Introducing the CDISC Analysis Data Model (ADaM) Implementation Guide

Michael Nessly

Director, Global Biostatistics

Shire Specialty Pharma

WHO Needs Standards ?

Important Caveats

Data Standards are not solely a FDA thing

The value is not only in submission

There are efficiency gains attainable from end to end in the data flow

The value in data exchange in partnerships, outsourcing, joint ventures, acquistions, mergers is considerable

Background

CDISC: Clinical Data Interchange Standards Consortium

SDTM: Study Data Tabulation Model

standard for interchange of ***collected*** data

submitted in Item 11

ADaM: Analysis Data Model

standard for interchange of ***analysis*** data

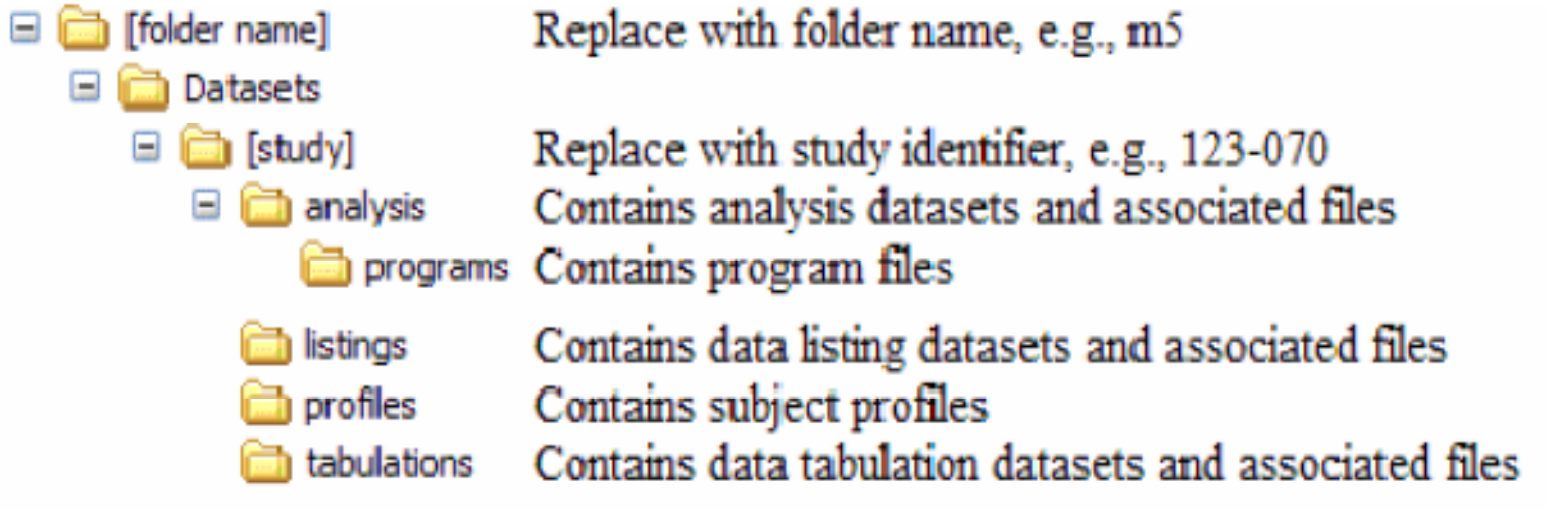
submitted in SRA (Statistical Review Aid)

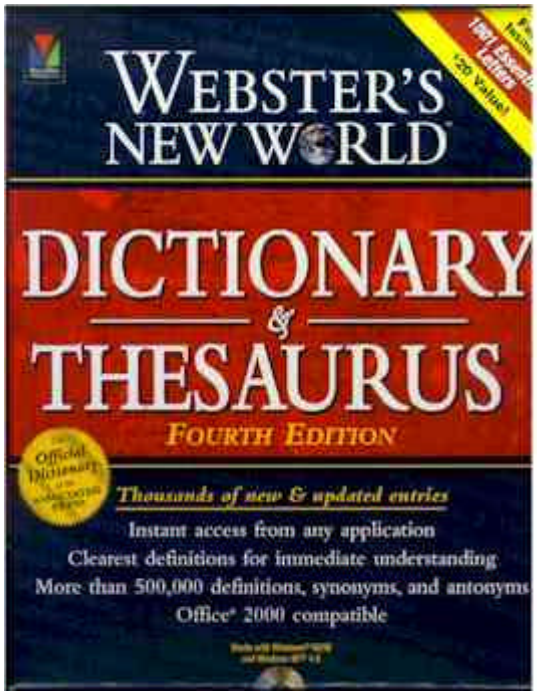
ADaM in eCTD

Submitted data are classified into four types:

(1) Data tabulations; (2) Data listings;

(3) Analysis datasets; (4) Subject profiles.

