PharmaSUG2010 – Paper CD04

The CDISC/FDA Integrated Data Pilot: A Final Summary of Findings, Reviewer Feedback, and Recommendations Implementing CDISC Standards Within and Across Studies

Chris Decker, d-Wise Technologies

ABSTRACT

The CDISC/FDA Integrated Data Pilot was initiated in 2008 with the mission of demonstrating that a data submission created using the CDISC standards will meet the needs and expectations of FDA reviewers in conducting an integrated review of data from multiple studies and compounds. The Pilot was completed recently and the results are currently being compiled. This paper will provide an overview of the original goals of the Pilot, the major findings across the CDISC models, the feedback from the FDA participants, and the timeline for publishing the final report and submission package. The session will present a review of experiences implementing the CDISC models, FDA reviewer feedback, and provide the industry with the Pilot team's experiences with the new CDISC models and their use for data integration.

INTRODUCTION

At the beginning of 2008, CDISC released a Technical Road Map providing an overview of activities regarding the development and harmonization of current and future CDISC technical products over the next three years. One of the roadmap's key objectives is to "execute pilots with regulatory authorities to gain a better understanding of the needs of regulators and industry". The CDISC/FDA Integrated Data Pilot was initiated in the spring of 2008 with the mission of demonstrating that a subject data submission created using CDISC standards will meet the needs and expectations of FDA reviewers in conducting an integrated review of data from multiple studies.

The goals of the Pilot project were to expand on the work of the SDTM/ADaM Pilot conducted from 2005 to 2007 involving a single study. The primary focus of the Pilot was to implement the newest standards models and explore the capability of the standards to review integrated data across several studies with different designs. The objectives of the pilot included the ability to:

- Assess the applicability of the CDISC models to integrate data
- Validate the components of the CDISC models can be effectively used together
- Evaluate and address any issues or questions with the most current CDISC models including SDTM, ADaM, Define.xml, and Trial Design
- Support the critical path initiatives around data standards, integrated databases, standard data collection, and studies of special populations

The Team members included over 20 participants from pharmaceutical and biotechnology companies, vendors, NIH and FDA. The Pilot team has been engaged with a number of medical and statistical reviewers across both CDER and CBER to identify efficiencies and explore limitations with using the CDISC standards.

During the course of the Pilot, the team has addressed a number of challenges with implementing the most current CDISC standards to review data within and across studies. This paper will provide the following:

- An overview of the Pilot including the mission, goals, and objectives
- A summary of the findings implementing SDTM 3.1.2 across studies
- A summary of the methodology and findings implementing the new draft ADaM 2.1 model and 1.0 Implementation Guide

- An overview of the draft extensions for define to support value level metadata and analysis metadata results
- Summary of FDA feedback in reviewing the components delivered by the Pilot team
- A timeline for completing the final Pilot Report and delivery of the report and submission package

PILOT PLAN

PILOT TEAM STRUCTURE

Due to the breadth and scope of work targeted within this Pilot, the Team was led by three co-leaders. In addition to managing the overall Pilot, each co-leader managed a subgroup.

- Programming group – Responsible for managing the metadata, performing all SDTM and ADaM transformations, and generating the summary tables
- Analysis and Design group – Responsible for writing the analysis plans including table shells, defining the ADaM specification, and summarizing the results within abbreviated clinical study reports
- Packaging group – Responsible for organizing the eCTD, bookmarking the submission, and implementing the define.xml and rendering of the outputs

In addition, a FDA review team was identified consisting of both medical and statistical reviewers across both CDER and CBER who were responsible for providing an ongoing review of the Pilot deliverables.

Managing a team with this much effort is much like running a small company and everyone on the team put forth tremendous effort over the course of the last two years.

CDISC MODELS

During the initial planning of the Pilot, the CDISC leadership and Pilot team leaders discussed which models to test given the challenges of ongoing development of different versions. The team decided to take on the challenge of testing the most current models, even though they were under development and continuously changing. This included SDTM 3.1.2, ADaM 2.1, and collaborating with the XML Technologies team to develop new strategies for supporting added functionality to the define.xml. This decision was important for meeting the strategic vision for CDISC; however, it created quite a challenge for the team as they “Were breaking new ground” with every discussion and decision.

SOURCE DATA

The data used within the Pilot came from a collection of studies that were standardized as part of a project between the FDA and SAS Institute. As part of this project, safety data from 30 studies were standardized to a common data model. After this work was completed, the FDA agreed to provide the data to the CDISC Pilot team after a rigorous de-identification process that included altering the data to ensure confidentiality.

The Pilot team defined a subset of eight studies that were in ‘better’ shape and provided a robust, but manageable set of data for a volunteer effort. The studies chosen included eight studies across three different compounds and two different study designs. The study designs included a double blind dose escalation design and an open label pharmacokinetic design. Due to confidentiality reasons, the team was supplied with only the following components:

- One page generic summary of each study
- SAS transport files for each study

Since the goal of the original project was to review safety, the domains included supported a safety profile. The following domains were provided to the Pilot team:

- DM – Demographics
- AE – Adverse Events
- DS – Disposition

- MH – Medical History
- EX – Exposure
- VS – Vital Signs
- LB – Labs
- CM – Concomitant Medications
- SC – Subject Characteristics
- PC – Pharmacokinetic Concentrations

The drugs studied within the Pilot data were indicated for the treatment of pediatric hypertension. Due to the limited information provided to the Pilot team, the team had to make a number of assumptions in working with the data, some of which are described later on.

PROJECT SCOPE

The team spent a long time defining the final deliverables the Pilot. During the planning phase the decision was made to delivering the following components:

- Analysis plans for each study, each integrated compound, and an overall composite integration
- SDTM domains for each study
- ADaM domains for each study, each integrated compound, and an overall composite integration
- Define files for each study, each integrated compound and an overall composite integration
- Clinical study reports for each study, each integrated compound and an overall composite integration

An important decision made early on was that all analyses would be conducted on the ADaM domains and not SDTM. The team decided to deliver SDTM only for the individual studies and not for any of the integration components. This decision will be evaluated during the FDA review to see if this meets the needs of both the medical and statistical reviewers.

During the project, the team realized how much resource and effort were going to be needed so the scope of the deliverables was scaled back to include the following:

- SDTM and ADaM data sets for four studies within a single compound
- Summary tables and Clinical Study Report for two studies
- Integrated data for one compound including integration for similar study designs as well as an overall compound level integration
- One integrated Clinical Study Report including all the integration buckets

The team felt this reduced scope would still meet the objectives of the Pilot project.

