UC DAVIS HEALTH

Data Management Considerations for Clinical Trials

CLINICAL AND TRANSLATIONAL SCIENCE CENTER

Brad Pollock, M.P.H., Ph.D.

Department of Public Health Sciences

The UC Davis CTSC receives support from the NIH National Center for Advancing Translational Sciences (award TR001860).

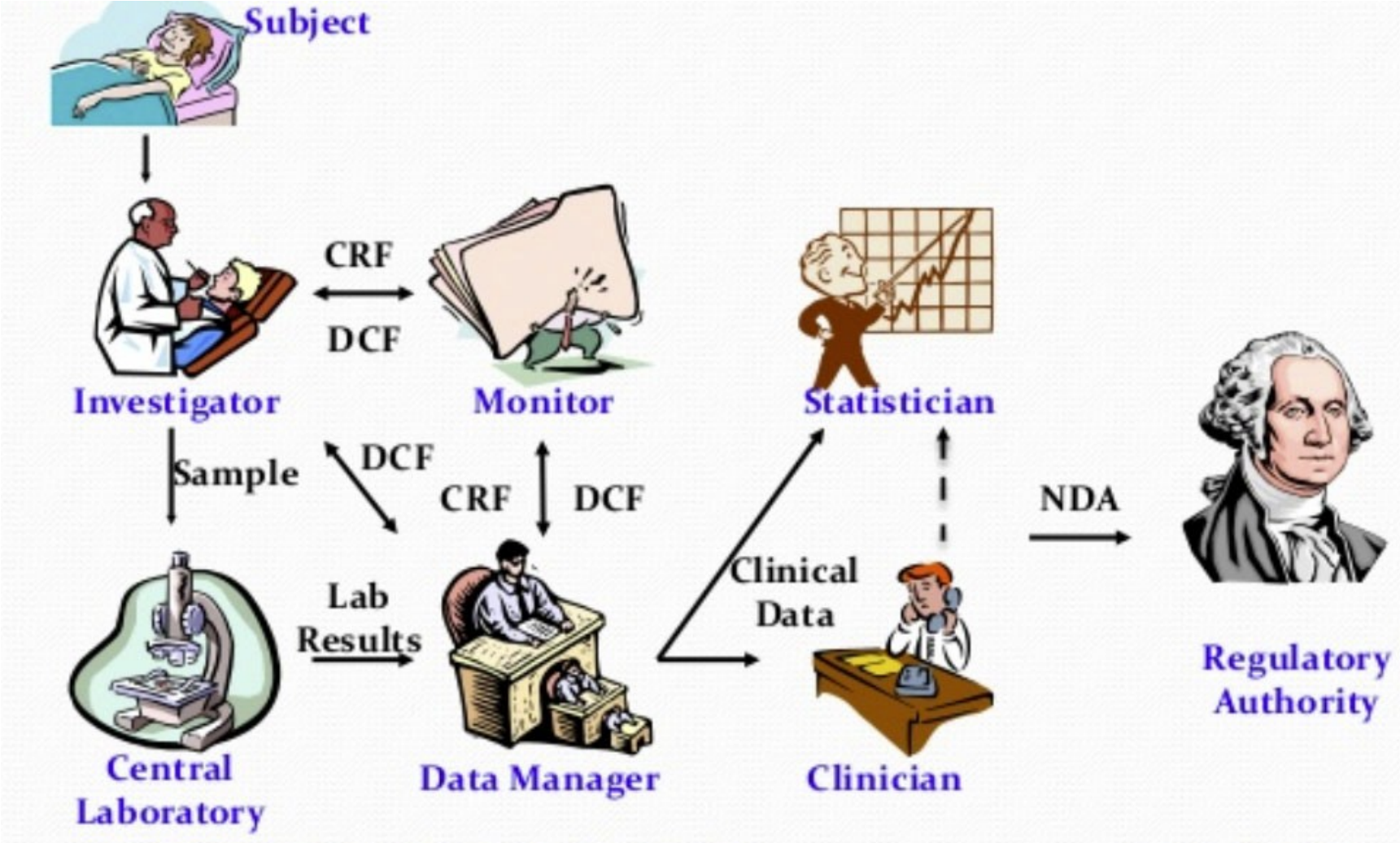
Topics

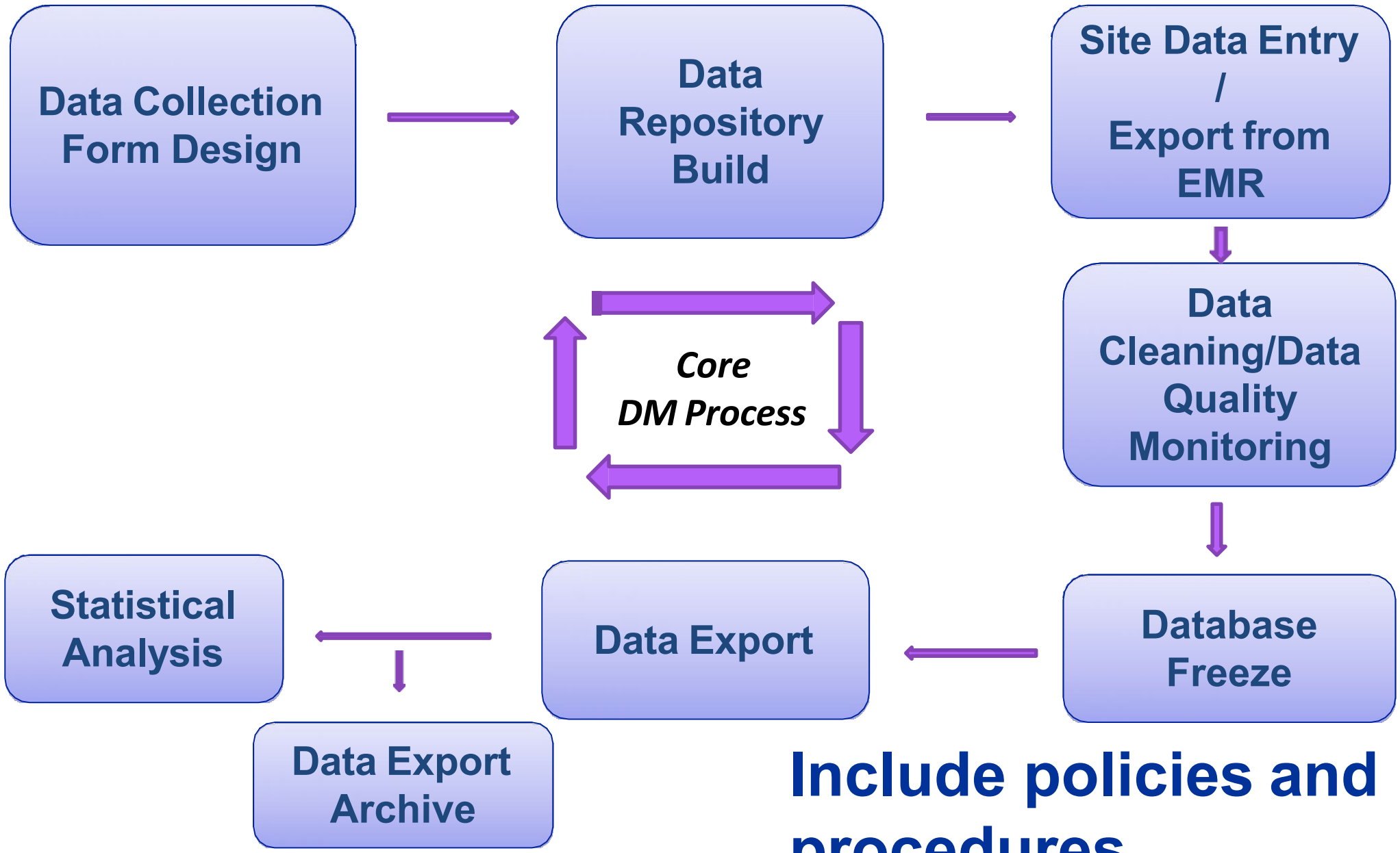
- Data operations
- Databases
- Software
  - Spreadsheets
  - Database management systems
  - Clinical trials management systems
- Other considerations

Common Terms

Abbreviation	Definition
CRF	Case Report Form
DB	Database
QC	Quality Control
DMP	Data Management Plan
CSR	Clinical Study Report
DCF	Data Clarification Form

Data Management Overview for Clinical Research





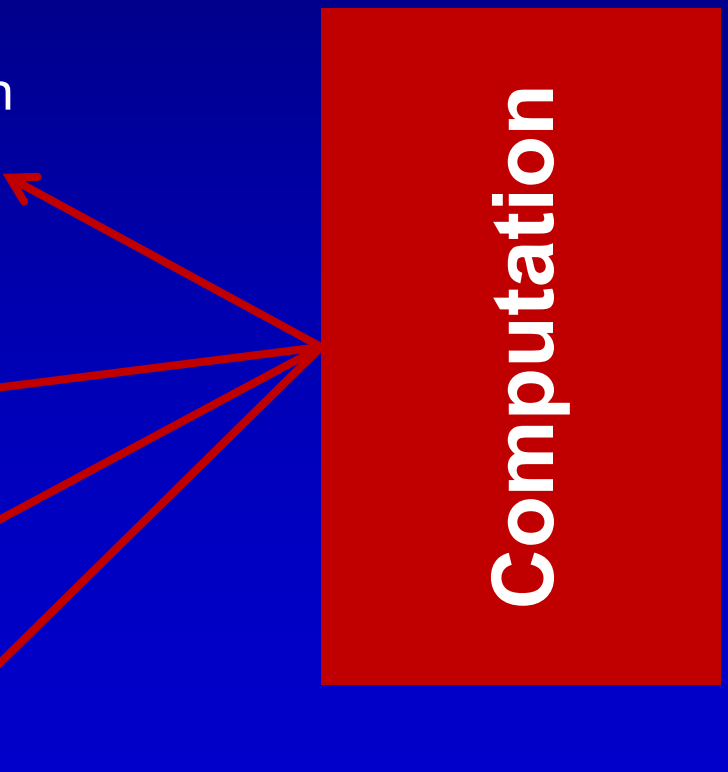
Include policies and procedures

My Background

- Biostatistics and epidemiology: oncology, HIV, clinical & translational research
- Biostatistics core director:
  - Cancer Center BSR
  - CTSA BERD
  - CTSA Informatics Cores
- National CTSA activities:
  - Chair of the BERD Key Function Committee
  - Co-Chair, Methods & Processes Domain Task Force
- PI, intervention trials and etiology studies

Biostatistics Core Functions

- Design studies
  - Clarify hypotheses and objectives
  - Define endpoints
  - Select study/experimental design
  - Sample size/power calculations
  - Develop analytic plans
- Monitor studies
  - Efficacy/futility
  - Safety
- Analyze studies
  - Statistical analysis
  - Writing reports/manuscripts



Why Talk About Data Management in a Biostatistics Seminar Series?