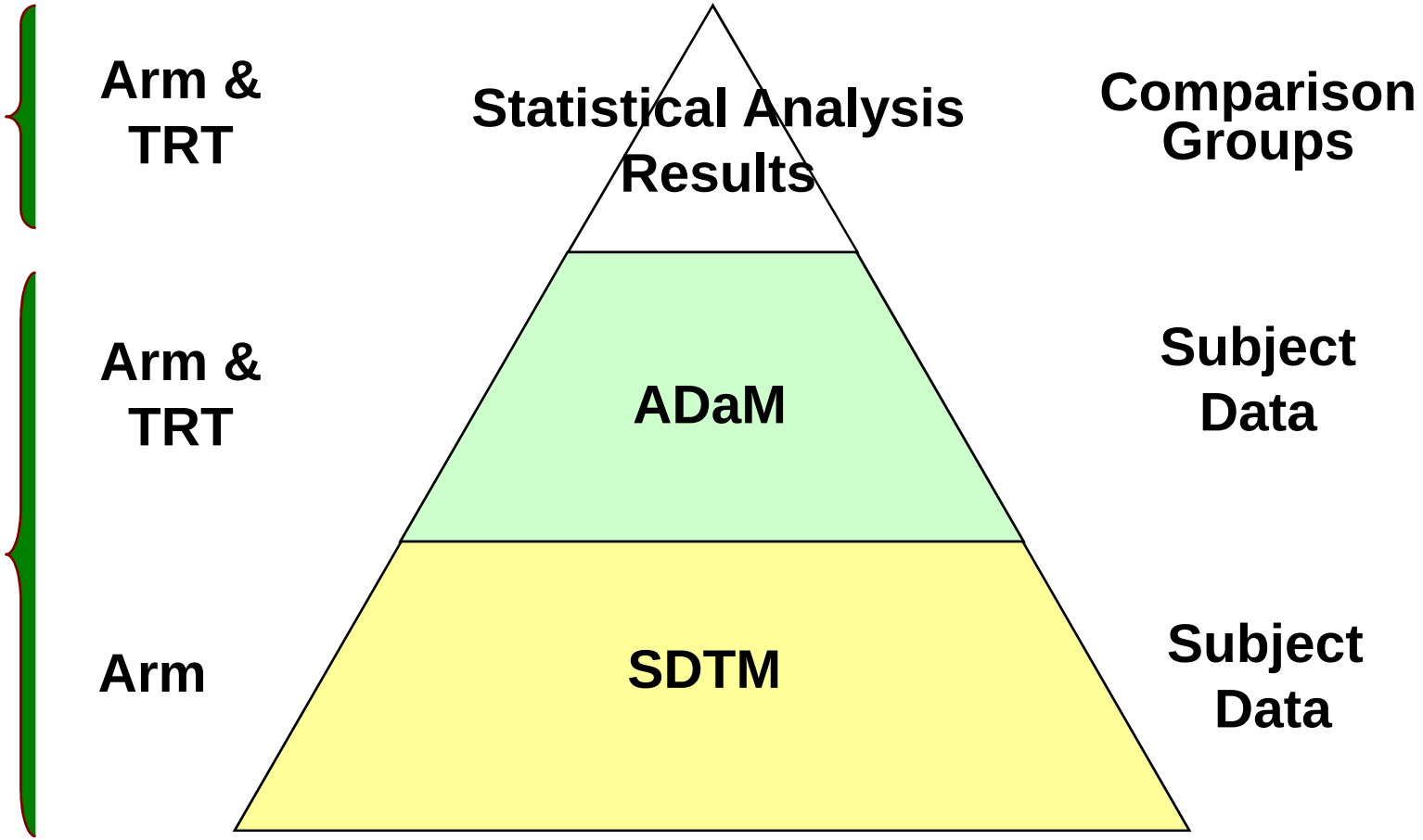




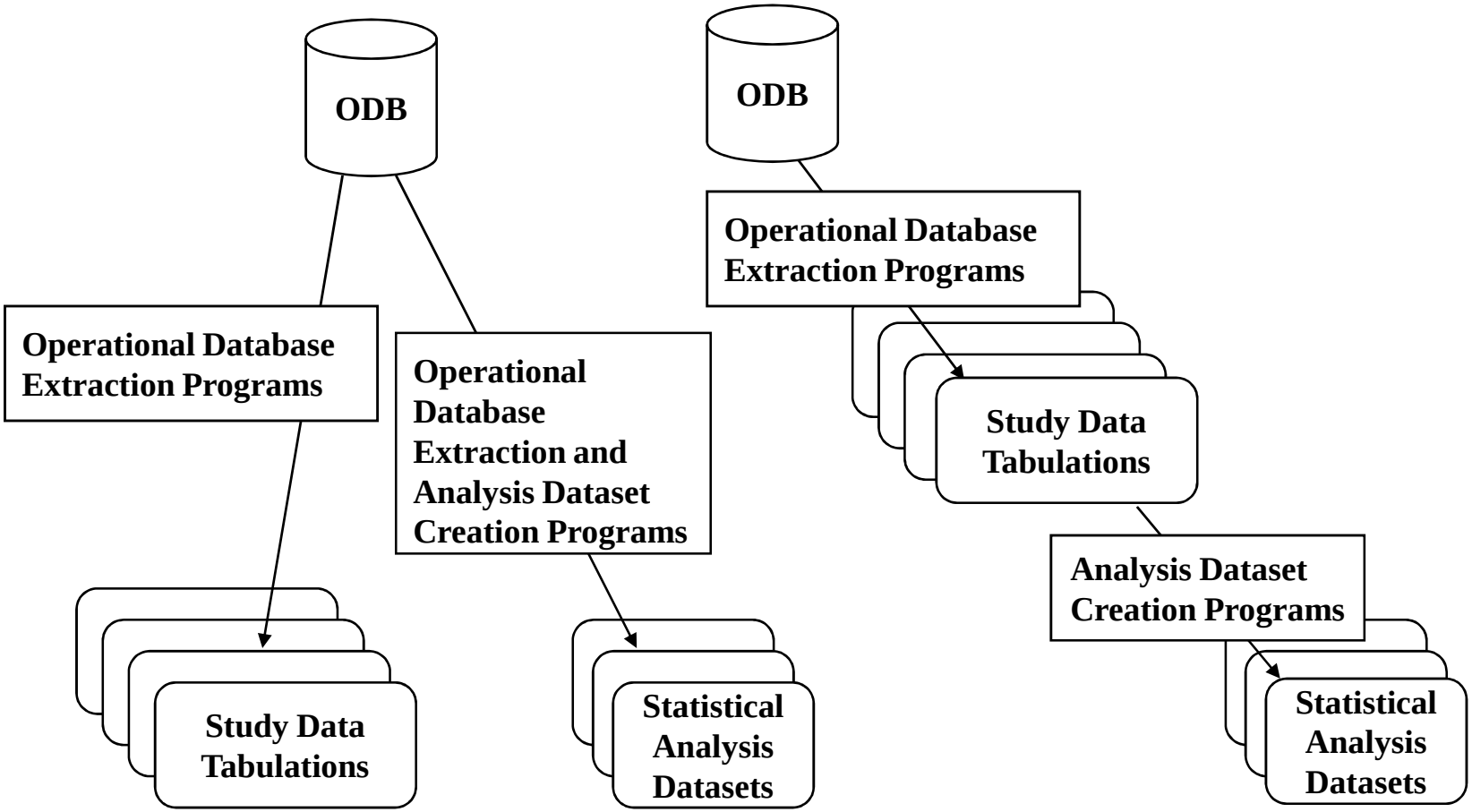
Upon this gifted age
in its dark hour
Falls from the sky
a meteoric shower
Of facts.
They lie, unquestioned,
uncombined Wisdom enough to
leach us of our ills
is daily spun
But there exists no loom
to weave it into fabric.

