

# Managing the CRO Relationship to Effectively Request and Receive CDISC STDM and ADaM Deliverables

Kenneth Eng, Shire Pharmaceuticals  
David C. Izard, Shire Pharmaceuticals  
Terek J. Peterson, Shire Pharmaceuticals

## ABSTRACT

The FDA has recently retracted the "1999 Guidance" that effectively took the "bring your own data schema" for regulatory submissions off the table. In its place, the "2005 Guidance", revised in 2006 [1], specifies the incorporation of STDM (required) and ADaM (strongly suggested) datasets within the eCTD structure. Use of these is mandatory on or after January 1st, 2008. A close examination of the contents of these deliverables, however, reveals that the contents of each of these data schemas does not lend itself to the traditional roles and responsibilities at a CRO for producing these items. This paper explores how to effectively specify these deliverables to facilitate the accurate and timely development, implementation and delivery.

## INTRODUCTION

Shire Pharmaceuticals is the epitome of a virtual company. Almost every operational task is outsourced, from discovery to medical monitoring to data management to statistics to medical writing. With the exception of our genetic therapy division, every product we develop has been purchased from an outside source, via in-licensing or acquisition. The key aspect of this largely outsourced environment is that we have very little infrastructure in terms of systems or data expectations, so to shift from one paradigm to another is not a big leap compared to what it might be at an organization that has a significant internal architecture.

This minimal infrastructure is a double-edged sword when it comes to industry paradigm shifts. On the one hand, little or no effort goes into changing our internal infrastructure to receive information in a new format; on the other hand, there is little or no human resource available internally to change legacy assets into useful and/or compliant entities to support future regulatory work. With the onset of the eCTD and CDISC being at the core of US Food and Drug Administration (FDA) regulatory submissions, we see the need to evaluate our portfolio of products and determine at what point we are willing to invest both time and money into migrating existing assets into CDISC complaint formats. The remaining domain, studies or development programs that will start from this point forward, are also evaluated for suitability for CDISC implementation based on likely future use of the trial assets once the development is complete.

The move to CDISC and the push to do it now rather than later has a number of supporting factors at Shire. By investing now in the implementation of CDISC based data structures and data capture tools, we can use this as the standard when putting data management, biostatistical and statistical programming work out to bid. Up front utilization of the standard will lead to reduced time and cost downstream when combining and organizing assets for regulatory submission. With in-licensing virtually all compounds comes added timeline pressure. The average time from our acquisition of a compound to filing an NDA is on the order of 2 years over the past three calendar years; we typically acquire compounds that have just initiated or completed Phase IIa clinical trials.

## BACKGROUND

Ever since the FDA included basic instructions back in 1988 [2], the world of electronic submissions has been both evolutionary and heuristic. The FDA, sponsor companies and service providers dabbled in many experiments, learning from each one. First came the CANDAs (Computer Assisted New Drug Applications), basically loading up enhanced electronic versions of your paper documents onto computers and literally delivering the hardware along with a manual to the agency instructing them on how to navigate this custom application. From this the FDA learned that hyper linking related documents and domains together accelerates a review but there was a high cost involved on both the sponsor preparing a CANDA and the FDA learning a new environment for each submission. This led to "bridge guidance", CDER's 1997 guidance on electronically archiving submissions previously produced on paper [2b], followed two years later by the infamous "1999 Guidance", the first attempt at specifying electronic deliverables that is prepared on the sponsor / service provider platform and subsequently delivered to the FDA electronically to run on their platform, including data and documentation.

The FDA raised the bar again in 2005 (draft, current "final" version dated April 19<sup>th</sup>, 2006) [1], issuing its guidance on submitting electronically to the FDA using the eCTD format. This guidance provides a framework to interpret ICH's CTD guidance, adding specifics of what the FDA wants and a mapping from the old NDA structure to the eCTD. Specific to the FDA implementation, as of now, is the requirement to submit datasets (Sic., Section III.E.4, pg. 15) and directing you to the companion document [2d] which contains specific reference to the use of CDISC SDTM to format your datasets as well as other components of CDISC required for implementation. This guidance, coupled with the announcement of withdrawal of the "1999 Guidance" and subsequent updates to the 1999 Guidance in 2002 and 2003 on September 29, 2006 [2e] with removal

of references to these guidance documents from the FDA docket on December 31, 2007, forces sponsors and service providers to deliver data electronically to the agency on or after January 1, 2008 within the eCTD structure using CDISC SDTM as the basis for organizing your data.

