra benefits included a revision of routines, streamlining procedures and general cleanup, but the implementation has also created new tasks to be done.

Furthermore this paper will detail how Ferring has defined the analysis dataset ADSL.

2.0 INTRODUCTION

The main reason for implementing ADaM at Ferring were to have standardized analysis datasets to support regular trial reporting, and for the standards to facilitate cross-trial analysis, aside from having analysis data as close to submission standards as possible.

Some standardization on analysis datasets has been conducted over the past years in the Statistical Programming group at Ferring Biometrics, but due to a lack of resources no industry standards were taken into consideration.
With the implementation of SDTM and also moving forward with standard CRFs, it was time to initiate the standardization process again. It was decided to go the ADaM way in respect to analysis data – starting with two of the metadata components described in CDISC Analysis Data Model version 2.1: Analysis dataset metadata and Analysis variable metadata. In our case the metadata is not in a submission ready format yet, but we have an analysis datasets specification in a word document which fully describes the variables within the analysis datasets.
Due to many different trial designs across our therapeutic areas and issues related to the conduct of the trials, it is not possible to create a complete specification of all analysis datasets. With the company standards in mind – and not a submission – we have created a “F-ADaM” (Ferring Analysis Data Model set of specifications) which is a maximum of what can be standardized for any trial across all our therapeutic areas – the common denominators.

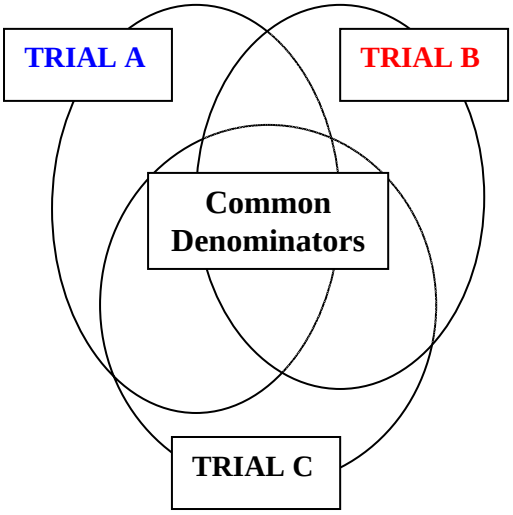

Fig. 1

Early in the process we realized that we cannot implement ADaM without a deeper knowledge of SDTM. ADaM will never be as structured as SDTM. Every time we were in doubts regarding what to call a variable or how to define it, we turned to the SDTM IG v.3.1.2 or discussed it with colleagues with greater knowledge of SDTM.

3.0 Ferring ADaM

Approximately 80% of the datasets we create within a trial can be standardized with respect to a large core of variables. For these we created F-ADaM, which is very operational, especially compared to V-ADaM (Vanilla ADaM), the structure to use for a submission.
Besides the required domains of V-ADaM, the F-ADaM includes additional operational domains with simple derivations added to SDTM domains (SDTM+) such as; baseline values, changes from baseline, flags etc. as opposed to adding the information as extra records. In some cases we have datasets with a combination of vertical structure but add variables in a horizontal structure (see fig. 2).

VISITNUM	VSTESTCD	VSSTRESN	VSSTBASE	VSSTCHG
1	DIABP	90	90	.
1	HEIGHT	180	180	.
1	PULSE	75	75	.
1	SYSBP	170	170	.
1	WEIGHT	90	90	.
2	DIABP	90	90	0
2	PULSE	74	75	-1
2	SYSBP	170	170	0
3	DIABP	81	90	-9
3	PULSE	82	75	7
3	SYSBP	149	170	-21
4	DIABP	74	90	-16
4	PULSE	73	75	-2
4	SYSBP	123	170	-47
5	DIABP	79	90	-11
5	PULSE	79	75	4
5	SYSBP	146	170	-24

Fig. 2

In fig. 2 the variables VISITNUM, VSTESTCD and VSSTRESN are in a vertical structure, coming directly from SDTM, while VSSTBASE and VSSTCHG are variables created in ADaM and added in a horizontal structure.