ANALYSIS

ANALYSIS METHODOLOGY

To address the need to demonstrate the feasibility of integrating data within the limits of what was provided, the Team developed an approach that is different from conventional analyses. An integrated efficacy analysis was not possible because the original protocols with eligibility criteria, primary outcome variables, and statistical analytic plan were not available, efficacy data was not provide and the data had been altered. An integrated safety analysis was not possible for similar reasons. In addition, safety is always assessed relative to efficacy and cannot be evaluated as an absolute, but.

The approach chosen was to use the data sets for what they were – a collection of values in different domains – and perform analyses of the general subject experience. The general subject experience means the series of events and observations that occurred for each subject and subject population. Not having baseline values for the particular

subject populations, the Team referenced published age calibrated normal values to determine if any given event or value was within or outside the range of normal.

The output is then a series of individual subject profiles that are integrated within a study to provide a descriptive study population profile. Across studies, the population is defined as all subjects because everyone received at least one dose of study drug. That profile is broken down further on the basis of demographic variables such as gender and age and study specific variables such as epoch within each study and across studies to test exposure to treatment. The rationale for performing an integrated analysis across the compound is to describe a profile for the entire population of subjects with hypertension. This profile could be used as a reference framework, and future analyses comparing individual compounds or even individual studies to a total disease or condition centric reference frame may be informative.

To summarize, due to various limitations, a conventional analytic approach was not possible. So an approach was used that would include all of the technical requirements and tools to perform an integrated analysis across studies, but would not specifically map to the more conventional and familiar analytic approaches. Consequently, conventional expectations should not be applied to this exercise. Instead, the exercise should be evaluated on its own merits as a demonstration of proof of concept for the use of data standards and a structured approach to integrate data and perform analyses.

ANALYSIS PLAN

Based on the methodology above, the analysis subgteam wrote abbreviated Statistical Analysis Plans for each type of study design as well as an integrated plan for a single compound.

Some of the specific approaches used were to convert and reconcile all the data fields and values to conform to the most recent version of the CDISC Study Data Tabulation Model (SDTM). Subsequently, ADaM datasets were created using the SDTM datasets as inputs based on the most current draft ADaM Implementation Guide. Due to evolving standards in both SDTM and ADaM, the process involved careful reconciliation and alignment. Once all the nomenclature, logical decisions and structure were agreed upon, some of the variable definitions were quantified.

As part of the exercise paradigm, the Team decided that all clinical events, vital signs and laboratory values that were outside the range of normal would be considered an abnormal event. Recorded adverse events, by virtue of being recorded, were automatically incorporated.

The major variable affecting laboratory values in children is age. Laboratory values were compared to age adjusted reference standards from the National Institutes of Health Clinical Center. Because the values are periodically updated, the normal ranges were selected as of a particular date to normalize all study values.

Pediatric normal vital signs, due to the changing proportions of height, weight, body surface and body volume as a whole and also based on anatomic regions, for example relative head size, are complex to calculate. For each study participant, the age, gender, height and weight were factored into normal distribution curves to determine abnormal events within vital signs. The Team used formulas derived from the Centers for Disease Control to support the programming decisions and logic to make the calculations.

The selection of tables, as exemplified by the table shells in the SAPs, was based on variables that plausibly may be informative such as age, gender, study epoch and exposure to test compound. Other analyses are also plausible and possible, but the Team chose to focus on the selected outcomes because they each demonstrated a different facet of the datasets and together provide a composite proof of concept for the integrated approach. Since a formal statistical analysis was not be performed, the decision was made to provide descriptive statistics and an associated summary with no conclusions. It will be the responsibility of the FDA members and other reviewers to perform an independent assessment.

METADATA FRAMEWORK

The first step in managing the information across studies was to define a metadata framework to collect and manage the metadata for both SDTM and ADaM. The team used a framework that was defined as part of the ADaM/SDTM Pilot which allowed the programming group to enter metadata within Excel spreadsheets in a standard format. This metadata was then processed within the framework and a metadata model was built to support the entire process including the development of SDTM and ADaM domains, the publishing of define.xml, and the validation of both the SDTM and the ADaM data.

A separate sheet was defined for tables (Figure 1), columns (Figure 2), and codelists (Figure 3).

	C	D	E	F	G	H	I	J	K	L
1	tlabel	torder	type	tdescription	location	structure	repeating	reference_	purpose	class
2	Adverse Events	1	TABLE	One record per adverse event per subject	../../study1/tabulations/ae.xpt			No	Tabulation	Events
3	Demographics	2	TABLE	One record per subject	../../study1/tabulations/dm.xpt			No	Tabulation	Special Purp
4	Disposition	1	TABLE	One record per disposition status or protocol milest	../../study1/tabulations/ds.xpt		No	No	Tabulation	Events
5	Exposure	1	TABLE	One record per constant dosing interval per subject	../../study1/tabulations/ex.xpt			No	Tabulation	Interventions

Figure 1. Table Metadata

	A	B	D	E	F	G	H	I	J	K	P	Q	R	S
1	table	column	cpkey	corder	clabel	clabellonc	ctype	clength	cformat	cformatflag	cdescription	ROLE	computation_method	
2	AE	AEACN		10	Action Taken with Stu		C	20	AEACNF	2	AE_AEACN	Record Qualifier		
3	AE	AEBODSYS		7	Body System or Orga		C	40			AE_AEBODSYS	Record Qualifier		
4	AE	AEDECOD		6	Dictionary-Derived T		C	40			AE_AEDECOD	Synonym Qualifier	PT from MedDRA 10.1 c	
5	AE	AEDUR		25	Duration of Adverse E		C	10			AE_AEDUR	Timing		
6	AE	AEENDTC		22	End Date/Time of Ad		C	10			AE_AEENDTC	Timing		

Figure 2. Column Metadata

	A	B	C	D	E	F
1	format	start	end	flabel	flabello	rorder
15	EPOCH	DOUBLE BLIND				2
16	EPOCH	PLACEBO WASHOUT				1
17	EPOCH	RANDOMIZED WASHOUT				3
18	EXADJ	CONTINUING				2
19	EXADJ	ESCALATING				1
20	EXDOSFRM	SOLUTION				1
21	EXDOSFRM	TABLET				2

Figure 3. Codelist Metadata

In addition to capturing table, column, and codelist information, the metadata framework was extended to capture value level metadata. An example of this metadata is below.

table	column	param	paramrel	ptype	plength	pformat	pformatfla	pdescription	computation_method
ADLB	PARAMCD	ALB	AVAL	N	8	4.1	1	ADLB_AVAL	Equal to LB.LBSTRESN
ADLB	PARAMCD	ALB	LBCAT	C	20	LBCATBC	2	ADLB_LBCAT	Equal to LB.LBCAT
ADLB	PARAMCD	ALB	PARAM	C	20	LBTALB	2	ADLB_PARAM	Equal to LB.LBTEST Plus units MM

Figure 4. Value Level Metadata

In this example, specific metadata is defined within the ADLB domain where the value of PARAMCD='ALB'. In the first row metadata is defined for the value of AVAL when PARAMCD='ALB' including a length of 8 and a format of 4.1. This metadata is used later during the creation of the define.xml file.

