How Good is Your SDTM Data? Perspectives from JumpStart

Mary Doi, M.D., M.S.

Office of Computational Science
Office of Translational Sciences
Center for Drug Evaluation and Research
US Food and Drug Administration

JumpStart Service

Purpose

1. Assess and report on whether data is fit for purpose
   - Quality
   - Tool loading ability
   - Analysis ability
2. Load data into tools for reviewer use and run automated analyses that are universal or common (e.g., demographics, simple AE)
3. Provide analyses to highlight areas that may need a focus for review

Benefits for the Reviewer

1. Understand their data, analyses that can be performed, and identify potential information requests for the sponsor
2. Improves the efficiency of their review by setting up tools and performing common analyses, which provides them with time to focus on more complex analyses
3. Points them to a possible direction for deeper analysis

JumpStart uses SDTM data

- JumpStart uses the following SDTM domains:
  - Data Fitness session analyzes all submitted datasets
  - Safety Analysis session focuses on DM, DS, EX, AE, LB, and VS
- JumpStart uses SDTM data in the following ways:
  - To run automated analyses
  - Create any derived variables that are needed
  - Check data against values in the submitted clinical study reports

Data from Applications in JumpStart

Data from a total of 34 applications:

- Consisting of 58 individual studies
- Across 14 review divisions in OND
- Both NDAs and BLAs

Data Quality Issues

- Define File
- Study Data Reviewer's Guide
- Disposition (DS) Domain
- Items in Technical Conformance Guide (TCG)
- Other Issues

Define File

- Use Define.xml v2.0
  - Lacking a complete Define file greatly increases the amount of time reviewers spend understanding an application
- Include detailed description of data elements:
  - Detailed, reproducible computational algorithms for derived variables
  - Code lists that describe categories, subcategories, reference time-pts
  - Applicable value level metadata & description of SUPPQUAL domains
  - Explanations of sponsor-defined identifiers (e.g., –SPID, -GRPID)
- Provide separate unit code lists for each domain

Study Data Reviewer's Guide (SDRG)

- Provide SDRG for each data package with each section populated (TCG 2.2)
  - **Missing (or incomplete) in 35% of applications**
  - Lacking a complete SDRG greatly increases the amount of time reviewers spend understanding an application
- Fix all possible issues identified by FDA Validation Rules
- Include clear and detailed explanation for all "non-fixable" issues (TCG 2.2)
- Provide Data Flow diagram that shows traceability between data capture, storage, and creation of datasets

Disposition (DS) Domain

- Include time-point information about the event in Disposition records, not just when the event was recorded (include start dates for Disposition events)
- Include records regarding subject study completion and last follow-up contact with subject in DS domain
- Accurately code reasons why subject did not complete the study or study treatment

Additional Items in TCG

- Provide Trial Design domains that are complete and accurate (TCG 4.1.1.3)
  - Trial Summary domain missing in 15% of applications
  - Down to 8% (2015, n=24) from 30% (2014, n=10)
- Include EPOCH variable in all appropriate domains (AE, LB, CM, EX, VS, DS) (TCG 4.1.4.1)
  - Missing in 79% of applications
- Use Baseline Flags in LB and VS (TCG 4.1.4.1)
  - Missing baseline tests or flags in 29% of applications

Items in TCG (continued)

- Use CDISC controlled terminology variables when available (TCG Section 6)
  - Controlled terminology issues in 62% of applications
- Include Seriousness Criteria for all serious adverse events (TCG 4.1.1.3)
  - Missing (or with inconsistencies) in 50% of applications
  - Important to independently verify that AE was serious
- Include study day variable for all observational datasets (TCG 4.1.4.1)

Other Issues

- Remove duplicate records
  - 59% with duplicate issues
  - Duplicate records in the LB, VS, and EG domain with potentially contradictory information make it difficult to summarize results
- Provide AE Treatment Emergent Flag in SUPPAE domain
  - 68% missing AETRTEMFL
- Convert to standard units consistently for laboratory data
  - 21% missing (or inconsistent) Standard Units for labs when original units were given
- Provide all AEDECODs
  - 13% missing at least one AEDECOD

Other Issues

- Ensure consistent subject death information across all datasets (DM, AE, DS, and CO)
  - DM death flags consistent between DS and AE records and consistent with death information located in CO domain
  - Important so that pertinent death information not missed if relying only on certain flags or indicators

Other Issues

- Properly use Actual Arm (ACTARM)
- Populate Reference End Date (RFPENDTC) according to SDTM guidance
- Provide additional MedDRA information that facilitates harmonization of versions across studies

Thank You

ADaM:
A Perspective

Steve Wilson, Dr.P.H., CAPT USPHS

Director, Office of Biostatistics, Division of Biometrics III

FDA/CDER/OTS

FDA/PhUSE Computational Science Symposium

March 14, 2016

Silver Spring Civic Center, Silver Spring, Maryland, USA

Disclaimer

This presentation reflects the views of the speaker and should not be construed to represent FDA's views or policies

Outline

- The Purpose of ADaM
- CDER/OTS/Office of Biostatistics
  - Data Review Committee
  - 1530 Statisticians -- Programmers
- Working with OCS to Validate ADaM Submissions
- ADaM and Therapeutic Area Standards
- Getting the View Right – Division Specific Requirements