It should be noted that, while the electronic submission of data could be universal, it has been to date an FDA centric activity as other regulatory agencies have not yet dictated their own standard or endorsed the use of an established standard as part of the regulatory submission process. Even geographic areas and countries such as the EU and Japan, members of the ICH governing body, have not delineated a location within the eCTD to provide data and don't intend on doing so in the near future. There are limited places where data is submitted to a regulatory agency, such as the submission of biopharmaceutical data for bioequivalence studies included in an NDS to Health Canada [3], but these are special purpose submissions that do not constitute the basis for a comprehensive evaluation of a pharmaceutical or medical device application.

## **APPROACH**

### **TRADITIONAL IMPLEMENTATION PLAN**

The traditional method of database deliverables starts with Shire Data Management providing CRO Data Management database specifications in forms of database specifications and/or proc contents based on previous studies. CRO Management programmer generally designs the database using internal standards or specifications in CRO data management system (i.e. Oracle Clinical, ClinTrial etc.). The programmer generates a conversion specifications document describing how the final clinical database structure, based on database specifications provided by Shire, will be created from CRO's clinical database. Once this document is reviewed and approved, the CRO programmer will generate and validate the programs. The locked clinical database includes CRF data and third party vendor data (i.e. Labs, ECGs, etc) plus coding. The locked database is secured for use by CRO Biostatistics Group. CRO Biostatistics produces Analysis Dataset Specifications, Statistical Analysis Plan, and Programming Rules Documents to detail how CRO Biostatistics Programmer generates derived datasets.

### **CDISC IMPLEMENTATION PLAN**

Data Manager, Statistician and Statistical Program conducted a thorough review of the contents of SDTM and AdaM datasets to develop dataset specifications for all domains that would be utilized by Shire. These dataset specs will be sent to partner CRO for development of CDISC compliant database with minor updates specific to a particular study. An example of AE dataset specifications is shown in Table 1. The CRO Database Programmer generates a conversion specifications document describing how the final clinical database structure, based on Shire's specifications. Once this document is reviewed and approved, the CRO programmer will generate and validate the programs. The locked clinical database includes CRF data and third party vendor data (i.e. Labs, ECGs, etc) plus coding. The clinical database only includes any SUPP QUAL data specific to the study. The locked database, called SDTM minus, is secured for use by CRO Biostatistics Group. CRO Biostatistics produces Analysis Dataset Specifications, Statistical Analysis Plan, and Programming Rules Documents to detail how to generate SDTM data that includes basic derivations and Trial Design dataset if applicable

- A method of approaching this problem that will allow you to highlight the issues within each data schema, identify the individuals who are most capable of providing the information correctly, and a business process for making it happen.

This figure (note: COPYRIGHTED...):