SDTM **ADaM**

From Data As Collected to Conclusions



Analysis Dataset Creation: Parallel vs Sequential Data Flow



Caveat

The statistical analysis data models presented here represent consensus across a large number of reviewers experienced in regulatory submissions. While these examples are presented as models, it must be recognized that these examples do not represent the only or even necessarily the best approach for all situations. While different data structures may be better suited for other situations, the standards represented herein should be adhered to whenever possible. We cannot emphasize enough that early and effective cross-communication between regulatory reviewers and sponsors is requisite for mutual success.

ADaM Key Principles

Shire

Analysis datasets should:

- facilitate unambiguous communication and provide a level of traceability
- be linked to machine-readable metadata
- be useable by currently available tools
- be analysis-ready

ADaM Datasets

Shire

Analysis datasets must:

include **subject-level** analysis dataset "ADSL".

comprise the **optimum number** of analysis datasets needed to allow analysis and review with little or no additional programming or data processing.

be named using the convention "**ADxxxxxx**."

follow naming conventions for datasets and variables that are **applied consistently** across studies, across a given submission or across multiple submissions for a product.

The ADaM Standard Analysis Dataset Structures

Shire

Subject Level Data (ADSL)

one record per subject

Basic Data Structure

one or more records

per subject,

per analysis parameter,

per analysis timepoint

ADaM Variable Names

Any SDTM variable name in ADaM:

"same name, same meaning, same values"

- When ADaM column has an SDTM name, values must be copied and not altered from SDTM

If a standard ADaM variable exists, then one must use the ADaM variable name for that concept

Obey SAS V5 transport file naming and labeling conventions

ADaM Variable Name Fragments

*N – a numeric version of a variable named *

*GRP – a grouping of a variable named *

*GRPN – a numeric version of a grouping of a variable named *

RACE, RACEN, RACEGRP, RACEGRPN

*FL – character flag

*FN – numeric version of character flag

ADaM Variable Name Fragments

TRT* – a treatment variable

*DY – a relative day (no day 0)

*DT, *TM, *DTM – numeric date, time, datetime

*DTF – date imputation flag (Y, M, D, null)

*TMF – time imputation flag (H, M, S, null)

ADaM "Core" Definitions

SDTM: Required, Expected, Permissible

ADaM:

Req	**Required**
Cond	**Required if applicable; conditionally required**
Perm	**Permissible**

Examples: Color Scheme

ADaM Core	Background color used in examples in this presentation
Req	Required standard ADaM variable
Req	Required ADaM variable copied from SDTM
Cond	Conditionally required standard ADAM variable
Perm	Permissible standard ADaM variable
Perm	Permissible / recommended variable copied from SDTM
Perm	Permissible user-created variable

Basic Data Structure Example

Row	USUBJIB	PARAMCD	AVISIT	AVAL	DTYPE	ANLFL	ITTFL	TRT1P
1	2782	SYSBP	Screening	120		Y	Y	Soma 5 mg
2	2782	SYSBP	Run-In	116		Y	Y	Soma 5 mg
3	2782	SYSBP	Week 0	114		Y	Y	Soma 5 mg
4	2782	SYSBP	Week 2	118		Y	Y	Soma 5 mg
5	2782	SYSBP	Week 2	126			Y	Soma 5 mg
6	2782	SYSBP	Week 4	122		Y	Y	Soma 5 mg
7	2782	SYSBP	Week 8	122	LOCF	Y	Y	Soma 5 mg
8	2782	SYSBP	Week 8	126	WOCF	Y	Y	Soma 5 mg

Analysis Parameter Variables

PARAM uniquely describes AVAL.
PARAM has no qualifiers – so is different from SDTM - -TEST.
PARAMCD and PARAMN are each 1:1 maps to PARAM.

Row	PARAM	PARAMCD	PARAMN	PARAMCAT	PARAMTYP	AVAL
1	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		120
2	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		116
3	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		114
4	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		118
5	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		126
6	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		122
7	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		122
8	Sitting Systolic BP (mm Hg)	SYSBP	4	VITALS		126

Analysis Parameter Variables

PARAM uniquely describes AVAL.
PARAM has no qualifiers – so is different from SDTM - -TEST.
PARAMCD and PARAMN are 1:1 maps to PARAM.

