Standardization in Medical Research Data: Improving Data Displays and Data Traceability

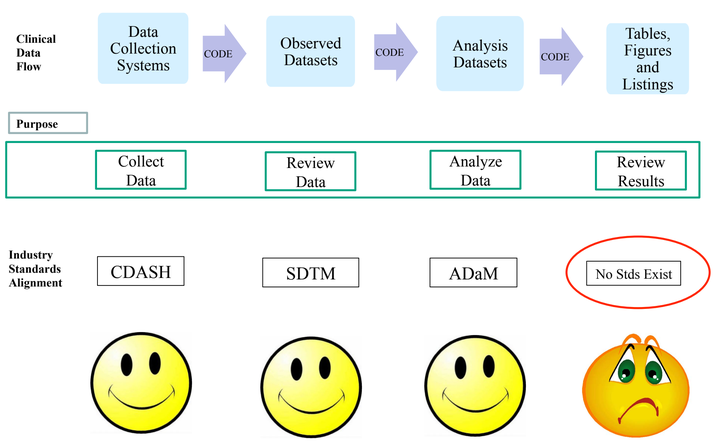
Mary Nilsson[1], Scott Getzin[1] , Michelle Barrick[1]

[1]Eli Lilly and Company, Indianapolis, IN, USA

ABSTRACT

Industry standards have evolved over time for data collection (CDASH), observed data (SDTM), and analysis datasets (ADaM). Although substantial progress has been made, additional standardization can improve product development. This poster will cover two specific needs. First, further advancement is needed to develop industry standards for analysis and reporting. Development of standard Tables, Figures, and Listings (TFLs) and associated analyses will lead to improved standardization from collection through data storage. (You need to know how you want to analyze and display results before finalizing how to collect and store data.) The development of standard TFLs will also lead to improved product lifecycle by ensuring reviewers receive the desired analyses for the consistent and efficient evaluation of patient safety and drug effectiveness. Second, additional improvement is needed with respect to data traceability from collection to analysis. This poster will display how Eli Lilly & Company plans to address this gap.

Status of Industry Standards



Why Standardize Tables/Figures/ Listings?

Need to know how the data will be displayed to ensure proper collection and storage

Ensures reviewers receive the desired displays

Enables consistent and efficient evaluation of patient safety and drug effectiveness

Facilitates the incorporation of advice from subject matter experts

- Creates forum to communicate improved methodology
- For safety, leads to improved signal detection

What TFLs Can Be Standardized?

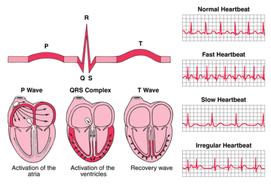
Demographics

Adverse Events

- Create consistent definitions of adverse events and treatment emergent adverse events

Vital Signs, Laboratory Measures, ECGs

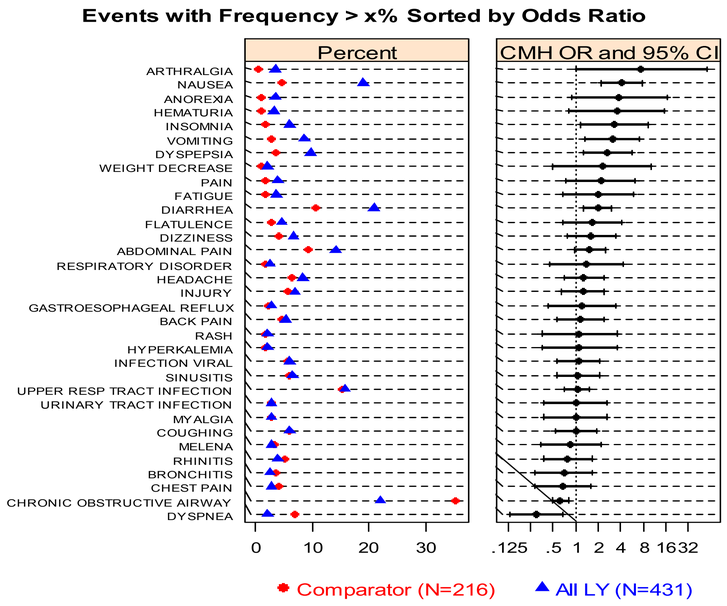
- Consistent approach for central tendency
- Consistent approach for outliers/shifts
- Consistent advice on reference ranges or thresholds