- You have learned a lot about biostatistics, but for most statisticians, the drudgery and hard work is getting and preparing study data for statistical analysis.
- 90/10 Rule

Clinical and Translational Research

- Purpose of clinical and translational research is to discovery new ways to improve the health of individuals and populations
- We do this by conducting research studies:
  - Hypothesis-generating studies
  - Hypothesis-testing studies*

*includes clinical trials, intervention trials, etc.

Clinical and Translational Research
(continued)

- Regardless of type of study, the most eloquently designed study is only as good as its data.
- Strength of evidence depends on *complete* and *valid* data:

Data → Information → Knowledge

Clinical and Translational Research
(continued)

- Data completeness and quality are critical for scientific discovery:
  - Good data with a bad design are worthless
  - Bad data with a good design is even worse
- Many investigators armed with an electronic spreadsheet think they have they need to conduct reproducible clinical/translational research

Wrong!

Clinical and Translational Research
(continued)

- What's sexier?
  - Statistical methods
  - Data management (DM)
- Data management is easily one of the most overlooked, underappreciated aspects of clinical and translational research

Note: For our discussion, a clinical trial is a specific study design within a range of clinical/translational research study types

Journal of Clinical and Translational Science

ACTS
ASSOCIATION FOR CLINICAL AND TRANSLATIONAL SCIENCE

EDUCATION
SPECIAL COMMUNICATION

Statistical competencies for medical research learners: What is fundamental?

Felicity T. Enders[1*], Christopher J. Lindsell[2], Leah J. Welty[3], Emma K. T. Benn[4], Susan M. Perkins[5], Matthew S. Mayo[6], Mohammad H. Rahbar[7], Kelley M. Kidwell[8], Sally W. Thurston[9], Heidi Spratt[10], Steven C. Grambow[11], Joseph Larson[1], Rickey E. Carter[1], Brad H. Pollock[12] and Robert A. Oster[13]

[1] Division of Biomedical Statistics & Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[2] Department of Emergency Medicine and Center for Clinical and Translational Science and Training, University of Cincinnati, Cincinnati, OH, USA
[3] Department of Preventive Medicine, Northwestern University, Evanston, IL, USA
[4] Department of Population Health Science and Policy, Center for Biostatistics, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5] Department of Biostatistics, School of Medicine, Indiana University, Indianapolis, IN, USA
[6] Department of Biostatistics, School of Medicine, University of Kansas Medical Center, Kansas City, KS, USA
[7] Department of Internal Medicine, McGovern Medical School, Division of Clinical and Translational Sciences, University of Texas Health Science Center at Houston, Houston, TX, USA
[8] Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
[9] Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA
[10] Department of Preventive Medicine and Community Health, The University of Texas Medical Branch, Galveston, TX, USA
[11] Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA
[12] Department of Public Health Sciences, University of California, Davis, Davis, CA, USA
[13] Department of Medicine, Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

Journal of Clinical and Translational Science (2017), 1, pp. 146–152 doi:10.1017/cts.2016.31

Statistical Competencies

6. Understand the value of **data quality and data management**.
7. Understand the reasons for performing **research that is reproducible from data collection** through publication of results.
9. Distinguish between **variable types** (e.g. continuous, binary, categorical) and understand the implications for selection of appropriate statistical methods. Extensively covered by required coursework.
12. Understand issues relating to generalizability of a study, including sampling methods and the amount and type of **missing data**.
16. Understand the need to address **loss to follow-up**.
21. Understand the purpose of **data and safety monitoring plans**.

Ender et al, *J Clin Trans Res*, 2017, 1:146–152

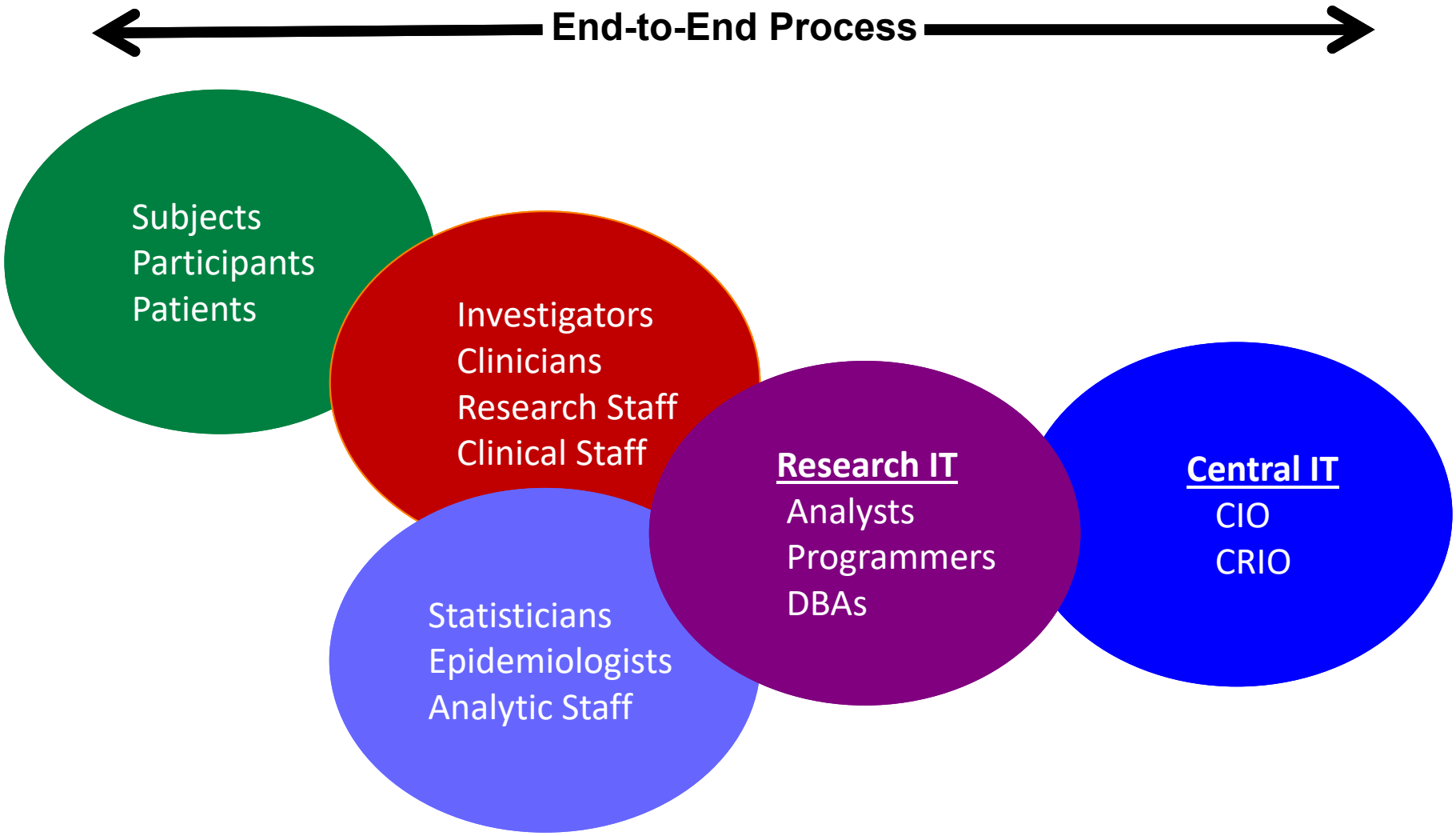
DATA MANAGEMENT

What is Data Management?

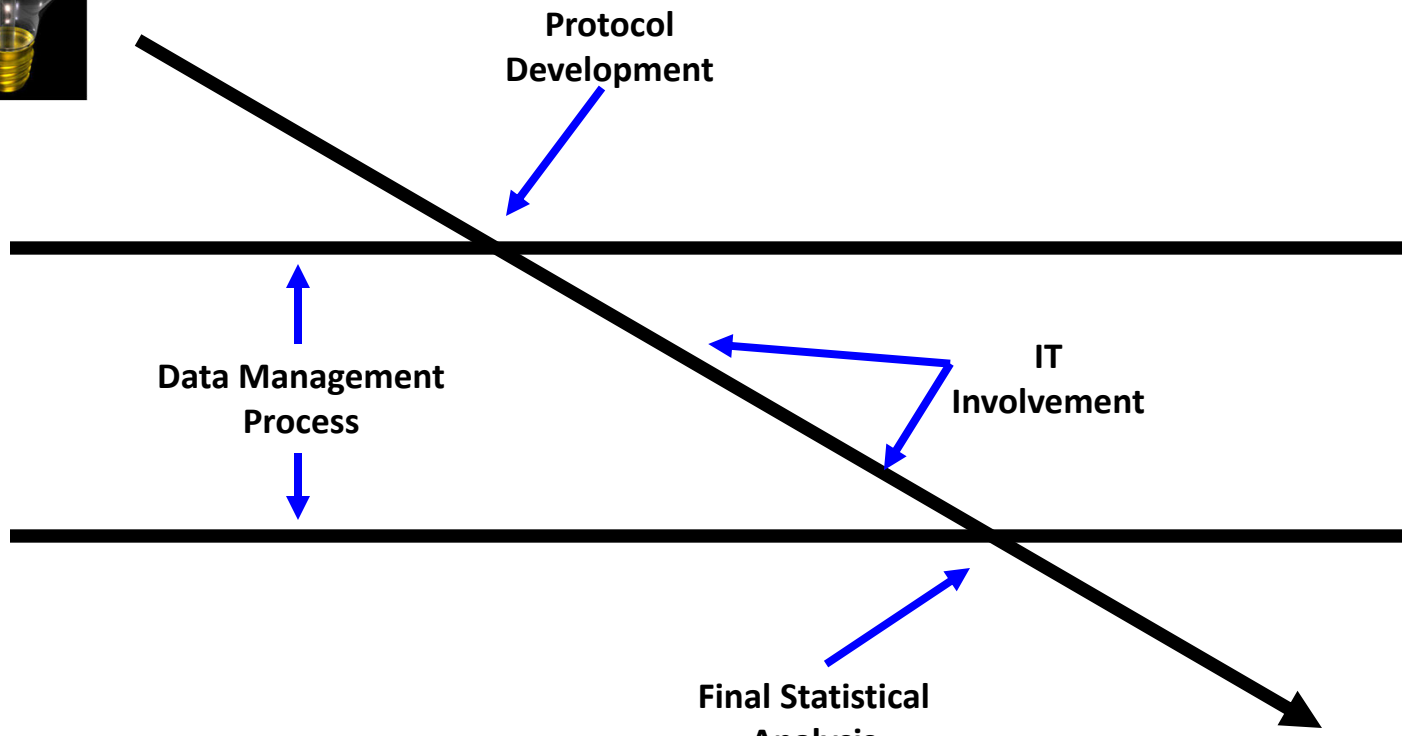
- The development, execution and supervision of plans, policies, programs and practices that control, protect, deliver, and enhance the value of data and information assets*