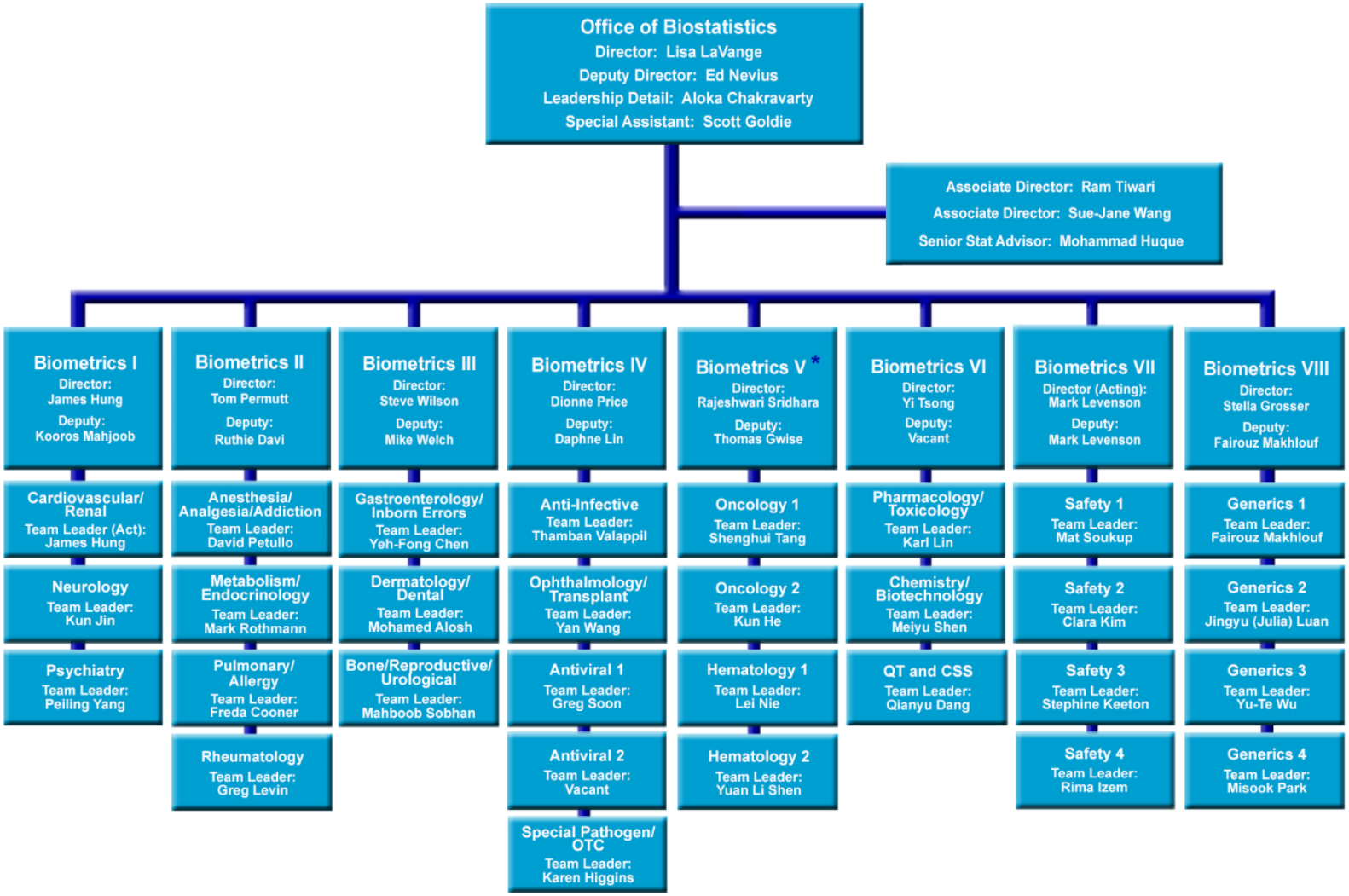
The Purpose of ADaM

- Both the SDTM and ADaM standards were designed to support submission by a sponsor to a regulatory agency such as the United States Food and Drug Administration (FDA).
- The origin of the fundamental principles of ADaM is the need for transparency of communication with and scientifically valid review by regulatory agencies.

Implementing Data Standards

Building the Plane As We Fly It

CDER/OTS/Office of Biostatistics



Office of Biostatistics (OB)
Statistical Review Staff

Reviewers and Data Standards

Being a passenger on the plane while it is being built.

