Data Management: The Clinical Research and Regulatory Perspective

NIH BECON/BISTIC Symposium
21 June 2004

Wayne Kubick

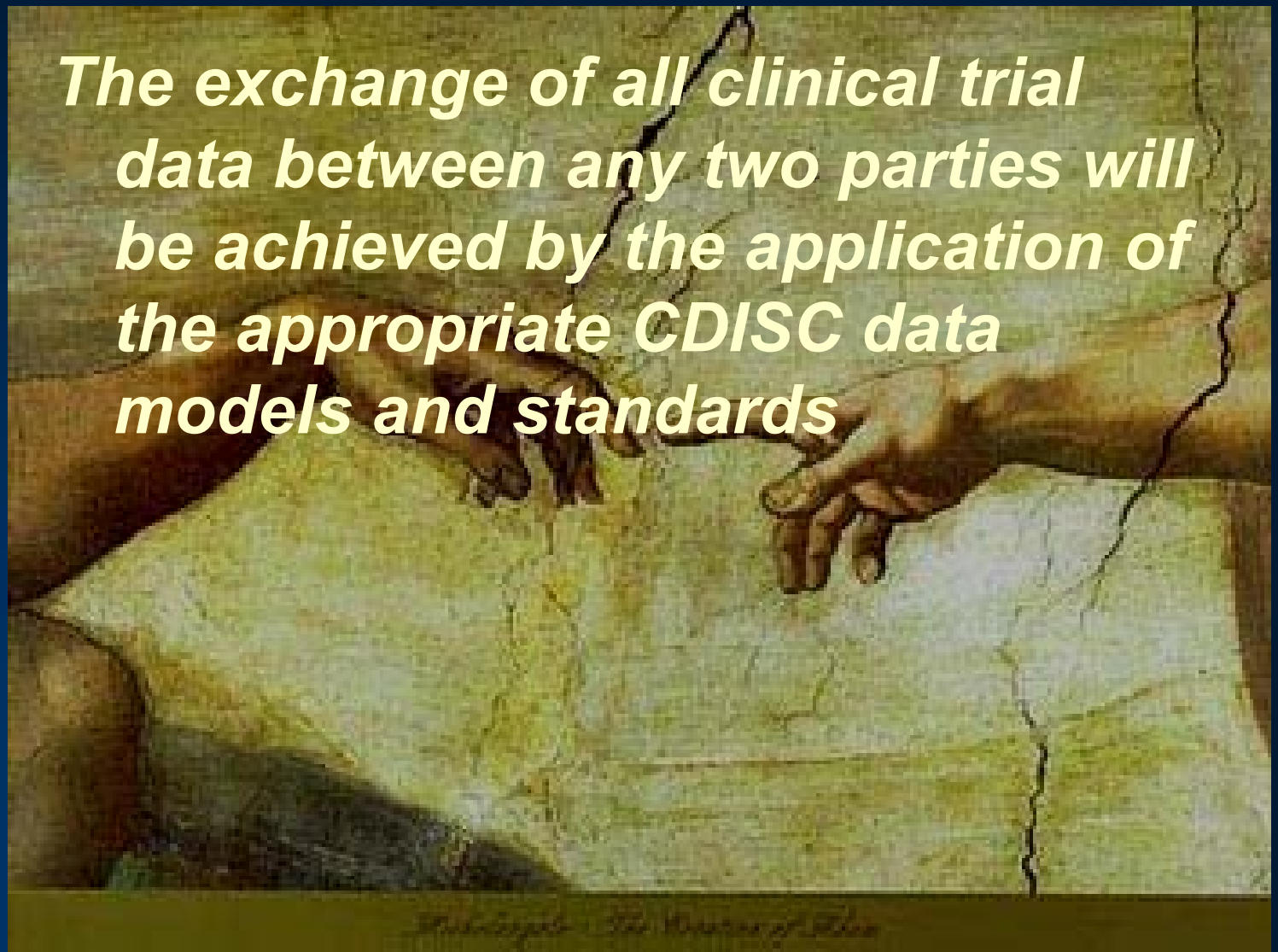
Lincoln TECHNOLOGIES, INC.

CDISC

Topics

- The Legacy of Clinical Trials Data
- CDISC and HL7
- Basic Concepts of the CDISC Study Data Tabulation Model
- The FDA Janus Data Warehouse
- Tools Enabled by Standard Data

CDISC Vision



CDISC and HL7 RCRIM TC Project Topics

CDISC

- **Study Data Tabulation Model** (Submission Data Standards)
- Operational Data Model
- Data Description Specification
- Lab Data Model
- Analysis Data Models
- Standard for Exchange of Nonclinical Data

HL7 RCRIM (with CDISC)

- ECG Waveform Message*
- LAB Data Message*
- Structured Protocol
- Structured Product Label
- Stability Data Model
- Integrated Clinical Safety Report Message

* ANSI - approved Normative V3 Message Standards

Basic Concepts of the SDTM: General Observation Classes

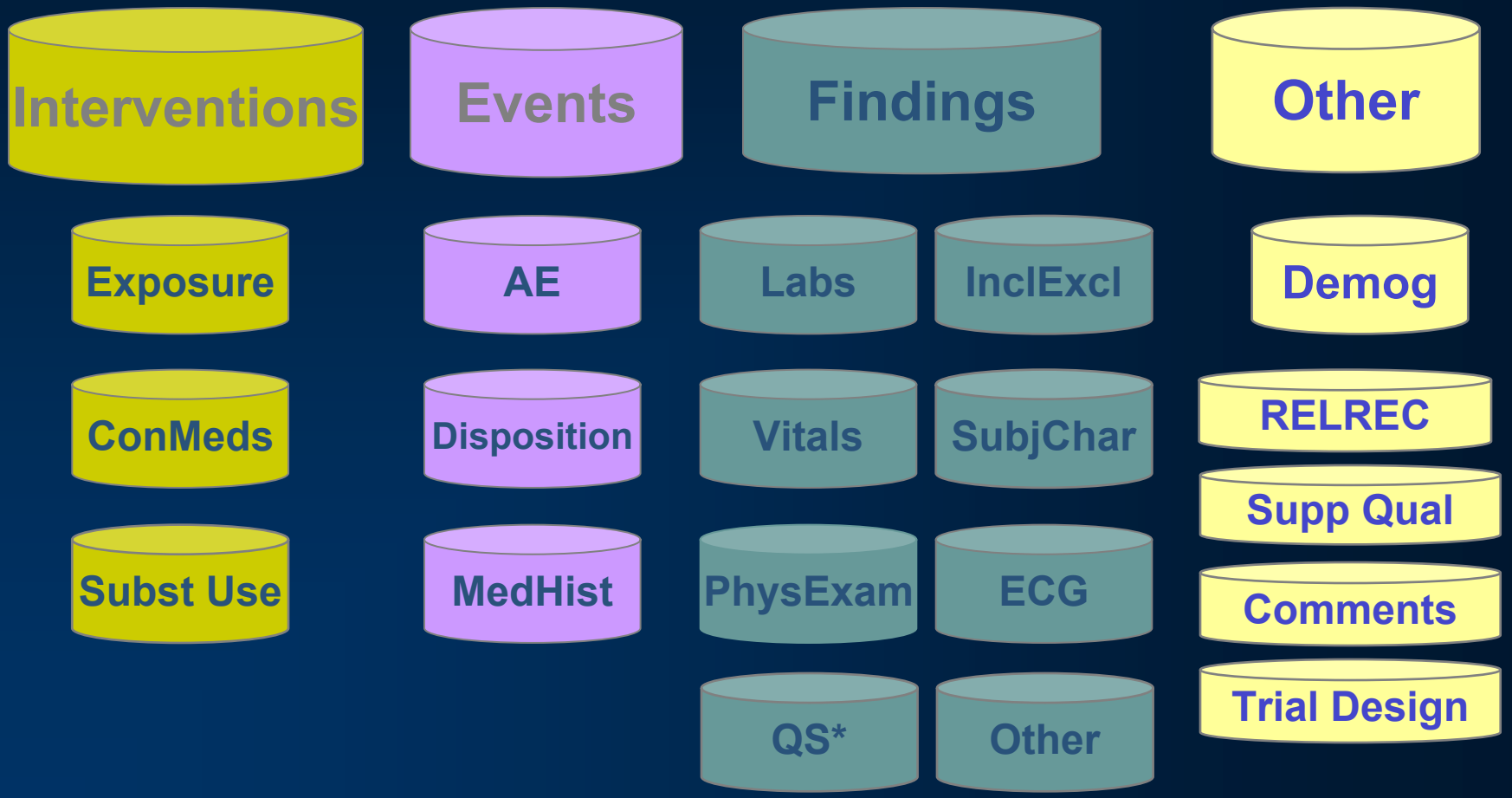
- SDTM: A fundamental model for organizing data collected in clinical trials that will be submitted to regulatory authorities
- All data stored as a series of observations in ***Domains***
- Demographics - Basic subject-level information
  - Used to select, sort and group data in all domains
- Interventions - Treatments or procedures administered
- Events - Things that just happen (AEs, Med History)
- Findings – General subject observations, such as questions and tests
  - *>80% of data will likely be placed in findings*
- *Other special purpose datasets are also defined for special purposes.*