*Data Management Association, Data Management Body of Knowledge (DAMA-DMBOK), 2008

Who is Involved in Data Management?



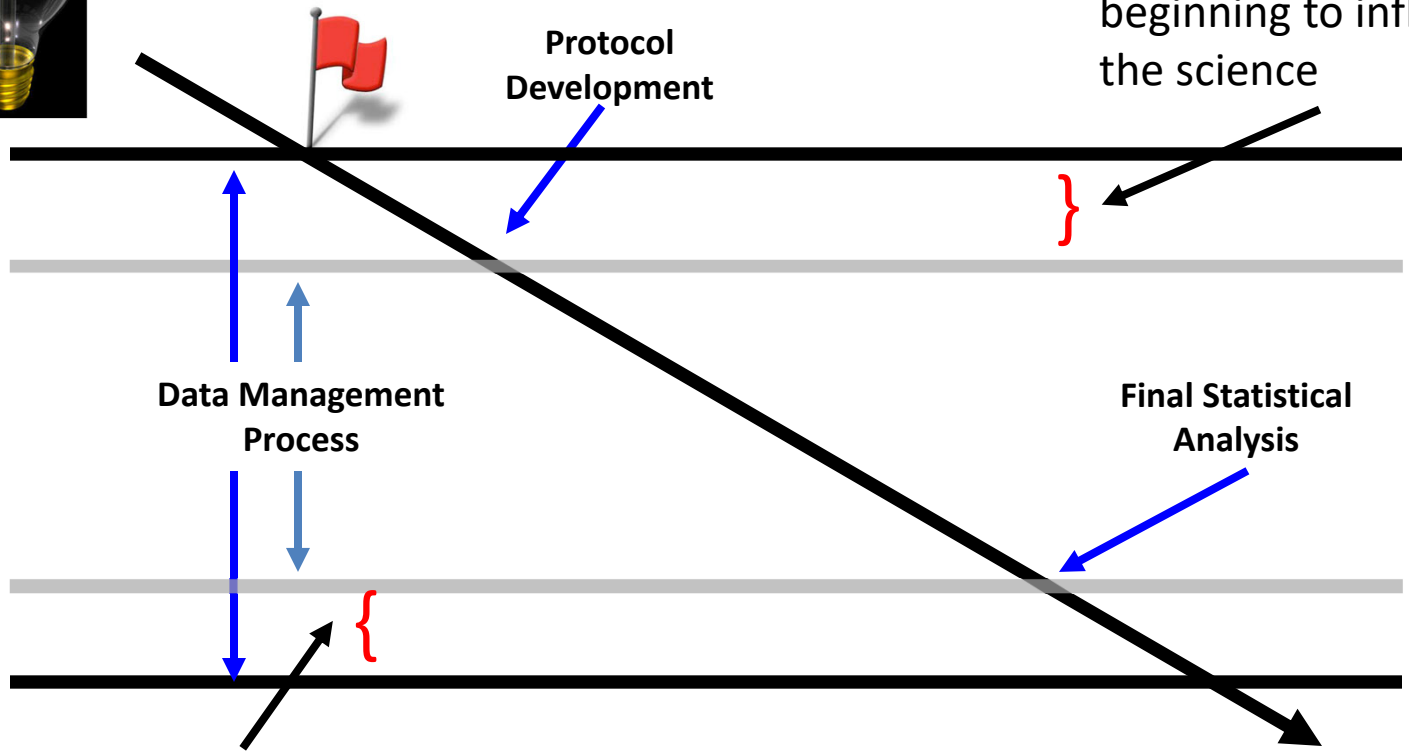
Data Management within the Research Process



JAMA®
The Journal of the American Medical Association

Data Management Changing Within the Research Process

Data management considerations are beginning to influence the science



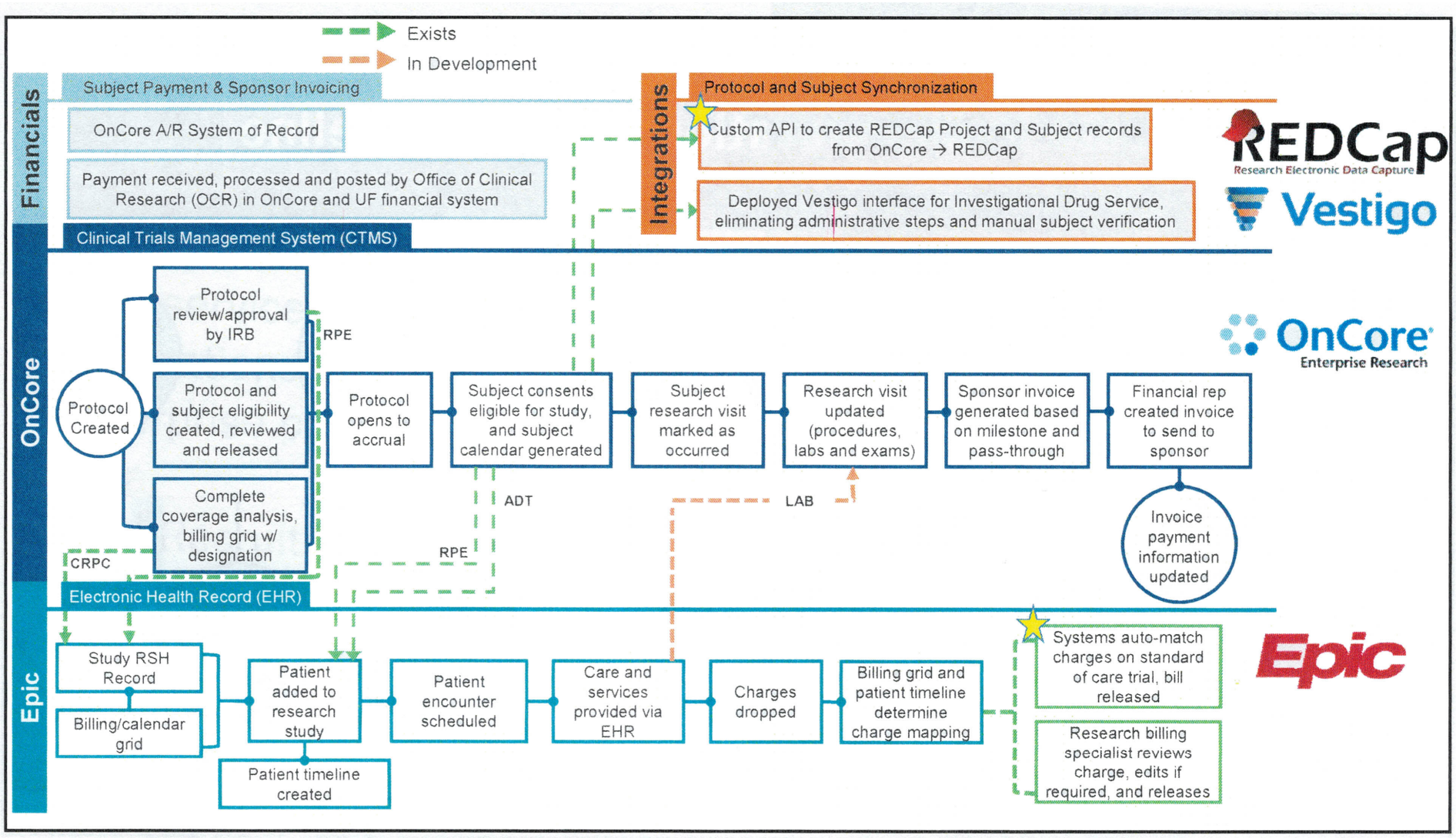
Storage and long-term utilization affect the data long after the protocol's final analysis

JAMA®
The Journal of the American Medical Association

Data Management Elements

- Need to maintain functional, flexible, scalable, cost-efficient set of resources to handle a variety of data:
  - Demographic
  - Clinical/laboratory and -omics
  - Environmental

  Exposome
- Data quality and compliance with regulatory requirements
  - HIPAA, 21 CFR Part 11, FISMA
- Prospective planning for:
  - Long time horizons
    - Environmental Influences on Child Health Outcomes (ECHO)
  - Interoperability and federation
    - OnCore CTMS Enterprise Research with EPIC and REDCap