Office of Biostatistics (OB) Statistical Policy Council (SPC) Data Review Committee (DRC) – OB/SPC/DRC

- Assess OB practice for the handling of review data
- Evaluate compliance with CDER policies
- Collaborate with internal (i.e., CDER and FDA) data standards committees and working groups
- Liaise with outside organizations (e.g., CDISC, PhUSE, PharmaSUG, TransCelerate, CPath Institute, CFAST, DIA, HL7, SCDM and ASA)
- Produce SOPs/guidelines
- Help Design Training for Reviewers

Applied Statisticians (GS-1530) at CDER/OB – A New Job Category

- DUTIES
- The incumbent works with the team leader and the statistical reviewer to provide statistical programming and data management support for review of the evidence related to an applicant's claims regarding safety, efficacy and benefit-risk in New Drug Applications (NDAs), Biologic License Applications (BLAs), and supplemental and abbreviated applications (sNDAs; ANDAs);
- … assesses the quality and completeness of submissions;
- … prepares clinical trial analysis datasets;
- … interprets the applicant's programs;
- … assists with communication with applicants regarding specifications for regulatory information requests;
- ….validates sponsor results

Https://www.usajobs.gov/GetJob/ViewDetails/430760100/

USAJOBS "WORKING FOR AMERICA"

Keyword: Job Title, Skills, Agency (etc)

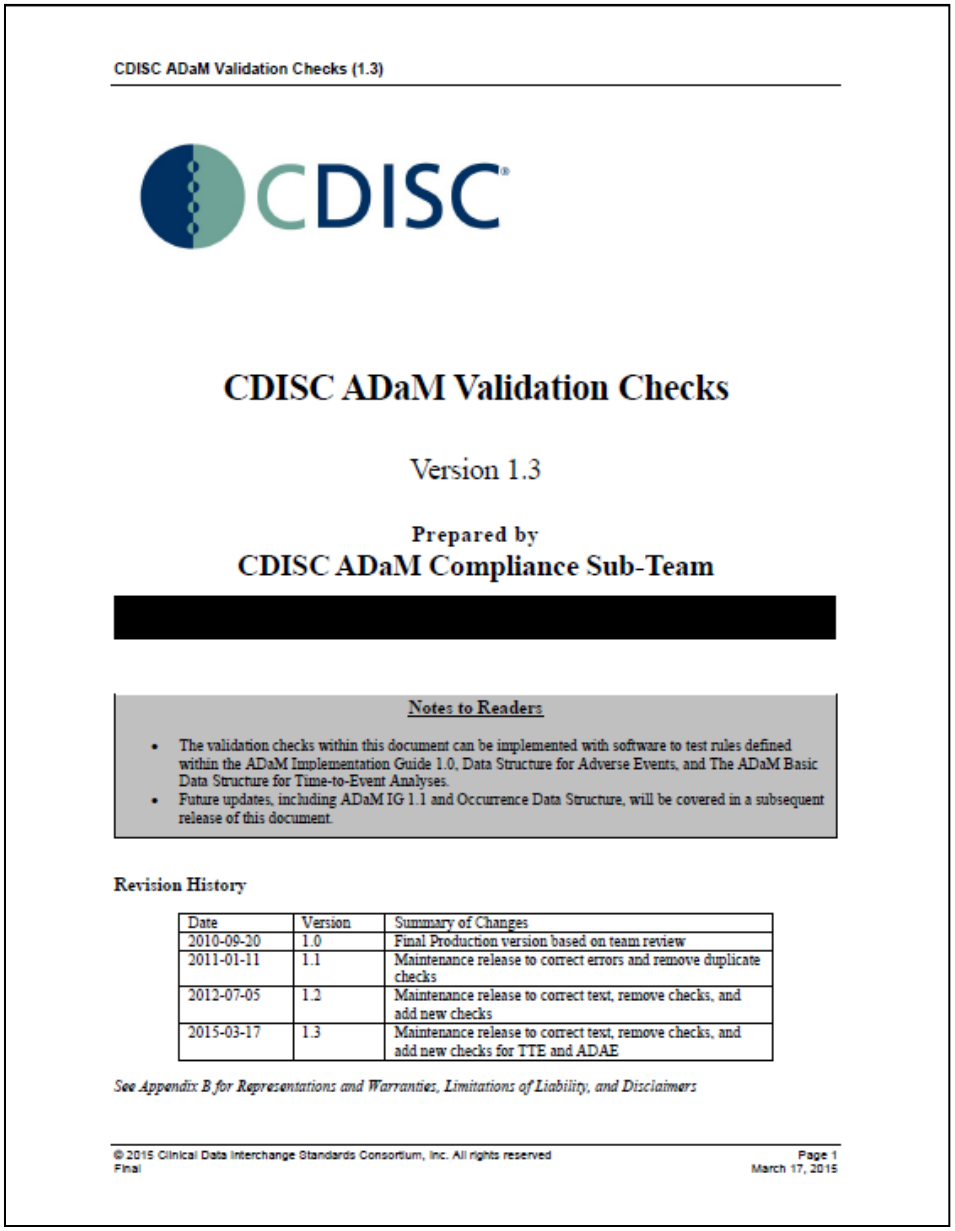
Loc

Overview | Duties | Qualifications & Evaluations | Benefits & Other Info | How to Apply

United States Department of Health & Human Services
Leading America to Better Health, Safety and Well-Being

Job Title: Statistician, GS-1530-12
Department: Department Of Health And Human Services
Agency: Food and Drug Administration
Job Announcement Number: FDA-CR-16-DE-1637185

ADaM Validation



CDISC

CDISC ADaM Validation Checks

Version 1.3

Prepared by

CDISC ADaM Compliance Sub-Team

[redacted]

Notes to Readers

- The validation checks within this document can be implemented with software to test rules defined within the ADaM Implementation Guide 1.0, Data Structure for Adverse Events, and The ADaM Basic Data Structure for Time-to-Event Analyses.
- Future updates, including ADaM IG 1.1 and Occurrence Data Structure, will be covered in a subsequent release of this document.