A detailed discussion of the metadata framework can be reviewed within a paper by Steffens, Fleming (SAS Global Forum, 2009) referenced below.

SDTM IMPLEMENTATION

The first step in defining the SDTM specification for the eight studies was to perform a gap analysis of the data to evaluate what additional work would need to be done to conform to the specification and support the analysis plans.

The data supplied to the team was supposedly already in SDTM 'like' format so the team thought this would be a fairly straightforward process. While the work was not as complicated as it might be in real world context, it was still more extensive than originally expected. This finding supports some of the challenges of interpreting SDTM and the fact that users can apply the standard differently. The team divided the original ten domains across multiple team members and performed an analysis to address any gaps between the data and the analysis plan needs as well as the SDTM 3.1.2 implementation guide.

There were a variety of issues identified during this gap analysis but most can be summarized in one of three categories. The first category is the most common problem working with legacy data – BAD DATA. There were a number of areas where data discrepancies existed. Where possible, the team made assumptions and 'fixed' the data. The second bucket involved compliance with SDTM. Even though the original data was supposed to be SDTM 'like' there were a number of instances where the input data did not meet the SDTM specification. This is a great example of the challenges with interpreting standards. The final bucket involved issues around what derived data should be included within the SDTM domains. The team evaluated how much derived data to include and decided to only include it where absolutely necessary, such as calculation of age or baseline flags. Instead the team planned for this type of data within ADaM.

This section contains a number of issues identified and how they were addressed. These are just a few examples, whereas all issues will be documented within the final Project report. Again, due to the limitations on the documentation provided the team had to make some 'assumptions' regarding the data. Most of the issues identified were typical challenges in working with legacy data.

SPECIAL PURPOSE DOMAINS

The DM domain was the only special purpose domain implemented. The DM domain was fairly straightforward with few issues. The primary issues identified are included in Table 1.

Table 1. Summary of Special Purpose Domain Issues

Domain	Issue	Solution
DM	All character variables were $200 including those defined as being limited within SDTM (e.g. ARMCD)	Looked across studies and set length appropriate to the data collected or IG requirements. This was an issue identified throughout all the domains.
DM	SUPPDM is not supplied which would include population flags	Population flags captured within ADaM ADSL and not included in SDTM
DM	SITEID was missing in some studies and is required	Created a dummy SITEID based on investigator information

INTERVENTIONS DOMAINS

The intervention domains originally supplied included CM and EX. After a thorough review of the original CM data the team decided to remove it from the list due to the sparseness of the data as well as not needing it for the planned analyses. Table 2 identifies issues addressed within the EX domain.

Table 2. Summary of EX Domain Issues

Domain	Issue	Solution
EX	The original data had what appeared to be records at each visit with the same dose but different dates. How should the team apply the comment "This domain should contain one record per constant dosing interval per subject" found within the IG	Decided to collapse records where dosing was consistent and also corrected overlapping records where applicable.

Domain	Issue	Solution
EX	EPOCH was added to the SDTM 3.1.2 IG.	Added EPOCH to the EX domain and defined the phases of the trial to match the trial design domains.
EX	EXDOSE was equal to 0 for records where EXTRT was Placebo. This example was removed from the SDTM 3.1.1 IG	Team decided to set value to 0 for those records. No definitive definition in the IG
EX	EXDUR was not collected in the required ISO format	Fixed to meet ISO requirements

FINDINGS DOMAINS

The findings domains included LB, VS, SC, and PC and involved the majority of the issues discussed by the team.

SC Domain

The SC domain was fairly straightforward and only includes 3-4 SCTESTCD values depending on the study design. There was one issue identified within SC. The RACEOTH value of SCTESTCD was collected within this domain but the SDTM 3.1.2 IG now places this within SUPPDM. The decision was made to leave in SC and not create a new domain for one value.

LB and VS Domains

The LB and VS domains contained the majority of challenges in transforming the data to SDTM. Below is a summary of issues the team dealt with.

Table 3. Summary of LB and VS Domain Issues

Domain	Issue	Solution
LB	The laboratory test codes were not standardized and included verbatim text	A standardized set of test codes were defined and the existing terms were mapped manually
LB	Units were not consistent across sites and included verbatim text	The team had to research the verbatim text and standardize the units reported within the data.
LB	Urinalysis data had very inconsistent values, units, and terms and a substantial amount of missing data	Given the effort it would take to clean up this data, the team decided to drop the Urinalysis from the data
LB	The laboratory values had a variety of different measurements within a single parameter	The team converted all laboratory values to a standard measure per parameter and stored the data within the *STRES* variables
LB	Normal ranges were scarce and not consistent	Team obtained published laboratory ranges by age and gender and use those values
VS	VS domain does not usually contain normal ranges	Since the team added abnormal vital signs to the analysis, published vital signs ranges were used and included in the domain
LB/VS	No baseline flags were included in the raw data and no analysis plans were available to determine an algorithm to define flags	The team discussed whether to define baseline flags within SDTM or ADaM. Team decided to create their own baseline flags within the SDTM LB and VS domains
LB/VS	All character variables were $200 including those defined as being limited within SDTM (e.g. LBTESTCD)	Looked across studies and set length appropriate to the data collected or IG requirements.

PC Domain

The PC domain was a new domain in SDTM 3.1.2 and the team had to evaluate it based on a draft status. Since a number of variables and associated metadata changed between the initial PC draft domain and the final SDTM 3.1.2 version, the team had to update those variables. The original data provided collected both planned and actual time variables within PC – variables not really defined within the SDTM specification.

EVENT DOMAINS

The event domains included AE, DS, and MH. Below is a summary of the issue identified within these domains.

Table 4. Summary of Event Domain Issues

Domain	Issue	Solution
AE	The adverse event terms were coded using a older coding system	The team recoded the adverse events using MedDRA 10.1 and stored the additional coding within SUPPAE
AE	The AE domain contained both regular adverse events as well as some abnormal laboratory events. The abnormal lab records did not contain other AE data (e.g. AEOUT)	Left records within AE domain with missing data but added values to AECAT to distinguish between records
DS	DSSTDTC was missing values but DSSTDY is present. Data just not available	Team decided to back populate the DSSTDTC variable to have consistent data
DS	DSTERM had a wide range of verbatim text	Created controlled terminology and manually mapped verbatim text where appropriate
MH	Very little data present with the exception of verbatim medical history term	The team mapped what they could. Domain has limited use.
ALL	All character variables were $200 including those defined as being limited within SDTM (e.g. LBTESTCD)	Looked across studies and set length appropriate to the data collected or IG requirements.