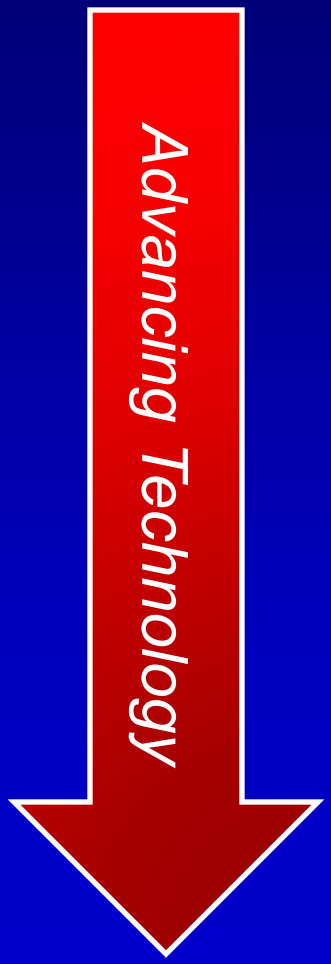

Exists
In Development
Financials
Subject Payment & Sponsor Invoicing
OnCore A/R System of Record
Payment received, processed and posted by Office of Clinical Research (OCR) in OnCore and UF financial system
Integrations
Protocol and Subject Synchronization
Custom API to create REDCap Project and Subject records from OnCore → REDCap
Deployed Vestigo interface for Investigational Drug Service, eliminating administrative steps and manual subject verification
REDCap
Research Electronic Data Capture
Vestigo
OnCore
Clinical Trials Management System (CTMS)
Protocol Created
Protocol review/approval by IRB
Protocol and subject eligibility created, reviewed and released
Complete coverage analysis, billing grid w/ designation
RPE
CRPC
Protocol opens to accrual
Subject consents eligible for study, and subject calendar generated
ADT
RPE
Subject research visit marked as occurred
Research visit updated (procedures, labs and exams)
LAB
Sponsor invoice generated based on milestone and pass-through
Financial rep created invoice to send to sponsor
Invoice payment information updated
OnCore Enterprise Research
Epic
Electronic Health Record (EHR)
Study RSH Record
Billing/calendar grid
Patient added to research study
Patient timeline created
Patient encounter scheduled
Care and services provided via EHR
Charges dropped
Billing grid and patient timeline determine charge mapping
Systems auto-match charges on standard of care trial, bill released
Research billing specialist reviews charge, edits if required, and releases
Epic

Database Management Functions

- Database design
  - Data elements
  - Relationships (data model)
  - Access control/security/integrity
- Application development
  - Data capture
  - Data curation
  - Querying
  - Reporting
  - Audit
- Database operations

How Data Are Handled?

- Paper forms (CRFs) and keypunch
- Client-server DBMS and networked DBMS
- Web-front end DBMS
  - Pediatric Oncology Group replaced paper in 1998
    - Web front-end
    - Oracle back-end
- Clinical Trials Management System (CTMS)



Databases

- Data elements

Data Elements

- Common Data Elements (CDE)
  - Try to use standards with ontologies
    - Common Terminology Criteria for Adverse Events (CTCAE)
    - Patient-Reported Outcomes Measurement Information System (PROMIS)
    - International Classification of Diseases for Oncology (ICD-O)
  - Data dictionaries
  - Case Report Forms (eCRFs)
  - Map/link to other information systems (biorepository, EHR)
- Specialized (study-specific data elements)

Building and Adolescent and Young Adult Oncology Research Database

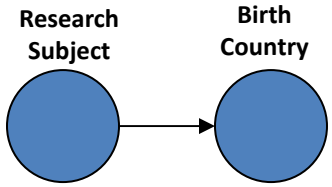
- Demo

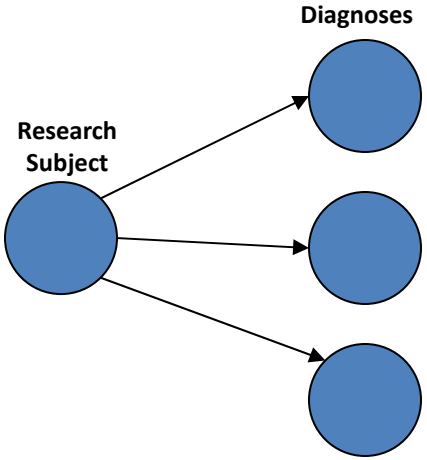
Databases

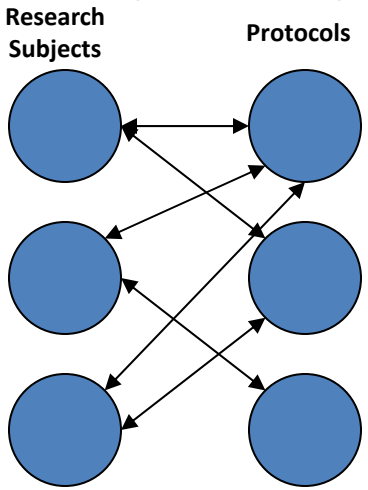
- Data elements
- Database models

Database Model: Data Relationships

- Three types of relationships:
  - One to One
  - One to Many
  - Many to Many







- The relationships of the data reflect the rules of the system (your protocol) and not all potential possibilities
  - NOTE: One of the most expensive things to change once underway

Databases

- Data elements
- Database models
- Validation
  - Part of the data plan, multiple methods
- Curation
  - Goal is to maintain the value of the data over time
  - Organization, annotation, revisions/audit log
  - Reuse, future proofing

Software

“Database Management” Software

Microsoft Excel

Excel Characteristics

- Advantages
  - Easy to work with → Quick start up, low costs
  - Potentially you can force data types
- Disadvantages
  - Easy to work with →

    No requirement to clearly define needs
  - Will “interpret” data entries for you
    - Will not allow you to automatically override

Examples of Good and Bad Variable Names

good name	good alternative	avoid
Max_temp_C	MaxTemp	Maximum Temp (°C)
Precipitation_mm	Precipitation	precmm
Mean_year_growth	MeanYearGrowth	Mean growth/year
sex	sex	M/F
weight	weight	w.
cell_type	CellType	Cell type
Observation_01	first_observation	1st Obs.

A spreadsheet with inconsistent date formats

	A	B	C
1	Date	Assay date	Weight
2		12/9/05	54.9
3		12/9/05	45.3
4	12/6/2005	e	47
5		e	45.7
6		e	52.9
7		1/11/2006	46.1
8		1/11/2006	38.6

Examples of spreadsheets that violate the ‘‘no empty cells” recommendation

A

	A	B	C
1	id	date	glucose
2	101	2015–06–14	149.3
3	102		95.3
4	103	2015–06–18	97.5
5	104		117.0
6	105		108.0
7	106	2015–06–20	149.0
8	107		169.4

B

	A	B	C	D	E	F	G	H	I
1		1 min				5 min			
2	strain	normal		mutant		normal		mutant	
3	A	147	139	166	179	334	354	451	474
4	B	246	240	178	172	514	611	412	447

A tidy version of the above data

	A	B	C	D	E
1	strain	genotype	min	replicate	response
2	A	normal	1	1	147
3	A	normal	1	2	139
4	B	normal	1	1	246
5	B	normal	1	2	240
6	A	mutant	1	1	166
7	A	mutant	1	2	179
8	B	mutant	1	1	178
9	B	mutant	1	2	172
10	A	normal	5	1	334
11	A	normal	5	2	354
12	B	normal	5	1	514
13	B	normal	5	2	611
14	A	mutant	5	1	451
15	A	mutant	5	2	474
16	B	mutant	5	1	412
17	B	mutant	5	2	447

BHP1

Spreadsheets with nonrectangular layouts

A

	A	B	C	D	E	F
1						
2		101	102	103	104	105
3	sex	Male	Female	Male	Male	Male
4						
5		101	102	103	104	105
6	glucose	134.1	120.0	124.8	83.1	105.2
7						
8		101	102	103	104	105
9	insulin	0.60	1.18	1.23	1.16	0.73

B

	A	B	C	D	E	F	G
1	1MIN						
2			Normal			Mutant	
3	B6	146.6	138.6	155.6	166	179.3	186.9
4	BTBR	245.7	240	243.1	177.8	171.6	188.1
5							
6	5MIN						
7			Normal			Mutant	
8	B6	333.6	353.6	408.8	450.6	474.4	423.8
9	BTBR	514.4	610.6	597.9	412.1	447.4	446.5

C

	A	B	C	D	E	F	G
1							
2	Date	11/3/14					
3	Days on diet	126					
4	Mouse #	43					
5	sex	f					
6	experiment		values			mean	SD
7	control		0.186	0.191	1.081	0.49	0.52
8	treatment A		7.414	1.468	2.254	3.71	3.23
9	treatment B		9.811	9.259	11.296	10.12	1.05
10							
11	fold change		values			mean	SD
12	treatment A		15.26	3.02	4.64	7.64	6.65
13	treatment B		20.19	19.05	23.24	20.83	2.17