Standardization and horizontal data structure often do not fit well together. Therefore many of the analysis datasets were in a vertical structure like in SDTM, which means that ”one proc away” for analysis will probably be “a couple of proc’s away”. However this approach eases maintenance and the ability to combine trial databases.

Usually for a submission all variables not required in SDTM will go into the Supplemental Qualifiers (SUPPQUAL) datasets. These additional variables are kept in the related domains in F-SDTM, instead of splitting them up in two domains this makes the database more operational. We find that very convenient and it makes the programming easier. From here they are collected for use in F-ADaM analysis datasets unchanged. For a submission these additional variables will be moved to the appropriate SUPPQUAL datasets in V-SDTM, and there will be some additional effort required when the V-ADaM datasets are to be created for the submission. Another solution could be to rename or map the variables into new variables in the analysis datasets from the beginning.

Besides the rules and definitions already specified in the ADaM/SDTM documentation, Biometrics at Ferring has also built their own set of rules upon ADaM/SDTM regarding

- EPOCH
- Imputation rules for missing/partial values, which require SAS dates derived from ISO character date/time (--DTC)
- Ferring controlled terms additional to those published by CDASH (Clinical Data Acquisition Standards Harmonization)

There is no “right” answer to how many datasets and variables should be submitted.
Instead, we chose to create all the ADaM analysis datasets which contain the analysis variables used for the analyses specified in the SAP and reflecting the protocol. This means that they can also be used for a submission.

3.1 The implementation

To successfully conduct this implementation of ADaM, it was of upmost importance that resources were allocated and support from managers and colleagues was present. The brainstorming within the department and discussion/interpretation of the ADaM implementation was been very important too. Extra benefits included a revision of routines, streamlining procedures and general cleanup.

We did not have an implementation plan or an ADaM team, just a strategy to follow the linear development path.

Raw Data → F-SDTM → F-ADaM

We began by reading the documentation and implementation guidelines for SDTM and ADaM. Afterwards we defined each analysis dataset needed for reporting a trial. During the process of naming new variables created in ADaM it was beneficial not only to look in the ADaM documentation, but also to look in CDISC SDTM Implementation Guide (version 3.1.2.) appendix D: CDISC variable-naming fragments, which define a standard list of fragments to use when naming variables.

To conduct a trial in parallel with development of Ferring ADaM standards as opposed to choosing a pilot trial *after* finalization proved to be very beneficial. It was possible to catch some discrepancies or inconsistencies between our definitions and the ones stated in the SDTMIG and ADaMIG and also solve some difficulties within the interpretation of the ADaMIG when defining variables in a “real” trial. A lot was learned when using ADaM in practice.

The interpretation of the ADaMIG (and SDTMIG) was at times very difficult and in these cases SDTM-knowledgeable colleagues from Data Management were of great help. Discussion of these interpretations was very beneficial. For example, whether the definition of study day ADY/--DY should be the ADaM definition or reflecting the trial protocol, or how to use the variables BASE, CHG, PCHG, the use of flags etc.

3.1.1 Types of standardized analysis datasets

To begin with, the analysis datasets that we found could be standardized across all therapeutic areas were identified and fully defined. These were analysis datasets such as ADRD, ADDV, ADXP (Ferring domain for analysis populations), ADAE, ADLB, ADSL and ADSLSF. Most of the variables are defined as in F-SDTM e.g. used without any change and with the same name, label, format and content. For almost all trial designs it would not be necessary to remove or add any variables for these analysis datasets, of course there are always exceptions!

A second group of analysis datasets were defined. These were analysis datasets that could be partly standardized across all trial designs, such as ADVS, ADPE, ADEG, ADCM and ADSV. The most common variables which are (almost) always present in the domains were defined. Most of the variables were defined as in F-SDTM, that is, used without any change and with the same name, label, format and content and with a few calculated variables specified according to ADaMIG v.1.0 and Analysis Data Model v 2.1. For certain therapeutic areas it may be necessary to add domain specific variables.