OVERALL CHALLENGES

Controlled Terminology

Throughout the domains there were differences in the controlled terminology within the values of a variable both across studies and in comparison to SDTM controlled terminology. Where possible, the team used the CDISC controlled terminology and mapped the raw terms or codes to those values. This information was documented as part of the metadata collected within the Pilot.

Mapping Raw Variables to SDTM

Throughout the raw data there were numerous occasions where the data expected within SDTM either didn't map very well or just didn't exist. For example, within Concomitant Medications instead of having separate raw variables that easily mapped to dose, frequency and unit, the raw data captured this information within a single text field with no standard structure. This made it nearly impossible to effectively map to the standard SDTM variables and made the domain somewhat useless. In this case the analysis group made the decision to not include this domain since it was not necessary for the final analysis. There were more examples similar to this issue and are common problems with the conversion of legacy data.

Derived Data

A core discussion that always arises when developing SDTM and ADaM domains is how much derived data should be maintained within SDTM domains. This discussion has been ongoing for many years as the SDTM model matured and pieces were added to support the medical review. This discussion continues as the FDA and CDISC try to ensure they are delivering what both the medical and statistical reviewers need to perform a complete review.

Within this Pilot, the team initially made the decision to limit derived data within SDTM as much as possible. However, as the Pilot work continued, the team decided that some derived data had to be present to ensure a complete review.

TRIAL DESIGN

One of the objectives of this Pilot was to implement the Trial Design domains, something that had not been done in the previous Pilot. The different trial designs within the Pilot and the complexity of the dose escalation multi phase trials made this an excellent sample to test the creation and use of the Trial Design domains. The double blind studies had three phases and within the double blind phase, there was a three day dose escalation followed by a fixed dose for the remainder days within this phase. A team member worked with the CDISC trial design group to review this study design and developed a robust set of trial design domains. The decisions made during this process were carried over to the development of the other domains (e.g. Epoch) which provided consistency. The use of the Trial Design domains will be evaluated during the FDA review of the Pilot.

ADAM IMPLEMENTATION

After completing the SDTM domains the next step was to define the ADaM specifications for each individual study as well as the integration of studies. The team used the ADaM implementation guide 1.0 which was in an early draft format when the project began. The team was fortunate enough to have a few team members who also served on the ADaM team allowing for an easy flow of questions and information between teams.

The team initiated the process of defining the ADaM specification by identifying the variables needed to support the analysis within the abbreviated analysis plans. As these discussions continued within the team and among other groups, the team realized the Pilot analysis data sets might be used for other intentions beside the somewhat narrowly focused analysis and would need to support other efforts. Given this realization, the team included additional variables within the analysis domains.

During the process of defining the ADaM domains, the team ran into a lot of the same challenges the ADaM team was struggling with in finalizing the ADaM 1.0 Implementation Guide. The team defined a subset of domains to use for the ADaM implementation which included ADSL, ADAE, ADLB, ADVS, and ADEX.

ADSL DOMAIN

The first step was to define the ADSL domain. This was the fundamental domain that would be used throughout the analysis and used to define the subset of variables which would be included in other domains. The definition of the standard variables expected in ADSL (e.g. RACE, AGE, SEX, etc.) was relatively straightforward. After this set of variables the process of defining additional variables became a bit more subjective and led to a significant amount of discussion.

Basic Variables

The team discussed what variables should be copied over from SDTM. The required variables were obvious such as SITEID, SUBJID, and AGE. After those required variables, it became very subjective as to which SDTM variables needed to be included. The discussion focused on whether the variables were used in the analysis, supportive, or the same as an analysis variable that was needed. In the end, the team probably included more variables than needed from SDTM – better to be safe than sorry.

In addition to the SDTM variables, other basic variables included the following:

- Numeric version of the SDTM variables (e.g. RACEN, SEXN)
- Specific baseline and endpoint subject level variables such as HEIGHTBL and HEIGHTEN
- Additional subject level variable from other domains such as DSDECOD from the DS domain and TANNER (a disease specific measurement) from the SC domain

While not all these variables were used in the final planned analyses they were examples of standard variables and associated naming conventions that might be used in ADaM.

Analysis Subgroups

The next set of variable defined within ADSL was a complete set of subgroup variables needed to support the summaries defined within the analysis plans. This was an iterative process as the team continued to refine and update the target analysis. Most of the discussion focused on how to capture a complete set of variables for each subgroup including the cumulative dose category, maximum dose category, and age group subgroups. Questions discussed within the Team included:

- What ranges to use for the various subgroups?
- How to name the various subgroup variables?
- Does each subgroup need the actual value, numeric group value, and character group value?

These questions, in addition to others, led to many different opinions on how to implement these variables. The final decision was to implement two variables for every subgroup analysis, a character variable containing the group and an associated numeric variable for sorting purposes. In parallel to the Pilot project, the ADaM team was changing the way grouping variables were defined from a GRP suffix to a GRy suffix where 'y' was numeric. In addition, the Pilot team wanted to have variables that were more descriptive. Based on these ongoing changes the team ended up with the following subgroup variables:

- AGEGRP and AGEGRPN – Age Group category variables
- DOSCUMGP and DOSCUMGN – Cumulative dose category variables
- DOSMAXGP and DOSMAXGN – Maximum dose category variables

Now that the ADaM IG has been released, these variables should probably be renamed to something like AGEGR1, DOSCMGR1, and DOSMXGR1. The detailed algorithms for these variables were collected in the metadata and captured within define.xml.

Treatment Variables

The team spent significant time discussing the best way to implement the treatment group variables. Some questions raised during this discussion included:

- Does a TRTxP variable need to exist for every phase of a trial or only those of interest in the analysis?
- Should the TRTSEQP variable include a combination of the TRTxP variable or display some other combination of values?

This was another example of the ADaM Implementation Guide not providing a precise definition on how to define these variables. In the Pilot a TRTxP variable was added for every phase whether or not it was used in the analysis. In addition the numeric version (TRTxPN) of the treatment was added and the associated start and stop date variables (TRTxSTDT and TRTxSPDT). In the case of the double blind dose escalation studies there were three phases within the study so nine variables were included in ADSL.

There was also discussion of the use of TRTSEQP. The implementation guide indicated it is required when there is a sequence of treatments that are analyzed, and if applicable, is required. There wasn't a clear definition of what should be contained in this variable which implies a sponsor definition. The team defined it as the concatenation of the TRTxP variables.

ADSL Flag Variables

Another challenge the team faced was the definition and implementation of analysis flags at the subject level. First, the basic population flags were included for the completers and safety population (COMPFL and SAFFL). Since all

subjects were included in the analysis no additional population flags were needed. In addition, specific analysis flags had to be defined for each subject. These flags are listed in Table 5.