D

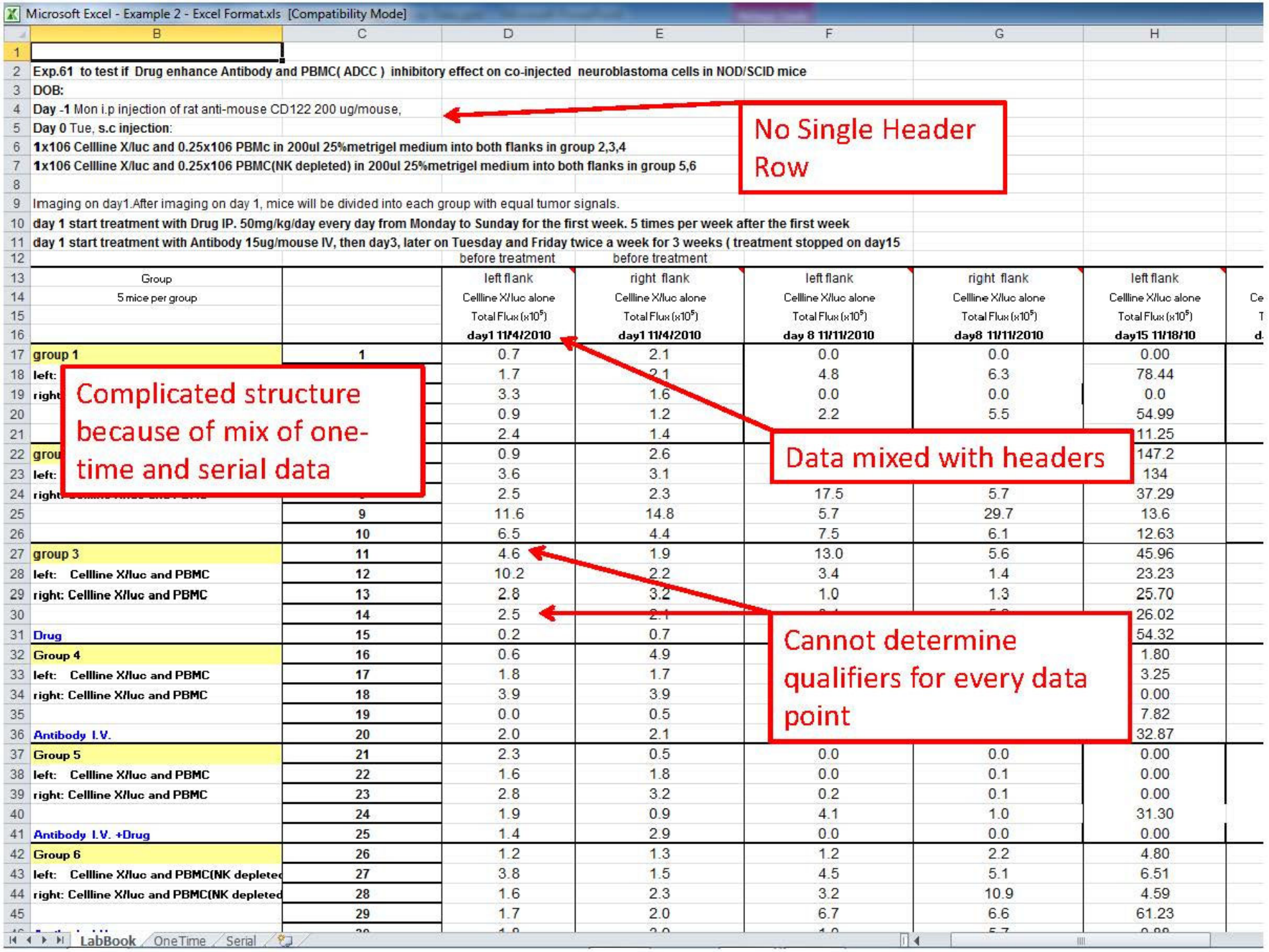
	A	B	C	D	E	F
1		GTT date	GTT weight	time	glucose mg/dl	insulin ng/ml
2	321	2/9/15	24.5	0	99.2	lo off curve
3				5	349.3	0.205
4				15	286.1	0.129
5				30	312	0.175
6				60	99.9	0.122
7				120	217.9	lo off curve
8	322	2/9/15	18.9	0	185.8	0.251
9				5	297.4	2.228
10				15	439	2.078
11				30	362.3	0.775
12				60	232.7	0.5
13				120	260.7	0.523
14	323	2/9/15	24.7	0	198.5	0.151
15				5	530.6	off curve lo

Slide 37

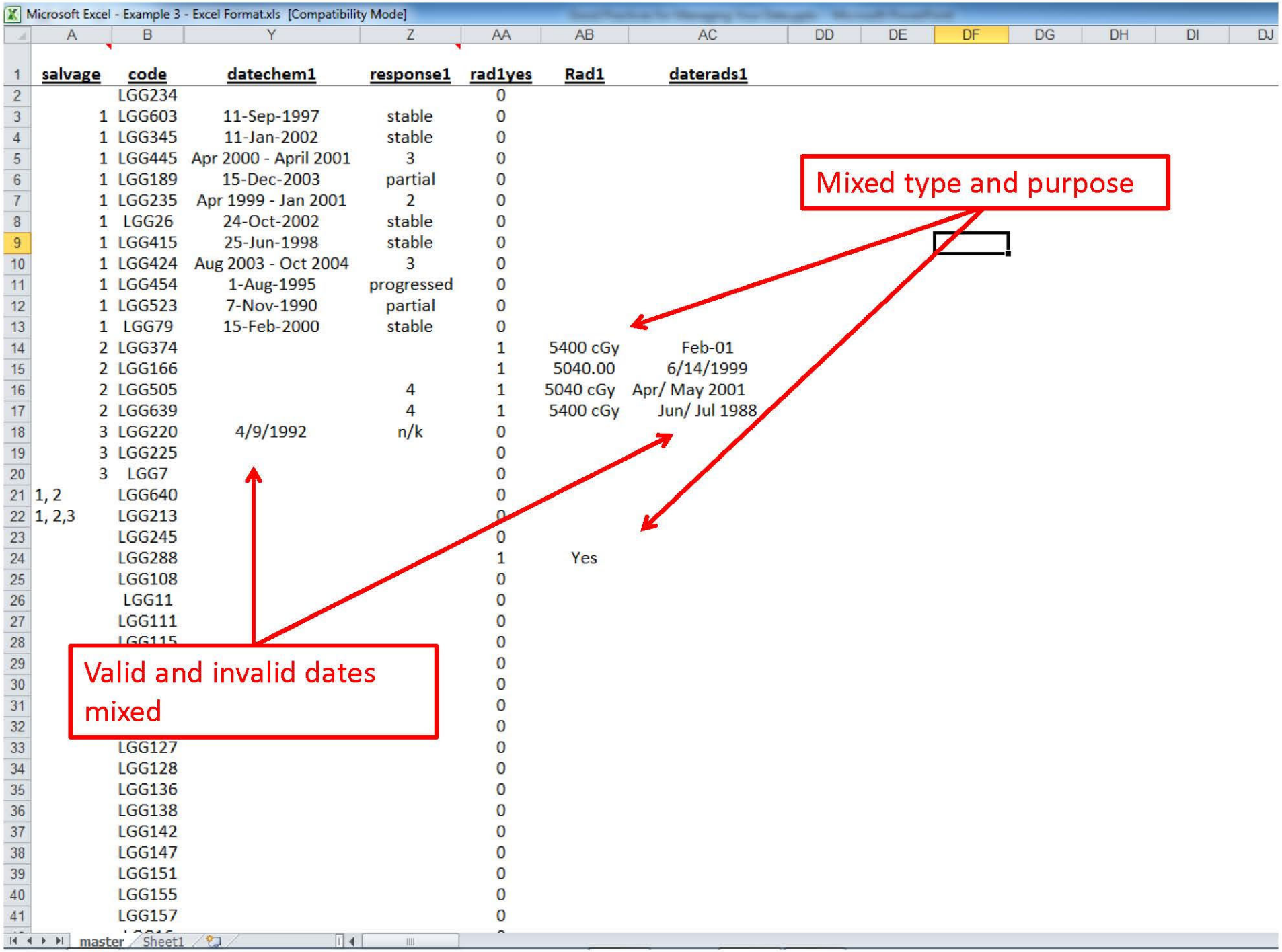
BHP1 Brad H. Pollock, 3/9/2020

spreadsheet with a rectangular layout

	A	B	C	D	E
1	id	sex	glucose	insulin	triglyc
2	101	Male	134.1	0.60	273.4
3	102	Female	120.0	1.18	243.6
4	103	Male	124.8	1.23	297.6
5	104	Male	83.1	1.16	142.4
6	105	Male	105.2	0.73	215.7

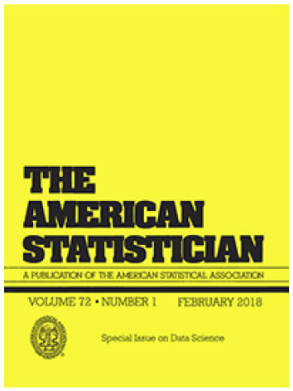


			before treatment	before treatment			
Group		left flank	right flank	left flank	right flank	left flank	
5 mice per group		Cellline X/luc alone	Cellline X/luc alone	Cellline X/luc alone	Cellline X/luc alone	Cellline X/luc alone	Ce
		Total Flux (x10^5)	Total Flux (x10^5)	Total Flux (x10^5)	Total Flux (x10^5)	Total Flux (x10^5)	T
		day1 11/4/2010	**day1 11/4/2010**	**day 8 11/11/2010**	**day8 11/11/2010**	**day15 11/18/10**	**d**
group 1	1	0.7	2.1	0.0	0.0	0.00	
left:		1.7	2.1	4.8	6.3	78.44	
right		3.3	1.6	0.0	0.0	0.0	
		0.9	1.2	2.2	5.5	54.99	
		2.4	1.4			11.25	
grou		0.9	2.6			147.2	
left:		3.6	3.1			134	
right:		2.5	2.3	17.5	5.7	37.29	
	9	11.6	14.8	5.7	29.7	13.6	
	10	6.5	4.4	7.5	6.1	12.63	
group 3	11	4.6	1.9	13.0	5.6	45.96	
left: Cellline X/luc and PBMC	12	10.2	2.2	3.4	1.4	23.23	
right: Cellline X/luc and PBMC	13	2.8	3.2	1.0	1.3	25.70	
	14	2.5	2.1			26.02	
Drug	15	0.2	0.7			54.32	
Group 4	16	0.6	4.9			1.80	
left: Cellline X/luc and PBMC	17	1.8	1.7			3.25	
right: Cellline X/luc and PBMC	18	3.9	3.9			0.00	
	19	0.0	0.5			7.82	
Antibody I.V.	20	2.0	2.1			32.87	
Group 5	21	2.3	0.5	0.0	0.0	0.00	
left: Cellline X/luc and PBMC	22	1.6	1.8	0.0	0.1	0.00	
right: Cellline X/luc and PBMC	23	2.8	3.2	0.2	0.1	0.00	
	24	1.9	0.9	4.1	1.0	31.30	
Antibody I.V. +Drug	25	1.4	2.9	0.0	0.0	0.00	
Group 6	26	1.2	1.3	1.2	2.2	4.80	
left: Cellline X/luc and PBMC(NK depleted	27	3.8	1.5	4.5	5.1	6.51	
right: Cellline X/luc and PBMC(NK depleted	28	1.6	2.3	3.2	10.9	4.59	
	29	1.7	2.0	6.7	6.6	61.23	