CDISC SDTM Domain Concept

EG.xpt, ECG — Findings, Version 3.1, June 3, 2004. One record per ECG observation per time point per visit per subject , Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**EG	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
EGSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
EGGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to link together a block of related records for a subject in a single domain.	Perm	SDSIG 2.1; SDTM 2.2.4
EGREFID	ECG Reference ID	Char		Sponsor Defined or Derived	Identifier	Internal or external ECG identifier. Example: UUID for external ECG Waveform File.	Perm	
EGSPID	Sponsor ID	Char		Sponsor Defined or Derived	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number from the ECG page.	Perm	SDTM 2.2.4
EGTESTCD	ECG Test or Examination Short Name	Char	**	CRF or Derived	Topic	Short name of the measurement, test, or examination described in EGTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in EGTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g.'1TEST'). EGTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: PR, QT, FIND, INTP.	Req	
EGTEST	ECG Test or Examination Name	Char	**	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in EGTEST cannot be longer than 40 characters. Examples: PR Interval, QT Interval, ECG Finding, etc.	Req	
EGCAT	Category for ECG	Char	*	Sponsor Defined	Grouping Qualifier	Used to categorize ECG observations. Examples: MEASUREMENT or FINDING.	Perm	SDSIG 2.1
EGSCAT	Subcategory for ECG	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the ECG. Example: monitoring (1lead), 12-lead.	Perm	SDSIG 2.1

Domains by Observation Class



SEND Domains for NonClinical Data

A collection of observations with a topic-specific commonality:

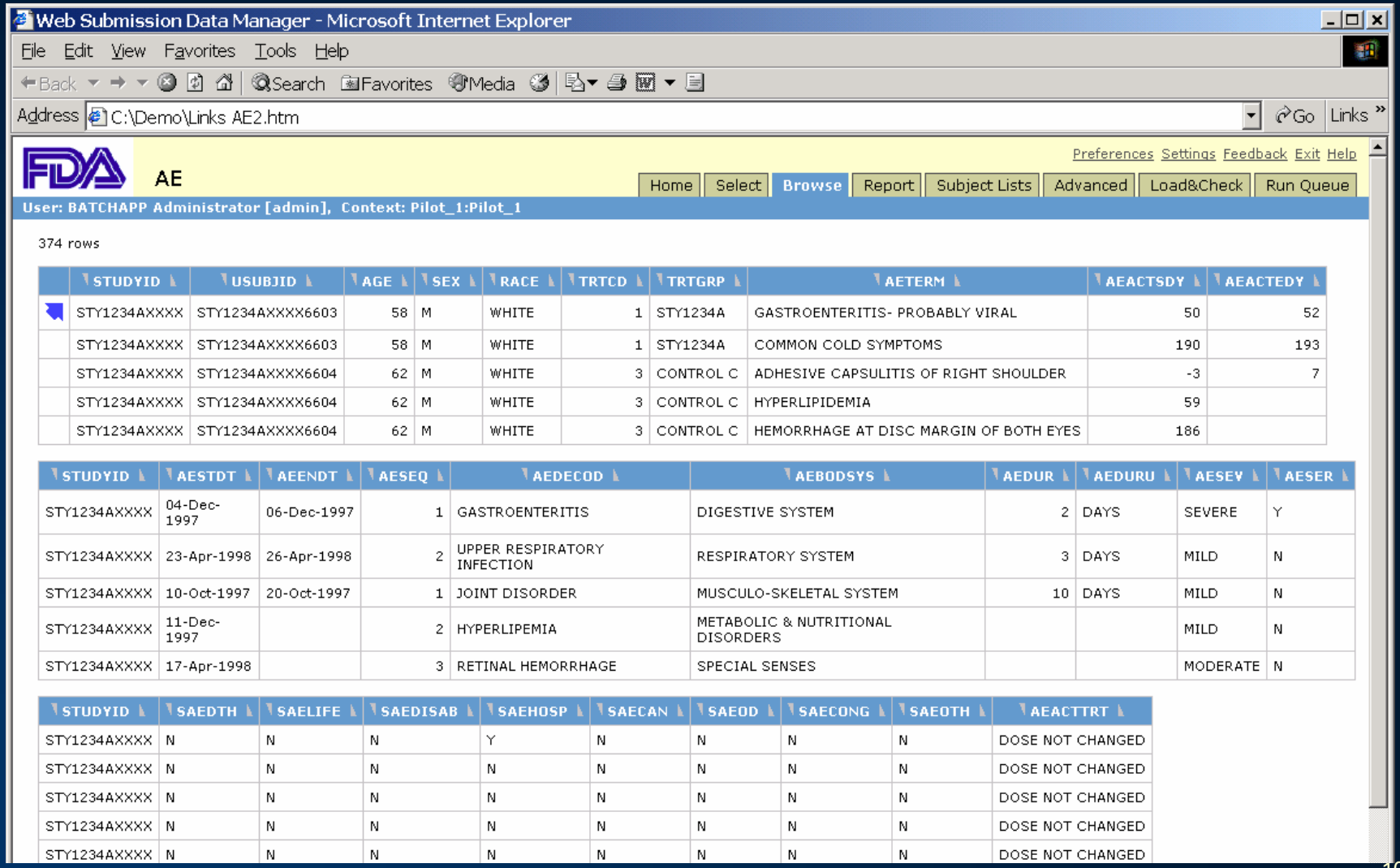
(Interventions, Findings, and Special)

- Animal Characteristics
- Animal Disposition
- Body Weights
- Clinical Pathology
- Clinical Signs
- Drug/Metabolite Levels
- **Exposure**
- Food Consumption
- Fetal Data
- Female Fertility
- Group Characteristics
- Group Observations
- Macroscopic Findings
- Male Fertility
- Microscopic Findings
- Ophthalmoscopic Findings
- Organ Weights
- Rodent Micronucleus
- Study Summary
- **Study Timing**
- Tumor Analysis
- Water Consumption

Variable/Item Roles

- Unique Identifiers - Study, subject, domain, sequence, IDs
- Topic - Focus of observation (Treatment, Term, Test)
- Timing - When the observation occurred
- Qualifiers - Additional attributes/traits of observation
  - Lab results, units, decodes, position, severities, etc.
  - Grouping, Result, Synonym, Record, Variable sub-types
- Each domain must include:
  - Unique identifiers
  - A topic variable
  - A set of timing variables and qualifiers
- *SDS domain models define the expected set of timing variables and qualifiers for each domain.*