Table 5. List of ADSL Subject Flags

Description	Variable
Flag if the subject had at least one serious adverse event	AESERANY
Flag if the subject had at least one adverse event	AEEVNTFL/ AEANY
Flag if the subject had at least one abnormal vital sign event	VSEVNTFL
Flag if the subject had at least one abnormal laboratory event	LBEVNTFL
Flag if the subject had at least one abnormal event (including adverse events, abnormal vital signs, or abnormal laboratory events)	ABEVNTFL

The team attempted to follow the implementation guide and use the FL suffix to define each of these flags. However, there was no need for the numeric equivalent so FN variables were not included. In addition, some 'ANY' suffix variables snuck into ADSL for adverse events. In the example above, two variables were accidently defined for the same "at least one adverse event" flag. Retrospectively, only one variable should have been included and the FL naming convention should have been used for all flags.

Core Variables

Another discussion the team had was around what variables from ADSL would be designated core variables and be carried over to all other ADaM domains. Even though it was determined that only variables used in analysis should be carried over, the actual implementation carried over more than that because the team wasn't sure if additional analyses might added or reviewers might want to perform ad-hoc analysis. This is a decision that a sponsor needs to make in collaboration with their FDA reviewer.

ADLB/ADVS DOMAINS

ADLB and ADVS were the two basic data structure (BDS) domains included in the ADaM specifications. Where possible the team used the general BDS rules and naming conventions. Along with ADSL, these two domains were discussed more than any other component within the Pilot.

Basic Data Structure Variables

The team included a number of standard variables outlined in the IG within these two domains. Some of these included:

- AVAL – Analysis value
- BASE – Baseline value within a parameter
- CHG – Change from baseline
- PCTCHG – percent change from baseline

The team also defined a shift variable that captured the change in vital signs and laboratory values from baseline to other time points. The team used the SHIFT variable within the initial draft IG which has now been changed to SHIFTy where y is a grouping.

In general, the implementation of the standard variable within the BDS structure was fairly straightforward. It was just a matter of deciding what needed to be included. The next two sections identify some lengthy discussion the team had regarding implementing certain concepts.

Record Level Flags

Within both ADVS and ADLB multiple analysis flags had to be defined to support the analyses within the analysis plans. The first flag was the record within a window that should be analyzed. The Pilot team had similar challenges that the ADaM team had regarding whether a record should be flagged based on whether it was analyzed or whether

it could be analyzed. At the end of the day the Pilot team made the decision to flag the record that was analyzed and chose the observation closest to the middle of the window. This variable was defined as ANLFL as per the draft IG. When the ADaM team released the final IG, they changed the naming algorithm so the name would be ANLzzFL where zz is 2-digit padded number to support multiple flags. The Team left the variable as ANLFL since it had already implemented and used within various analyses.

Another analysis flag had to be defined at the record level that flags a specific shift from baseline to that record. In this case, only certain shifts (e.g. low to high, middle to high) were identified as significant and needed to be flagged. This analysis flag, which was named ANL2FL, has a very different definition then the first record level flag. Again, if the team was complying with the final IG, this variable would be named ANL02FL.

Two record level flags were identified that were very different in their definition. In both cases the algorithms used for these two variables were captured as part of the metadata. The team did discuss the use of the *CRIT* variables to describe the second analysis flag since it seemed to fit the definition within the implementation guide. The ANLzzFL section within the implementation guide refers to ANLzzFL flags as flagging records as analyzable whereas the second flag variable the team defined was more of a criteria definition. The team decided to leave the variable as an ANLzzFL even though the CRIT variables should probably be implemented in this example.

Phases, Visits, and Windows

The team had a considerable discussion regarding the implementation of phase, visit, and visit windows. The draft IG did not clearly define how to capture phases of a trial within ADaM. It references the use of EPOCH from SDTM but does not address the need for an ADaM specific analysis phase. Initially the team defined an APHASE variable to capture this information but after discussions, APHASE was removed and replaced with EPOCH from SDTM. After further discussion the team realized that the Pilot studies had an analysis phase that did not really match the EPOCH from SDTM, and therefore defined AEPOCH (Analysis Epoch) to capture the required information. The ADaM team also had to address this issue and defined the following set of variables and associated date and time variables within the final IG.

- APERIOD – The numeric value characterizing the period to which the record belongs. The value of APERIOD must be consistent with the xx value in TRTxxP, TRTxxA, and all variables whose names begin with TRxx and APxx.
- APHASE – Generally, a higher-level categorization of APERIOD. Does not replace APERIOD, because APERIOD provides the indexing for the TRxx and APxx variables.

The team had a similar discussion defining visit variables. The question raised was whether the visit variables should be copied over and used if there are no differences in the algorithms for generating visit or should a new variable be created. The team decision was to append an ‘A’ to the visit variables and define the algorithm in the metadata as a 1-1 mapping. The ADaM IG provides an explanation of the AVISIT variable.

ADAE DOMAINS

The ADAE domain was implemented with some differences. In addition to the requirements of the analysis plans, the team identified a number of other variables that could be needed for reviewers as referenced in the safety review guidance. Based on this, ADAE had a number of additional variables. In addition, since ADAE doesn’t fit the BDS structure the team used a very early draft of the ADAE domain ADaM is currently working with. All the MedDRA coding terms including in SUPPAE were included as variables within the ADAE domain. In addition, two variables under consideration for the ADaM ADAE model BODSYCUR (Current Version Body System Organ Class) and DECODCUR (Current Version Dictionary-Derived Term) were included. These variables were meant to capture the most current coding system used if the study was coded in an older version of MedDRA.

ADEX DOMAINS

Just like ADAE there is no standard for capturing exposure information within ADaM so the team had to be more flexible in defining these domains. In general the team used the ADaM rules and guidelines for defining the exposure domain. In this case three exposure domains were defined to support the analysis. They included:

- ADEX (Exposure) – Contained one record per constant dosing interval per subject and was very similar to SDTM
- ADEXS (Exposure Summary) – Rolled up to contain one record per EPOCH per subject
- ADEXSSM(Study Exposure Summary) – Contained one record per subject with exposure summary information

In addition, these domains provided another challenge of how to define multiple sets of relative day variables where each set was relative to a different time point. The ADaM IG does not deal with this example so the team had to extend the general ADaM rules and take a best guess.

OVERALL CHALLENGES

Naming Conventions

One of the biggest challenges during the definition of the ADaM specifications involved those variables not clearly defined within the IG and the relationship between SDTM and ADaM. The ADaM IG is designed to allow for flexibility; however, this also can lead to challenges in defining variable name and roles and determining the relationship between SDTM and ADaM.

