

## Using CDISC Models for the Analysis of Safety Data

Susan J. Kenny, Inspire Pharmaceuticals Inc. Research Triangle Park, NC, USA

Edward D. Helton, SAS Institute, Cary, NC, USA

### ABSTRACT

The collection of data for the assessment of safety during a clinical trial is a standard practice. These data are used to analyze and explore the relationship between drug exposure and measures of safety. These analyses can range from simple univariate statistics, such as unadjusted incidence rates and measures of central tendency, to more sophisticated methods such as survival analysis and multivariate analysis of laboratory data. This paper will describe the use of CDISC standards to facilitate the analysis of safety. We will demonstrate the evolution from SDTM domains to ADaM analysis data sets that are designed to support the clinical and statistical review of safety data. Some of these models were developed as part of the CDISC SDTM/ADaM Pilot Project and used by FDA reviewers as part of a mock review.

### INTRODUCTION

The Clinical Data Interchange Standards Consortium (CDISC) has published standards for the organization, structure, and content of clinical data. The Study Data Tabulation Model (SDTM) specifies the structure and metadata for the study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the US Food and Drug Administration (FDA). These SDTM models can be generalized to accommodate derived data but the structure of the models is not always conducive to statistical analysis. The Analysis Data Models (ADaM) provide metadata models and examples of analysis datasets used to generate the statistical results for a clinical study or submission. The key principles of ADaM are to provide clear and unambiguous communication to the reviewer as to how analysis results were obtained and to have the analysis datasets with a structure and content to allow the reviewer to use the datasets to reproduce analyses with minimum programming.

Many sponsor companies are strategizing on how to implement CDISC standards within the conduct of a clinical trial. Ideally, SDTM domains and ADaM datasets should be mutually supportive. One method of implementing CDISC standards is to build the ADaM datasets from the SDTM domains. By doing so, both the clinical and the statistical reviewer have access to complimentary information and efficiencies can be gained by the sharing of metadata between the two sets of data. This paper provides an example of building ADaM datasets from SDTM domains. These models were used as part of the recent CDISC SDTM/ADaM Pilot project and were developed to support the review of safety data.

When analyzing data from a clinical trial, we often tend to concentrate on the efficacy data. However, safety data is of equal concern, especially in recent years with the spate of compounds that have been withdrawn from the market due to safety issues. Trying to understand the potentially complex association between a pharmaceutical compound and indicators of patient's health can be difficult. One analysis approach that is often used to provide a measure of safety risk is to use routinely collected laboratory data to determine if there is any risk of liver problems. Acute liver failure is often relational to drug induced liver injury (DILI). It often occurs at incidence rates less than  $10^{-3}$  which makes it difficult to ascertain in pre-market approval study populations. Medical reviewers are often left to ponder if the DILI is dose proportional or simply idiosyncratic or if in fact the DILI would be dose proportional in an idiosyncratic population. Many times the mechanism of the hepatotoxicity is not understood and it is initially viewed as being idiopathic but when viewed over time, the association becomes clearer.

Liver injury is indicated by raised serum alanine aminotransferase [ALT] activity and/or raised aspartate aminotransferase (AST) and reduced liver function is evidenced by increased total serum bilirubin [TBL] concentration. When these three different indicators of liver function are observed simultaneously in a person, it is both a sensitive and highly specific test for potentially serious DILI. The concurrent analysis of these laboratory measures is referred to as Hy's Law. One example of applying Hy's Law is to identify observations where values of AST or ALT are 1.5 times or greater the upper limit of normal and the value of TBL is 1.5 times or greater than the upper limit of normal over the time period of exposure. The value used as the cut point to compare laboratory measures (1.5 in this example) is variable and depends on the disease under investigation and the patient population.

## PhUSE 2006

Hy's Law is used extensively in the review of clinical trial data since it provides a relatively simple mechanism to detect a signal for liver injury.