Row	PARAM	PARAMCD	PARAMN	PARAMCAT	PARAMTYP	AVAL
1	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.079
2	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.064
3	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.057
4	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.072
5	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.100
6	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.086
7	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.086
8	Log10(Sitting Systolic BP (mm Hg))	LSYSBP	27	VITALS	DERIVED	2.100

Analysis Timepoint Variables

Row	AVISIT	AVISITN	VISIT	VISITNUM	VSSEQ	ADY	AWTARGET	AWTDIFF	ANLFL
1	Screening	-4	VISIT 1	1	3821	-30	-28	2	Y
2	Run-In	-2	VISIT 2	2	3822	-16	-14	2	Y
3	Week 0	0	VISIT 3	3	3823	-2	1	2	Y
4	Week 2	2	VISIT 4	4	3824	13	14	1	Y
5	Week 2	2	VISIT 4 UNSCHEDULED	4.1	3825	17	14	3	
6	Week 4	4	VISIT 5	5	3826	23	28	5	Y

Analyzed Record Flag ANLFL

Row	AVISIT	AVISITN	VISIT	VISITNUM	VSSEQ	ADY	AWTARGET	AWTDIFF	ANLFL
1	Screening	-4	VISIT 1	1	3821	-30	-28	2	Y
2	Run-In	-2	VISIT 2	2	3822	-16	-14	2	Y
3	Week 0	0	VISIT 3	3	3823	-2	1	2	Y
4	Week 2	2	VISIT 4	4	3824	13	14	1	Y
5	Week 2	2	VISIT 4 UNSCHEDULED	4.1	3825	17	14	3	
6	Week 4	4	VISIT 5	5	3826	23	28	5	Y

Record Derivation Type DTYPE

Row	AVISIT	AVISITN	VISIT	VISITNUM	VSSEQ	AVAL	DTYPE	ANLFL
1	Screening	-4	VISIT 1	1	3821	120		Y
2	Run-In	-2	VISIT 2	2	3822	116		Y
3	Week 0	0	VISIT 3	3	3823	114		Y
4	Week 2	2	VISIT 4	4	3824	118		Y
5	Week 2	2	VISIT 4 UNSCHEDULED	4.1	3825	126		
6	Week 4	4	VISIT 5	5	3826	122		Y
7	Week 8	8	VISIT 5	5	3826	122	LOCF	Y
8	Week 8	8	VISIT 4 UNSCHEDULED	4.1	3825	126	WOCF	Y

Record Derivation Type DTYPE

Row	AVISIT	AVISITN	VISIT	VISITNUM	VSSEQ	AVAL	DTYPE	ANLFL
1	Screening	-4	VISIT 1	1	3821	120		Y
2	Run-In	-2	VISIT 2	2	3822	116		Y
3	Week 0	0	VISIT 3	3	3823	114		Y
4	Week 2	2	VISIT 4	4	3824	118		Y
5	Week 2	2	VISIT 4 UNSCHEDULED	4.1	3825	126		
6	Week 4	4	VISIT 5	5	3826	122		Y
7	Week 8	8	VISIT 5	5	3826	122	LOCF	Y
8	Week 8	8	VISIT 4 UNSCHEDULED	4.1	3825	126	WOCF	Y

Baseline Record Flag ABLFL

Row	PARAMCD	AVISIT	AVISITN	AVAL	ABLFL	BASE	DTYPE	ANLFL
1	SYSBP	Screening	-4	120		114		Y
2	SYSBP	Run-In	-2	116		114		Y
3	SYSBP	Week 0	0	114	Y	114		Y
4	SYSBP	Week 2	2	118		114		Y
5	SYSBP	Week 2	2	126		114		
6	SYSBP	Week 4	4	122		114		Y
7	SYSBP	Week 8	8	122		114	LOCF	Y
8	SYSBP	Week 8	8	126		114	WOCF	Y

Baseline Type BASETYPE

Row	PARAMCD	AVISIT	AVISITN	AVAL	AVALC	ABLFL	BASE	BASEC	BASETYPE
1	SYSBP	Screening	-4	120			114		Base Study
2	SYSBP	Run-In	-2	116			114		Base Study
3	SYSBP	Week 0	0	114		Y	114		Base Study
4	SYSBP	Week 2	2	118			114		Base Study
5	SYSBP	Week 2	2	126			114		Base Study
6	SYSBP	Week 4	4	122			114		Base Study
7	SYSBP	Week 8	8	122			114		Base Study
8	SYSBP	Week 8	8	126			114		Base Study

Parameter-Invariant Functions of AVAL and BASE

Row	PARAMCD	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	PCHG	R2BASE
1	SYSBP	Screening	-4	120		114	6	5.26	1.053
2	SYSBP	Run-In	-2	116		114	2	1.75	1.018
3	SYSBP	Week 0	0	114	Y	114	0	0	1
4	SYSBP	Week 2	2	118		114	4	3.51	1.035
5	SYSBP	Week 2	2	126		114	12	10.52	1.105
6	SYSBP	Week 4	4	122		114	8	7.02	1.070
7	SYSBP	Week 8	8	122		114	8	7.02	1.070
8	SYSBP	Week 8	8	126		114	12	10.52	1.105