Revision History

Date	Version	Summary of Changes
2010-09-20	1.0	Final Production version based on team review
2011-01-11	1.1	Maintenance release to correct errors and remove duplicate checks
2012-07-05	1.2	Maintenance release to correct text, remove checks, and add new checks
2015-03-17	1.3	Maintenance release to correct text, remove checks, and add new checks for TTE and ADAE

See Appendix B for Representations and Warranties, Limitations of Liability, and Disclaimers

© 2015 Clinical Data Interchange Standards Consortium, Inc. All rights reserved

Working with OCS to Validate ADaM Submissions

- Process that attempts to ensure that submitted data are both compliant and useful.
  - **Compliant** (conformance) - Data conform to the applicable and required data standards.
  - **Useful** (quality) - Data support the intended use (i.e., regulatory review and analysis).

Adapted from Fitzmartin, 2015

Therapeutic Area Standards

CFAST

Program Overview – February 2016

CDISC

Therapeutic Area	Charter Approved	Check of Concepts Completed	Posted for Internal Review	Posted for Public Review	Projected Publication
Breast Cancer v1	Oct 14	Oct 14	Mar 15	Nov 15	*Q116*
Diabetic Kidney Disease v1	May 15	Aug 15	Jan	Mar	*Q216*
Tuberculosis v2	Apr 15	Apr 15	Sep 15	Oct 15	*Q116*
Rheumatoid Arthritis v1	Jun 15	Oct 15	Jan	Mar	*Q216*
CV Imaging v1	May 15	Jul 15	Dec 15	Mar	*Q316*
Prostate Cancer v1	Nov 15	Mar			*Q316*
Major Depressive Disorder v1**	Dec 15	Feb			*Q316*
Kidney Transplant v1	Jan	Feb			*Q316*
Colorectal Cancer v1	Q116				*Q416*
Vaccines v1*	Q414	Oct 15	Feb	Q116	*Q216*
Ebola v1*	Sep 15	Q116	Q216		*Q316*
Malaria v1*	Oct 15	Q116	Q216		*Q316*
Nutritional Standards v1*	Mar 15	Q116	Q216	Q316	*Q416*
Cardiovascular Concepts in Traditional Chinese Medicine v1*	Q116	Q216			*Q416*

Key| Stage completed | Stage ongoing | All months reflect when stage is, or is projected to be, completed

*Project duration dependent on volunteer resourcing

**General Anxiety Disorder and Bipolar Disorder pending resource availability

ADaM & Therapeutic Area Standards

- PhRMA in the Docket for CDER TA Project Plan.
- More emphasis on [ADaM] given that the "TA standards development initiative focuses primarily on requirements for the efficacy review and evaluation of new medical products."
- At CDER (and at this FDA/ PhUSE Meeting):
  IBM Consultants to the Office of Strategic Programs:
  Susan Kenny & Alisa Khomyanina
- Assignment of expert stat reviewers for Therapeutic Area Users Guides (TAUG) development and acceptance
- More need for industry/academic/government statisticians to get involved in development -- volunteer for TAUG Working Groups

Therapeutic Area Standards

CFAST

Program Overview – February 2016

CDISC

Therapeutic Area	Charter Approved	Check of Concepts Completed	Posted for Internal Review	Posted for Public Review	Projected Publication
Breast Cancer v1	Oct 14	Oct 14	Mar 15	Nov 15	*Q116*
Diabetic Kidney Disease v1	May 15	Aug 15	Jan	Mar	*Q216*
Tuberculosis v2	Apr 15	Apr 15	Sep 15	Oct 15	*Q116*
Rheumatoid Arthritis v1	Jun 15	Oct 15	Jan	Mar	*Q216*
CV Imaging v1	May 15	Jul 15	Dec 15	Mar	*Q316*
Prostate Cancer v1	Nov 15	Mar			*Q316*
Major Depressive Disorder v1**	Dec 15	Feb			*Q316*
Kidney Transplant v1	Jan	Feb			*Q316*
Colorectal Cancer v1	Q116				*Q416*
Vaccines v1*	Q414	Oct 15	Feb	Q116	*Q216*
Ebola v1*	Sep 15	Q116	Q216		*Q316*
Malaria v1*	Oct 15	Q116	Q216		*Q316*
Nutritional Standards v1*	Mar 15	Q116	Q216	Q316	*Q416*
Cardiovascular Concepts in Traditional Chinese Medicine v1*	Q116	Q216			*Q416*

Key| Stage completed | Stage ongoing | All months reflect when stage is, or is projected to be, completed

*Project duration dependent on volunteer resourcing

**General Anxiety Disorder and Bipolar Disorder pending resource availability

Improving the Statistical Review of Confirmatory Clinical Trials

The Elephant in the Conference Room -- Reviewer Needs

FDA Effort to Standardize Data Submission – Antiviral Experience

Wen Zeng, Ph.D. and Greg Soon, Ph.D.
Mathematical Statistician
Division of Biometrics IV, CDER
U.S. Food and Drug Administration

DIA 2011
Chicago, Illinois
DIA®
www.diahome.org

Anti-viral team's effort: HIV case

Why HIV: Currently, there are close to 30 drugs (belongs to 6 classes) approved for HIV treatment.