An example of the use of safety data to evaluate risk of DILI is the recent analysis of data collected for the compound Rezulin (troglitazone). This compound was marketed in 1997 for the treatment of Type 2 diabetes and early post-marketing data suggested that there may be a risk for liver failure. Determining for certain whether there is a risk is not a simple process. Typical research questions included the following. Is the number of cases of clinically significant DILI is disproportionate to other adverse events? What is the range of clinical severity of liver injury? What is the relationship between drug dose, duration of exposure, and patient susceptibility to liver damage? Is there a signal of liver injury in the clinical trial database based on imbalances between placebo and active drug arms, using Hy's Law to detect a signal? Are there any patterns of liver injury associated with some subject demographics, an underlying disease, concurrent illness or concurrent medications? Answering these types of questions requires a reorganization of the clinical trial data to create analysis data sets which are designed to address these specific questions. Once the data is in a structure that is conducive to in-depth review and analysis, answers to important research questions can be obtained. It was this type of complex restructuring of data and subsequent analysis that resulting in the determination that Rezulin did pose an increased risk for DILI and in 2000, the Food and Drug Administration (FDA) removed the drug from the market.

### USING ADAM MODELS FOR SAFETY ANALYSIS

Any one who has generated tables for a clinical study report understands the complexity that is often associated with getting the clinical data ready to produce a quality tabular presentation with a programming language such as SAS® or S-Plus®. These study report tables are used by clinicians and medical reviewers to draw conclusions pertaining to the safety and efficacy of the drug compound. The responsibility of a regulatory reviewer is to not only understand how an analysis was performed but to be able to reproduce the results with the submitted data. If an analysis required complex restructuring of the clinical data, imputation of data values, or determination of windowed visits, for example, the reviewer will be challenged to reproduce the analysis using only the collected data in SDTM format. A good rule of thumb is if multiple programming steps were used to manipulate data and/or create derived variables in order to produce a statistical result, then the submission of ADaM datasets is warranted. Procedures such as transposes or many to many merges are not always easy for researchers who are not familiar with programming languages to execute. The adage of 'one proc away' supported by ADaM should be used as a measure of success for whether the structure and content of the analysis data set will meet the need of reviewers.

Consider three types of safety analyses that are routinely presented in a clinical study report: unadjusted incidence rates of adverse event, time to event for first occurrence of an adverse event of special interest, and determination of the risk for hepatotoxicity using Hy's Law. The ADaM datasets that are built for these analyses demonstrate the continuum of complexity in building analysis ready data sets. The analysis dataset that supports the calculation of incidence rates is created by the simple addition of variables to an SDTM domain yet the analysis dataset to support the investigation of hepatotoxicity requires a complete retransformation of data from several SDTM domains.

### ADVERSE EVENT RATES

At first glance, the typical adverse event summary table looks very simple. Adverse events are grouped by coded terms and the number and percent of subject experiencing each unique event is presented. Often, the total number of occurrences of each type of events is also of interest. P-values may be presented to aid in the determination of whether adverse event rates for subjects exposed to study drug differ from those for subjects exposed to placebo.

In order to produce this adverse event analysis using the SDTM AE domain, the following steps need to be taken.

1. Identify treatment emergent events
2. Obtain the counts of subjects experiencing each event at least once (numerator)
3. Obtain the number of subjects exposed (denominator)
4. Obtain the counts of the total number of occurrences for each type of event (prevalence)
5. Calculate the percentage.
6. Obtain p-value
7. If safety analyses were to be performed by a subgroup, such as by gender, then additional steps are needed to identify subjects within the subgroup and the appropriate denominator.

This is one of the simplest types of summaries for adverse event data. More sophisticated analyses may require person-time denominators, measures of drug exposure, or the use of Bayesian hierarchical models.