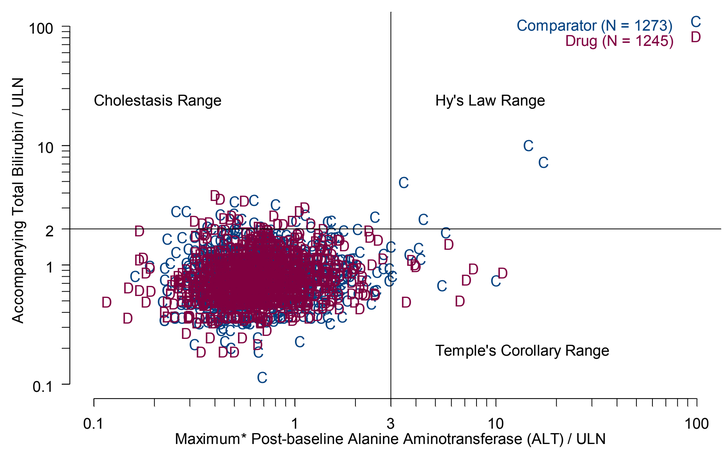
TFLs for Common Events of Special Interest

- Hepatotoxicity
- Hypersensitivity

Example Adverse Event Figure

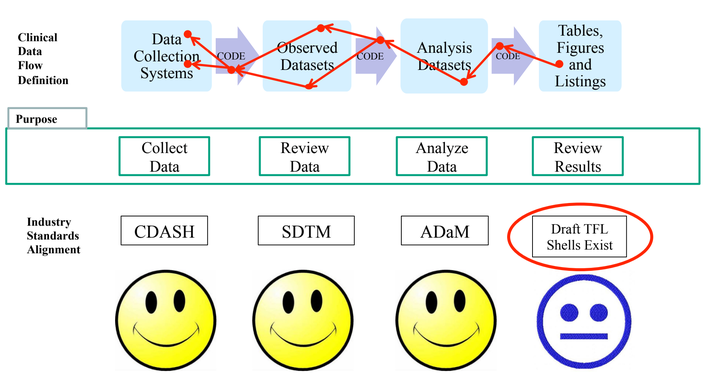


Example Hepatotoxicity Figure



ULN = upper limit of normal of the reference range
* If a patient has any total bilirubin >= 2x ULN then maximum ALT amongst the total bilirubin >= 2x ULN is plotted. Otherwise, the overall maximum ALT for the patient is plotted.

Progress of Lilly Data Standards Improving Traceability



Traceability Example of Draft Data Standards Definition

TFL Display

Demographic Summary — Page x of y
<Enrolled / Randomized> Population — hh:mm ddmmmyyyy
<Study Name> — production status code

		Treatment A			Treatment B			Total		
		Female	Male	Total	Female	Male	Total	Female	Male	Total
Demographic		(N=XXX) DN2	(N=XXX) DN3	(N=XXX) DN1	(N=XXX)	(N=XXX)	(N=XXX)	(N=XXX)	(N=XXX)	(N=XXX)
Age (yrs) AR1	n	xx	xx	xx	xx	xx	xx	xx	xx	xx
	Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
	Median	xx	xx	xx	xx	xx	xx	xx	xx	xx
	Min, Max	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)	(xx, xx)