Event Structure – 1 record per AE

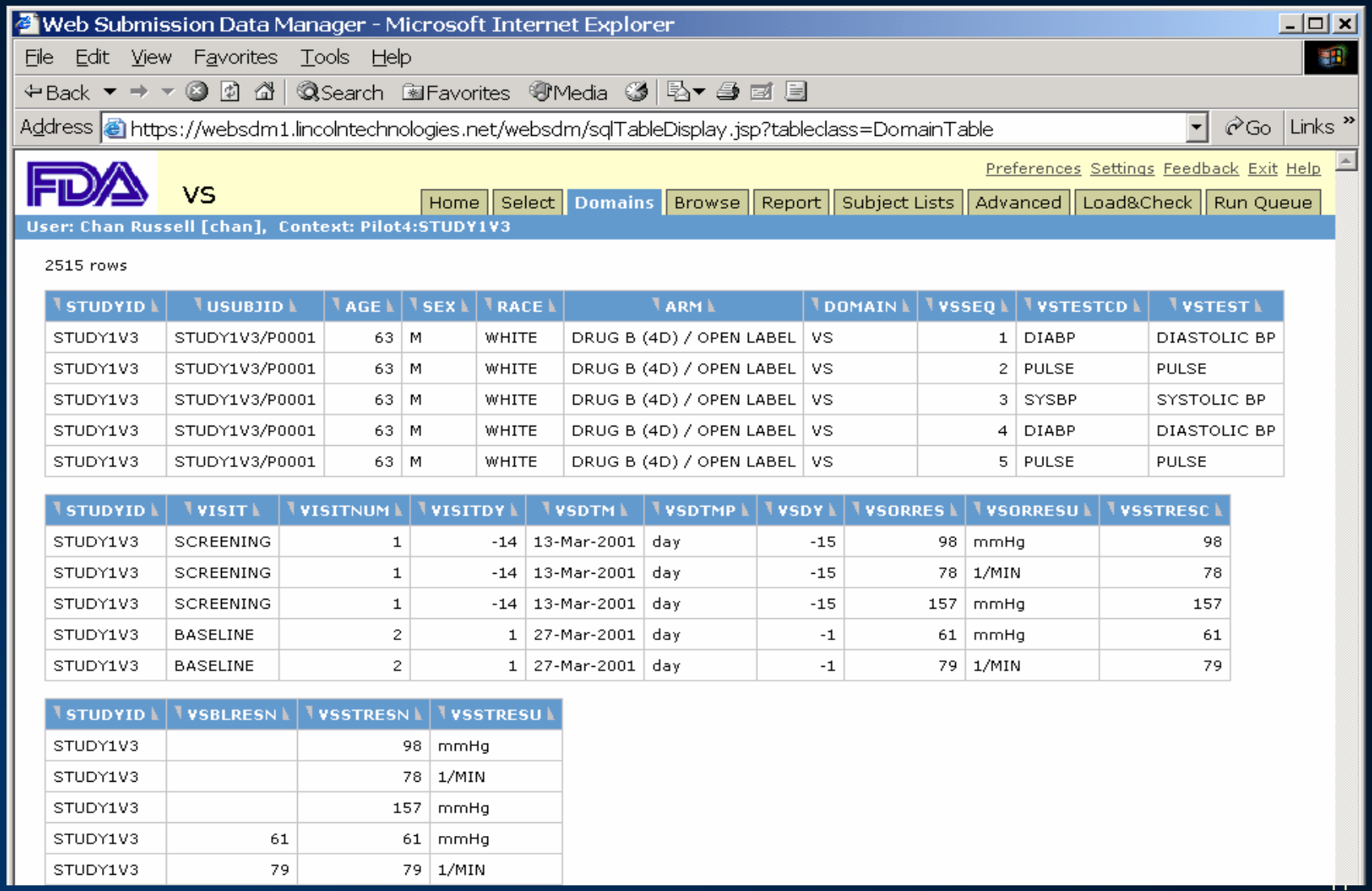


STUDYID	USUBJID	AGE	SEX	RACE	TRTCD	TRTGRP	AETERM	AEACTSDY	AEACTEDY
STY1234AXXXX	STY1234AXXXX6603	58	M	WHITE	1	STY1234A	GASTROENTERITIS- PROBABLY VIRAL	50	52
STY1234AXXXX	STY1234AXXXX6603	58	M	WHITE	1	STY1234A	COMMON COLD SYMPTOMS	190	193
STY1234AXXXX	STY1234AXXXX6604	62	M	WHITE	3	CONTROL C	ADHESIVE CAPSULITIS OF RIGHT SHOULDER	-3	7
STY1234AXXXX	STY1234AXXXX6604	62	M	WHITE	3	CONTROL C	HYPERLIPIDEMIA	59	
STY1234AXXXX	STY1234AXXXX6604	62	M	WHITE	3	CONTROL C	HEMORRHAGE AT DISC MARGIN OF BOTH EYES	186	

STUDYID	AESTDT	AEENDT	AESEQ	AEDECOD	AEBODSYS	AEDUR	AEDURU	AESEV	AESER
STY1234AXXXX	04-Dec-1997	06-Dec-1997	1	GASTROENTERITIS	DIGESTIVE SYSTEM	2	DAYS	SEVERE	Y
STY1234AXXXX	23-Apr-1998	26-Apr-1998	2	UPPER RESPIRATORY INFECTION	RESPIRATORY SYSTEM	3	DAYS	MILD	N
STY1234AXXXX	10-Oct-1997	20-Oct-1997	1	JOINT DISORDER	MUSCULO-SKELETAL SYSTEM	10	DAYS	MILD	N
STY1234AXXXX	11-Dec-1997		2	HYPERLIPEMIA	METABOLIC & NUTRITIONAL DISORDERS			MILD	N
STY1234AXXXX	17-Apr-1998		3	RETINAL HEMORRHAGE	SPECIAL SENSES			MODERATE	N

STUDYID	SAEDTH	SAELIFE	SAEDISAB	SAEHOSP	SAECAN	SAEOD	SAECONG	SAEOTH	AEACTTRT
STY1234AXXXX	N	N	N	Y	N	N	N	N	DOSE NOT CHANGED
STY1234AXXXX	N	N	N	N	N	N	N	N	DOSE NOT CHANGED
STY1234AXXXX	N	N	N	N	N	N	N	N	DOSE NOT CHANGED
STY1234AXXXX	N	N	N	N	N	N	N	N	DOSE NOT CHANGED
STY1234AXXXX	N	N	N	N	N	N	N	N	DOSE NOT CHANGED

Finding Structure: 1 Rec. per Vital Sign



Web Submission Data Manager - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address https://websdm1.lincolntechnologies.net/websdm/sqlTableDisplay.jsp?tableclass=DomainTable Go Links

FDA VS

Preferences Settings Feedback Exit Help

Home | Select | Domains | Browse | Report | Subject Lists | Advanced | Load&Check | Run Queue