### TIME TO EVENT ANALYSIS

Another frequently used method for the analysis of safety data is to investigate whether there is a treatment effect for the time it takes an event to first occur. There are a number of statistical methods that can be used and one of the

## PhUSE 2006

most common is the Kaplan Meier product limit method. Using SAS, this analysis can be produced by PROC LIFETEST. However, for the general case, that data input into this proc must have one record per subject and an indication of whether they had the event of interest and a measure of the duration of observation time. For subjects with an event, this would be the duration of time between, for example, the start of treatment to the start of the event of interest. For subjects without an event, this duration may be the start of treatment to the end of the study (or early discontinuation). In the example presented here, the event of interest is the occurrence of any adverse events that is classified as one of 'special interest'. Special interest adverse events are predefined and may include events that span across different body systems.

To use PROC LIFETEST to produce an analysis of time to event of any special interest adverse event, it is necessary to first:

1. Identify adverse events that are included in the definition of 'special interest'
2. If a subject has at least one event of special interest, determine when the first event occurred first and calculate the time, from start of the study (for example), to this event
3. If no events of special interest were observed for a subject or if the subject had no adverse events, then calculate the duration of time they were observed, during which the event may have occurred. This subject is defined as a censored observation.
4. Create censor flags so that PROC LIFETEST can distinguish between subjects who had the event and those that were censored.

### IDENTIFYING SUBJECTS AT RISK FOR HEPATOTOXICITY

One example of applying Hy's Law is to identify observations where values of AST or ALT are 1.5 times or greater the upper limit of normal and the value of TBL is 1.5 times or greater than the upper limit of normal over the entire time period of exposure. The size of the magnitude of difference in the observed laboratory measures versus the upper limit of normal (1.5 in this example), may vary depending on the nature of the safety investigation. For example, in a very sick population of subjects where many laboratory values are expected to be out of range, a higher threshold value, such as 3.0, may be more appropriate. Subjects are then classified according to these findings.

To identify subjects at risk for DILI using laboratory data from the SDTM LB domain, the following steps would need to be taken:

1. Identify laboratory measures that were obtained after the first exposure to study medication and for a pre-determined amount of time after exposure ended.
2. Compare each laboratory measure of AST, ALT and Bilirubin against the upper limit of normal range recorded for each laboratory measure and classify whether these values are out of range, given the cut point of interest.
3. Review the laboratory data collected at each time point in the study and compare the on-treatment measures of hepatotoxicity with baseline measures. Use these measures to classify each subject as to whether they are at risk for DILI at each time point.
4. Thoroughly investigate other covariates, such as subject's demographics, concurrent diseases, length of time of exposure to study medication, to better understand the potential etiology of DILI.

### EXAMPLES OF ADAM MODELS

These three types of safety analysis provide useful examples to illustrate the relationship between SDTM and ADaM. The metadata for each of these analysis data sets will be described below. The models presented below are based on analysis data sets that were created as part of the CDISC Pilot Project. This project was a collaborative effort between ODM, SDTM, ADaM and FDA representatives. The goal of the project was to create a mock submission to the FDA using de-identified clinical trial data that was submitted using CDISC standards. The objective of the Pilot project was to determine whether the existing CDISC standards met or exceeded the regulatory agencies needs. This submission consisted of a single study report, including statistical tables and results, SDTM domains, ADaM analysis data set domains, and associated define.xml which contained detailed metadata for all domains. When the Pilot project is completed, the team reports will be made publicly available to the CDISC community.

This paper does not attempt to describe all of the ADaM recommendations. For example, the models below reference a subject-level analysis dataset, termed ADSL. This is a standard ADaM dataset that is a one record per subject analysis dataset and is designed to contain all of the variables that describe individual subjects, such as population flags, demographic characteristics, key milestone dates, etc. The metadata for the ADSL analysis data set is described in the ADAM Analysis Data Model V2.0 document. Readers are encouraged to review the published ADaM standards and provide comments on new ADaM standards as they are released for public comment.

### CAVEATS

The models that are described in this paper are based on the ADaM analysis data sets that were submitted as part of the Pilot project submission package. These models closely follow the draft models for adverse events and time to event that the ADaM team is developing. However, these models should not be considered to be final ADaM

## PhUSE 2006

recommendations, especially attributes such as variable names. The ADaM team is rapidly working toward developing standard models for the structure of analysis files and standard variable names for common analysis variables. These standards will be released by the ADaM team as an Implementation Guide. At the time of the writing of this paper, the ADaM recommendations have not been finalized and so the information presented here should be used as examples and not standards.