Maps to Analysis Results Metadata — Maps to Analysis Results Metadata

Analysis Results

Analysis Results Identifier	Analysis Result Description	Type	Source Data	Analysis Variable	Selection Criteria	Study Treatment	Documentation
Unique identifier	Description of result	Required or Optional	Data set(s) used to produce result	Analysis variable being analyzed	Selection criteria for analysis subset and/or numerator	Study treatment variable to use in analysis	Textual description of analysis performed
DN1	Denominator - Total number of subjects in the population within the treatment group	Required	ADSL	<ADSL.ENRLFL / ADSL.RANDFL>	<ADSL.ENRLFL / ADSL.RANDFL>='Y'	ADSL.TRT01P	Calculate the total number of subjects in the population within each treatment group.
DN2	Denominator - Total number of female subjects in the population within the treatment group	Required	ADSL	<ADSL.ENRLFL / ADSL.RANDFL>	<ADSL.ENRLFL / ADSL.RANDFL>='Y' and ADSL.SEX='F'	ADSL.TRT01P	Calculate the total number of female subjects in the population within each treatment group.
AR1	A set of summary statistics of age of subjects .Age is age in yrs at entry to the study.	Required	ADSL	ADSL.AGE	<ADSL.ENRLFL / ADSL.RANDFL>='Y'	ADSL.TRT01P	For each treatment group and subgrouped into Female, Male and Total, calculate the following statistics: **n, mean, standard deviation, median, minimum, and maximum**. n is the number of subjects contributing non-missing data.

Trace back to ADaM

ADaM.ADSL

TABLE	COLUMN	DIRECTSOURCE	ORDER	KEY	LABEL	REVIEW_ALGORITHM
ADSL	STUDYID	SDTM.DM.STUDYID	1	Y		
ADSL	USUBJID	SDTM.DM.USUBJID	2	Y		
ADSL	AGE	SDTM.DM.AGE	4			
ADSL	AGEU	SDTM.DM.AGEU	5			
ADSL	ENRLFL		800		Enrolled Population Flag	If ADSL.ICDT is not missing then set ENRLFL=Y, else ENRFL=N
ADSL	RANDFL		801		Randomized Population Flag	If ADSL.RANDDT is not missing then set RANDFL=Y, else RANDFL=N

Trace back to SDTM

SDTM.DM

TABLE	COLUMN	DIRECTSOURCE MODEL	DIRECTSOURCE ID	KEY	LABEL	SASTYPE	SASLENGTH	ORIGIN	BUSINESS_RULE
DM	STUDYID			Y	Study Identifier	C	50	Protocol	
DM	DOMAIN				Domain Abbreviation	C	2	Assigned	Set to DM domain
DM	USUBJID			Y	Unique Subject Identifier	C	50	Derived	
DM	BRTHDTC	ODM	DM1001.BRTHDAT		Date/Time of Birth	C	50	CRF Page xx	Year as reported on CRF page. (Excludes paediatric studies, where full date of birth will continue to be collected)
DM	AGE				Age	N	8	Derived	Where only Year present in BRTHDTC, Day and Month defaulted to 01 July (e.g. 1974-07-01). Age=floor((intck('month',brthdtc,rfstdtc)-(day(rfstdtc) < day(brthdtc)))/12)
DM	AGEU				Age Units	C	50	Assigned	For paediatric subject reporting, Age should be reported as required

Trace back to CRF

CRF

: Demographics (DM) [frmDM1001]

[frmDM1001]

1.* What is the subject's date of birth? [Birth Date] — (BRTHDAT) Req (2012-2014)

CDASH name on the Data Element Standard, which includes SDTM mapping

Data Element Standard

Data Standard	OID	Name	SAS Field Name	Datatype	Length	Significant Digits	CodeListRef	Measurement Unit	Completion Required
Std00065	BRTHDAT	Birth Date	BRTHDAT	partialDate					X
Question:	What is the subject's date of birth?								
Description:	A subject's date of birth. The complete Date of Birth is made from the temporal components of Birth Year, Birth Month, and Birth Day.								
Alias:	**Context:** Lilly Legacy DES **Name:** BIRTHDTYY								
Alias:	**Context:** SDTM 3.1.2 **Name:** BRTHDTC								
Origin:	CRF								
Date Class:	Y Required								
Standard Type:	Standard								
Standard Ownership:	Core								
Item Prompt:	Birth Date								
Comment:	Only Birth Date - Year should be collected. Prior approval must be obtained from Global Privacy Office to collect "Birth Date - Day" and "Birth Date - Month".								
Business Rules:	Every effort should be made to obtain a subject's birth year, estimated if necessary. An estimated birth year is likely to be more accurate than one that is determined by a system algorithm.								

CDASH name — SDTM name

CONCLUSIONS

- There is a HUGE opportunity for increased efficiency by standardizing common TFLs across the industry
- Standardizing common TFLs for safety may improve safety signal detection
- Improving data traceability will facilitate the review process