User: Chan Russell [chan], Context: Pilot4:STUDY1V3

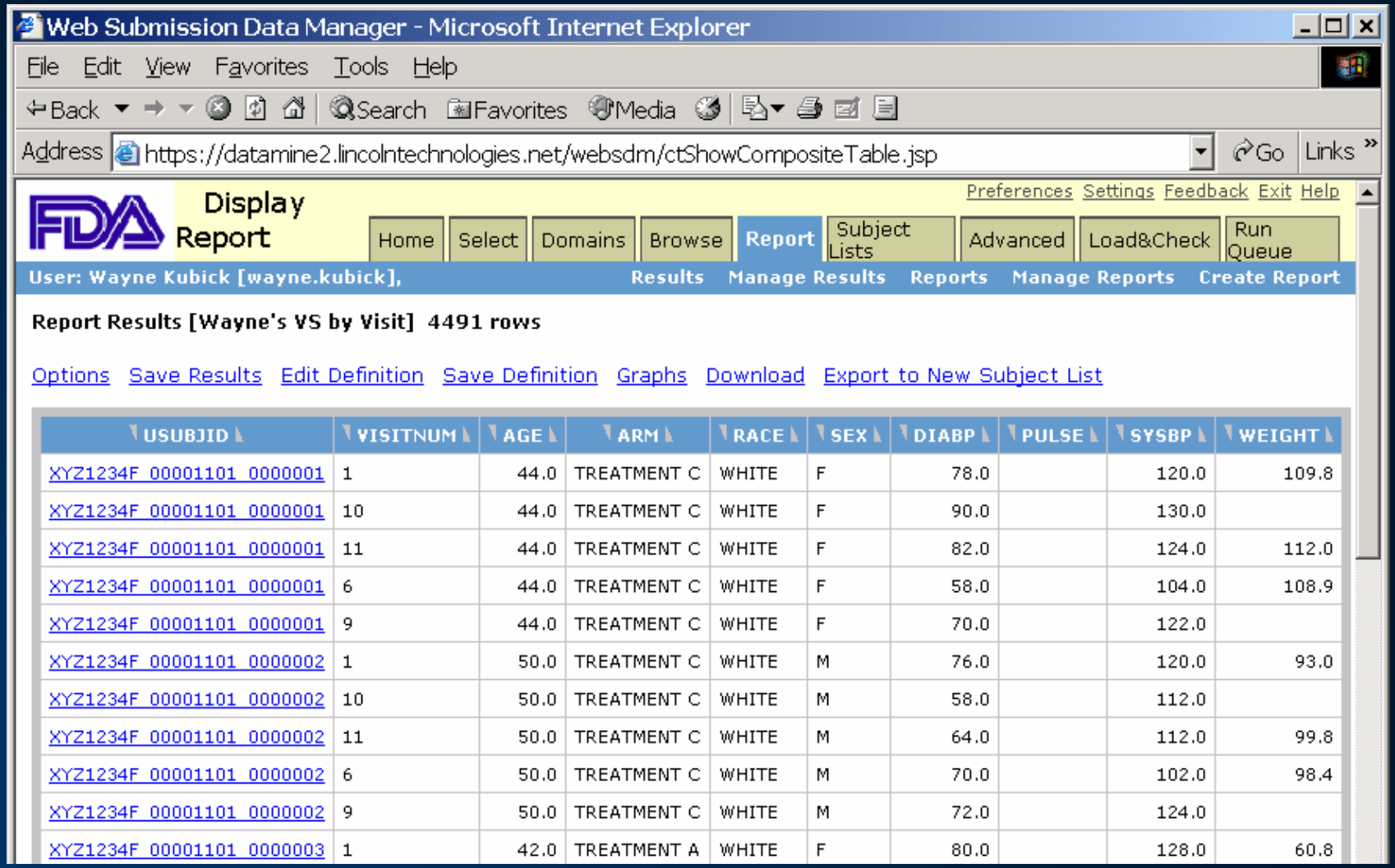
2515 rows

STUDYID	USUBJID	AGE	SEX	RACE	ARM	DOMAIN	VSSEQ	VSTESTCD	VSTEST
STUDY1V3	STUDY1V3/P0001	63	M	WHITE	DRUG B (4D) / OPEN LABEL	VS	1	DIABP	DIASTOLIC BP
STUDY1V3	STUDY1V3/P0001	63	M	WHITE	DRUG B (4D) / OPEN LABEL	VS	2	PULSE	PULSE
STUDY1V3	STUDY1V3/P0001	63	M	WHITE	DRUG B (4D) / OPEN LABEL	VS	3	SYSBP	SYSTOLIC BP
STUDY1V3	STUDY1V3/P0001	63	M	WHITE	DRUG B (4D) / OPEN LABEL	VS	4	DIABP	DIASTOLIC BP
STUDY1V3	STUDY1V3/P0001	63	M	WHITE	DRUG B (4D) / OPEN LABEL	VS	5	PULSE	PULSE

STUDYID	VISIT	VISITNUM	VISITDY	VSDTM	VSDTMP	VSDY	VSORRES	VSORRESU	VSSTRESC
STUDY1V3	SCREENING	1	-14	13-Mar-2001	day	-15	98	mmHg	98
STUDY1V3	SCREENING	1	-14	13-Mar-2001	day	-15	78	1/MIN	78
STUDY1V3	SCREENING	1	-14	13-Mar-2001	day	-15	157	mmHg	157
STUDY1V3	BASELINE	2	1	27-Mar-2001	day	-1	61	mmHg	61
STUDY1V3	BASELINE	2	1	27-Mar-2001	day	-1	79	1/MIN	79

STUDYID	VSBLRESN	VSSTRESN	VSSTRESU
STUDY1V3		98	mmHg
STUDY1V3		78	1/MIN
STUDY1V3		157	mmHg
STUDY1V3	61	61	mmHg
STUDY1V3	79	79	1/MIN

Transformed VS: 1 Record per Visit

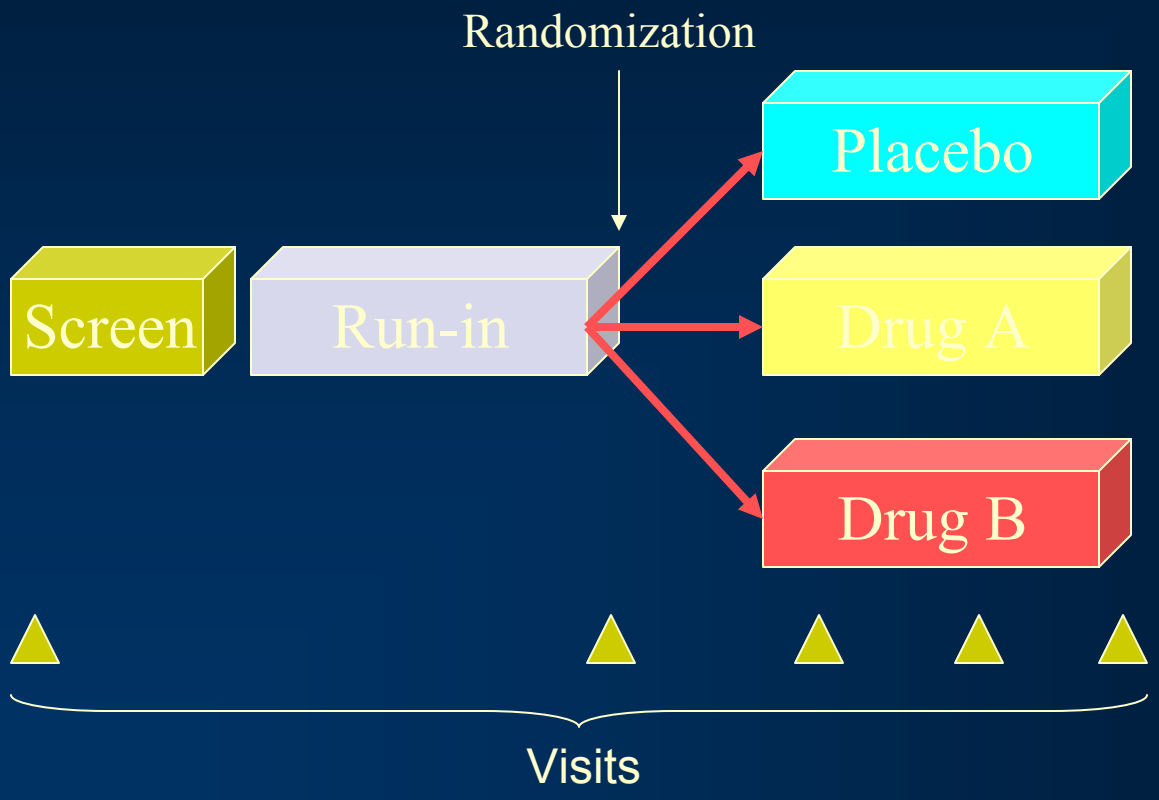


USUBJID	VISITNUM	AGE	ARM	RACE	SEX	DIABP	PULSE	SYSBP	WEIGHT
XYZ1234F 00001101 0000001	1	44.0	TREATMENT C	WHITE	F	78.0		120.0	109.8
XYZ1234F 00001101 0000001	10	44.0	TREATMENT C	WHITE	F	90.0		130.0	
XYZ1234F 00001101 0000001	11	44.0	TREATMENT C	WHITE	F	82.0		124.0	112.0
XYZ1234F 00001101 0000001	6	44.0	TREATMENT C	WHITE	F	58.0		104.0	108.9
XYZ1234F 00001101 0000001	9	44.0	TREATMENT C	WHITE	F	70.0		122.0	
XYZ1234F 00001101 0000002	1	50.0	TREATMENT C	WHITE	M	76.0		120.0	93.0
XYZ1234F 00001101 0000002	10	50.0	TREATMENT C	WHITE	M	58.0		112.0	
XYZ1234F 00001101 0000002	11	50.0	TREATMENT C	WHITE	M	64.0		112.0	99.8
XYZ1234F 00001101 0000002	6	50.0	TREATMENT C	WHITE	M	70.0		102.0	98.4
XYZ1234F 00001101 0000002	9	50.0	TREATMENT C	WHITE	M	72.0		124.0	
XYZ1234F 00001101 0000003	1	42.0	TREATMENT A	WHITE	F	80.0		128.0	60.8