SDTM to ADaM

Similar to the discussion within the SDTM section regarding derived data and the issues already discussed, the same issue resurfaced during the definition of the ADaM Specifications. The team had a lot of the discussion focused on what variables should be copied from SDTM to ADaM, should the variable names and labels change, and how and when to store imputed records. Within the implementation guide it describes the possible need to copy over SDTM variable for supportive purposes. However, what if the SDTM variables are actually used for analysis? Should they copied over or should they be renamed or have a prefix of ‘A’ added?

Some team members thought the SDTM variables should be copied over if used in the analysis and maintain all attributes, while other members thought variables from SDTM should be very limited and analysis variable names should be used even if the values of the variables were just copied with no derivation. In the end, the team ended up with a mix of both options which they felt met the analysis needs.

Relative Day Variables

The ADaM IG gives some general guidelines for defining relative days but can lead to differences based on each person’s interpretation. SDTM defines relative day variables as well as ADaM, each of which could have a different definition. The question raised was whether SDTM variables should be copied over as is, or whether new variables should be defined even if the definition of the variable is the same. In addition, the team struggled to define the appropriate naming conventions given the implementation guide only defines that the variable must have a suffix of *DY. The Team finally made the decision to not copy over any SDTM relative day variables, and use the naming convention of including an ‘A’ in front of the relative day to indicate an analysis relative day (e.g. ADY, ASTDY, AENDY).

In addition to the examples above, the team had to decide how to define additional relative day variables that were based on other milestones independent of the start of study (e.g. EPOCH) as described in the ADEX section above.

ADAM INTERPRETATION

As the team worked through the implementing the ADaM specifications, a number of analysis concepts were identified that could not be answered by the existing IG. Either the concepts were not clearly defined or were not defined at all. The examples already provided give a flavor of some of the challenges the team faced when defining the analysis variables.

The primary lessons learned during the development of the ADaM specifications was that the ADaM IG doesn't answer all your questions, has some area of confusion, and might not cover all the questions users might have in defining an analysis standard. However, given the overall mission of ADaM and the fact that that team was using a draft IG some of these challenges can be expected. A lot of the issues addressed during this Pilot were resolved in the final ADaM IG. While the process was tedious it did lead to a very robust and standard set of analysis data sets.

INTEGRATION

After the team went through the extensive process of discussion and implementing the ADaM IG for the individual studies the next step was to define the specifications of the integrated data. The purpose of this exercise was to determine if using the ADaM standard at the study level made the process of defined integrated standards more efficient.

STUDY DESIGN

Within the Pilot data, half the studies were dose escalation studies with three epochs and half the studies were open label pharmacokinetic studies with one epoch. After significant discussions, the Analysis group made the decision that it didn't make sense to combine studies of different designs and use the same algorithms. Therefore, the studies with similar designs were combined first using algorithms described above. In the case of combining different designs slightly different algorithm had to be used.

One of the main analyses included a summary of abnormal events by phase. Since the study designs were different, when the data was combined across those different study designs, the team could not use the same epoch information. Instead, a simple summary of epoch was developed that only included before treatment and on treatment.

REDEFINING METADATA

In order to deal with limited resources, the decision was made to identify a small subset of integration components to address, instead of the originally proposed target deliverables. The focus was to redefine the analysis subgroups for the integrated analysis. Most of these changes were due to the differences between trial designs. These differences led to changes in how the subgroups for both the integrated data within a compound and across compounds would be defined.

Within the similar study designs the team focused on redefining two subgroups within the integration as well as addressing minor differences in the data formats. Both the Cumulative Dose Group and the Maximum Dose Group were redefined as quartiles based on the integrated data and not the quartiles as originally defined in the individual studies. In addition the following minor differences were identified within the data.

- One study only collected a date field while the other study collected datetime information. A 00:00:00 was added to the first study and an imputation flag was included as per the ADaM IG.
- AVISIT variables were adjusted slightly to deal with an additional visit within one study.

Besides these minor changes the merging of the ADaM domains was very transparent within the same study design.

The integration across all studies presented a different set of changes. One change was defining a new AEPOCH as described above. In addition, some minor changes had to be made to adjust for variables that had both data and datetime formats. The primary issue identified in the overall integration was how to deal with the TRT variables. The

double blind studies had defined three treatment periods and the associated variables while the open label studies only had one epoch. In this case TRT1P from each study did not mean the same thing as one represents a placebo run-in epoch and the other represents the double blind epoch. The team readjusted the TRT variables so there was some continuity in the variables.

ASSESSMENT

The team is currently developing the integrated study report. Again, due to constrained resources the team is writing one integrated report which will contain each of the similar study design integrations as well as the overall integration. Data and summary tables have been produced for each of three integrations. While the team is still completing the integrated study report, the standards have provided significant benefit in integrating the analysis data. The only caveat to this conclusion is that the team was using legacy data for this exercise so making the studies look 'similar' was an easier process then ongoing projects within a company's organization.

Even though the process of moving from individual studies to the integrated component was easier, there were a number of issues that were identified and had to be addressed. Therefore, while the use of standard made the process more efficient the ADaM domains from the individual studies could not just be 'set' together. The team did have to go through a number of discussions on how to address the issues described above.

SUBMISSON DEVELOPMENT

After the development of all the pieces and parts the final steps included the generation of the final reports and the compilation of all the components into an eCTD structure for delivery to the FDA. This work involved the development of the abbreviated clinical study reports, definition of the eCTD structure, and the creation of the define.xml.

ANALYSIS COMPONENTS

The Analysis group worked on developing an abbreviated Clinical Study Report (CSR) using the ICH guideline for the structure of content of a clinical study report. While the standard template was used a lot of sections were not applicable to the Pilot. While the content was limited, the team wanted to provide a structure familiar to the FDA. Each CSR has summary tables listed and the appropriate linking between the define file and the CSR.

ECTD STRUCTURE

Within the first Pilot only one study was submitted and therefore the process of developing the eCTD, defining the appropriate links between components, and creating the final package was fairly straightforward. In this case, this task became more complicated due to multiple studies and integrated components. Most of the discussion revolved around two general challenges.

The first challenge was to define how to link the Integrated data and associated define files within the individual studies. After much discussion, the team decided there was not an efficient way of providing this linkage so default standard eCTD rules were used for individual studies.

The second challenge was where to store the define files and associated style sheets. Normally the define files and style sheets are stored within the data. However, the team wanted to avoid copying style sheets in every directory. After consulting the team members with eCTD experience, it was determined that a higher level folder did not 'break' any eCTD rules, so a top level folder was created containing one set of style sheets. The define files were stored with the data as required.

DEFINE.XML

One of the most critical components of delivering the metadata for the Pilot was to ensure the define.xml provided a clear path of traceability from the SDTM domains for each study to the ADaM domains for each study to the overall integration.

Value Level Metadata

An gap identified within the first Pilot with the FDA was the inability to capture value level metadata within define.xml – information about one variable based on the value of another variable. An example would be how the variable VSPOS would be defined if the value of VSTESTCD="SYSBP". Value level metadata helps define the attributes of VSPOS based on the value of VSTESTCD.