FIGURE 1

# Traditional Implementation Plan

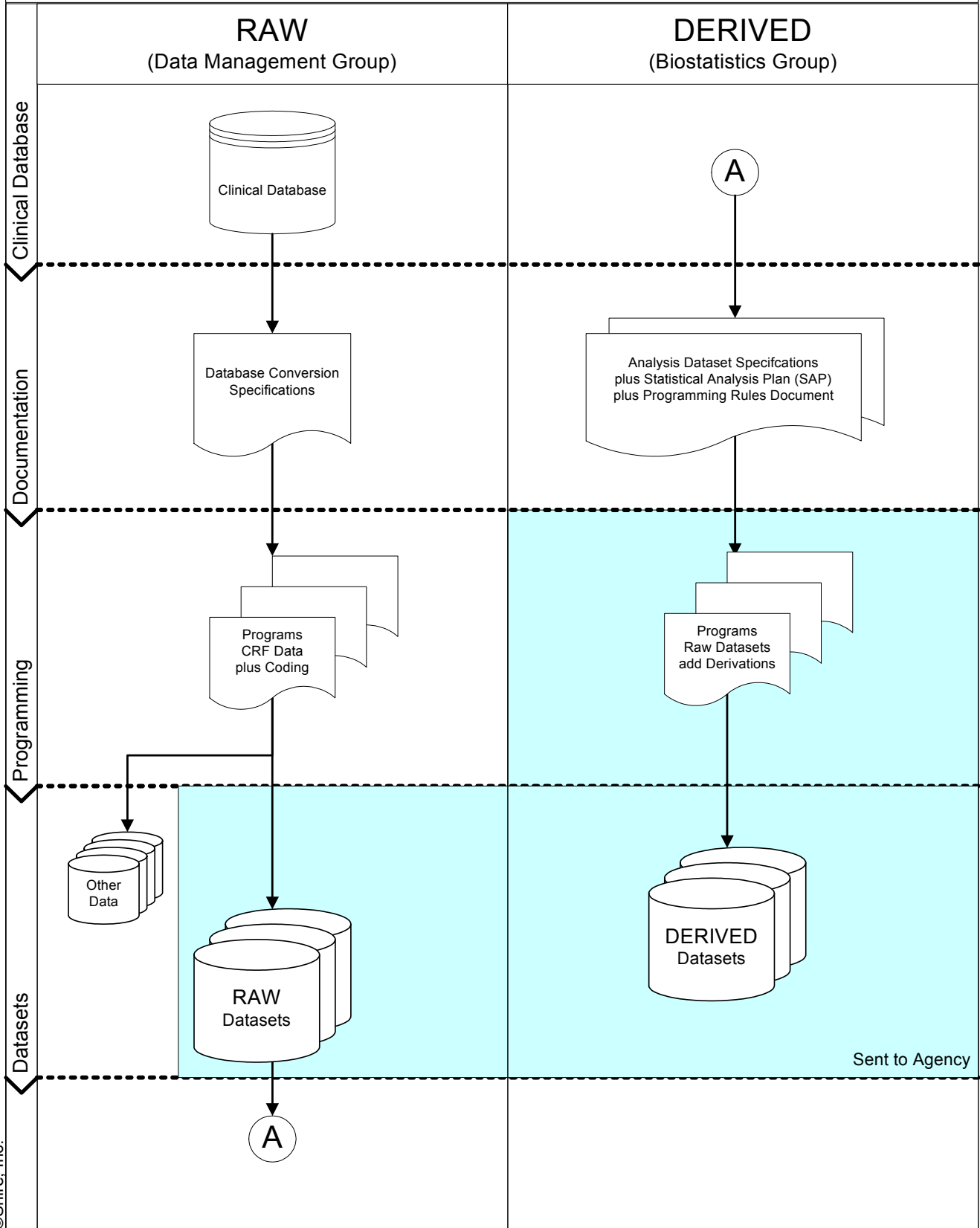
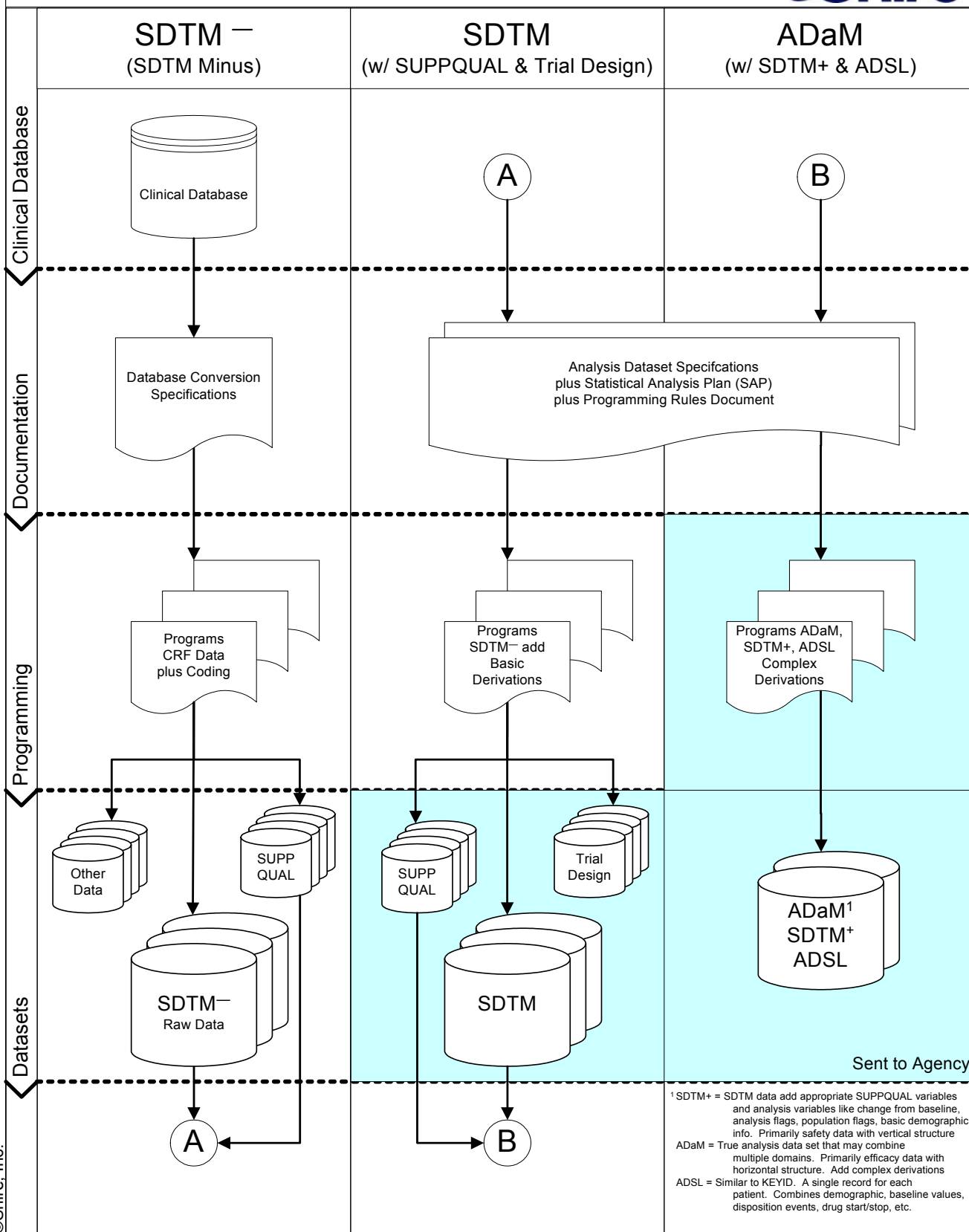