	A	B	Y	Z	AA	AB	AC	DD	DE	DF	DG	DH	DI	DJ
1	**salvage**	**code**	**datechem1**	**response1**	**rad1yes**	**Rad1**	**daterads1**							
2		LGG234			0									
3	1	LGG603	11-Sep-1997	stable	0									
4	1	LGG345	11-Jan-2002	stable	0									
5	1	LGG445	Apr 2000 - April 2001	3	0									
6	1	LGG189	15-Dec-2003	partial	0									
7	1	LGG235	Apr 1999 - Jan 2001	2	0									
8	1	LGG26	24-Oct-2002	stable	0									
9	1	LGG415	25-Jun-1998	stable	0									
10	1	LGG424	Aug 2003 - Oct 2004	3	0									
11	1	LGG454	1-Aug-1995	progressed	0									
12	1	LGG523	7-Nov-1990	partial	0									
13	1	LGG79	15-Feb-2000	stable	0									
14	2	LGG374			1	5400 cGy	Feb-01							
15	2	LGG166			1	5040.00	6/14/1999							
16	2	LGG505		4	1	5040 cGy	Apr/ May 2001							
17	2	LGG639		4	1	5400 cGy	Jun/ Jul 1988							
18	3	LGG220	4/9/1992	n/k	0									
19	3	LGG225			0									
20	3	LGG7			0									
21	1, 2	LGG640			0									
22	1, 2,3	LGG213			0									
23		LGG245			0									
24		LGG288			1	Yes								
25		LGG108			0									
26		LGG11			0									
27		LGG111			0									
28		LGG115			0									
29					0									
30					0									
31					0									
32					0									
33		LGG127			0									
34		LGG128			0									
35		LGG136			0									
36		LGG138			0									
37		LGG142			0									
38		LGG147			0									
39		LGG151			0									
40		LGG155			0									
41		LGG157			0									

master Sheet1



Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Data Organization in Spreadsheets

Karl W. Broman & Kara H. Woo

To cite this article: Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1, 2-10, DOI: 10.1080/00031305.2017.1375989

Dangers of Spreadsheets

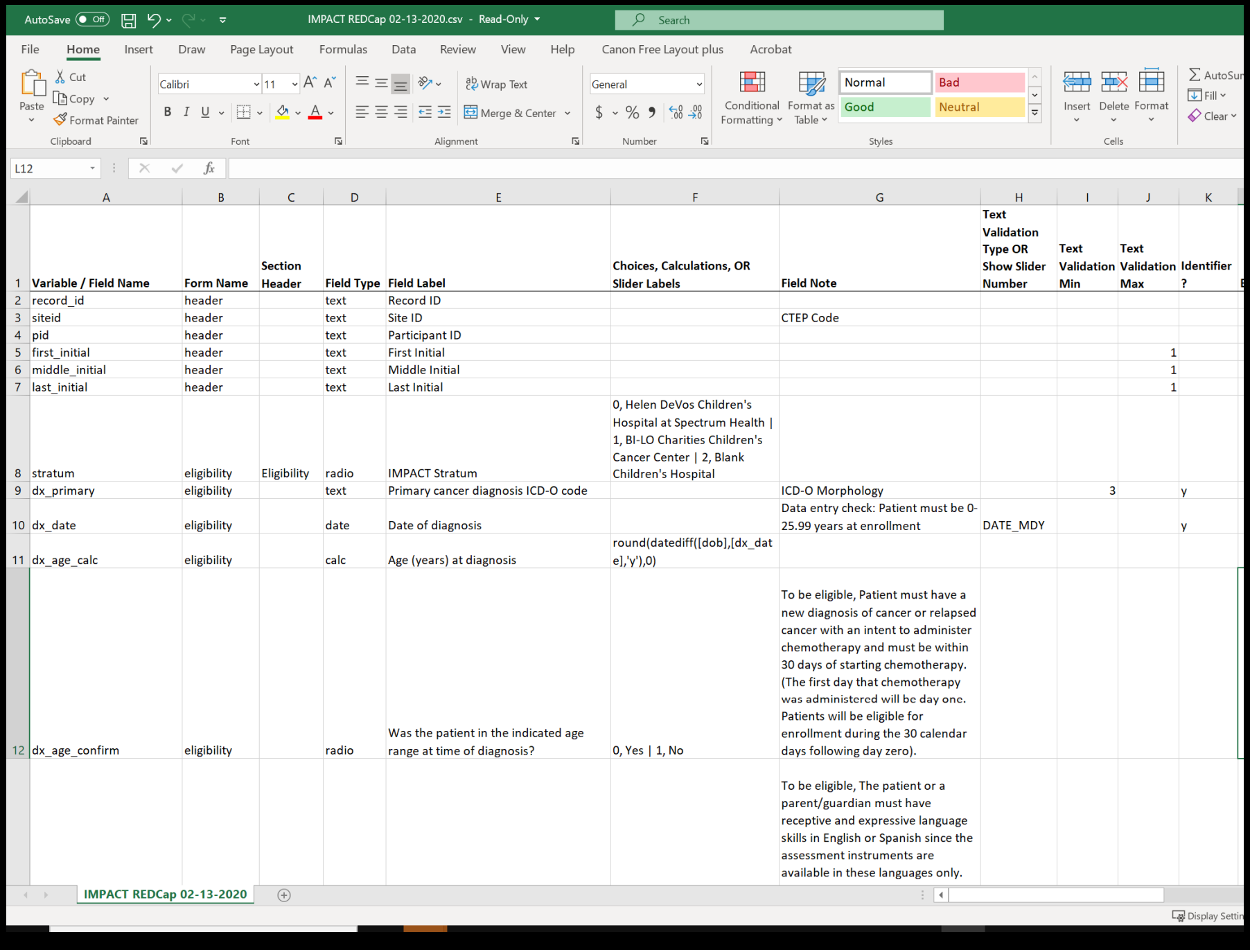
- The dangers are real
  - European Spreadsheet Risks Interest Group keeps a public archive of spreadsheet "horror stories" (http://www. eusprig.org/horror-stories.htm).
- Many researchers have examined error rates in spreadsheets
  - Panko* (2008) reported that in 13 audits of real-world spreadsheets, an average of 88% contained errors.
- Popular spreadsheet programs also make certain types of errors easy to commit and difficult to rectify.
  - Excel converts some gene names to dates and stores dates differently between operating systems, causing problems in downstream analyses (Zeeberg et al. 2004; Woo 2014).

*http://panko.shidler.hawaii.edu/SSR/Mypapers/whatknow.htm

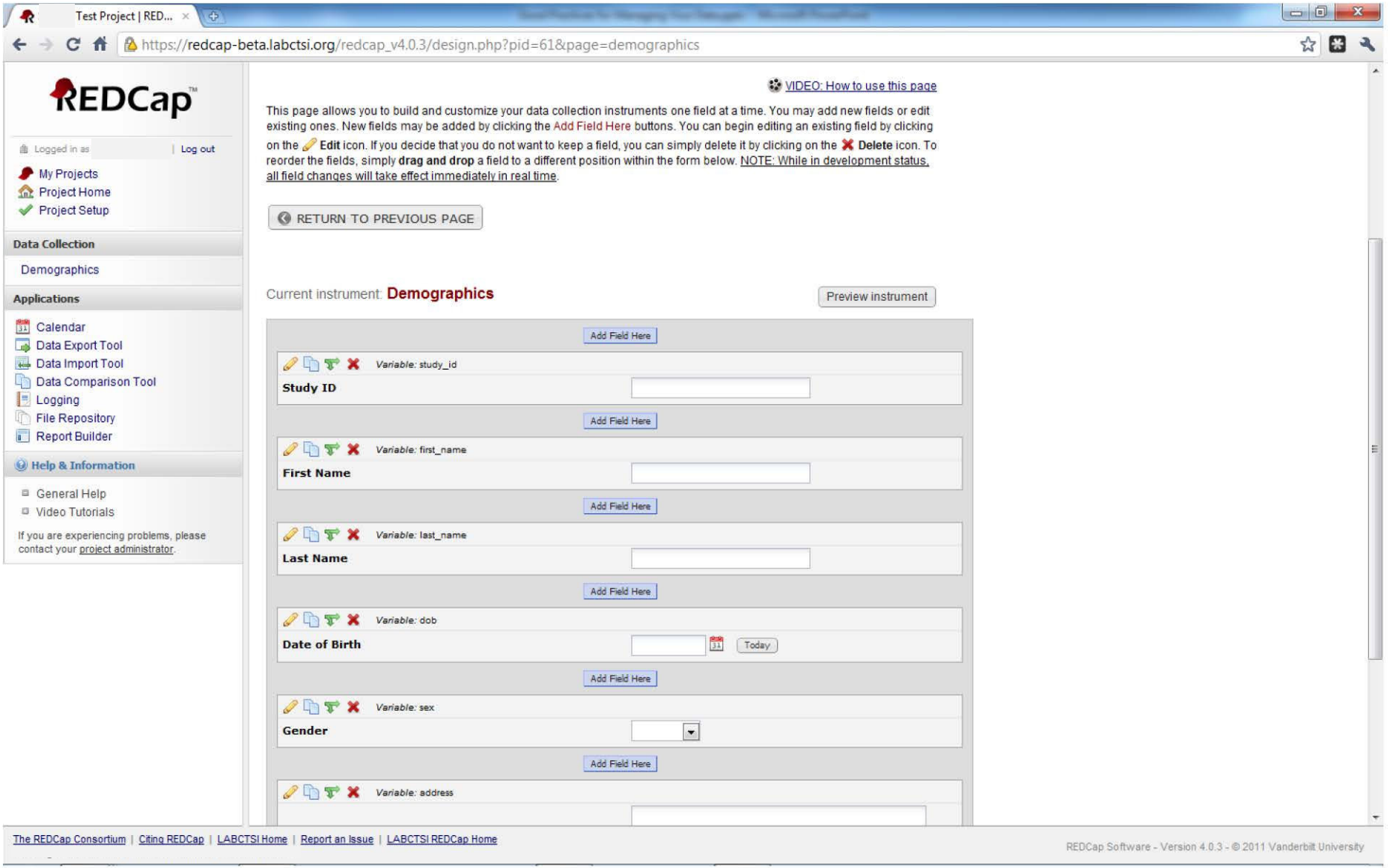
Dangers of Spreadsheets (continued)