The third group of analysis datasets is the least specified due to the impact of the trial design. It was the exposure analysis dataset ADEX and the efficacy analysis dataset(s) ADEF/ADEFn. They contain specified variables which are required as a minimum in the analysis datasets, but for these datasets we knew that for each trial it is necessary to add trial specific variables. Occasionally, it would be necessary to remove some of the specified variables.

Finally, there will always be trial specific analysis datasets which will only be used in one therapeutic area or trial and never in others. They were not specified, only the rules and definitions according to ADaMIG v.1.0 and Analysis Data Model v 2.1 of how to specify these trial specific analysis datasets were given.

3.1.2 ADSL

The only requirement of ADaM datasets for a submission is the subject-level analysis dataset ADSL. There are some variables which are required with standard variable names and definitions, but how the ADSL dataset is structured may vary. We have made some modifications in ADSL which we found appropriate for our standardization. Our ADSL definition still lives up to the requirements but we excluded some variables and added others, so we have an ADSL dataset which is operational and standardized.

Previously, we had divided demographic data and disposition data into two datasets ADDS and ADSL. Since some of the disposition variables are required in ADSL, we decided to add all ADDS variables to ADSL and do not define a separate analysis dataset for ADDS. We still have one record per subject in a horizontal structure. Furthermore, we have added a duration variable (days since first dose at discontinuation).

Since it is not recommended that screening failures should be included in ADSL, we created the dataset ADSLSF, only containing screening failures, with the exact same structure as ADSL including some variables with missing content not relevant for screening failures. In some trials ADSLSF might end up having 0 records.

3.1.3 TRTxP

Again, we had the conflict regarding submission versus standardization with TRTxP. One of the issues was in terms of whether to add TRTxP or not in ADSL. According to the ADaMIG, TRTxP is a required variable for a submission, but it would conflict with our standardization. This is because adding all the treatment periods in ADSL horizontally will cause the standard structure across all trials to be lost. Since we were not building this standard for a submission only, we decided to exclude TRTxP from ADSL and only keep the information in the exposure dataset (ADEX).

EXTRT	TRTP	EXEPOCH	EXTSTDT	EXTENDT	EXPSTDT	EXPENDT	EXSTDT
BLINDED	Degarelix	Treatment (1)	05JUN2007	10SEP2007	05JUN2007	09JUN2007	05JUN2007
BLINDED	Placebo	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	09JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	09JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	09JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	10JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	11JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	12JUN2007
HP-hMG	HP-hMG	Treatment (2)	05JUN2007	10SEP2007	09JUN2007	14JUN2007	13JUN2007
BLINDED	Degarelix	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	14JUN2007
HP-hMG	HP-hMG	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	14JUN2007
HP-hMG	HP-hMG	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	15JUN2007
HP-hMG	HP-hMG	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	16JUN2007
HP-hMG	HP-hMG	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	17JUN2007
HCG	HCG	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	18JUN2007
UTROGESTAN	UTROGESTAN	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	21JUN2007
UTROGESTAN	UTROGESTAN	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	22JUN2007
UTROGESTAN	UTROGESTAN	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	23JUN2007
UTROGESTAN	UTROGESTAN	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	24JUN2007
UTROGESTAN	UTROGESTAN	Treatment (3)	05JUN2007	10SEP2007	14JUN2007	10SEP2007	25JUN2007

Fig. 3

Then the other issue was to decide how to add TRTxP in ADEX, and again due to standardization across all trials, we have defined the treatment periods in the exposure dataset in a vertical structure (fig.3). Furthermore, we have tried to find a meaningful name for the exposure variables that reflects the content. For example, EXPSTDT: EX for domain, P for period, ST for start and DT for date. For a better overview of both structure and variable names see fig. 4.