The team worked in collaboration with the CDISC XML Technologies team to identify a method for capturing this additional metadata which was critical for providing a robust set of information about the data. The team initially came up with an extension to define.xml that seemed to support the requirements. However, after going through a prototype implementation, the team determined the extension did not support all the metadata required. The team went back to the drawing board and was able to define an approach that would not require any additions to the define.xml schema. Below is an example of the rendering as well as the define.xml. This is only a **draft** and has not been finalized.

This figure shows the reference to the LBTESTCD variable. Note the dynamic link to LBTESTCD.

Variable	Label	Key	Type	Length	Code List / Controlled Terms	Derivation	Origin	Role
USUBJID	Unique Subject Identifier	1	text	22				Identifier
VISITNUM	Visit Number	2	float	8				Timing
LBDTC	Date/Time of Specimen Collection	3	text	19				Timing
LBTESTCD	LAB Test or Examination Short Name	4	text	8	LBTESTCD	LB_LBTESTCD		Topic

This link then displays the value level metadata for LBCAT when LBTESTCD='ALB' and has links to an associated virtual code list and derivation.

Parameter Value	Parameter Related Variable	Type	Length	Format	Code List / Controlled Terms	Derivation
Context: In LB where LBTESTCD is ALB	LBCAT	text	20		LBCATBC	LB_LBCAT_PV

Finally, the virtual code list and derivation can also be displayed.

LBCATBC, Reference Name (CODELISTC14)
Valid Values
BLOOD CHEMISTRY

LB_LBCAT_PV	Computation	Equal to LBNORMALS.LBCAT merged on LBTESTCD

Below is the actual define.xml text being proposed to trace the metadata down to the value level.

```
** Variable ItemDefs **
<ItemDef OID="COL5" Name="VSTESTCD" DataType="text" Length="8" Origin="Derived">
     <CodeListRef CodeListOID="CODELIST2"/>
     <cp02:ValueListRef ValueListOID="VALUELIST1"/>
</ItemDef>
<ItemDef OID="COL7" Name="VSPOS" DataType="text" Length="200" Origin="CRF”>
     <cp02:ValueListRef ValueListOID="VALUELIST2"/>
</ItemDef>
```

```
** Codelist for the VSTESTCD variable **
<CodeList OID="CODELIST2" Name="VSTESTCD" DataType="text" SASFormatName="VSTESTCD">
  <EnumeratedItem CodedValue="WEIGHT" />
  <EnumeratedItem CodedValue="HEIGHT" />
</CodeList>

** ValueLists for VSTESTCD/VSPOS.  Note the order relative to the VSTESTCD Codelist **
<cp02:ValueListDef OID="VALUELIST1">
   <ItemRef ItemOID="PRELCOL1" Mandatory="No" OrderNumber="1"/>
   <ItemRef ItemOID="PRELCOL2" Mandatory="No" OrderNumber="2"/>
</cp02:ValueListDef>
<cp02:ValueListDef OID="VALUELIST2">
  <ItemRef ItemOID="PRELCOL3" Mandatory="No" OrderNumber="1" MethodOID="VLMETHOD001"/>
  <ItemRef ItemOID="PRELCOL4" Mandatory="No" OrderNumber="2" MethodOID="VLMETHOD002"/>
</cp02:ValueListDef>

**Computational algorithms for VSPOS when VSTESTCD equals a certain value **
<MethodDef OID="VLMETH001" Name="VSPOS_HEIGHT_PV" Type="Computation">
   <Description>
     <TranslatedText xml:lang="en">Here is derivation one </TranslatedText>
   </Description>
</MethodDef>

<MethodDef OID="VLMETH002" Name="VSPOS_WEIGHT_PV" Type="Computation">
   <Description>
     <TranslatedText xml:lang="en">Here is the derivation two</TranslatedText>
   </Description>
</MethodDef>
```

Analysis Metadata Results

Within the first Pilot, the concept of Analysis Metadata Results was introduced and well received by the FDA reviewers. This includes metadata about the analysis tables which can be included within the define.xml and provides a flow from the summary table to the Clinical Study Report, the input tables, and the associated algorithms. Within the first pilot, this was implemented as a one off extension to the define specification. Based on this prototype the XML Technologies team, in collaboration with the ADaM team, has defined a formal draft specification which the Pilot team has implemented and tested. The example below contains both the implemented XML and the rendering of the proposed analysis results metadata. This example takes on a more tabular format as defined by the ADaM team but can rendered in any format based on a custom style sheet.

```
<Study OID = "ADaM Results Standard Example1">
   <GlobalVariables>
       <StudyName>Define Analysis Results Metadata Example</StudyName>
       <StudyDescription>Example of ADaM Analysis Results Metadata</StudyDescription>
       <ProtocolName>ADaM Standard</ProtocolName>
   </GlobalVariables>
   <MetaDataVersion OID = "CDISC.ADAM.MDV1" Name = "ADaM_Standard, Data Definitions"
      Description = "Study ADaM_Standard, Data Definitions"
      crt:DefineVersion="1.0.1" crt:StandardName="CDISC SDTM"
      crt:StandardVersion="3.1.1">
   <adamref:AnalysisResults DisplayIdentifier = "Table 14-3.01" DisplayName = "Primary
       Endpoint Analysis:ADAS Cog(11) - Change from Baseline to Week 24 - LOCF"
       ResultIdentifier = "Analysis of dose response" Reason = "Pre-specified in
       Protocol">
       <adamref:Parameter ParamCD = "ACTOT11" Param = "ADAS-Cog (11) Total Score"/>
       <adamref:AnalysisVariable Name = "CHG"/>
```

```
    <adamref:AnalysisDataset Name = "ADQSADAS"/>
    <adamref:SelectionCriteria>EFFFL='Y' and ITTRFL='Y' and AVISIT='Week 24' and
          PARAMCD='ACTOT11'</adamref:SelectionCriteria>
    <adamref:Documentation> <TranslatedText xml:lang = "en">SAP_SEC_10.1.1,
    SAP_TEMP_5. Linear model analysis of dose response for the  ADAS-Cog(11) total
    score change from baseline at Week 24 - missing values imputed using LOCF,
    Efficacy population. Used PROC GLM in SAS to produce p-value (from Type III SS
    for treatment dose); Independent terms in model are TRTDOSE (randomized dose: 0
    for placebo; 54 for low dose; 81 for high dose)  SITEGRP (site group, as a
    class variable) and BASE (baseline ADAS-Cog score).</TranslatedText>
    </adamref:Documentation>
    <adamref:ModelSpecification>PROC MIXED; CLASS USUBJID SITEGRP AWEEK TRT1P;
    MODEL CHG = TRT1P SITEGRP AWEEK TRT1P*AWEEK BASE BASE*AWEEK / OUTP=PRED
    DDFM=KR; REPEATED AWEEK / SUBJECT=USUBJID TYPE=UN; LSMEANS TRT1P / DIFF CL;
    RUN;</adamref:ModelSpecification>
</adamref:AnalysisResults>
```