Selection of Records for Analysis

Row	PARAMCD	AVISIT	AVISITN	AVAL	BASE	CHG	DTYPE	ANLFL	ITTFL	TRTPN
1	SYSBP	Screening	-4	120	114	6		Y	Y	2
2	SYSBP	Run-In	-2	116	114	2		Y	Y	2
3	SYSBP	Week 0	0	114	114	0		Y	Y	2
4	SYSBP	Week 2	2	118	114	4		Y	Y	2
5	SYSBP	Week 2	2	126	114	12			Y	2
6	SYSBP	Week 4	4	122	114	8		Y	Y	2
7	SYSBP	Week 8	8	122	114	8	LOCF	Y	Y	2
8	SYSBP	Week 8	8	126	114	12	WOCF	Y	Y	2

Selection Criteria for ITT analyses of Change from Baseline in SYSBP at week 8

	PARAMCD = 'SYSBP' and AVISITN = 8 and ITTFL = 'Y' and
Data as Observed	DTYPE = '' and ANLFL = 'Y'
LOCF	(DTYPE = '' or DTYPE = 'LOCF') and ANLFL = 'Y'
WOCF	(DTYPE = '' or DTYPE = 'WOCF') and ANLFL = 'Y'

Alternative: create an analyzed record flag ANLxFL for each of n analyses, x=1 to n

Crossover Example

USUBJID	AVISIT	AVISITN	VISITNUM	DTYPE	ANLFL	TRT1PN	TRT2PN	TRTPN	PERIOD	TRTSEQPN	AVAL
3984	Screening	-4	1		Y	2	1			2	16
3984	Week -2	-2	2		Y	2	1			2	16
3984	Week 0	0	3		Y	2	1			2	18
3984	Baseline	-8888		AVERAGE	Y	2	1			2	17
3984	Week 4	4	4		Y	2	1	2	1	2	14
3984	Week 8	8	4.1			2	1	1	2	2	10
3984	Week 8	8	5		Y	2	1	1	2	2	12
3984	Endpoint	9999		ENDPOINT	Y	2	1	2	1	2	14
3984	Endpoint	9999		ENDPOINT	Y	2	1	1	2	2	12

Time to Event Example

Analyze Time to First Hypertension Event:
SYSBP > 130, DIABP > 90, or Hospitalization

DATA AS FOUND IN SDTM VS DATASET						
USUBJID	VISITNUM	VSSEQ	VSDTC	VSDY	VSTESTCD	VSSTRESN
2010	1	22	2004-08-05	1	SYSBP	115
2010	1	23	2004-08-05	1	DIABP	75
2010	2	101	2004-08-12	8	SYSBP	120
2010	2	102	2004-08-12	8	DIABP	90
2010	3	207	2004-08-19	15	SYSBP	135
2010	3	208	2004-08-19	15	DIABP	92

DATA AS FOUND IN SDTM DS DATASET					
USUBJID	DSSEQ	DSSTDTC	DSSTDY	DSDECOD	DSTERM
2010	25	2004-08-05	1	RANDOM	Subject Randomized
2010	99	2004-08-13	9	HOSPSTRT	Subject Hospitalized
2010	140	2004-08-15	11	HOSPEND	Subject Discharged from Hospital
2010	301	2004-08-26	22	COMPLETED	Subject Completed

Time to Event Example

DATA AS FOUND IN SDTM VS DATASET						
USUBJID	VISITNUM	VSSEQ	VSDTC	VSDY	VSTESTCD	VSSTRESN
2010	1	22	2004-08-05	1	SYSBP	115
2010	1	23	2004-08-05	1	DIABP	75
2010	2	101	2004-08-12	8	SYSBP	120
2010	2	102	2004-08-12	8	DIABP	90
2010	3	207	2004-08-19	15	SYSBP	135
2010	3	208	2004-08-19	15	DIABP	92

DATA AS FOUND IN SDTM DS DATASET					
USUBJID	DSSEQ	DSSTDTC	DSSTDY	DSDECOD	DSTERM
2010	25	2004-08-05	1	RANDOM	Subject Randomized
2010	99	2004-08-13	9	HOSPSTRT	Subject Hospitalized
2010	140	2004-08-15	11	HOSPEND	Subject Discharged from Hospital
2010	301	2004-08-26	22	COMPLETED	Subject Completed

USUBJID	PARAM	PARAMCD	AVAL	SRCDOM	SRCVAR	SRCSEQ
2010	Time to Hospitalization (Days)	HOSPSTRT	9	DS	DSSTDY	99
2010	Time to SYSBP > 130 mm Hg (Days)	SBP	15	VS	VSDY	207
2010	Time to DIABP > 90 mm Hg (Days)	DBP	15	VS	VSDY	208
2010	Time to First Hypertension Event (Days)	HYPEREVT	9			

When to Derive Rows vs. Columns

1. A parameter-invariant function of AVAL and BASE on the same row that does not invalidate the description in PARAM should be added as a new column.

   e.g., CHG, PCHG, R2BASE, R2ULN, ...

2. A transformation of AVAL that necessitates a new description in PARAM should be added as a new parameter, and AVAL should contain the transformed value.

   e.g., log of AVAL

When to Derive Rows vs. Columns

Shire

3. A function of multiple rows within the same parameter for the purpose of creating an analysis timepoint should be added as a new row for the same parameter.

    e.g., baseline is an average

4. A function of multiple rows within a parameter that invalidates the description in PARAM should be added as a new parameter.

    e.g., cumulative sum of AVAL

When to Derive Rows vs. Columns

Shire

5. A function of more than one parameter should be added as a new parameter.

    e.g., a ratio of two parameters

6. When there is more than one definition of baseline, each additional definition of baseline requires the creation of its own set of rows.