Use of information in the exposure data

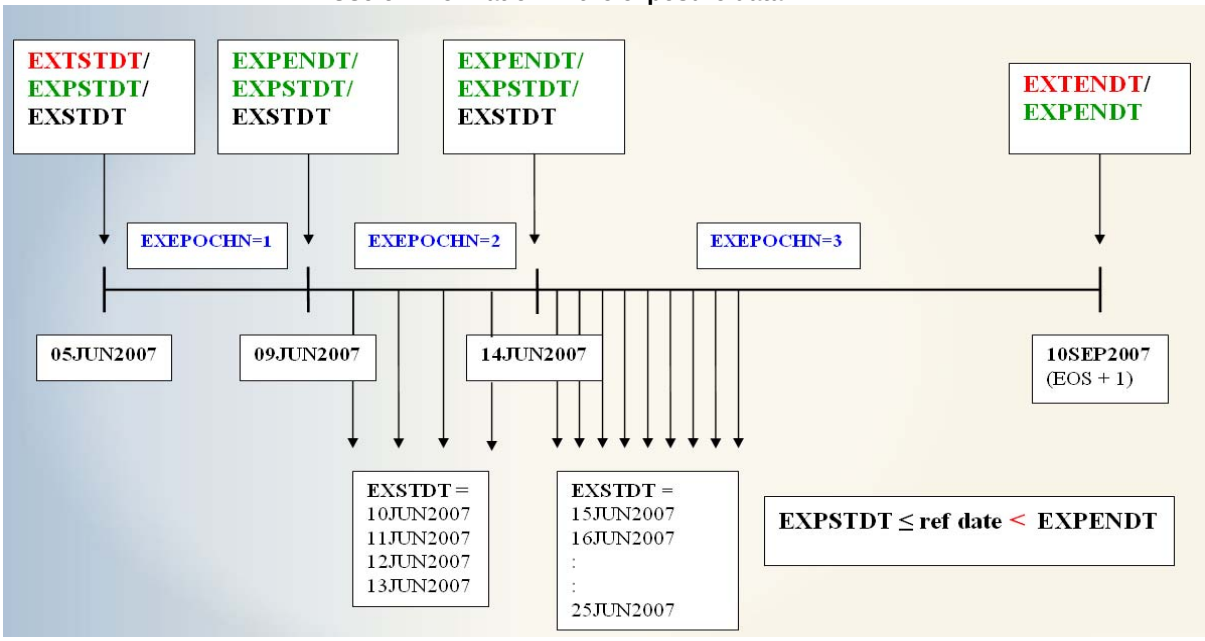

Fig. 4

It should be noted that ADEX was one of the more difficult analysis datasets to define. It is very trial specific and therefore difficult to standardize, but that was the rationale of using a vertical structure.

3.1.4 Traceability

As analysis datasets are derived from SDTM datasets most of the time, some level of traceability between SDTM datasets and analysis datasets must exist. So whenever possible, variables to facilitate traceability should be included. Typical SDTM variables for traceability are --SEQ, VISIT and VISITNUM. The --SEQ variable is required in SDTM, whereas VISIT is only a permissible variable in a submission. However, since --SEQ is only a supportive variable we decided not to use in F-SDTM. Instead we decided to include VISITNUM in F-ADaM to support traceability, since it is a more meaningful variable which reflects the protocol and trial design. In some domains (FINDINGS) VISITNUM is expected, but in the special purpose domain SV it is a required variable.

3.1.5 Discrepancies – what to do?

According to Analysis Data Model Version 2.0 section 4.2.2, you can create an analysis study day variable ANLDY, which is allowed to contain 0. However, according to the ADaMIG version 1.0 section 3.3 timing variables all variables ending in DY “Relative Day of…” are NOT allowed to contain 0. This is a discrepancy which needed to be resolved. We chose to create our own study day variable, which can contain 0, naming it VISITDAY, to be able to fully reflect the trial protocol. Furthermore we decided to create a duration variable used in many different analysis datasets defined as (reference date – date of first dose). Since the variable is very close to the --DY variable, we called the variable --DU to avoid any misunderstanding, knowing that the fragment --DUR is reserved in SDTM for durations collected in the eCRF.