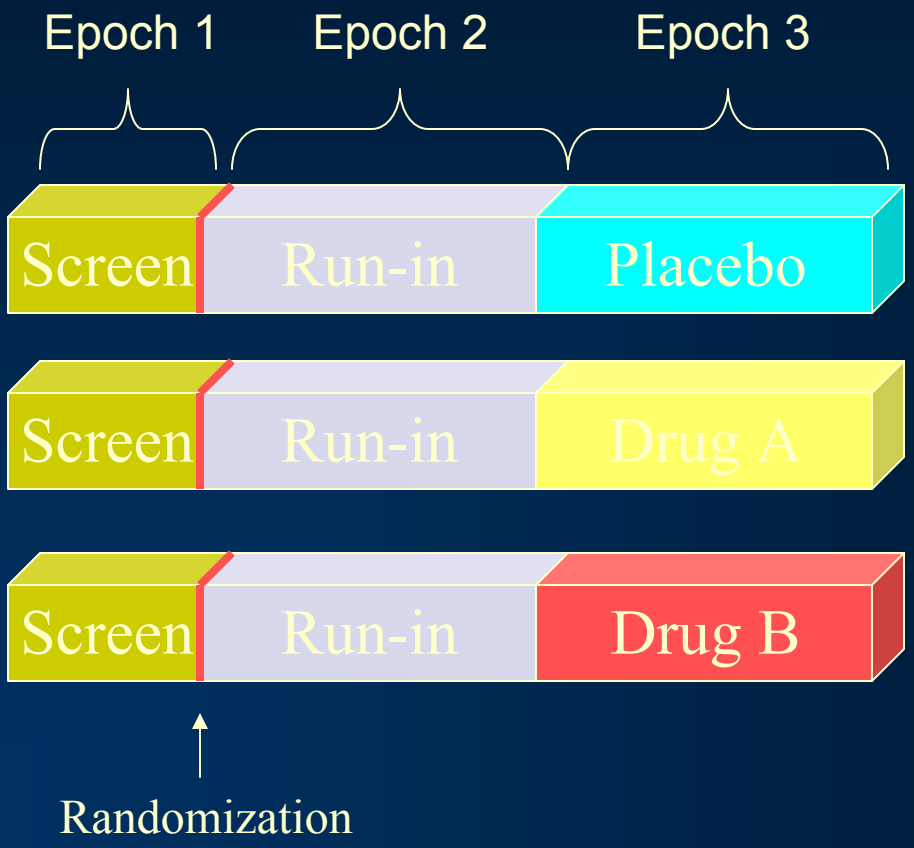
SDTM Concepts: Trial Design

- Model for defining aspects of a trial design
  - Epoch: a time interval in the planned conduct of a study
  - Element: a particular time interval within an Arm
  - Arm: an ordered sequence of Elements
    - Equivalent to a planned treatment group
  - Visit: a clinical encounter
- Domains for:
  - Planned Elements, Arms, Visits, IE Criteria, Interventions and Assessments
  - Actual Subject Elements and Visits