FIGURE 2

# CDISC Implementation Plan



## SDTM STRUCTURE FOR RAW CRF DATA

The end goal of a clinical trial is to submit to the FDA datasets that conform to the SDTM structure. As a rule of thumb when delivering CDISC datasets all required and expected data variables should be delivered. The permissible variables will be provided when appropriate or specified by Data Management or Biostatistics. According to the SDTM the concept of core variable is used both as a measure of compliance and to provide general guidance.

### SDTM <sup>-</sup> (SDTM Minus)

What is SDTM <sup>-</sup> (SDTM Minus)? Basically SDTM <sup>-</sup> dataset domains are SDTM domains less derived variables. SDTM <sup>-</sup> dataset domains should contain all the CRF data plus any coding (i.e. MedDRA, WhoDrug). These are the responsibilities traditionally completed by a data management group. As stated earlier all remaining obligatory and appropriate permissible variables will later be added by an additional programmer, traditionally in the biostatistics group, to create a complete SDTM package.

As can be seen in the appendix graph titled "CDISC Implementation Plan", the creation of SDTM <sup>-</sup> from the clinical database to Point A would traditionally be created by a data management group. The CRF data is entered into a clinical database. Database Conversion Specifications are written to describe how datasets are converted, coded, and created from the clinical database. Programs or systems are put in place to handle the dataset creation. From these programs or systems three primary groups of datasets are created:

1. **SDTM <sup>-</sup>** raw dataset domains
2. **SUPPQUAL** datasets will be provided for data domains as needed
3. **OTHER** datasets maybe needed during the conduct of the study but will not be submitted to the FDA

These three primary sets of datasets will be delivered to Data Management for review. This set of data should be sufficient to complete any scheduled data reviews during the conduct of the study. These datasets will not be submitted to the FDA but are the building blocks for the creation of the SDTM dataset domains.

## SDTM STRUCTURE FOR ANALYSIS DATA

Remember the end goal of a clinical trial from a data perspective is to submit to the FDA datasets that conform to the SDTM structure. At this point all remaining obligatory and appropriate permissible variables will be added to the SDTM <sup>-</sup> datasets to create a complete SDTM package including the trial design datasets. Only basic derivations should be needed to create the remaining variables. Traditionally, the most appropriate staff to do this is in the biostatistics/statistical programming group. This is due to the fact the programming logic to create even basic derivations, such as baseline value or study day, will be defined in a SAP, Analysis Dataset Specifications, and/or a Programming Rules Document.