We were also in doubt of what to do regarding the variables BASE (baseline value), CHG (Change from baseline) and PCHG (percentage change from baseline). In the ADaM documentation they are defined in relation to the analysis parameter PARAM and AVAL. In our case we use and calculate change from baseline in several analysis datasets such as ADLB and ADVS. Change from baseline would not be the same for both domains, since it is not the same value for a certain record for a certain subject, and therefore we could not use the same name BASE, CHG and PCHG in both analysis datasets. Our

solution was to name the variables from their domains and origin variable and putting BASE, CHG and PCHG on afterwards (e.g. LBSTCHG, VSSTCHG).

Another example of discrepancies in the documentation of ADaM is indicator variables or flags. For example, “Names of all character flag variables end in FL” (CDISC ADaM Implementation Guide version 1.0, Section 3.7 Indicator variables), but the flags defined for the timing variables *DTF and *TMF only end on “F”. We chose the naming of timing flags as defined for timing variables, and overruled the FL-rule. Another issue we noticed was that the population variable SAFETY was name SAFFL rather than SAFETYFL. Many similar questions arose during this work.

When defining the treatment periods in ADEX, we had to be careful. In SDTM the variable EPOCH is already defined as the epoch associated with the elements in the ARM. We wanted to define the treatment period variable different than in SDTM, and we decided to name the variable EXEPOCH, where EX is for the actual domain.

3.2 Trial specific variables and datasets

We tried to standardize as much as possible across the therapeutic areas, but there will always be trial designs, statistical methods, analysis situations where it is not possible to follow the definitions outlined here. For these analysis variables it is necessary to define rules for them, then the next step will be to clearly define as many as possible of the trial specific variables to be added. To do that requires a good understanding of the principles of ADaM and SDTM and it has to be done according to the structure, definitions and rules described in Analysis Data Model v2.1 and ADaMIG v1.0.

- Any SDTM variable copied directly to ADaM the name, meaning and value are not to be altered.
- If a standard ADaM variable already exists then the variable name must be used for that concept.
- If a variable is not specified in either SDTM or ADaM, name the variable according to naming fragments of CDISC SDTMIG version 3.1.2

This is one of the new tasks the implementation of ADaM has created.

4.0 CONCLUSION

Looking back at the entire process, we are surprised how long time it actually took to get where we are now. The review–feedback on review–review again process is long, but very important, educational and necessary.

During the process I had to deal with a lot of challenges. It was not possible to mention them all here and some are more important to mention than others. In this paper I have covered a selection of the more serious/tricky issues encountered.

The ADaM implementation together with the standardization is a ground breaker for Ferring and has created new tasks to do, e.g:

- Standardization of trial/indication specific variables and/or datasets
- Analysis value level metadata
- Analysis result metadata

It is one thing to set standards, it is another thing to enforce and maintain standards.
Thus “tools” need to be developed to facilitate this and to ease the work within Ferring Biometrics and the collaboration with CROs when outsourcing. We are very early in the process and we are still in the middle of the pilot trial, therefore it is difficult to give too many details about the results, possible changes to the current version of Ferring ADaM or to identify additional areas where further reviews will be necessary.

What is next? As mentioned above some tasks are still pending regarding a “full” implementation of ADaM, it is just not planned yet when to start these next steps. At the moment we are experiencing the challenge, that an update/change in F-SDTM is to come, and further more some issues are found during the conduct of the ADaM pilot trial. When is it appropriate or necessary to update our ADaM specification and how will it affect the trials we are in the middle of right now? And how to deal with the CRO’s? In our situation with an ongoing process of implementing ADaM in a standardization process, these issues are probably more present as opposed to dealing with a submission. These are the tasks we a going to deal with in the near future,

5.0 ACKNOWLEDGEMENTS

I would like to express my biggest thanks to all of my good colleagues in Biometrics for feed back, brainstorming, comments and always setting aside time to answer my questions. A special thanks to my colleague, Gitte Dam Jepsen, for her always constructive comments, commitment and eager to help. I couldn't have done this without her/their help.

6.0 CONTACT INFORMATION

Karin Fleischer Steffensen
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