The ADaM models are described below by presenting the variable level metadata which would be provided as part of the DEFINE.XML file. In the interest of space for this paper, the columns for 'Type' and 'Source' are not shown. In order to present more information in this paper, the column 'Computational Algorithm or Method' is used for two purposes. Normally this column would be used to describe the analysis rules and methods used to create the derived variable. In this paper, any text that is in italics are notes relative to the model and should not be interpreted as text that would normally be present in this column. Additionally, variables of similar nature, such as variables inherited straight from an SDTM domain, are grouped together on one line but normally each variable would have its own line of metadata.

### ADAM ADAE MODEL

The structure of the ADaM analysis dataset for adverse events, ADAE, is identical to the structure of the SDTM AE domain, which is one record per reported adverse event. It is the addition of variables to the AE domain that results in an analysis ready dataset. The added variables can be classified into 4 types:

1. Subject-level variables that exist within other SDTM domains and are added to the rows of the AE file. Examples of this include subject information, such as demographic information (age group, gender, race), population flags, important dates (start and stop date of study medication), measures of exposure (total exposure to study medication over the course of the trial), or disposition variables (indicators of whether the subject completed the study, discontinued due to an adverse event).
2. Record-level variables that may or may not exist within other SDTM domains. Examples of this include a flag to indicate whether an adverse event was treatment emergent. If such a variable is present in SDTM, it would be located in the supplemental qualifier domain for AE (SUPPAE). Examples of these variables include the dose at the time of the adverse event. Such information may be important for a dose-escalating study where safety analysis is to be performed by the dose at the time of the event. Other variables such as the derived dichotomous variable to indicate whether an event was 'Severe' or 'Not Severe' (the combination of mild+moderate) may be necessary.
3. Record-level variables that are represent variables obtained from a coding dictionary. These dictionary variables may be new dictionary terms obtained from a recoding, or 'dictionary leveling', of adverse events that often occurs during the creation of an integrated summary of safety.
4. Indicator variables that can be used to easily identify individual subjects who had each type of event and indicator variable to describe whether they had an event of special interest. These indicator variables can be used to rapidly calculate the numerator for the percentages of each level of event.

The following is an example of the variable level metadata that describes an ADAE analysis file.

Adverse Event Analysis Data Set (ADAE)			
Variable	Label	Controlled Terms or Format	Computational Algorithm or Method <i>Informational notes provided in italics</i>
USUBJID	Unique Subject Identifier		
SITEID	Study Site Identifier		
SEX	Sex	M, F, U	<i>Subject level variables, generally obtained from the core analysis data set, ADSL, should be added to facilitate subgroup analysis and by-patient investigations or listings. Both character and numeric versions of the same variable are helpful and aid both visual and analytical review of the data</i>  <i>Recommendations provided by ICH, FDA, and other regulatory agencies should be used as a guideline for what types of subject level variables to include.</i>
SAFETY	Safety Population Flag	Y,N	
RACE	Race	CAUCASIAN, AFRICAN AMERICAN, HISPANIC, OTHER	
RACEN	Race, Numeric	1,2,3,4	
AGE	Age in AGEU at RFSTDTC		
AGEGRP	Age Group	<=60, 60-80, >80	