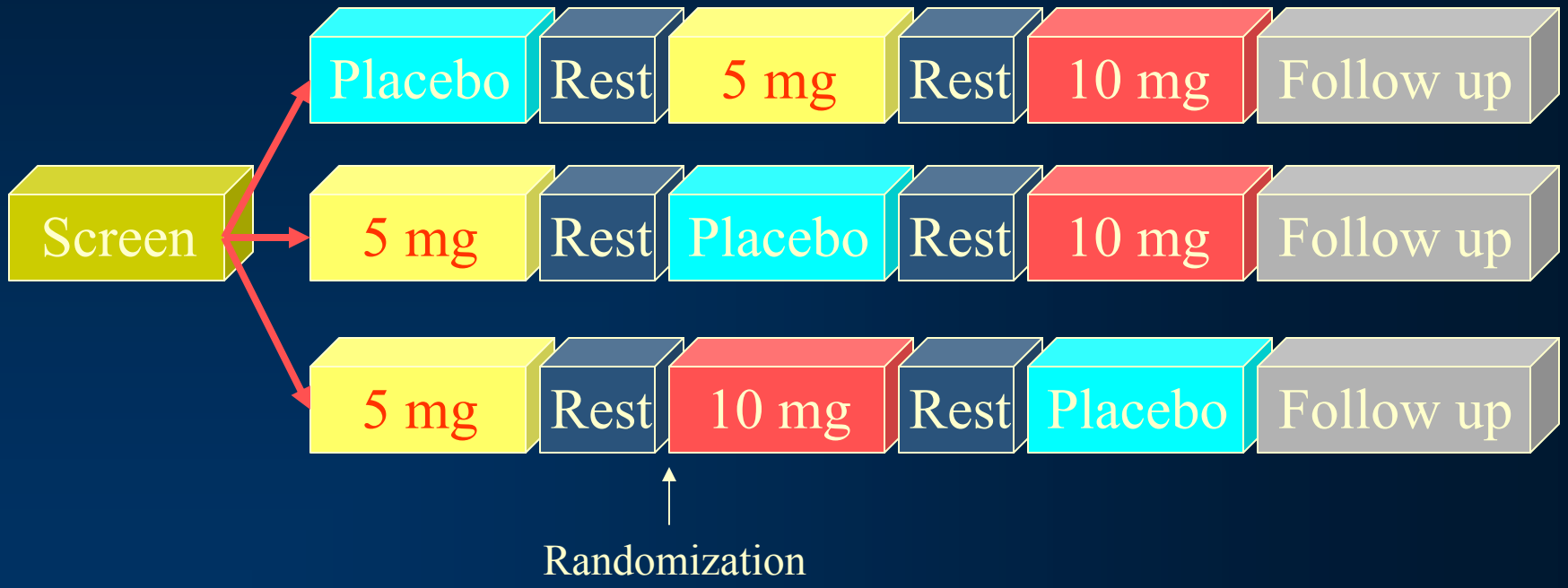
Parallel Trial Design Flowchart



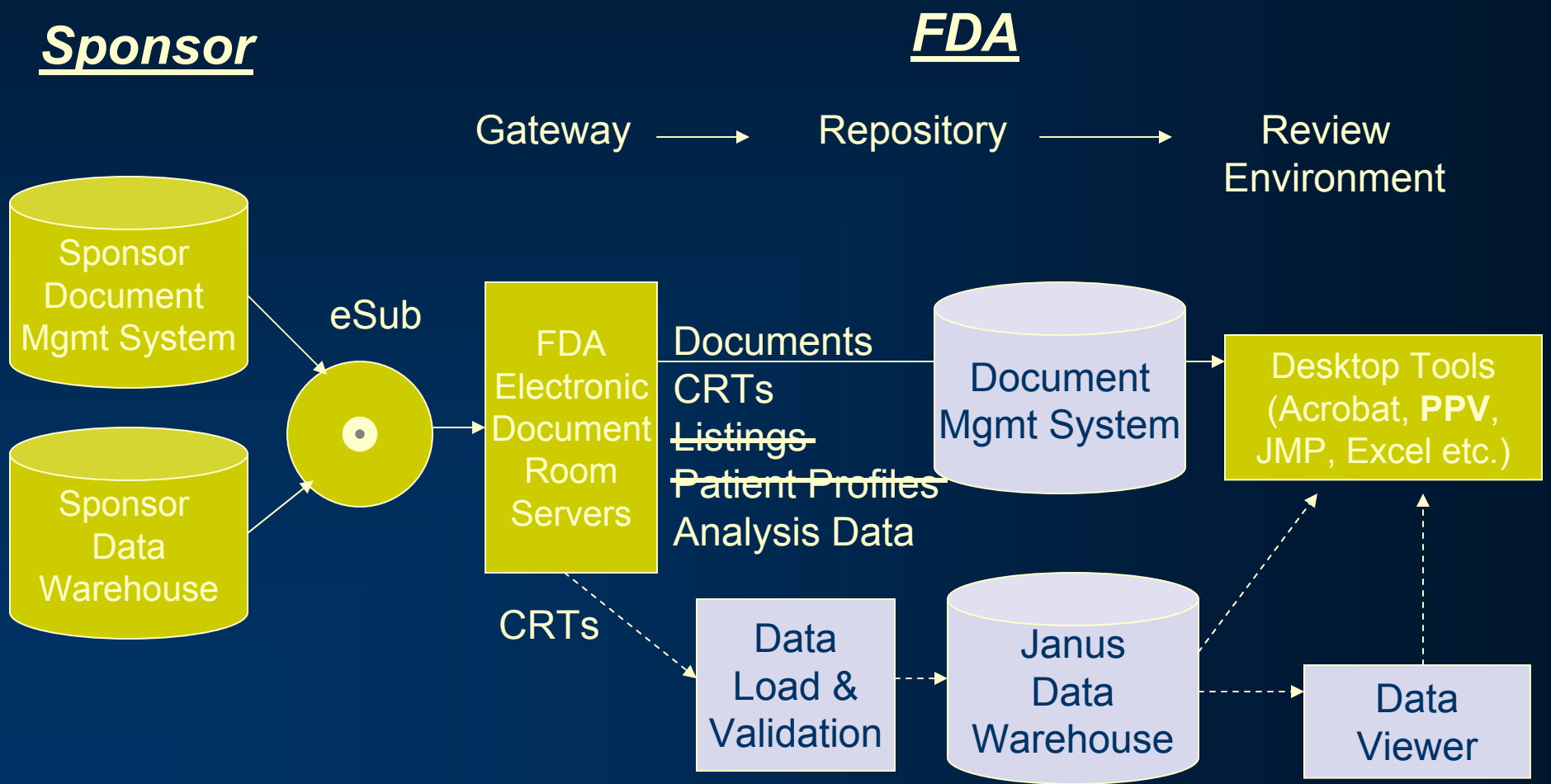
Parallel Trial Arms



Flowchart: Crossover Trial



Facilitating the FDA Review Process

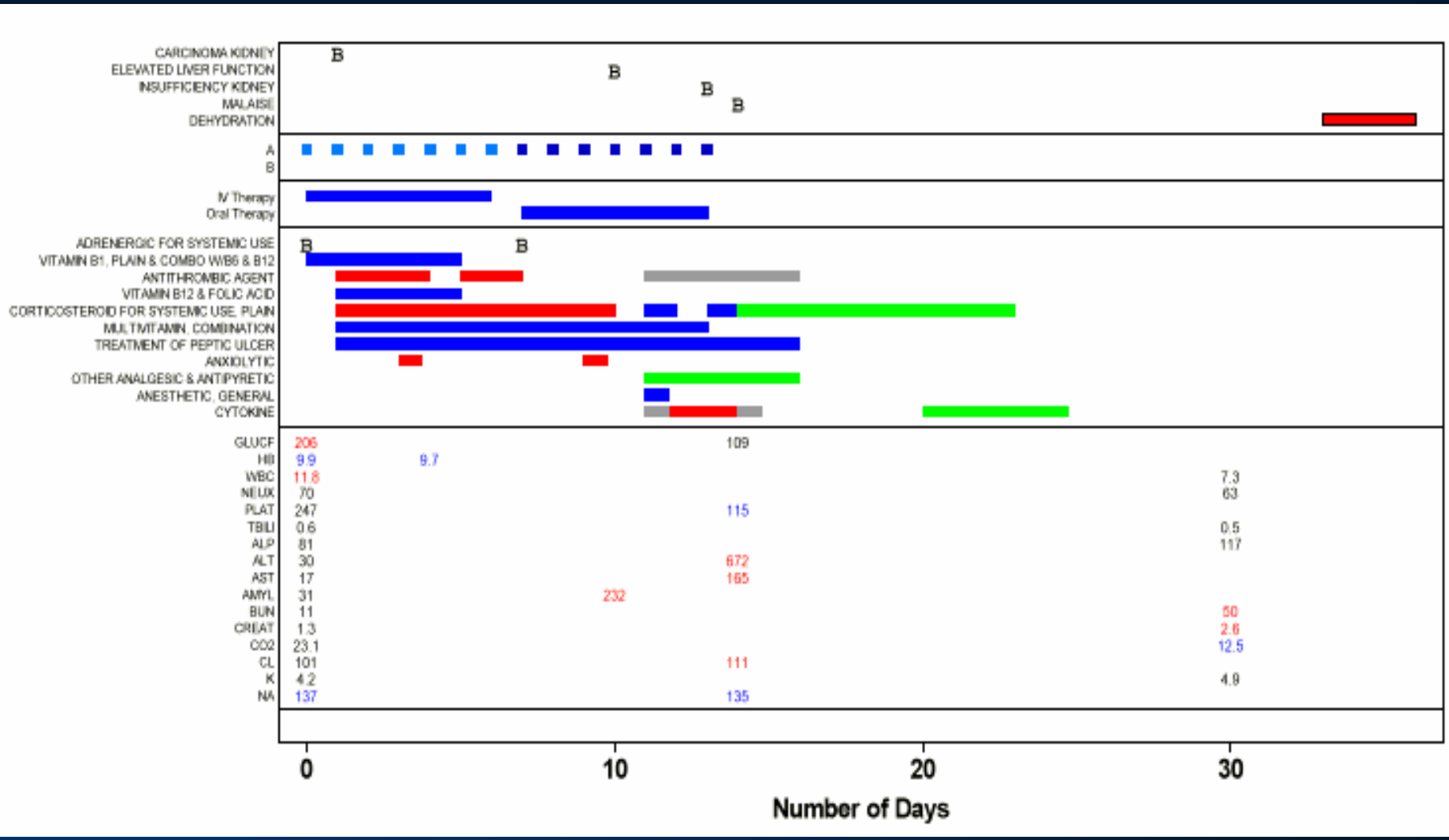


CRADAs Related to CDISC SDTM

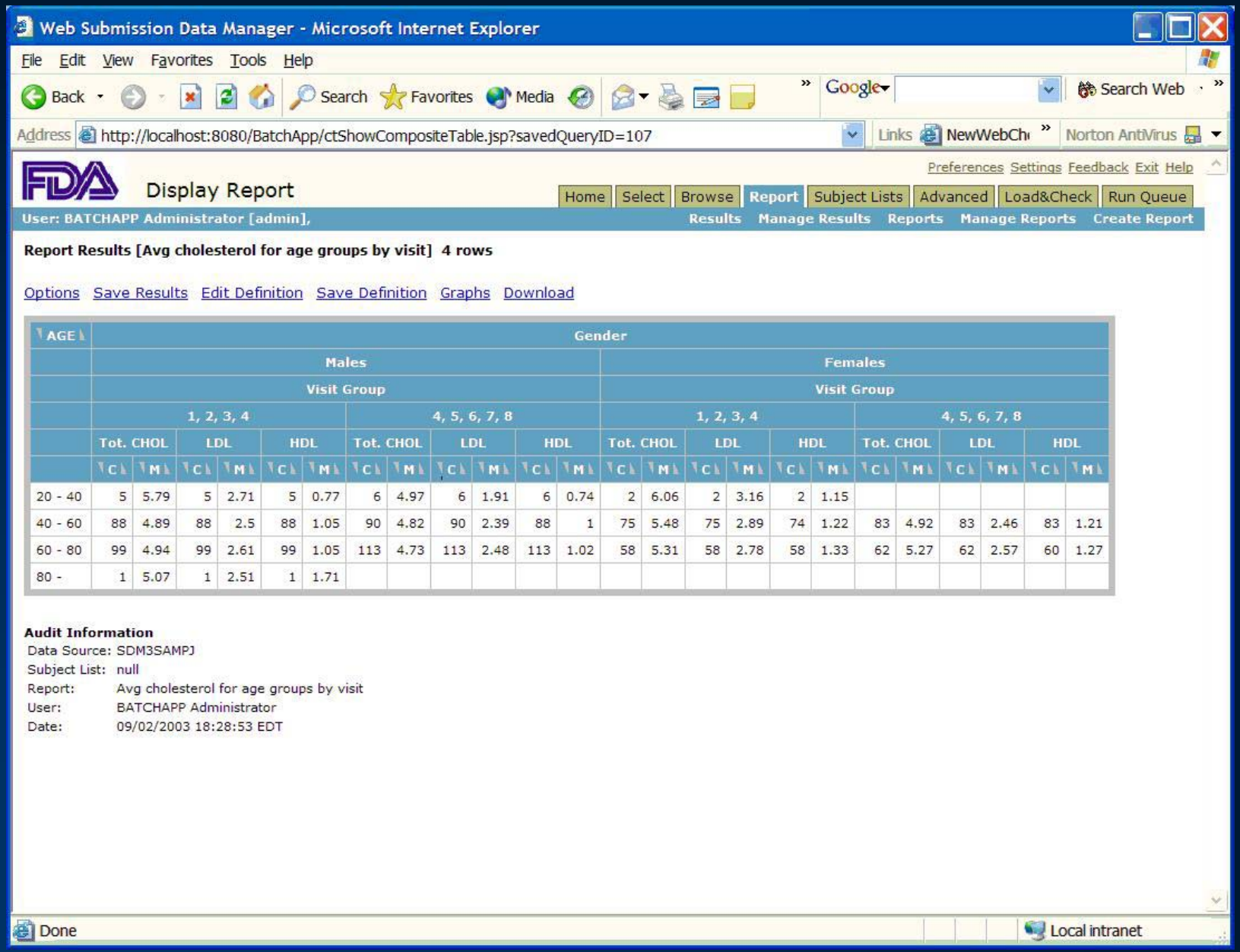
- Lincoln Technologies Web Submission Data Manager
- PPD Patient Profile Viewer
- IBM and Janus Data Warehouse
- Mortara ECG Data Warehouse/Viewer
- PharmQuest's ToxVision for Animal Data

Tools will continue to evolve as standard data becomes available.

Invoking Patient Profiles



Example Enhanced Report



AGE	Gender																							
	Males												Females											
	Visit Group												Visit Group											
	1, 2, 3, 4						4, 5, 6, 7, 8						1, 2, 3, 4						4, 5, 6, 7, 8					
	Tot. CHOL		LDL		HDL		Tot. CHOL		LDL		HDL		Tot. CHOL		LDL		HDL		Tot. CHOL		LDL		HDL	
	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M