### SDTM Including Trial Design

As can be seen in the appendix graph titled "CDISC Implementation Plan", the creation of SDTM from Point A to Point B would traditionally be completed by a biostatistics/statistical programming group. An analysis programmer would take the raw CRF datasets (SDTM <sup>-</sup> datasets), in conjunction with the SAP, Analysis Dataset Specifications, and/or a Programming Rules Document to create programs. These programs would be used to add basic derivations to the SDTM <sup>-</sup> datasets to create the SDTM dataset domains of special purpose, interventions, events, and findings.

There are various ways to create the trial design dataset domains. One technique is to create some of the trial design domains from spreadsheets. These dataset domains created from spreadsheets can be used in conjunction with the SDTM <sup>-</sup> datasets to create the trial design dataset domains containing subject level data.

From the programs that take the SDTM <sup>-</sup> datasets then add basic derivations and the programs that create the trial design dataset domains, three primary groups of datasets are created:

1. **SDTM** CDISC compliant dataset domains
2. **SUPPQUAL** datasets will be provided for data domains as needed
3. **TRIAL DESIGN** datasets to describe a the conduct of a study

These three primary groups of datasets will be delivered to a sponsors biostatistics group for review. These datasets will be ultimately submitted to the FDA and are critical building blocks for the creation of the ADaM datasets used for analysis.

## ADaM STRUCTURE FOR ANALYSIS DATA

The ADaM specifications state that analysis datasets should be submitted in a way that they can be analyzed with little or no programming or complex data manipulations. Such datasets are said to be “Analysis-ready” or “One Statistical Procedure Away” from the statistical results. This approach eliminates or greatly reduces the amount of programming required by the statistical reviewers. Analysis-ready does not mean that a formatted table can be generated in a single statistical procedure or PROC. Rather it means that each statistic in the table can be replicated by running a standard statistical procedure (SAS PROC, S-PLUS function...) using the appropriate analysis dataset as input. Reviewers can then replicate and explore these results with little or no data manipulation, allowing them to concentrate on the results, not on programming.

### ADaM, SDTM<sup>+</sup> (SDTM Plus), and ADSL

One approach to understanding ADaM is to see the analysis-ready datasets as three different groups: ADaM, SDTM<sup>+</sup> (SDTM Plus), and ADSL. All analysis datasets will begin with the letters “AD” and follow the CDISC ADaM specifications. However, the structure desired for different data are explained in further detail below.

In general, an ADaM dataset is a true analysis dataset that may combine multiple SDTM datasets. An ADaM dataset will generally be used for efficacy data (e.g. YMRS, MADRS, HAM-D) and will be horizontal in form. An ADaM dataset will need to be specified and agreed upon by the statisticians and FDA.

In order to simplify the creation of analysis-ready datasets that will be sent to FDA the concept of SDTM<sup>+</sup> (SDTM Plus) is introduced. A SDTM<sup>+</sup> dataset is a SDTM dataset with additional variables added. In general, a SDTM<sup>+</sup> dataset will generally be used for safety data and long efficacy questionnaires (e.g. LABS, VITALS, ECGS, SF36) and will be vertical in form. A SDTM<sup>+</sup> dataset can be created by adding appropriate SUPQUAL variables (if they exist) and analysis variables like change from baseline, analysis flags, population flags, and basic demographic information. Follow the ADaM specifications when naming variables.

The ADSL dataset is a single dataset that contains one record per patient. This subject-level analysis dataset will contain all of the variables that are important for describing the target population to whom the study results are generalizable. These variables will include data that either describe the subjects or events in a clinical trial prior to treatment, or that group the subjects or events in some way for analysis purposes. This will include critical demographic and baseline characteristics of the subjects, as well as other factors arising during the study that could affect response. Specifically include analysis population flags.