Table 14-3.2	Exposure by Age Category (Safety Population)	Summary Statistics of Maximum Dose	DOSMAX	Pre-specified in SAP	ADSL	SAFFL eq 'Y'; class AGEGRPN;
Table 14-3.3	Exposure by Gender (Safety Population)	Summary Statistics of Cumulative Dose	DOSCUM	Pre-specified in SAP	ADSL	SAFFL eq 'Y'; class SEXN;
Table 14-3.3	Exposure by Gender (Safety Population)	Summary Statistics of Maximum Dose	DOSMAX	Pre-specified in SAP	ADSL	SAFFL eq 'Y'; class SEXN;
Table 14-4.1.1	Total Treatment Emergent Events Based on Patients Experiencing the Event (Safety Population)	Frequency of Adverse Events	DECODCUR	Pre-specified in SAP	ADAE	"trtemrfl='Y' and saffl=""Y"""

Location

There has never been a clear recommendation on how to physically store the define files within the eCTD structure.

- Should you have one define file for each study containing all the metadata?
- Should you have separate files for SDTM and ADaM?
- Where should the analysis results metadata been contained?

Within the first pilot a number of issues were identified with placing all the metadata in one file. First, the rendering of the define file created errors when a certain maximum number of external links was reached. Second, the define file had to be copied and stored within both the tabulations and analysis folders.

Initially the packaging subgroup decided to create three define files: 1) SDTM metadata 2) ADaM metadata and 3) Analysis Results Metadata. After further discussion, the team realized the Analysis Results Metadata would have to be linked to the ADaM domains and it would be easier if this information was stored together. In addition, the eCTD does not allow two define files within the analysis folder. The final decision was to include one define file containing the SDTM metadata stored with the SDTM domains, and one define file containing ADaM metadata and the analysis results metadata.

In addition to the rendering of the define file using style sheets, a PDF rendition was also created to provide a printable version. This request came from the reviewers of the first Pilot who indicated they would like a printer friendly version.

FDA FEEDBACK

The Pilot team delivered the first interim package in December to a group of FDA reviewers who volunteered to provide feedback. This package contained the first two studies including SDTM, ADaM, and define.xml within an eCTD structure. Overall, the reviewers were able to easily navigate the submission and locate all the components.

They did provide comments regarding the limited amount of information available and thus did not reflect a true submission. The team had explained the limitations of the studies provided and reminded the reviewers of this challenge.

They had no technical issues opening and navigating the define.xml as well as the associated documents and data sets referenced. They did comment that the define.xml did not seem to provide enough detail and that they might need 'another define for the define'. The comment implies the define.xml should contain very clear and documented metadata. They also comments that the PDF rendition of the define.xml was very useful for printing out certain pieces they would want to read on paper.

Beyond those assessments, the reviewers have not had the opportunity to really spend time reviewing the data components in detail. They were able to load the data into their standard tools which expect SDTM. Once the SDTM data was loaded a few reviewers spent a bit of time reviewing the SDTM domains. Most of the comments provided were not specific about the Pilot but more general limitations they have identified with SDTM in the past. The primary issues identified included:

- Reviewing and working with the SUPPQUAL data sets can be challenging without the tools to support it
- Limited ability to support multiple baselines within the SDTM domains
- SDTM guidance for not supporting more verbatim text from investigators

The FDA reviewers are still working through an assessment of the ADaM domains. The only primary finding so far has been some confusion over the purpose and definition of the analysis flags within ADaM both at the subject and record levels. Once the team explained these concepts the reviewers had a better understanding of their use.

The Pilot team will continue to meet with FDA reviewers over the course of next month to gather more feedback on the current package as well as the integrated package the team plans to send in April. The team will continue to collate this information and publish a complete set of findings from the FDA team in the project report.

SUMMARY AND NEXT STEPS

The CDISC/FDA Pilot team has been working diligently over the last two years to work through a wide range of issues and questions encompassing the implementation of new and ongoing standards. The team completed the following deliverables during the course of the project:

- SDTM domains and ADaM for four of the studies
- Summary Tables and Clinical Study Reports for the first two studies
- Statistical Analysis Plans for the individual studies as well as the integrated studies
- Integrated ADaM domains for similar study designs as well as an overall integration within a compound
- Integrated Clinical Study Report summary all the integrations

The team delivered an interim package to the FDA in December of 2009 and held multiple meetings to gather FDA feedback. While this review is ongoing, the team built the integrated data as well as the integrated study report. The team plans to deliver this package to the FDA in early April for review.

The Pilot team hopes to make the entire package available to the public by July including an extensive report of the Pilot's findings. As with any volunteer effort, the ability to deliver in a timely fashion can sometimes be a challenge but the hope is that the information provided will help organizations improve their use of standards and avoid common pitfalls.

SUMMARY

In 2008 CDISC released the technical roadmap which included an objective to plan and implement pilots to gain a better understanding of the needs of both industry and regulatory groups. This Pilot was initiated to test the latest CDISC models and their ability to facilitate the review process and support the integration of data. While the Pilot is

ongoing, it has continued to provide benefit. Internal to CDISC, the Pilot has provided a real world environment to test the standards to ensure they are meeting the needs of all customers and has provided invaluable input back to the CDISC teams. With the collaboration between CDISC and the FDA it has provided the FDA an opportunity to assess the capability to use standards to improve the review process. However, for Pilots to be successful in the future, CDISC has to work towards defining smaller scale activities that pilot small goals within a more narrow scope.

The adoption of the CDISC standards will continue across the industry and within regulatory bodies to improve the drug development process, and pilots that help these two customers collaborate will only improve the efficiency of drug development.

REFERENCES

Gregory Steffens, Principal Research Scientist, Eli Lilly & Co.,Ian Fleming, Senior Statistical Programmer Analyst,Genentech." The Second CDISC Pilot Project: A Metastandard for Integrating Databases". SAS Global Forum, March, 2009; PharmaSUG, May, 2009. http://support.sas.com/resources/papers/proceedings09/171-2009.pdf

ACKNOWLEDGEMENTS

The co-leaders of the Pilot would like to thank all the Pilot team members for their ongoing and exemplary work on this project. Without their effort, CDISC and the Pilot would not be successful.

CONTACT INFORMATION

Chris Decker, Life Sciences Director
d-Wise Technologies
Work Phone: 919-600-6234
E-mail: cdecker@d-wise.com
Web: www.d-wise.com