- Researchers who use spreadsheets should be aware of these common errors and design spreadsheets that are tidy, consistent, and as resistant to mistakes as possible

“Database Management” Software



	A	B	C	D	E	F	G	H	I	J	K
1	**Variable / Field Name**	**Form Name**	**Section Header**	**Field Type**	**Field Label**	**Choices, Calculations, OR Slider Labels**	**Field Note**	**Text Validation Type OR Show Slider Number**	**Text Validation Min**	**Text Validation Max**	**Identifier ?**
2	record_id	header		text	Record ID						
3	siteid	header		text	Site ID		CTEP Code				
4	pid	header		text	Participant ID						
5	first_initial	header		text	First Initial					1	
6	middle_initial	header		text	Middle Initial					1	
7	last_initial	header		text	Last Initial					1	
8	stratum	eligibility	Eligibility	radio	IMPACT Stratum	0, Helen DeVos Children's Hospital at Spectrum Health | 1, BI-LO Charities Children's Cancer Center | 2, Blank Children's Hospital					
9	dx_primary	eligibility		text	Primary cancer diagnosis ICD-O code		ICD-O Morphology		3		y
10	dx_date	eligibility		date	Date of diagnosis		Data entry check: Patient must be 0-25.99 years at enrollment	DATE_MDY			y
11	dx_age_calc	eligibility		calc	Age (years) at diagnosis	round(datediff([dob],[dx_date],'y'),0)					
12	dx_age_confirm	eligibility		radio	Was the patient in the indicated age range at time of diagnosis?	0, Yes | 1, No	To be eligible, Patient must have a new diagnosis of cancer or relapsed cancer with an intent to administer chemotherapy and must be within 30 days of starting chemotherapy. (The first day that chemotherapy was administered will be day one. Patients will be eligible for enrollment during the 30 calendar days following day zero).				
							To be eligible, The patient or a parent/guardian must have receptive and expressive language skills in English or Spanish since the assessment instruments are available in these languages only.				


Test Project | RED...
https://redcap-beta.labctsi.org/redcap_v4.0.3/design.php?pid=61&page=demographics
REDCap
Logged in as
Log out
My Projects
Project Home
Project Setup
Data Collection
Demographics
Applications
Calendar
Data Export Tool
Data Import Tool
Data Comparison Tool
Logging
File Repository
Report Builder
Help & Information
General Help
Video Tutorials
If you are experiencing problems, please contact your project administrator.
VIDEO: How to use this page
This page allows you to build and customize your data collection instruments one field at a time. You may add new fields or edit existing ones. New fields may be added by clicking the Add Field Here buttons. You can begin editing an existing field by clicking on the Edit icon. If you decide that you do not want to keep a field, you can simply delete it by clicking on the Delete icon. To reorder the fields, simply drag and drop a field to a different position within the form below. NOTE: While in development status, all field changes will take effect immediately in real time.
RETURN TO PREVIOUS PAGE
Current instrument: Demographics
Preview instrument
Add Field Here
Variable: study_id
Study ID
Variable: first_name
First Name
Variable: last_name
Last Name
Variable: dob
Date of Birth
Today
Variable: sex
Gender
Variable: address
The REDCap Consortium | Citing REDCap | LABCTSI Home | Report an Issue | LABCTSI REDCap Home
REDCap Software - Version 4.0.3 - © 2011 Vanderbilt University

REDCap Features

- Good Points
  - Easy to set up, not resource intensive
  - Requires a real data dictionary
  - Central server engine (security & data integrity)
  - Easy access through web front-end
- Not so Great Points
  - Display interface not very customizable
    - Layout, limited skip patterns, etc.
  - Each application is a separate instance
  - Adverse events monitoring difficult
  - Not truly relational
  - No data curation, electronic data collection only

REDCap
(Research Electronic Data Capture)

- Online or offline use
- Regulatory compliance
  - HIPAA, 21 CFR Part 11, FISMA
- Features:
  - Customizable
  - Automated export procedures, built-in project calendar, scheduling module
  - Audit trails
  - Ad hoc reporting tools
  - Branching logic, file uploading, and calculated fields

“Database Management” Software

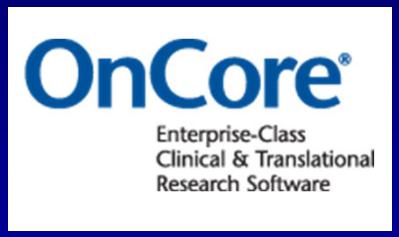
REDCap
Research Electronic Data Capture

SCIENCE TRAX™
Your Data, Our Passion, Accelerating Research.
Home | StudyTRAX | Events | News | Customers

qualtrics® sophisticated research made simple™

ORACLE®

Clinical Trials Management Systems (CTMS)



IMPACT® CTMS

Uses:

- Planning, preparation, monitoring and reporting of clinical trials
- Administrative/financial/portfolio management capabilities
- Electronic case report forms (eCRFs)
- ± Interoperate with other systems

Other Considerations for Data Operations

- Standard Operations Procedures (SOPs)
- Disaster recovery
- Version control (Surround SCM)
- Audit
- Separation of duties
  - DBAs, analysts, statisticians
- Electronic Sign-offs (Editor → Monitor → PI)
- Honest broker role (PHI-related)

How important are research IT/informatics solutions for novel clinical trial designs?

The Clinical Trial, Re-Imagined

The ground-breaking I-SPY 2 trial of neoadjuvant treatment for locally advanced breast cancer established a new benchmark for efficiency of phase II clinical trials. Widely regarded as a pioneer of the 'platform' trial, I-SPY 2's success continues to be a major influence on the development of next-generation trial designs in oncology and beyond.

I-SPY 2 TRIAL

(Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular Analysis 2)

I-SPY 2 is a clinical trial for women with newly diagnosed locally advanced breast cancer (neoadjuvant)

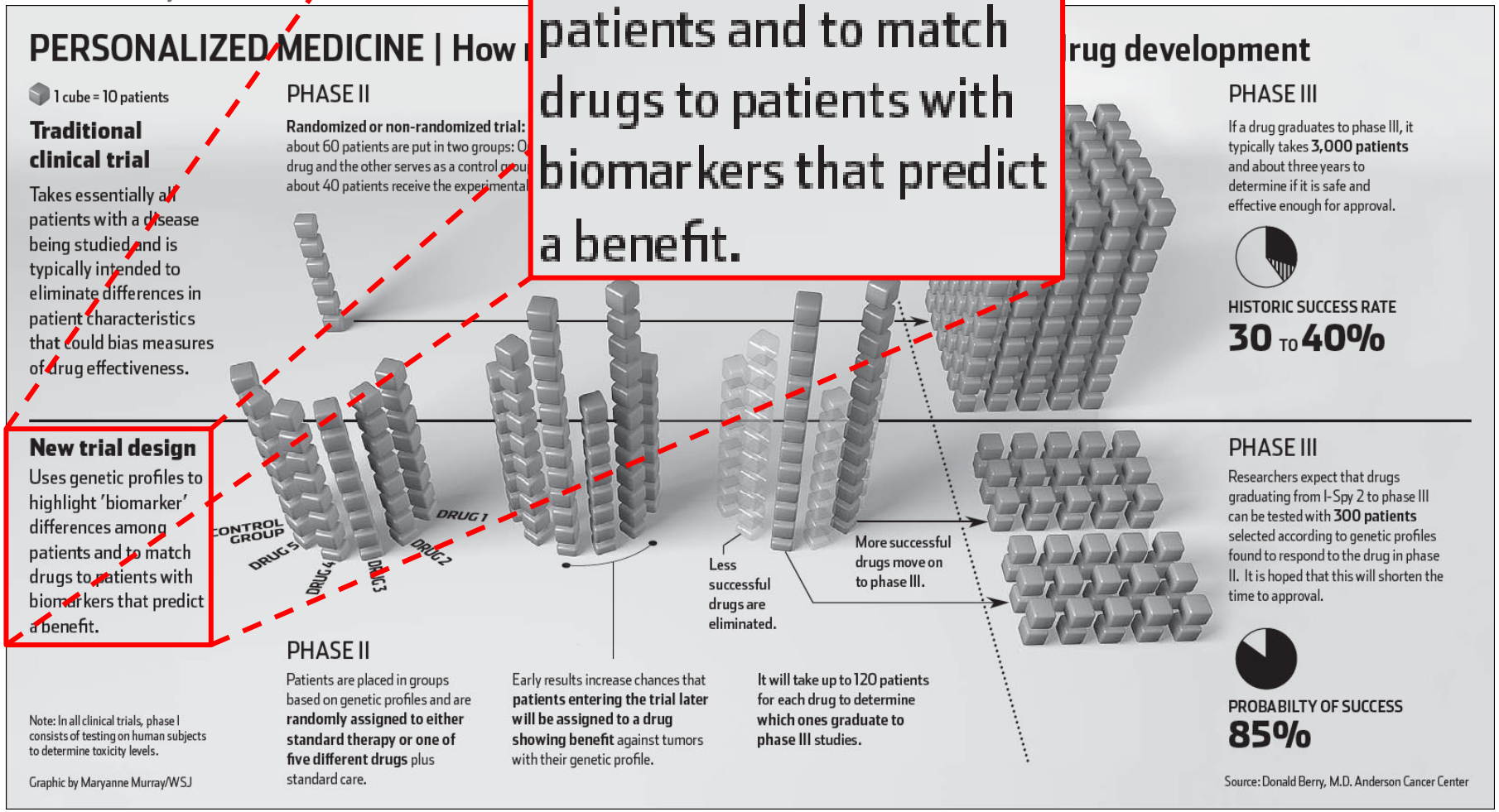
THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCT

A New Rx for Med

Fed up with slow drug trials, ca *ack to personalized treatments.*

By RON WINSLOW



New trial design
Uses genetic profiles to highlight 'biomarker' differences among patients and to match drugs to patients with biomarkers that predict a benefit.

