PhUSE 2015

Paper CD10

ADaM – Our Experience

Rob Wartenhorst, GSK Vaccines, Amsterdam, Netherlands

ABSTRACT

Last year, Legacy Novartis Vaccines (now part of GSK Vaccines) started a project to implement and roll out ADaM for all studies. We already had implemented SDTM 3 years earlier. Table generation was still done using a set of standard macros which had all the business logic built in. With the FDA now mandating, that studies starting end of 2016 have to be be supported by ADaM datasets, the time was right to change to ADaM as basis for our analysis.

This paper addresses the approach taken to define our internal company ADaM standard and the challenges we face(d) rolling out ADaM without disrupting the report creating process too much..

INTRODUCTION

Legacy Novartis Vaccines (LNVx) implemented CDASH and SDTM in 2011. For analysis we adapted our standard analysis software to use SDTM datasets as input instead of the old legacy standard datasets. Mid 2014 it was decided it was time to implement ADaM. By 2014 we had a very stable implementation of SDTM and a solid understanding of the contents. A project team was formed that consisted of a ADaM lead, project manager, a statistician, 3 programmers and later in the development of the standard a part time external CDISC expert..

DESIGN

We were fortunate to have in house a statistician who had worked with ADaM for many years in a previous job and who also had a full understanding of the analysis that were to be performed. Vaccines is definitely an area that has its own unique data structure and its own unique analysis requirements. The following elements were considered in the design:

1. How many ADaM datasets
2. How to incorporate protocol deviations
3. Define.xml
4. Specification collection
5. Programs for ADaM dataset creation
6. Programs for TFL generation

1. HOW MANY ADAM DATASETS

Choices identified were,

a) only datasets to support the primary analysis,
b) datasets to support all analysis,
c) datasets to support all analysis and listings.

First decision to make was if we would support all analysis or only the primary analysis. Given the standard nature of the data for the non-primary analysis the decision was to have ADaM datasets for all analysis. For example medical history and concomitant medications tables are standard and don't change from study to study. Hence ADaM specifications are standard as well.

Initially we choose to support all analysis as well as listings. However when implementing pilot studies it became clear that including data for listings required a significant effort in writing the specifications and at the same time remain compliant with the ADaM rules. As an example, in vaccines when reactogenicity data is collected the CRF asks for the first 7 days the severity of an event (PARAM). The values can be NONE, MILD, MODERATE, SEVERE. Then there is final question if the event continued after day 7 with value YES or NO. To combine this data in ADaM for the purpose of creating a listing, a separate PARAM would need to be defined for the after day 7 data as the data for the first 7 days is not categorized the same as the data for after day 7. So for ADaM you would first need to specify for each event (can be 20 events) a new PARAM and then when producing the listing write code to combine the first 7 day events with the after day 7 event.

The final decision was that we create ADaM datasets to support all analysis and for listings use a sensible approach where variables for listings are only included in the ADaM dataset if it requires minimal effort. I.e. a

direct copy from SDTM. Generally listings will be generated from a combination of ADaM datasets (at least ADSL) and SDTM data. This we found is the most effective way to create listings.
Currently we have a total of 19 standard ADaM datasets defined that support all the analysis in our standard Table of Contents.

2. HOW TO INCORPORATE PROTOCOL DEVIATIONS

Protocol deviations and exclusions are collected throughout the study. At the time of database lock a document is created describing all the deviations at subject level. If you would require only SDTM data as input for your ADaM, all these deviations would need to be specified again in the ADaM specifications which would be a) a lot of work to specify and b) duplicate the work already done to create the document with the deviations.

The final decision was that deviations document is supported by a listing of all the deviations and exclusions and the dataset that this listing is based on, is used as input for ADaM programming of population flags and population summary dataset. A BDS like structure for this dataset enables straight forward mapping of population flags. A proper setup of the deviation dataset is important to minimize time spent on the population flag definitions.

Protocol deviation finalization occurs around the same time as ADaM specifications writing. This helps to define the correct population flags in ADaM, but at the same time can distract from the ADaM specification process. It is important that sufficient resources are available during this time to avoid delay in the study analysis.

3. DEFINE.XML

A must have part of the project was the inclusion of the DEFINE.XML, including the analysis results metadata. The specifications document for the datasets has been designed with a dual purpose: a) As specification source for the ADaM dataset programmer and b) as direct input for the DEFINE.XML. One element is not specified, the variable length. When the datasets are finalized the maximum variable length is programmatically determined. The metadata from which the define.xml is built, is programmatically populated with this variable length.