- Rich datasets from previous NDAs
- Many sponsors
- Many questions to be answered from Meta-data analysis: NI margin, etc

Anti-viral team's effort: HIV case

We did two parts:

Part 1: Request HIV efficacy datasets for coming NDAs

Part 2: Try to convert old NDA datasets SDTM and ADaM

Anti-viral team's effort: Efficacy Dataset

Datasets additional to regular data submission.

Datasets

One record per subject (our "ADSL")

- Subject characteristics and subject-level covariates
- Efficacy outcomes and related covariates

Multiple records per subject (By visit and subject)

- Raw HIV viral load data
- Immunology data (CD4+ count)
- Safety data (Lipids, Liver, Renal, Hema., etc)

Anti-viral team's effort: Converting legacy NDAs into ADaM

10 ADaM datasets were created.

Name	*Description*
ADSL	Subject level Demographic (non-screening failure)
ADSL_SF	Subject level Demographic (screening failure)
ADBL	Baseline Characteristics
ADEX	Exposure
ADEF	Efficacy outcomes
ADCV	Covariates / Subgroups
ADDS	Disposition
ADAE	Adverse Events
ADSF	Safety information (lab)
ADFL	Population flags

Anti-viral team's effort: Converting legacy NDAs into ADaM

Variable Name (max=8)	Variable Label	Type	Length	Codes (example)	Comments
1. Demog (DM) -- ADSL (non screen failure) & ADSL_SF (screen failure)					
STUDYID	Study Identifier	Char	200		Unique identifier for a study.
USUBJID	Unique Subject identifier	Char	200		Unique among all patients submitted for the product.
SUBJID	Subject ID for the study	Char	200		Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.
SITEID	Study site identifier	Char	200		
SITEGRP	Pooled Site Group	Char	200		Same as Country
INVID	Investigator identifier	Char	200		
INVNAM	Investigator Name	Char	200		
4. Population Flag -- ADFL					
STUDYID	Study Identifier	Char	200		Unique identifier for a study.
USUBJID	Unique Subject identifier	Char	200		Unique among all patients submitted for the product.
ITTFL	Intent-to-treat population flag	Char	1	N, Y	ADaM character indicator for the ITT population: subjects randomized
ITTFLN	Intent-to-treat population flag, Num	Num	8	0, 1	ADaM numeric indicator for the ITT population: subjects randomized
SAFFL	Safety population flag	Char	1	N, Y	ADaM character indicator for the Safety population: subjects received at least one dose
SAFFLN	Safety population flag, Num	Num	8	0, 1	ADaM numeric indicator for the Safety population: subjects received at least one dose

Anti-viral team's effort: Converting legacy NDAs into ADaM

One example of ADaM Sponsor proposed.

Dataset	Description	Class	Structure	Purpose	Keys	Location
ADAE	Adverse Events Analysis	Analysis	One record per subject, per adverse event, per event start date	Analysis	STUDYID, USUBJID, AEDECOD, AETERM, ASTDTC, AENDTC	adae.xpt
ADCM	Antiretroviral and Concomitant Medications Analysis	Analysis	One record per subject, per medication source, per generic term, per sponsor ID, per medication start date	Analysis	STUDYID, SUBJID, SOURCE, WHOGEN, CMSPID, CMSTDTC	adcm.xpt
ADEFF	Efficacy Analysis	Analysis	One record per subject, per parameter, per analysis visit, per timepoint.	Analysis	STUDYID, USUBJID, PARAMCD, AVISITN, ADT, ATM	adeff.xpt
ADLB	Laboratory Test Results Analysis	Analysis	One record per subject, per lab group, per lab test, per analysis visit, per timepoint.	Analysis	STUDYID, USUBJID, PARCAT1, PARAMCD, AVISITN, ADT, ATM	adlb.xpt
ADPC	Plasma Concentration Analysis	Analysis	One record per subject, per analyte, per visit, per timepoint.	Analysis	STUDYID, USUBJID, PARAMCD, VISITNUM, PCTPTNUM	adpc.xpt
ADSL	Subject-Level Analysis Dataset	Analysis	One record per subject	Analysis	STUDYID, USUBJID	adsl.xpt
ADVS	Vital Signs Analysis	Analysis	One record per subject, per vital sign test, per visit, per timepoint.	Analysis	STUDYID, USUBJID, PARAMCD, AVISITN, ADT, ATM	advs.xpt
ADTTE	Time to Event Analysis	Analysis	One record per subject, per parameter	Analysis	STUDYID, USUBJID, PARAMCD	adtte.xpt
ADVIRO	HIV Resistance Data	Analysis	One record per subject, per isolate	Analysis	STUDYID, USUBJID, ISOLATE	adviro.xpt

Thinking About Analysis Review Data Sets and Programs

- Better understanding of the communication between industry and reviewers during the review process
- “Views” built on CDISC standards -- tailored to meet the needs of review.
- Finishing that plane: More thought about what a reviewer needs to do with submitted datasets – Data Standards & Good Review Practice
- There’s got to be a PhUSE Project Group in here somewhere!