PhUSE 2006

AGEGRP	Age Group, Numeric	1, 2, 3	
TRT	Description of Treatment Arm	Placebo 0.5 mg/kg/day 1.0 mg/kg/day	<i>Including all variations of treatment assignments allows maximum flexibility for the reviewer. TRT and TRTCD can be used for listings, and TRTN can be used for analyses that require ordered values.</i>
TRTCD	Treatment Code	P, 0.5, 1.0	
TRTN	Treatment Code, Numeric	0,1,2	
TRTSTDT	Start Date of Treatment	date	SV.SVDTC where SV.VISITNUM=3, converted to a SAS date
LSTDOSDT	Date of Last Dose	date	The date of final dose (from the CRF) is EX.EXENDTC on the subject's last EX record. If the date of final dose is missing for the subject and the subject discontinued after Visit 3, use the date of discontinuation as the date of last dose. Convert the date to a SAS date. <i>Note that this definition is specific to the CRF design used in this study. The derivation of this variable may be different depending on how dosing and exposure data is recorded.</i>
DSREASAE	Discontinued due to AE?	Y,N	<i>This variable would be obtained from either ADL, if included, or derived from an SDTM domain such as DS.</i>
AESEQ	Sequence Number		
AETERM	Reported Term for the Adverse Event		
AEBODSYS	Body System or Organ Class		<i>These variables represent the coded terms obtained from a coding dictionary. In this example, all 5 levels from MedDRA are included in the analysis file. Special attention needs to be given to these variables if multiple dictionaries have been used during the life of the project. It is advantageous to give the reviewer both the originally coded term(s) and the newly coded terms in the event that a newer dictionary version has been used for an integrated analysis.</i>
AEDECOD	Dictionary-Derived Term		
HLGTERM	High Level Group Term		
HLTERM	High Level Term		
LLTERM	Lower Level Term		
CODEINFO	Coding Dictionary Information		MedDRA Version 8 for all records <i>Including a variable that indicates the dictionary and version used for adverse event coding is helpful to the reviewer, especially if multiple studies are combined where different coding dictionaries were used.</i>
TRTEMFL	Treatment Emergent Flag	Y, N	If AESTDT >= ADSL.TRTSTDT then TRTEMFL='Y'. Otherwise TRTEMFL='N'.
AESTDT	Start Date/Time of Adverse Event	Date	AE.AESTDTC, converted to SAS date. Some events had missing month and day components. Partial dates for events that had a value of start year that was prior to treatment exposure were not imputed. Only those events that had the same year as treatment exposure were imputed if necessary. If day component was missing, a value of '01' was used. There were no events with missing month components.
AESTDTF	Imputed Date Flag	D, M, Y	D if the day value within the character date was imputed
ANLSTDY	Analysis Start Day		ADAE.AESTDT - ADSL.TRTSTDT +1 <i>Analysis start day differs from the SDTM variable for study day by the inclusion of day 0. For example,</i>

## PhUSE 2006

			<i>an event that occurred on the day prior to dosing would have ANLSTDY=0 where as the value of AE.AEDY would have a value of -1.</i>
AEENDT	End Date/Time of Adverse Event		AE.AEENDTC, converted to SAS date
AEENDY	Study Day of End of Adverse Event		ADAE.AEENDT-ADSL.TRTSTDT+1
AESER	Serious Event	Y, N	
AESERN	Serious Event, Numeric	1,0	If AESER='Y' then AESERN=1. If AESER='N' then AESERN=0.
AESEV	Severity / Intensity	Mild, Mod, Severe	
<i>Additional variables</i>	<i>SDTM AE variables such as AEREL, AESCAN, AESCONG, AEDISAB, AESDTH, AESHOSP, AESLIFE</i>		
AEDERMFN	Special Dermatologic-Related Event (1=Y)	1, or missing	if AE.AEDECOD contains any of the character strings of ('APPLICATION','DERMATITIS','ERYTHEMA','BLISTER') OR if AE.AEBODSYS='SKIN AND SUBCUTANEOUS TISSUE DISORDERS' but aeecod is not in ('COLD SWEAT','HYPERHIDROSIS','ALOPECIA') then AEDERMFN=1 Otherwise AEDERMFN=null <i>This variable is created to identify events in ADAE that are of special interest and to facilitate the creation of the analysis model for the time to event analysis</i>
TRTEMFL	Treatment Emergent Flag	Y or N	If AESTDT >=ADSL.TRTSTDT then TRTEMFL='Y'. Otherwise TRTEMFL='N'
AEFN	1st AE Occurrence (1=Yes)	1, or missing	AEFN=1 for the record with the first treatment emergent occurrence of any AETERM. Missing otherwise.
BODFN	1st Occurrence of SOC (1=Yes)	1, or missing	BODFN=1 for the record with the first treatment emergent occurrence of AEBODSYS. Missing otherwise.
DECODFN	1st Occurrence of preferred term (1=Yes)	1, or missing	DECODFN=1 for the record with the first treatment emergent occurrence of AETERM. Missing otherwise.
SAEFN	1st Serious Occurrence (1=Yes)	1, or missing	SAEFN=1 for the record with the first treatment emergent occurrence of AETERM where AESER='Y'. Missing otherwise.
SBODFN	1st Serious Occurrence of SOC (1=Yes)	1, or missing	SBODFN=1 for the record with the first treatment emergent occurrence of AEBODSYS where AESER='Y'. Missing otherwise.
SDECODFN	1st Serious Occurrence of Term (1=Yes)	1, or missing	SDECODFN=1 for the record with the first treatment emergent occurrence of AETERM where AESEV='Y'. Missing otherwise.
<i>Additional variables relating to exposure</i>	<i>Variables such as treatment duration, dose at the time of the adverse event, cumulative exposure at the time of the event may be important and useful for the review. Such variables may need to be derived from a variety of sources in SDTM. Subject level exposure variables that are used for multiple analyses should be placed in ADSL if appropriate.</i>		