20 - 40	5	5.79	5	2.71	5	0.77	6	4.97	6	1.91	6	0.74	2	6.06	2	3.16	2	1.15						
40 - 60	88	4.89	88	2.5	88	1.05	90	4.82	90	2.39	88	1	75	5.48	75	2.89	74	1.22	83	4.92	83	2.46	83	1.21
60 - 80	99	4.94	99	2.61	99	1.05	113	4.73	113	2.48	113	1.02	58	5.31	58	2.78	58	1.33	62	5.27	62	2.57	60	1.27
80 -	1	5.07	1	2.51	1	1.71																		

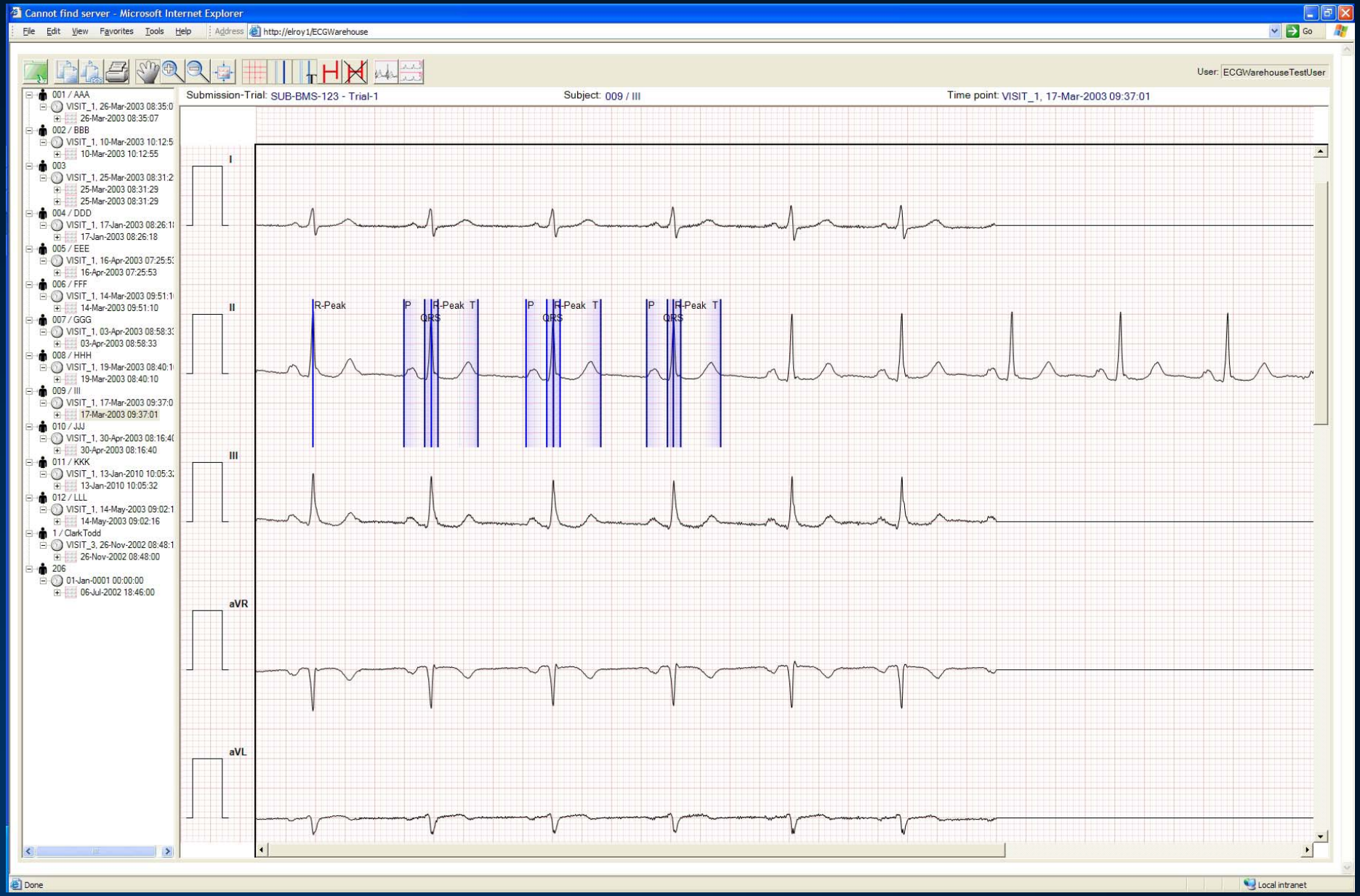
Audit Information

Data Source: SDM3SAMPJ
Subject List: null
Report: Avg cholesterol for age groups by visit
User: BATCHAPP Administrator
Date: 09/02/2003 18:28:53 EDT

Image Integration with SDTM Findings

- Findings data structure includes a variable to store UID for external reference files
  - ECG XML Waveforms in HL7 V3 messaging format
  - Medical Imaging files
- FDA standard review tools can invoke image viewing applications from FDA Janus Data Warehouse
- ECG and Image files will be stored in other repositories
  - Imaging formats and viewers are not yet defined.