When we see that smiling little kid (i.e., a CDER reviewers &/or an industry professional) … that should be all of the thanks we need

THANKS

stephen.wilson@cder.fda.gov

How good is your SEND data?

Timothy Kropp
FDA/CDER/OCS

How good is your SEND data?

Perspectives from…

- Data submitted with regulatory applications
- Data submitted with test applications
- Data loading into Janus

Current Status on SEND

- Receiving SEND 3.0 according to Data Standards Catalog
- Data validation rules available (http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM367129.xls)
- Using DataFit for clinical and nonclinical
- Nonclin SDRG in use and development (very helpful!)
- KickStart service for reviewer support

Items in TCG

- Encouragingly, we don't find many SEND issues that are related to the nonclinical-specific information in the TCG.
- Future updates to the nonclinical sections are expected – keep an eye on this important document.

Validation

- All regulatory submissions loaded for review.
- No critical data issues with regulatory submissions
- Test submissions, however, show issues.
- Most conformance issues are identical to clinical submissions (but with a lot fewer critical errors)
- There is also some sloppiness that you can't catch with a validator….

Inconsistencies in Trial Descriptions

STUDYID	DOMAIN	SETCD	SET	TXSEQ	TXPARMCD	TXPARM	TXVAL
XYZ123	TX	1	Vehicle Contro, Males	1	ARMCD	Arm Code	1
XYZ123	TX	1	Vehicle Contro, Males	2	GRPLBL	Group Label	Group 1, Control
XYZ123	TX	1	Vehicle Contro, Males	3	SPGRPCD	Sponsor-Defined Group Code	1
XYZ123	TX	1	Vehicle Contro, Males	4	TRTDOS	Dose Level	0
XYZ123	TX	1	Vehicle Contro, Males	5	TRTDOSU	Dose Units	mg/kg
XYZ123	TX	1	Vehicle Contro, Males	6	TCNTRL	Control Type	Vehicle
XYZ123	TX	1	Vehicle Contro, Males	9	SPLANSUB	Planned Number of Subjects	20
XYZ123	TX	1	Vehicle Contro, Males	10	PLANMSUB	Planned Number of Male Subjects	10
XYZ123	TX	1	Vehicle Contro, Males	11	PLANFSUB	Planned Number of Female Subjects	10
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	13	ARMCD	Arm Code	2
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	14	GRPLBL	Group Label	Group 2, 5 mg/kg
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	15	SPGRPCD	Sponsor-Defined Group Code	2
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	16	TRTDOS	Dose Level	5
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	17	TRTDOSU	Dose Units	mg/kg
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	21	SPLANSUB	Planned Number of Subjects	20
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	22	PLANMSUB	Planned Number of Male Subjects	10
XYZ123	TX	2	Low-Dose Males, 5 mg/kg Drug A once daily	23	PLANFSUB	Planned Number of Female Subjects	10
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	25	ARMCD	Arm Code	3
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	26	GRPLBL	Group Label	Group 3, 50 mg/kg
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	27	SPGRPCD	Sponsor-Defined Group Code	3
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	28	TRTDOS	Dose Level	50
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	29	TRTDOSU	Dose Units	mg/kg
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	33	SPLANSUB	Planned Number of Subjects	20
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	34	PLANMSUB	Planned Number of Male Subjects	10
XYZ123	TX	3	High-Dose Males, 50 mg/kg Drug A once daily	35	PLANFSUB	Planned Number of Female Subjects	10

Inconsistencies in Study Protocol and Dataset

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EXDOSFRQ	EXROUTE	EXLOT	EXMETHOD	EXTRTV	EXVAMT	EXVAMTU	EXSTDTC
XYZ123	EX	XYZ123-001	1	Drug A	0	mg/kg	LIQUID	ONCE	INTRAVENOUS		INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYZ123-002	2	Drug A	0	mg/kg	LIQUID	ONCE	INTRAVENOUS		INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYZ123-003	3	Drug A	0	mg/kg	LIQUID	ONCE	INTRAVENOUS		INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYZ123-004	4	Drug A	0	mg/kg	LIQUID	ONCE	INTRAVENOUS		INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYZ123-005	5	Drug A	0	mg/kg	LIQUID	ONCE	INTRAVENOUS		INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-006	6	Drug A	5	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-007	7	Drug A	5	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-008	8	Drug A	5	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-009	9	Drug A	5	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-010	10	Drug A	5	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-011	11	Drug A	50	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-012	12	Drug A	50	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-013	13	Drug A	50	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-014	14	Drug A	50	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00
XYZ123	EX	XYA123-015	15	Drug A	50	mg/kg	LIQUID	ONCE	INTRAVENOUS	12345	INFUSION	Saline	5	mL	2010-01-05T10:05:00

- The variable is populated with controlled terminology, however, the data between the study report are sometimes inconsistent.