The ADAE file has an identical structure and same number of records as the SDTM AE domain. It is the addition of variables, obtained from other SDTM domains, other analysis data sets such as ADSL, or derived specifically for the adverse event analysis that makes the analysis data set reviewer friendly. The indicator variables (AEFN, BODFN, DECODFN, etc.) are flags that quickly identify the subjects who experienced each type of event. Since most adverse event displays are presented by System Organ Class (body system) and Preferred Term, a flag variable is needed for each level. Similarly, other sets of indicator variables are needed to aid the summary of serious events (SAEFN, SBODFN, and SDECODFN). For the purposes of programming a standard adverse event table, these flag variables are not required since the counts of unique events can be easily obtained from simple SQL code. However, having these variables in an adverse event analysis data set helps reviewers who are not programming savvy to quickly

## PhUSE 2006

identify the events that comprise the numerator values.

Because the structure of ADAE is similar to SDTM AE, only subjects who experienced an adverse event are included in the analysis data set. This means that the denominator counts for each treatment group must be obtained from another source. The ADSL file would be a likely source for these denominators. Alternatively, providing the reviewer with an analysis file that contains just the denominator values for each treatment group and each treatment group by subgroup comparison may be helpful.

### ADAM TIME TO EVENT MODEL - ADTTE

In this example, the event of interest is any one of several types of adverse events. Therefore, it is reasonable to build this analysis file from the adverse event analysis data set rather than directly from the SDTM AE domain. Additional information will be needed from other sources, such as ADSL, that will aid the calculation of the duration of observation for censored subjects. The key characteristics of this analysis data set are:

1. There is one record per subject per event of interest. In this example, only one event is of interest ('a special interest ae') so, therefore, there is only one record per subject.
2. A variable for the censor flag. This is required by PROC LIFETEST and the censor flag is most often in the form 0 (censored) or 1
3. A variable to reflect the time to first event of special interest, for those subjects who had an event or the length of the observation time, for subjects who had no event of special interest.

The following is an example of the variable level metadata that describes an ADTTE analysis file.