ECG Viewer


Cannot find server - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://elroy1/ECGWarehouse
Go
User: ECGWarehouseTestUser
Submission-Trial: SUB-BMS-123 - Trial-1
Subject: 009 / III
Time point: VISIT_1, 17-Mar-2003 09:37:01
001 / AAA
VISIT_1, 26-Mar-2003 08:35:0
26-Mar-2003 08:35:07
002 / BBB
VISIT_1, 10-Mar-2003 10:12:5
10-Mar-2003 10:12:55
003
VISIT_1, 25-Mar-2003 08:31:2
25-Mar-2003 08:31:29
25-Mar-2003 08:31:29
004 / DDD
VISIT_1, 17-Jan-2003 08:26:1
17-Jan-2003 08:26:18
005 / EEE
VISIT_1, 16-Apr-2003 07:25:5
16-Apr-2003 07:25:53
006 / FFF
VISIT_1, 14-Mar-2003 09:51:1
14-Mar-2003 09:51:10
007 / GGG
VISIT_1, 03-Apr-2003 08:58:3
03-Apr-2003 08:58:33
008 / HHH
VISIT_1, 19-Mar-2003 08:40:1
19-Mar-2003 08:40:10
009 / III
VISIT_1, 17-Mar-2003 09:37:0
17-Mar-2003 09:37:01
010 / JJJ
VISIT_1, 30-Apr-2003 08:16:4
30-Apr-2003 08:16:40
011 / KKK
VISIT_1, 13-Jan-2010 10:05:3
13-Jan-2010 10:05:32
012 / LLL
VISIT_1, 14-May-2003 09:02:1
14-May-2003 09:02:16
1 / ClarkTodd
VISIT_3, 26-Nov-2002 08:48:1
26-Nov-2002 08:48:00
206
01-Jan-0001 00:00:00
06-Jul-2002 18:46:00
I
II
III
aVR
aVL
R-Peak
P
R-Peak
T
QRS
Done
Local intranet

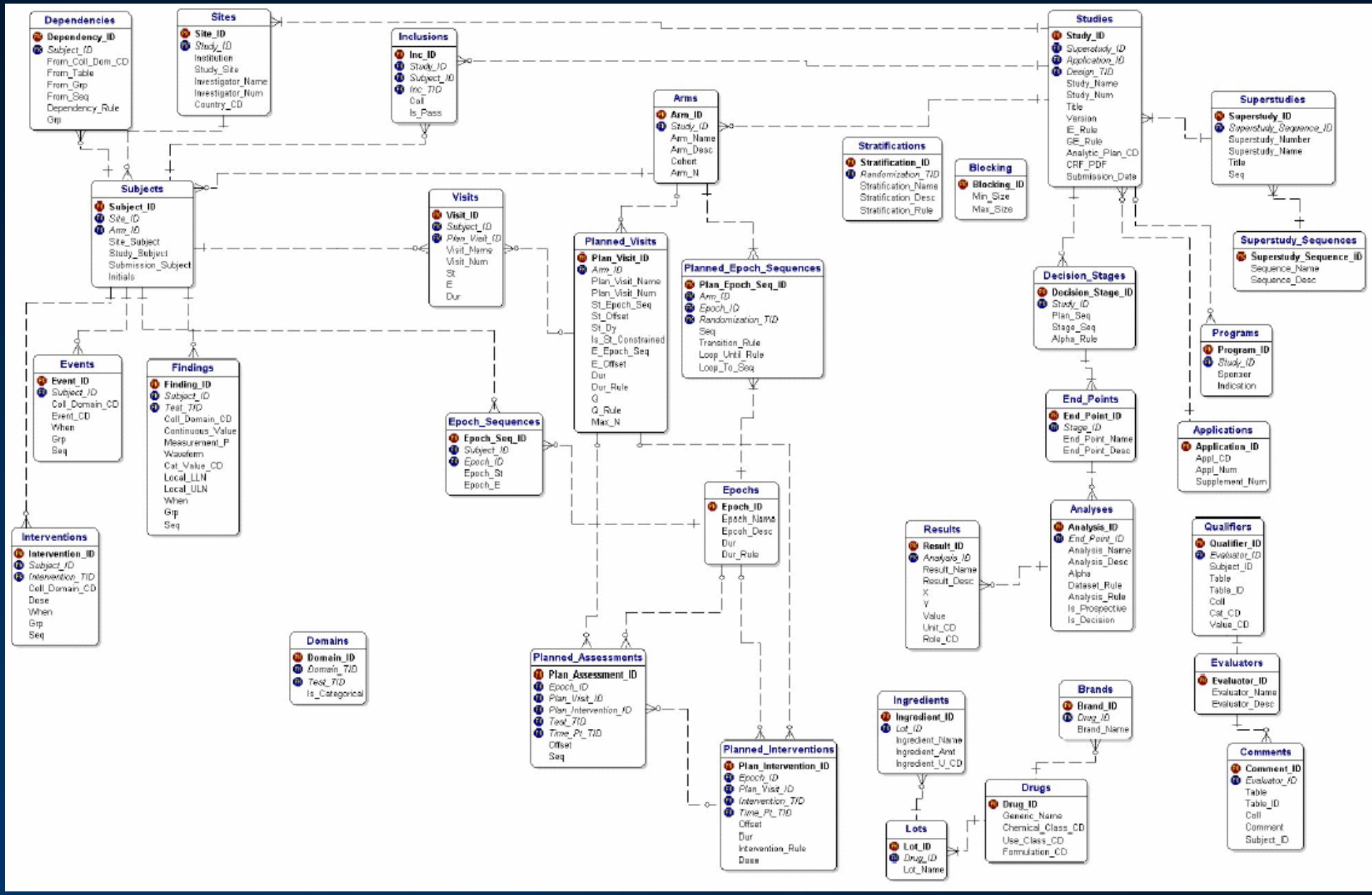
The FDA Janus Warehouse

Conceived and developed by Norman Stockbridge, MD (FDA) to improve reviews and conduct cross-study analysis

- Includes:
  - All clinical trial data
  - Protocol
  - Pharm/tox
  - *Analysis plan*
- Does not include:
  - Spontaneous reports
  - Study reports & Reviews
- Actual Data (submitted in standard SDS format):
  - Findings (LB, EG, VS, QS, PE, etc.)
  - Events (AE, MH, DS)
  - Interventions (EX, CM)
- Planned Protocol Definition (to support cross-study analysis):
  - Trial structure
  - Planned assessments
  - Planned interventions
  - Analytic plan

} Real vs. planned
Support cross-study analysis

Janus Basic Entity-Relationships



Janus Pre-Requisites

- Widespread adoption of the CDISC SDTM standard by industry
- Submission of trial design data for all studies
  - Still undergoing pilot testing and expansion
- Standardized vocabularies
  - Method for submitting sponsor-controlled vocabularies
- Method and syntax for defining analysis rules
- Specific methods for other FDA-requested features like data confidence flags, etc.

Tool Optimization and Validation

- Validation depends on ability of tools to reproduce same results as independent analyses by traditional methods
  - Currently only outputs are validated, not tools
- Tools optimization requires benchmark data sources
  - Pharmaceutical research data is highly confidential and proprietary
  - NCI or other NIH organizations could be a provider of public research data
- By comparison, public availability of AERS data has stimulated development of Bayesian data mining tools, repositories and applications
  - Mining of clinical research data will require different techniques.

The Clinical Research and Regulatory Perspective

For further information:

wayne.kubick@lincolntechnologies.com

www.cdisc.org

www.hl7.org

There are just two rules for success:

1. Never tell all you know.

   -- Roger H. Lincoln