- The study report PDF specifies that the animals were dose QD, but the exposure domain (EX) has the dosing frequency as ONCE!

Your CT error may indicate more than a CT issue

STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU
XYZ123	LB	XYZ123-005	1	CREAT	Creatinine	CLINICAL CHEMISTRY	55	umol/L
XYZ123	LB	XYZ123-006	2	CREAT	Creatinine	CLINICAL CHEMISTRY	46	umol/L
XYZ123	LB	XYZ123-007	3	CREAT	Creatinine	CLINICAL CHEMISTRY	62	umol/L
XYZ123	LB	XYZ123-008	4	CREAT	Creatinine	CLINICAL CHEMISTRY	CS	umol/L
XYZ123	LB	XYZ123-009	5	CREAT	Creatinine	CLINICAL CHEMISTRY	55	umol/L
XYZ123	LB	XYZ123-010	6	CREAT	Creatinine	CLINICAL CHEMISTRY	56	umol/L
XYZ123	LB	XYZ123-011	7	CREAT	Creatinine	CLINICAL CHEMISTRY	NS	umol/L
XYZ123	LB	XYZ123-012	8	CREAT	Creatinine	CLINICAL CHEMISTRY	51	umol/L
XYZ123	LB	XYZ123-013	9	CREAT	Creatinine	CLINICAL CHEMISTRY	CS	umol/L

- An original result does not always have an associated standardized result if the collection method uses the variable for more than one type of information.

- In this example, clotted sample (CS), is collected. Often this can be propagated into the STRESC, when it should be null with a Completion Status (–STAT) and Reason Not Done (–REASND) variables will be populated as “NOT DONE “and “Clotted sample”.

- If you have a CT issue, fix it. Also, please be aware that you may have more than a CT issue; you may also be putting information into the wrong place.

CTR Clinical and Nonclinical Critical Data Quality Issues

- Duplicates
  - Duplicate trial element codes
  - Duplicate trial inclusion exclusion criteria
  - Duplicate trial visits
  - Trial Summary duplicate records
- Structural
  - Duplicate STUDYID values
  - More than one record per subject/test/collection date
  - Incorrect data type
  - Special characters
  - Non ISO8601
- Missing Trial Summary information
- Missing units
- Invalid or incomplete define file

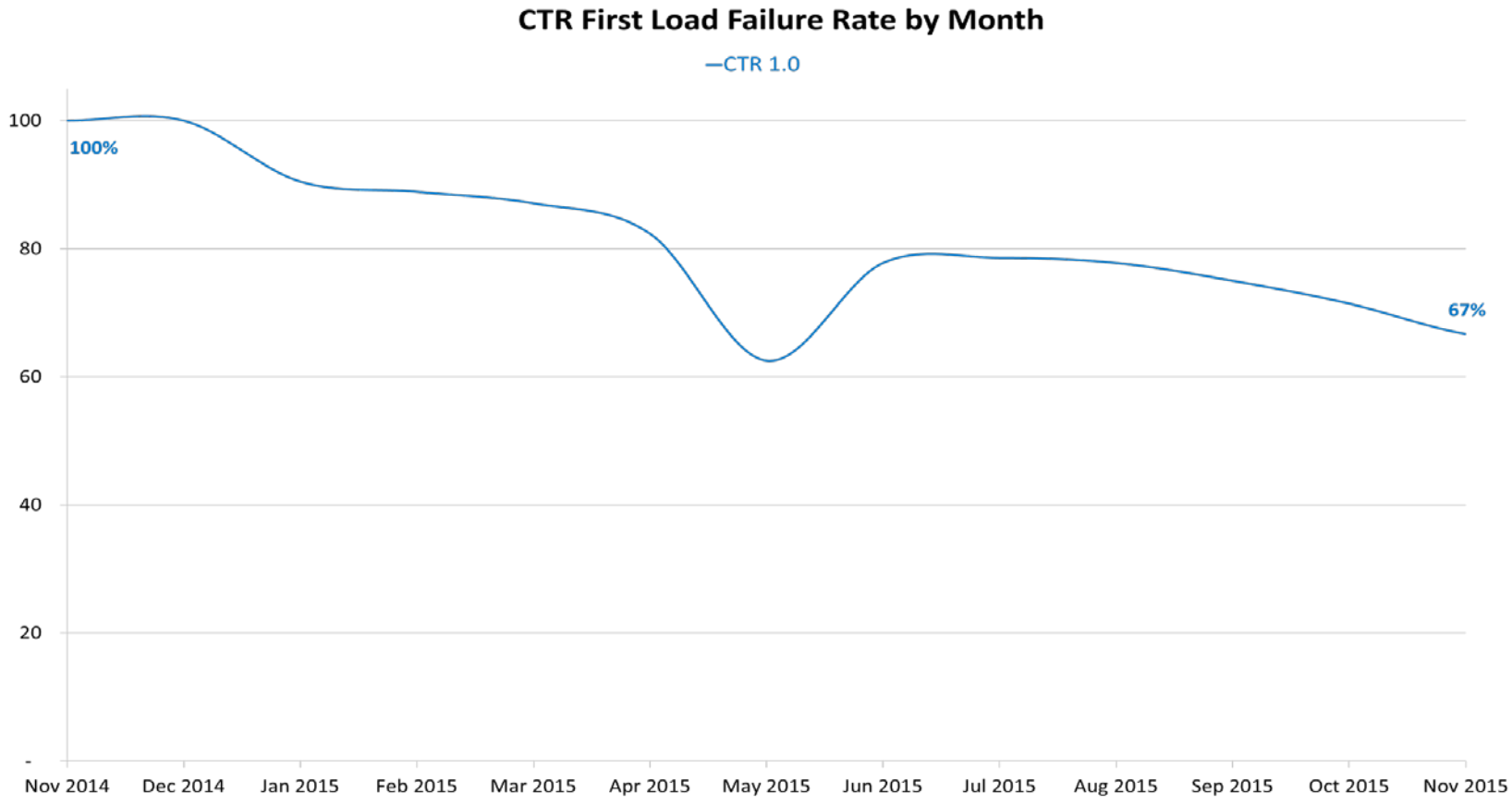
Perspectives from Data Loading and Processing