Time to Event Analysis Data Set (ADTTE)			
Variable	Label	Controlled Terms or Format	Computational Algorithm or Method <i>Informational notes provided in italics</i>
USUBJID	Unique Subject Identifier		
SITEID	Study Site Identifier		
AESEQ	Sequence Number		
TRT	Description of Treatment Arm	Placebo 0.5 mg/kg/day 1.0 mg/kg/day	<i>It is helpful to be consistent in the inclusion of treatment variables across all analysis datasets. In these examples, all 3 forms of treatment variables are included in each analysis dataset.</i>
TRTCD	Treatment Code	P, 0.5, 1.0	
TRTN	Treatment Code, Numeric	0,1,2	
<i>Additional variables</i>	<i>Additional subject level variables, such as sex, race, age group, discontinuation due to an adverse event, etc may be of interest. Consider including treatment start date and treatment stop date in all analysis data sets.</i>		
SAFETY	Safety Population Flag	Y or N	Y if ITT='Y' and TRTSTDT not equal to missing
AETERM	Reported Term for the Adverse Event		<i>Including the dictionary terms for adverse events of interest is helpful to the reviewer. In this example, the verbatim term plus the coded terms are provided. These variables would be blank for subject's who did not have any events of special interest.</i>
AEDECOD	Dictionary-Derived Term		
AEBODSYS	Body System or Organ Class		
TTEVALUE	Time to Event or End of Study		Where AEDERMFN=1, the value of TTE is the minimum value of ANLSTDY for records with AEDERMFN=1. Where AEDERMFN=missing, TTE = ADSL .ENDDT - ADSL.TRTSTDT+1.
TTECNSR	Censoring Indicator – Time to Event	1,0	TTECNSR=1 for any USUBJID that has at least one record where ADAE.AEDERMFN=1. TTECNSR=0 otherwise.

## PhUSE 2006

The ADTTE analysis data set has one record per subject per event of interest. In this example, there is only one event of interest ('special AE') so there is only one record per subject. If there were multiple events of interest that were being investigated using a time to event analysis, then each type of event would have a set of records. In this case, another variable would be needed to categorize these records. For example, the inclusion of another variable, for example 'TTEVAR', could be used to identify the category of event that is represented on the record (such as time to death, time to progression of disease, etc).

### ADAM MODEL FOR HY'S LAW– ADHY

The two examples of analysis data sets shown above built on the SDTM domains and kept the structure of the analysis file very similar if not identical to the source SDTM domain. In many circumstances the SDTM structure can support analysis provided additional columns and/or rows are added to the table. But for complex analyses, such as investigation of hepatotoxicity using Hy's Law, a structure that represents one record per subject per visit is warranted. This represents a transposition of the vertical structure of the LB SDTM domain to a more horizontal structure where the columns represent the values of important laboratory measures and their associated qualifier variables such as normal ranges. Additionally, there is the transposition of baseline data onto each record so that pre- to post-dose comparisons can be made. By having all laboratory values on one record, the reviewer can easily see the relationship between the laboratory measures and explore how they may be related to each other, either at that visit or compared to pre-dose baseline values.

For this paper, a description of the columns that would be present in a typical ADHY analysis data set is provided rather than a detailed metadata description of each individual variable. This is because the presentation of the metadata for the volume of potential variables would unnecessarily detract from the depiction of this analysis dataset. Instead the following classes of variables (columns) would be present in an analysis data set designed to support the investigation of potential DILI:

Class of Variable	Description	Example of Data Values or Algorithm
Classification of Lab Values		
ALT	ALT Upper Range of Normal	32
	ALT Lower Range of Normal	6
	Baseline ALT value	70
	Baseline value of ALT as a function of Upper range of Normal (ALT divided by upper range)	$70 / 32 = 2.2$
	Baseline value of ALT as a function of the Lower range of Normal (ALT divided by lower range)	$70 / 6 = 11.7$
	Indicator flag to categorize the baseline value of ALT relative to cut-point values of interest. Suggested flag values would be as H, L, N.	H (11.7 > cut point value of 1.5)
	Visit value of ALT	88
	Visit value of ALT as a function of Upper range of Normal	$88 / 32 = 2.75$
	Visit value of ALT as a function of the Lower range of Normal	$88 / 6 = 14.7$
	Indicator flag to categorize the visit value of ALT relative to cut-point values of interest. Flag values would be as H, L, N.	H
	Change from baseline in ALT measure	18
AST	.....Replicate the 11 columns above for measures of AST	
Total Bilirubin	.....Replicate the 11 columns above for measures of Total Bilirubin	
Indicator Flags to Identify Subjects at Risk		
	Did values of ALT or AST meet Hy's Law criteria at Baseline?	Y
	Did value of Total Bilirubin meet Hy's Law criteria at Baseline?	N
	Is there evidence that the subject is at risk for DILI at Baseline using the 2 levels of criteria?	N
	Did values of ALT or AST meet Hy's Law criteria at this Visit?	Y
	Did value of Total Bilirubin meet Hy's Law criteria at this Visit?	Y
	Is there evidence that the subject is at risk for DILI this Visit using the 2 levels of criteria?	Y