**TABLE 1**  
**Example Of Dataset Specifications Sent To The CRO**

| Domain | Name     | Label                               | Group    | Order | Type | Length | Origin  | Core | ControlledTerms                                  |
|--------|----------|-------------------------------------|----------|-------|------|--------|---------|------|--|
| AE     | STUDYID  | Study Identifier                    | SDTM-    | 1     | Char | 9      | CRF     | Req  |  |
| AE     | DOMAIN   | Domain Abbreviation                 | SDTM-    | 2     | Char | 8      | Derived | Req  | **AE   |
| AE     | USUBJID  | Unique Subject Identifier           | SDTM-    | 3     | Char | 15     | Derived | Req  |  |
| AE     | AESEQ    | Sequence Number                     | SDTM-    | 4     | Num  | 8      | Derived | Req  |  |
| AE     | AEGRPID  | Group ID                            | SDTM-    | 5     | Char | 100    | Derived | Perm |  |
| AE     | AEREFID  | Reference ID                        | NOT USED | 6     | Char | 40     | CRF     | Perm |  |
| AE     | AESPID   | Sponsor-Defined Identifier          | NOT USED | 7     | Char | 8      | CRF     | Perm |  |
| AE     | AETERM   | Reported Term for the Adverse Event | SDTM-    | 8     | Char | 200    | CRF     | Req  |  |
| AE     | AEMODIFY | Modified Reported Term              | NOT USED | 9     | Char | 200    | Derived | Perm |  |
| AE     | AEDECOD  | Dictionary-Derived Term             | SDTM-    | 10    | Char | 200    | Derived | Req  | **   |
| AE     | AECAT    | Category for Adverse Event          | NOT USED | 11    | Char | 100    | Derived | Perm | *  |
| AE     | AESCAT   | Subcategory for Adverse Event       | NOT USED | 12    | Char | 100    | Derived | Perm | *  |
| AE     | AEOCCUR  | Adverse Event Occurrence            | NOT USED | 13    | Char | 1      | CRF     | Perm | **Y, N or NULL                                   |
| AE     | AEBODSYS | Body System or Organ Class          | SDTM-    | 14    | Char | 100    | Derived | Exp  | **   |
| AE     | AELOC    | Location of the Reaction            | NOT USED | 15    | Char | 100    | Derived | Perm | *  |
| AE     | AESEV    | Severity/Intensity                  | SDTM-    | 16    | Char | 8      | CRF     | Perm | *MILD, MODERATE, SEVERE                          |
| AE     | AESER    | Serious Event                       | SDTM-    | 17    | Char | 1      | CRF     | Exp  | **Y, N   |
| AE     | AEACN    | Action Taken with Study Treatment   | SDTM-    | 18    | Char | 50     | CRF     | Exp  | *NONE, INTERRUPTED, DISCONTINUED, DECREASED DOSE |

|    |          |   |          |      |      |     |         |      |  |
|----|----------|---|----------|------|------|-----|---------|------|--|
| AE | AEACNOTH | Other Action Taken                      | SDTM-    | 19   | Char | 50  | CRF     | Perm | *NONE, PHARMACOLOGIC, NON-PHARMACOLOGIC, PHARMACOLOGIC/NO N-PHARMACOLOGIC      |
| AE | AEREL    | Causality                               | SDTM-    | 20   | Char | 15  | CRF     | Exp  | *RELATED, NOT RELATED  |
| AE | AERELNST | Relationship to Non-Study Treatment     | NOT USED | 21   | Char | 200 | CRF     | Perm |  |
| AE | AEPATT   | Pattern of Event                        | SDTM-    | 22   | Char | 20  | CRF     | Perm | *ONE EPISODE, INTERMITTENT, CONTINUOUS   |
| AE | AEOUT    | Outcome of Adverse Event                | SDTM-    | 23   | Char | 25  | CRF     | Perm | *RESOLVED, UNRESOLVED, SEVERITY CHANGE, RESOLVED WITH SEQUELAE, DEATH, UNKNOWN |
| AE | AESCAN   | Involves Cancer                         | NOT USED | 24   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESCONG  | Congenital Anomaly or Birth Defect      | NOT USED | 25   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESDISAB | Persist or Signif Disability/Incapacity | NOT USED | 26   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESDTH   | Results in Death                        | NOT USED | 27   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESHOSP  | Requires or Prolongs Hospitalization    | NOT USED | 28   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESLIFE  | Is Life Threatening                     | NOT USED | 29   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESOD    | Occurred with Overdose                  | NOT USED | 30   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AESMIE   | Other Medically Important Serious Event | NOT USED | 31   | Char | 1   | CRF     | Perm | **Y, N or NULL   |
| AE | AECONTRT | Concomitant or Additional Trtmnt Given  | NOT USED | 32   | Char | 1   | CRF     | Perm | **Y, N   |
| AE | AETOXGR  | Standard Toxicity Grade                 | NOT USED | 33   | Char | 200 | CRF     | Perm | *  |
| AE | AESTDTC  | Start Date/Time of Adverse Event        | SDTM-    | 34   | Char | 19  | CRF     | Exp  | ISO 8601   |
| AE | AEENDTC  | End Date/Time of Adverse Event          | SDTM-    | 35   | Char | 19  | CRF     | Exp  | ISO 8601   |
| AE | AESTDY   | Study Day of Start of Adverse Event     | SDTM     | 36   | Num  | 8   | Derived | Perm |  |
| AE | AEENDY   | Study Day of End of Adverse Event       | SDTM     | 37   | Num  | 8   | Derived | Perm |  |
| AE | AEDUR    | Duration of Adverse Event               | SDTM     | 38   | Char | 20  | Derived | Perm | ISO 8601   |
| AE | AEENRF   | End Relative to Reference Period        | SDTM     | 39   | Char | 20  | Derived | Perm | **BEFORE, DURING, AFTER, DURING/AFTER, U                                       |
| AE | AESTRF   | Start Relative to Reference Period      | ADaM     | 101  | Char | 20  | Derived |      | *BEFORE, DURING, AFTER, U  |
| AE | AEANY    | Did the subject have any AEs            | SUPPQUAL | 1001 | Char | 1   |         |      | *Y, N  |