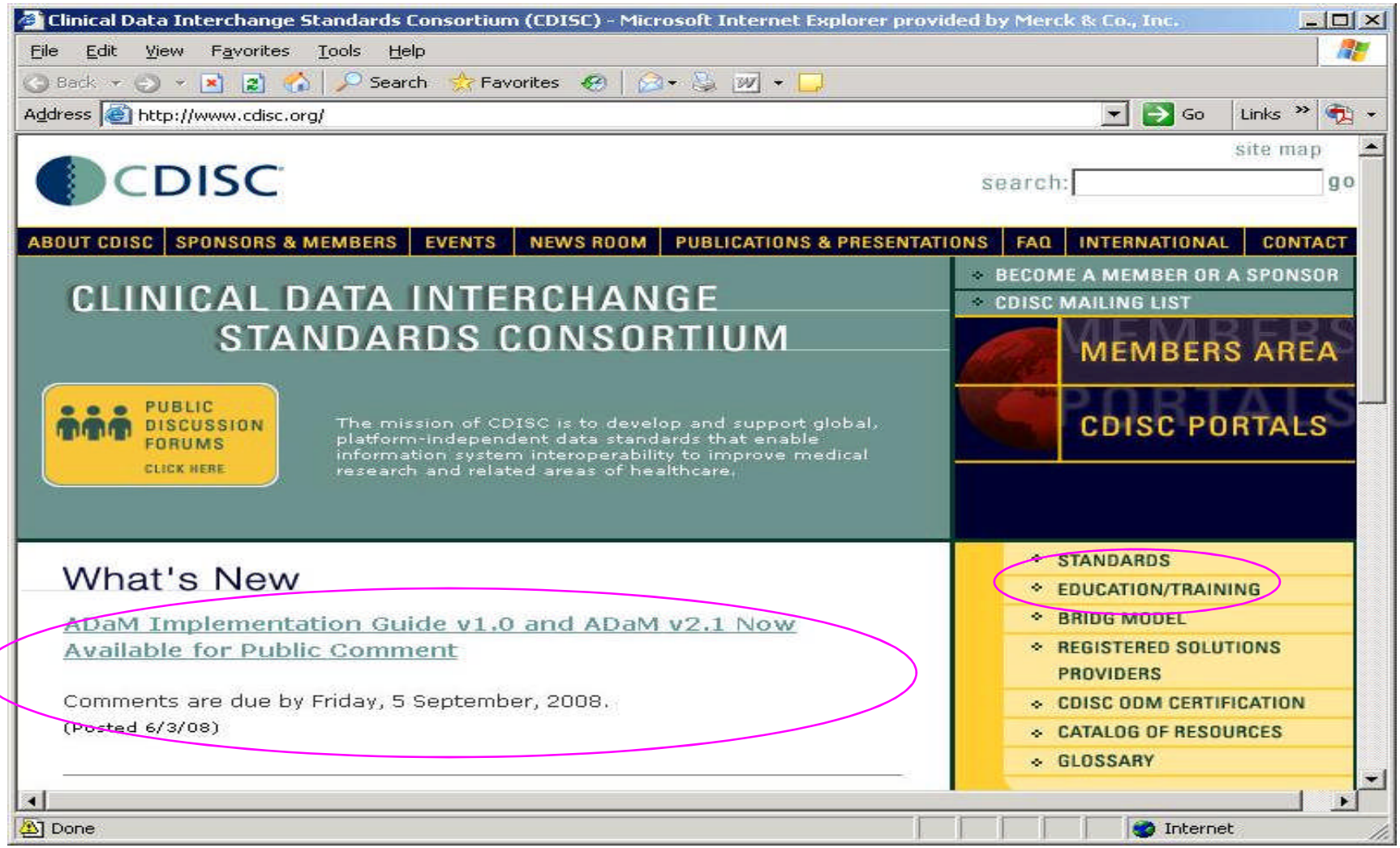
    e.g., base & extension study baselines