## PhUSE 2006

Visit / Timing Variables	
	Variables describing the visit, such as visit number, visit date, analysis study day, etc.
Subject Identifiers	
	Subject-level variables such as unique identifier, demographic characteristics, population flags, disposition events, treatment assignments
Subject Covariates	
	Clinical investigation of DILI often requires detailed information on many types of covariates, such as measures of exposure to study drug, dose (actual or cumulative) at the time of laboratory assessment, indicator variables for the presence or absence of potential confounders, whether any interventions were performed in response to liver injury, and resolution status of the subject.

The reader can see that an analysis dataset for Hy's Law may contain more than 50 variables. Whereas some of these variables may someday be created by software solutions or could easily be recreated by the reviewer (such as indicator flags for laboratory values), it is always better to err on the side of providing too much detail for the reviewer rather than not enough.

### CONCLUSION

This paper illustrates the use of ADaM models for the analysis of safety data. These models assume that SDTM domains are used as the source of data for the analysis datasets. The three models that were presented show the continuum of complexity in the creation of analysis datasets, with the simple adverse events analysis dataset being the addition of several columns to the complex Hy's Law analysis dataset requiring a complete restructuring of data from several domains and the creation of many derived variables. Regardless of the complexity of the analysis dataset, the fundamental principles of ADaM always apply. These principles are based on the premise that there needs to be a clear and unambiguous path to the derived data and eventual statistical analysis.

The advantage of creating analysis datasets for safety goes beyond the analysis of safety results for individual studies. If standard safety analysis data sets are created early in the development of a study compound, then the production of annual IND updates, periodic safety update reports, and other reports related to pharmacovigilance will be expedited. Medical and clinical reviewers within sponsor companies will also benefit by having analysis datasets that can be used for their regular review of patient safety data.

### ACKNOWLEDGMENTS

The authors wish to acknowledge the ADaM team for their efforts in developing standards for the generation and submission of analysis datasets. Additionally, we thank the CDISC Pilot team for their work in providing an end-to-end example of the use of CDISC standards for regulatory submission.

### RECOMMENDED READING

CDISC Standards (<http://www.cdisc.org/standards/index.html>)

- Analysis Data Model: Version 2.0
- SDTM Implementation Guide V3.1.1
- SDTM Study Data Tabulation Model 1.1

FDA and ICH Documents (this is not an exhaustive list)

Clinical White Paper (this paper discusses the use of Hy's Law for hepatotoxicity)  
<http://www.fda.gov/cder/livertox/clinical.pdf>

Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report of the Review (<http://www.fda.gov/cder/guidance/3580fnl.pdf>)

Data Elements for Transmission of Individual Case Safety Reports (E2B)  
<http://www.ich.org/LOB/media/MEDIA632.pdf>

Pharmacovigilance Planning (E2E) (<http://www.ich.org/LOB/media/MEDIA1195.pdf>)

Statistical Principles for Clinical Trials (E9) (<http://www.ich.org/LOB/media/MEDIA485.pdf>)

## PhUSE 2006

### **CONTACT INFORMATION**

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

Susan J. Kenny  
Inspire Pharmaceuticals Inc.  
4222 Emperor Boulevard  
Durham, North Carolina  
27703-8030  
Work Phone: 1-919-287-1279  
Email: skenny@inspirepharm.com

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 trademarks of their respective companies.