## OUR EXPERIENCE / CONCLUSIONS

A summary of our work with our CRO partners, looking at the process of doing this the first time [time extensions, cross department communication, buy-in by all parties], the savings gained by approaching the specification in a flexible fashion, and the

### Intrinsic Benefits / Learnings for Future Implementations

- **LEARNING:** Don't let typical timeline metrics impact the decision to attempt this, rather look at the opportunity to examine just how much time it took to do this the first time, compare to future implementations.
- **BENEFIT:** Work to truly develop the generic CRF pages, like the CDASH initiative is pursuing, is proceeding smoothly, nothing to debate in the underlying data structure.
- **BENEFIT:** Work to standardize TLF formats proceeding smoothly as well now that schemes for capturing study designs is standardized by the data structure.

- YET TO SEE: Reduction in timelines for future studies – studies that have started up since have been of different design (DB parallel vs. Phase I cross over), magnitude (2000 subjects versus 24 subjects), so comparative metrics are not yet available, but we look forward to them.

## REFERENCES

CDISC Submission Data Standards Team, (2005, August 26). Study Data Tabulation Model implementation Guide: Human Clinical Trials. Retrieved January 24, 2007, from Clinical Data Interchange Standards Consortium Web site: <http://www.cdisc.org/standards/index.html>.

Health Canada - Health Products and Food Branch, "DRAFT GUIDANCE FOR INDUSTRY: Preparation of Comparative Bioavailability Information for Drug Submissions in the CTD Format", issued May 12, 2004. Document can be found at [http://hc-sc.gc.ca/dhp-mps/alt\\_formats/hpfb-dgpsa/pdf/prodpharma/draft\\_ebauche\\_ctdbe\\_e.pdf](http://hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/draft_ebauche_ctdbe_e.pdf).

[1] US Health and Human Services - Food and Drug Administration, "Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications", Revision 1 issued April 19, 2006. Document can be found at <http://www.fda.gov/cder/guidance/7087rev.pdf>.

[2] [ieeexplore.ieee.org/iel2/463/6617/00262998.pdf?arnumber=262998](http://ieeexplore.ieee.org/iel2/463/6617/00262998.pdf?arnumber=262998)

United States Federal Register, "Guidances on Providing Regulatory Submissions in Electronic Format; Withdrawal of Guidances", Vol. 71, No. 189, pp. 57548-57549

## ACKNOWLEDGEMENTS

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

## CONTACT INFORMATION

Kenneth Eng  
Shire Pharmaceuticals  
725 Chesterbrook Boulevard  
Wayne, PA 19087  
(484) 595-8602  
(484) 595-8170  
keng@shire.com

David C. Izard  
Shire Pharmaceuticals  
725 Chesterbrook Boulevard  
Wayne, PA 19087  
(484) 595-8620  
(484) 595-8170  
dizard@shire.com

Terek J. Peterson  
Shire Pharmaceuticals  
725 Chesterbrook Boulevard  
Wayne, PA 19087  
(484) 595-8947  
(484) 595-8170  
tpeterson@shire.com