7. Analysis of a parameter in different units than the SDTM standardized units requires the creation of a new parameter.

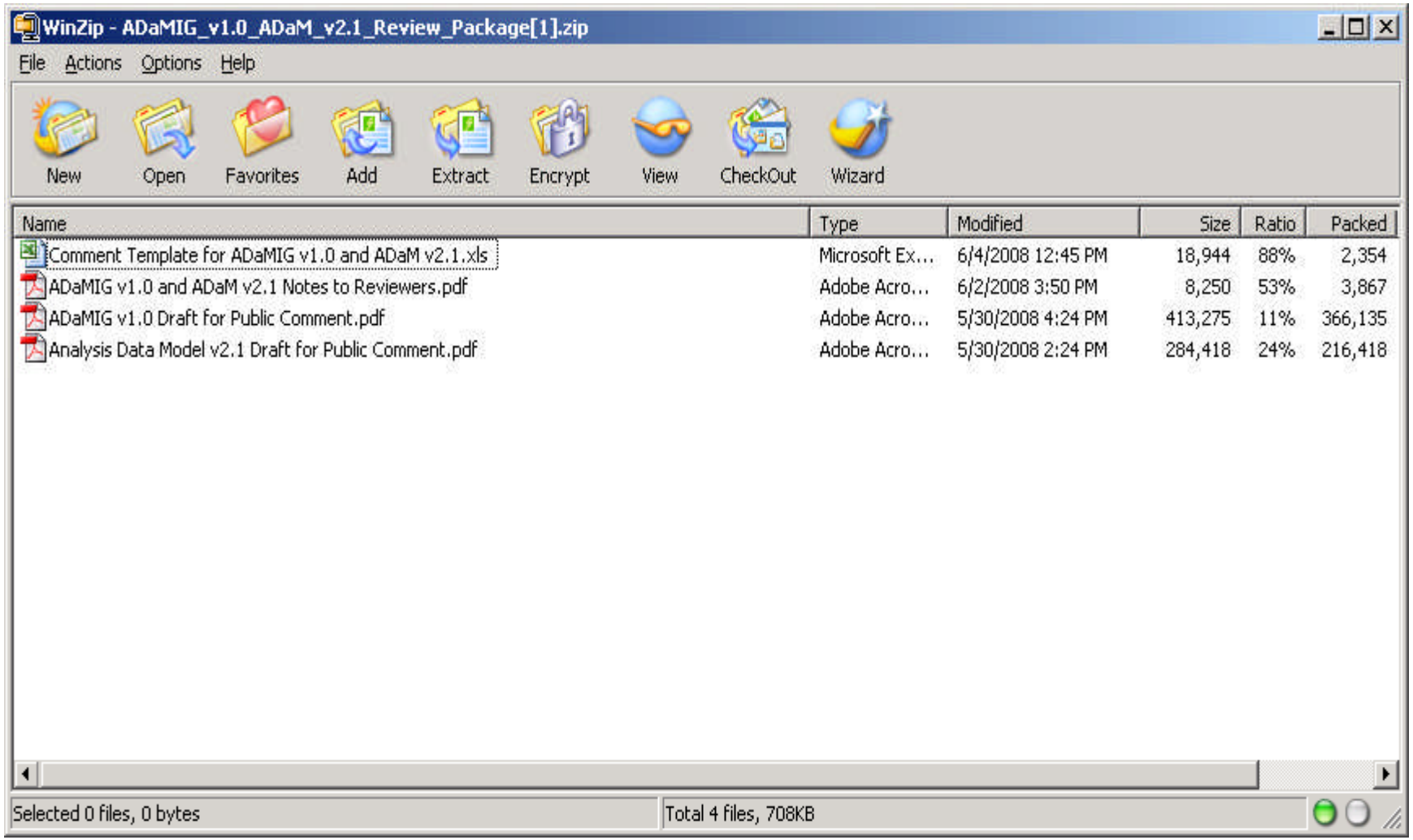
    e.g., SI & US units

View ADaMIG and ADaM announcement at Shire

WWW.CDISC.ORG

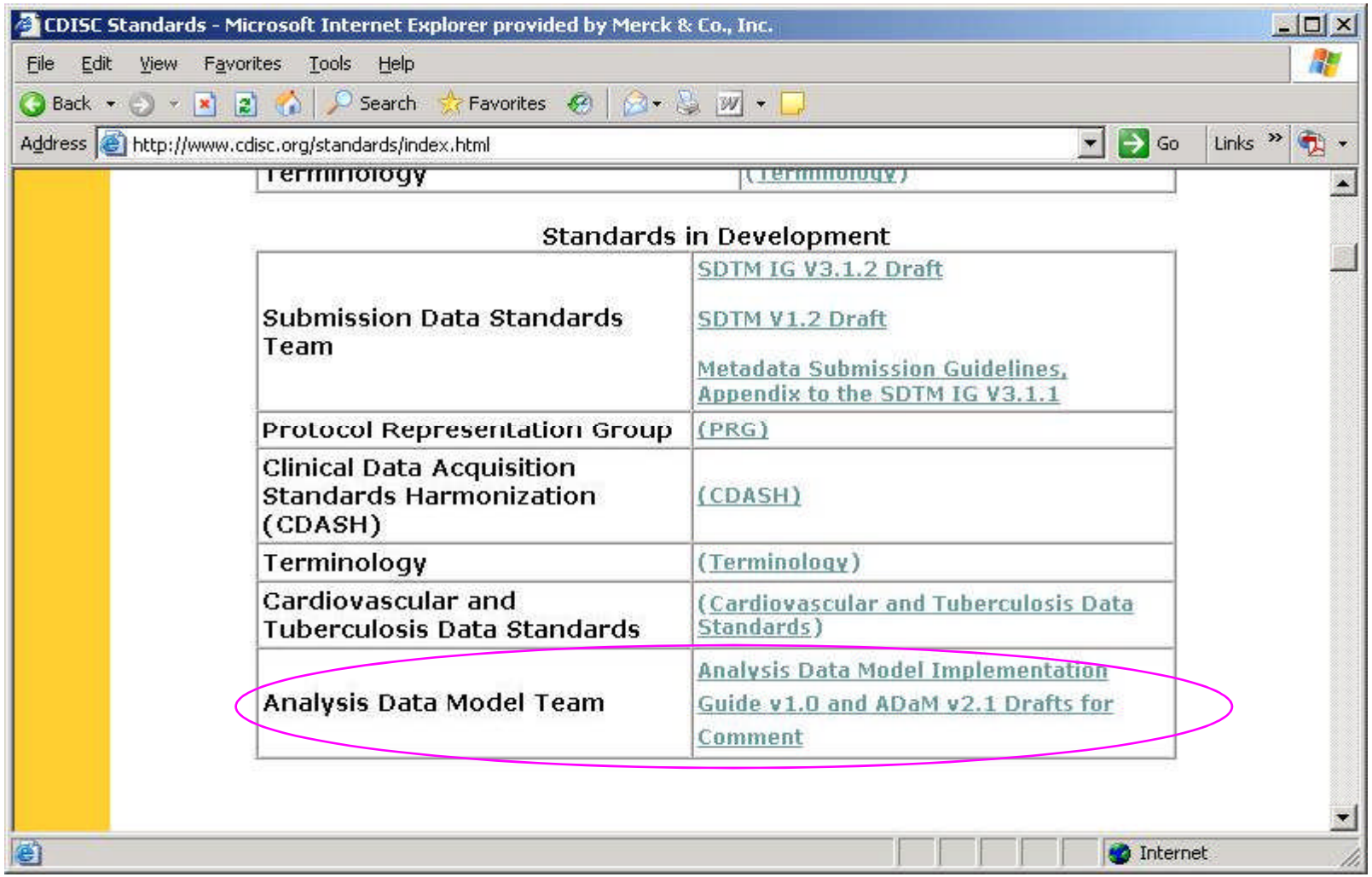


Unzip ADaMIG and ADaM Review Package



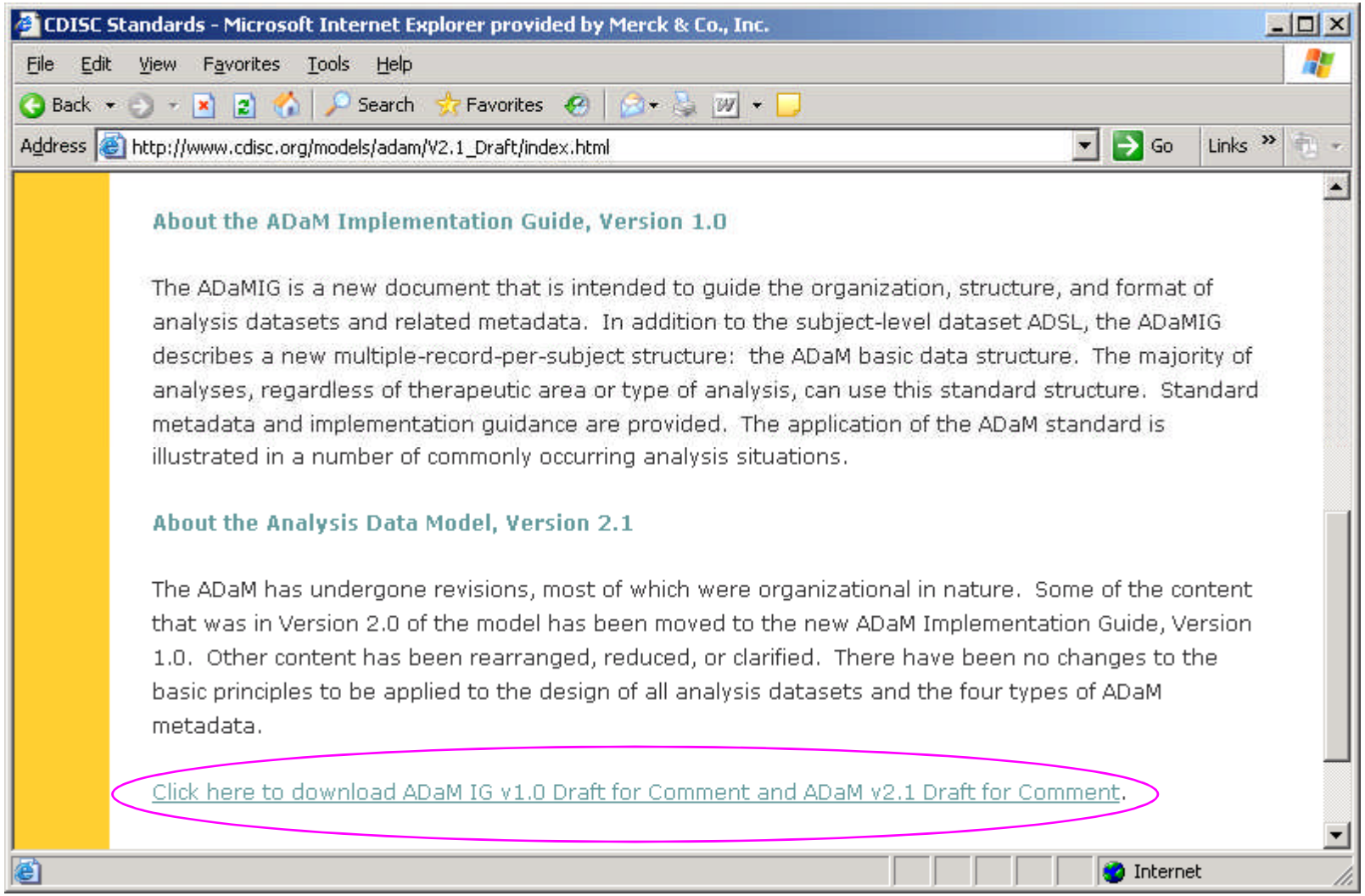
Access ADaMIG and ADaM documents at Shire

http://www.cdisc.org/standards/index.html



Download ADaMIG and ADaM documents at Shire

http://www.cdisc.org/models/adam/V2.1_Draft/index.html



ADaM Present Status

ADaM Implementation Guide 1.0 posted for public comment

- reviewing 858 comments

Analysis Data Model 2.1 posted for public comment

- reviewing 328 comments

ADaM Model and IG Enhancements in Progress

- Expanded coverage of metadata modeling and implementation

ADaM Present Status

Shire

ADaM Model and IG Enhancements in Progress

- Expanded coverage of Time to Event analysis in IG
- Fully worked examples of other common analyses in IG (Linear and Categorical Models)
- Analysis Data Model for Adverse Events

Integrated Safety Pilot

BRIDG DAM / Protocol Representation / SAP

Interaction with SDS team on derived data in SDTM

Training

Questions and Comments?

Thank you for your kind attention.

If you are too shy to speak up now, please forward your questions and comments to:

Michael Nessly

mnessly@shire.com