For the variables that are copied from SDTM, the contents and the attributes have to be the same in ADaM. Therefore the SDTM DEFINE.XML must be available before creating the ADaM DEFINE.XML. This requires coordination with the teams providing the SDTM data.

When we started designing the ADaM datasets, define V2 was expected to be released soon. Therefore we implemented DEFINE.XML V2. There were however some changes in the define V2 between the latest draft and the final version which required some updates to our define document. But compared to first implementing define V1 and then migrating to define V2 the effort for the updates was minimal.

4. SPECIFICATION COLLECTION

Like almost everyone else, we also have started with the ADaM specifications using Excel and we are still continuing to use it until a suitable metadata repository is in place. Excel was the product of choice to start with as it is standard available to all users. However it is not very well suited for an environment where you have multiple users working on the same study. Disadvantages of Excel are:

- Only one user at a time can work with an excel sheet
- Easy to make inadvertent changes
- Changes have to be manually highlighted
- Consistency checks cannot easily be implemented in Excel
- Users can change standard variables (e.g. change CRIT02FL name to CRIT01FL) and labels which may go unnoticed till late in the analysis process.

5. PROGRAMS FOR ADAM DATASET CREATION

The current design has a limited set of standard programs. One program to convert the excel sheet in metadata and a format library. A program to set up the ADaM dataset shells and a program to finalize the datasets (putting variables in order, apply labels, set lengths and create XPT version of the dataset). Programs are tested via double. This setup allows for rapid enhancements to AdaM dataset designs. The intention is to develop more standard utilities that are (partly) data driven once our ADaM standard has further matured. Currently setup of the programs for a study takes between 3 and 10 days, depending on the complexity of a study.

6. PROGRAMS FOR TFL GENERATION

In contrast to the old approach of a big standard program that would know everything and could do everything (within the limits of the standard) the new setup consists of a few standard building blocks:

- Setup block
- Block for the denominator
- Block for the counts and percentages
- Block for basic statistics
- Block for reporting

Any statistic that is not available in a standard building block is custom programmed. This allows for the flexibility that was previous not available in our standard tool set. The reporting block outputs the data both as a PDF/RTF as well as a dataset. This dataset is used for both comparing results from double programming as well as comparing results from earlier runs. Currently we double program all tables. For efficiency, programmers are encouraged to use the analysis results metadata where possible to automate the table generation. For certain types of analysis a program can be run multiple times referencing the selection criteria and analysis variables in the analysis results metadata, cutting down the development time per table for some of the analyses to minutes instead of hours.

The programs built from the building blocks are much smaller than the old standard programs. As a consequence they run much faster. During the final stages of a study running all programs in a batch has become less of an issue with respect to load on the SDD server.

Compared with the old set of standard programs, programmers need little training to start working with the building blocks and become productive quite quickly. As we progressed through more studies, more examples became available increasing productivity as the year progressed.

DEVELOPMENT

Development can be split in 2 categories. Tool development and ADaM standard development. Effort required for tool development is small compared to developing an ADaM company standard. For 3 studies that were already reported (based on SDTM), we redid part of the analysis to establish our version 1.0 of the standard. Since then we have done 6 studies using ADaM which brought us quickly to version 2.0 of our standard. The biggest challenge is to keep the number of flags down. Take for example the ADAE dataset following the occurrence structure. For each AE we have 3 flags. AOCCFL AOCCPFL and AOCCSFL. In the vaccines world we present many variations of the AE data. AE's leading to withdrawal, AE's of special interest etc.. All these variations are then also presented for a specific time period. It is easy to have 20 different selections presenting the AE's. This results already in 60 different occurrence flags. An alternative is to duplicate the AE records for each presentation and only have 3 flags. The implementation guide and examples only show the basic analysis and working out the most efficient analysis dataset takes a considerable amount of time. The challenges in developing the ADaM standard we encountered so far are:

- Population flags.
  - Only include overall population flags in ADSL
  - Implement population flags that are visit and/or parameter specific as record level flags.
  - Include overall population flags only if applicable to a dataset. For example the overall solicited safety population flag is not applicable to the lab data and should be left out. Too many population flags makes it difficult to work with the data in a data viewer
- Analysis flags
  - As in the ADAE example mentioned before, duplicate records using the same limited number of flags, rather than creating more flags. An alternative is duplicating datasets. But more datasets means also much more spec writing and a much larger DEFINE.XML with a lot of duplication. My preference is to have a bigger dataset with duplication of records. The implementation guide right now does not provide much guidance in naming of variables to separate these duplicates.
- Application of CDISC standard to vaccines
  - Vaccines studies typically consist of one or more vaccinations of healthy subjects. Analysis is typically based on actual or planned vaccine. There is no treatment in the sense of administering a drug on a regular interval. This leads to artificially adding treatment variables to the datasets with the only purpose to satisfy the CDISC requirements, but without benefit to the analysis.