- Clinical
  - Processing, validation, loading, and use
  - Janus repository, DataFit, various tools, and JumpStart
- Nonclinical
  - Processing, validation, loading and use
  - Janus repository, DataFit, various tools, and KickStart

JANUS Clinical Trials Repository

- Data repository for subject-level clinical and toxicology trial data
- Operational at FDA's White Oak Data Center since 1/15
- Currently ingesting SDTM and SEND data
- Errors for loading inform technical issues
- More experience with clinical but since issues are roughly the same, the outcomes likely will be as well.

CTR Data Loading Overview



CTR Clinical Loading Stats

- Loaded on 1st attempt: 22%
- Could not load to CTR: 4%
- Could not load on 1st attempt: 74%

CTR Clinical Loading Stats

Could not load into CTR:

- Missing Define files
- Other significant data issues
  - Cannot load without additional information from Applicant

CTR Clinical Loading Stats

Eventually loaded:

- Critical Data Quality Issues: 55%
  - 43% had issues within Define files
  - 49% had issues within TS domain
- Missing TS domain: 18%
- Other loading issues: 11%
  - Other domains or sizing issues

Examples of Define File Errors

Error	Description
Study Name element is same for two different studies under same product application number	Two different studies have same Study Name element in define.xml for same product application number
Invalid XSL file	The XSL file included with the study data is invalid. This results in an error in creating the define html
Datasets included in the study data folder but not described in define.xml	CTR uses define.xml to load study data. If a domain is not described in define.xml then it will not be loaded.
define.xml refers to a define.xsl file that does not exist	define.xml refers to a define.xsl file that does not exist in the study folder. CTR uses the referred define.xsl to render define.xml into html for display on the study summary application.
Invalid class for a custom domain	The define.xml specifies an invalid class such as Findings for a domain that should be loaded as an Events or Interventions domain.

Examples of Trial Summary Errors

Error	Description
Treatments not associated correctly with the dose and route data	Trial summary values for treatment need to be associated with the dose and the route trial summary values. If there are multiple values for treatment and dose and route then TSGRPID values should be used to associate which treatment goes with which dose. If multiple treatments are included in a study and this association is not provided then load software assumes that each dosage group is associated with each dosage group.
Trial summary records are not unique based on TSPARMCD and TSSEQ	Trial summary records should be uniquely identifiable using TSPARMCD and TSSEQ, There cannot be duplicate records for the combination of TSPARMCD and TSSEQ values
Trial Summary Value (TSVAL) for TSPARMCD AGEMIN uses invalid format	AGEMIN is required to be in the ISO 8601 format. If a single age is listed (i.e., 18), the age should be in the format of P##Y, where the ## are the age noted. In this example, 18 should be P18Y.
Trial Summary ACTSUB has invalid format	The value for parameter ACTSUB has to be a number or null
Trial Summary Value (TSVAL) for TSPARMCD AGEMAX uses invalid format	AGEMAX is required to be in the ISO 8601 format. Values such as UNLIMITED, NONE, No maximum age are invalid
Invalid format for TSPARMCD RANDQT	Randomization quotient needs to be a numerical value.
Trial summary parameters SSTDTC and SSENDTC are of invalid format	Trial summary parameters specifying date/time values need to be in proper date time format.

Examples of Other Errors

Error	Description
Duplicate records in observation domains	There are duplicate records with in any of the findings,events or interventions domains. Multiple records have the same sequence number value for the same USUBJID within a domain.
Special characters in data set files	There are special characters in the data set files: ex. □,β,™
Invalid Trial Design domain	Valid Trial Design domains based on the SDTM IG are TS, TE,TV, TA and TI. Any other domain with a class of Trial Design will result in load errors
Duplicate USUBJID in DM domain	Duplicate values for USUBJID in the demographic domain cannot be loaded. Each records in the DM domain is expected to identify a unique subject
Invalid date values	Date/time values in SDTM are required to follow the ISO 8601 date/time format. These invalid formats can result in invalid data in the CTR warehouse and cause some of the analysis panels to not work.
Non-standard variables in the submission data sets	SDTM does not allow the addition of new or unexpected variables

Summary