TRAINING

Our training consists of 2 elements. First is a classroom training in which the general principles of ADaM are taught, tailored to the vaccines requirements. Second is on the job training. We formed group called ADaM center of excellence (ACoE) which provides the following services:

- Assistance with specification writing.
- Programming of ADaM datasets.
- Support programmers in the use of the building blocks to create the tables figures and listings.
- Maintenance of the ADaM standard.

Currently statisticians are trained by having them write the ADaM specifications for at least 2 studies under supervision of the ACoE. We are currently evaluating if this is the best way to train statisticians on the ADaM standard.

Programming of the datasets is done by the ACoE for 3 reasons. First to provide quick feedback to the statisticians with respect to the specifications, second to govern the standard, and last, further develop the ADaM standard based on the experiences of each study.

TFL Programmers are trained on the concept of the analysis results metadata and the use of the building blocks. With a little mentoring from ACoE and/or programmers that have experience they then quickly start producing their first tables. Having examples available is very valuable in getting the TFL programmers getting to understand the metadata and use of ADaM with the building blocks.

The intention is that after the initial training the statisticians role will be limited to reviewing the specifications and ADaM specification writing will be the task of experienced programmers, with the ACoE providing support and governance of the standard. Also here we will, in the future, evaluate if this is the most efficient approach in ADaM dataset generation.

ROLLOUT

In legacy Novartis Vaccines a project manager keeps track of all studies. His role was expanded to project manage the rollout. A plan was made with studies, to analyze using ADaM, in the first year. In principle, from January 2015, all studies would be analyzed using ADaM, unless there was a good reason to analyze using the old tools. Acceptable reasons for not using ADaM were: ACoE resource constrains, majority of the work from an interim analysis can be reused for final analysis, the risk of delay in delivering study not acceptable. In the end, this year half of the studies will be analyzed using ADaM. Support from higher management was important to help convince study teams to accept the ADaM approach for their studies. Meaning if clinical teams asked that their study was to be exempt from ADaM analysis due to the risk of delay of analysis that management would not support that, but instead supported the analysis of the study using ADaM.

The way of working changes a little when using ADaM. Work has to start earlier due to the added work of ADaM specification writing and datasets programming. Having an independent project manager help the statistician to adjust to the added tasks was instrumental in still being able to complete studies on time.

After completion of each study, a survey was sent out and a lessons learned session held. The lessons learned sessions are a good tool to ensure that issues do not go unnoticed. Improvements are continuously made as we progressed through the year and this was reflected by the lessons learned results by better scores for more recent studies.

CONCLUSION

The ADaM implementation has been successful, as we are able to deliver most of the analysis on time, including the required CDISC documentation(DEFINE.XML). Minimal additional work is required after study delivery to provide the study in CDISC format. The additional work would be writing the data reviewers guide and placing the files in a submission folder structure. Defining the ADaM company standard is the most time consuming part of the implementation and we are still improving our standard.

The main factors contributing to the success are:

- Having at least one person with ADaM experience and vaccines study experience in the team.
- Formation of ADaM Center of Excellence to define/maintain ADaM standard and support study teams.
- Availability of project manager to adjust study teams into the new process.
- Support from top level management.
- Building blocks approach for table generation.

Challenges in the implementation are:

- Making statisticians available earlier in the process to start working on specifications.
- Availability of extra resources to compensate for the extra time needed to learn SDTM, ADaM and write more detailed specifications.

- Defining who should write the study ADaM specifications and who is the owner of the study ADaM specifications.
- Training everyone in the statistics organization in the CDISC standard.
- Finalization of protocol deviations around the same time as ADaM dataset creation and the dependency of the ADaM datasets on the protocol deviations dataset.
- Application of CDISC to vaccines studies.

From the lessons learned we can see:

- That over time, roles and responsibilities better defined and understood by teams and resource management has improved.
- That it is important that the Center of Excellence group has a good understanding of the studies for optimum support.
- That in the beginning we started the specification process too late with not enough details in the mock tables.
- That the quality of the output has improved due to transparency of the data.
- The need for more examples for study teams.
- Programmers quickly pick up the use of building blocks to create tables using the ADaM datasets.

CONTACT INFORMATION

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

Rob Wartenhorst
GSK Vaccines
Hullenbergweg 81-89
1101 CL Amsterdam
The Netherlands
Work Phone: +31 205640564
Email: Rob.x.wartenhorst@gsk.com