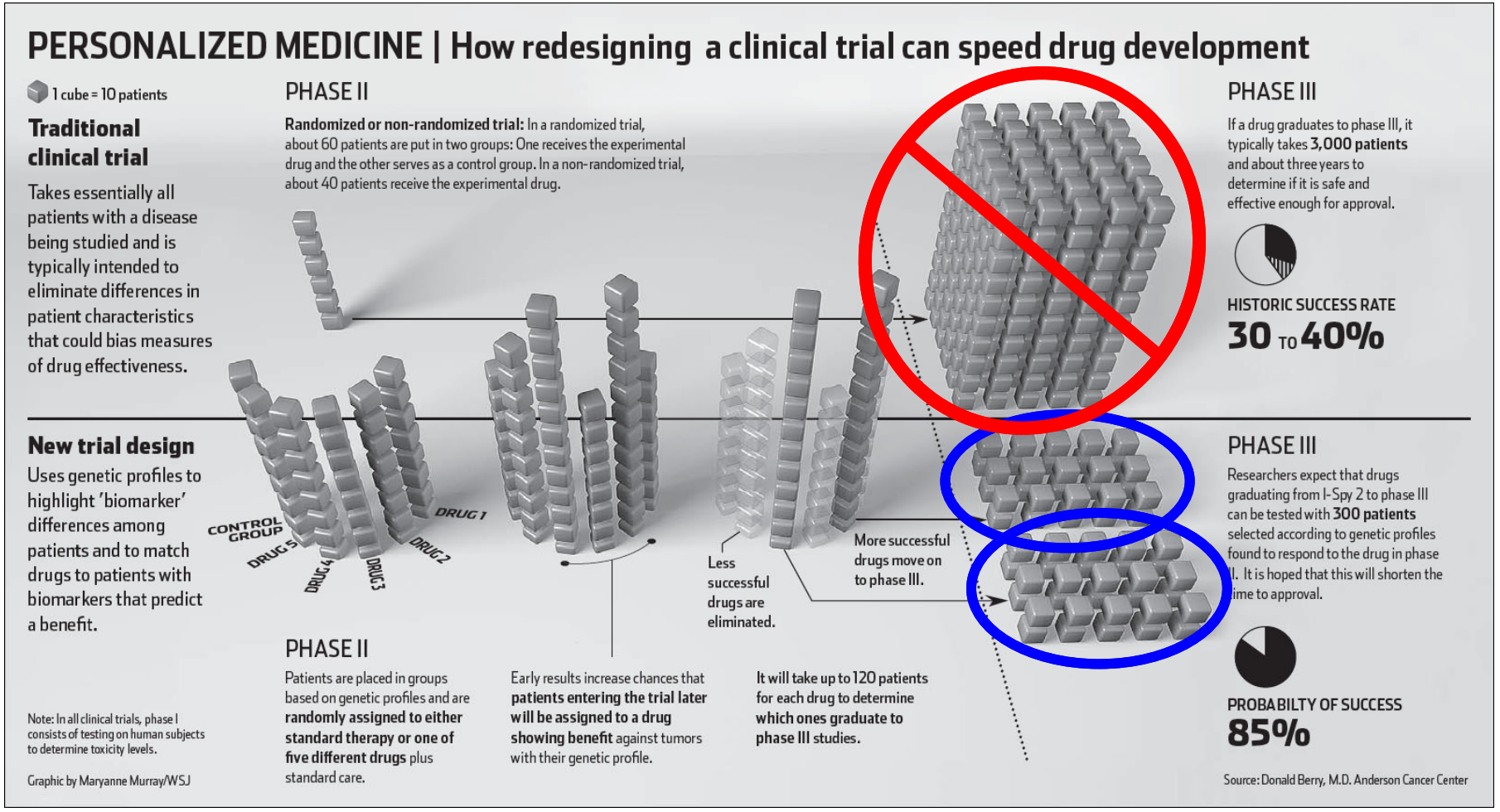
THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCTOBER 2, 2010

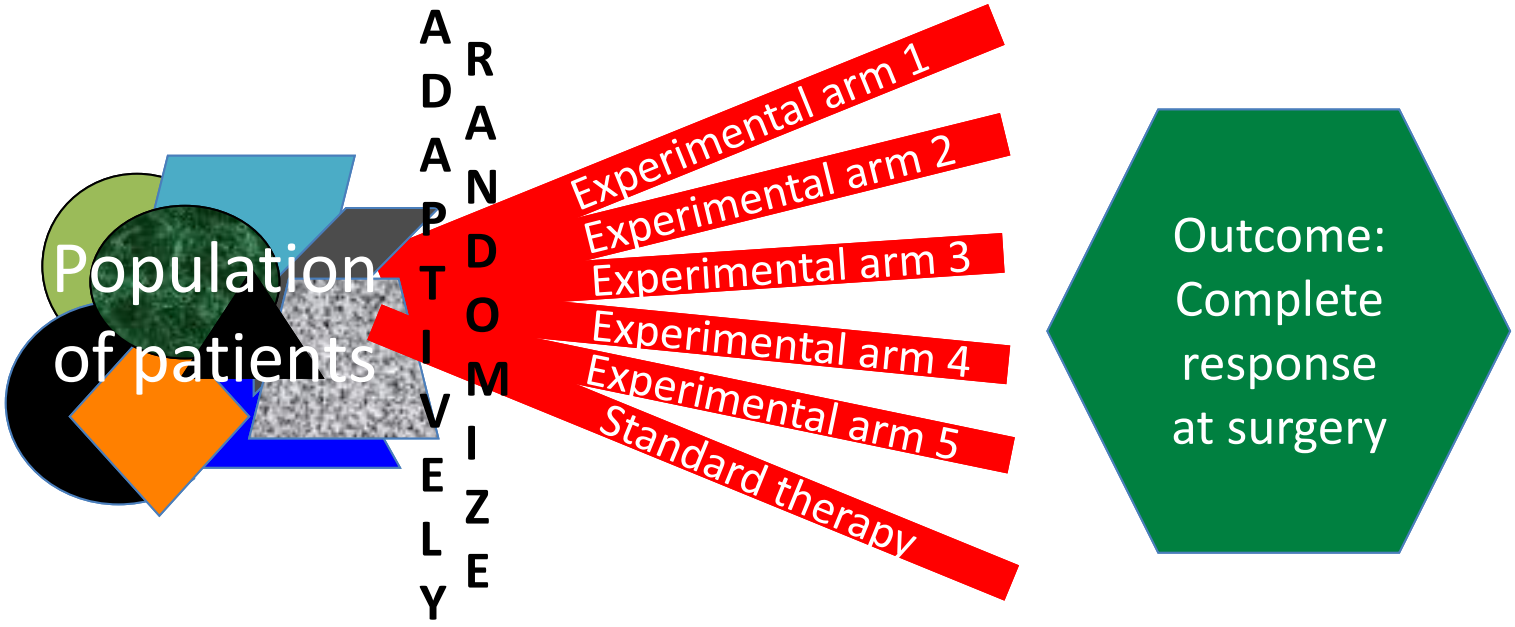
A New Rx for Medicine

Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSLOW



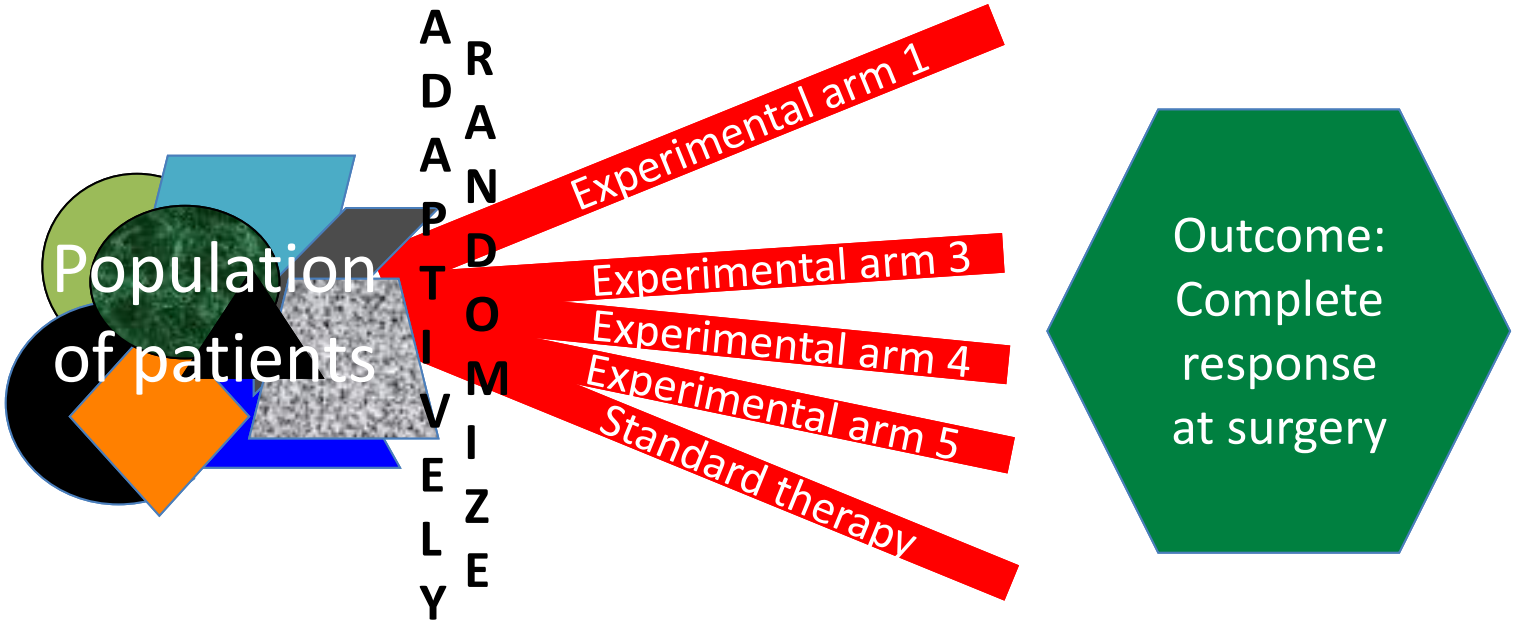
I-SPY2 TRIAL


Population of patients
ADAPTIVELY
RANDOMIZE
Experimental arm 1
Experimental arm 2
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy
Outcome: Complete response at surgery

I-SPY2 TRIAL

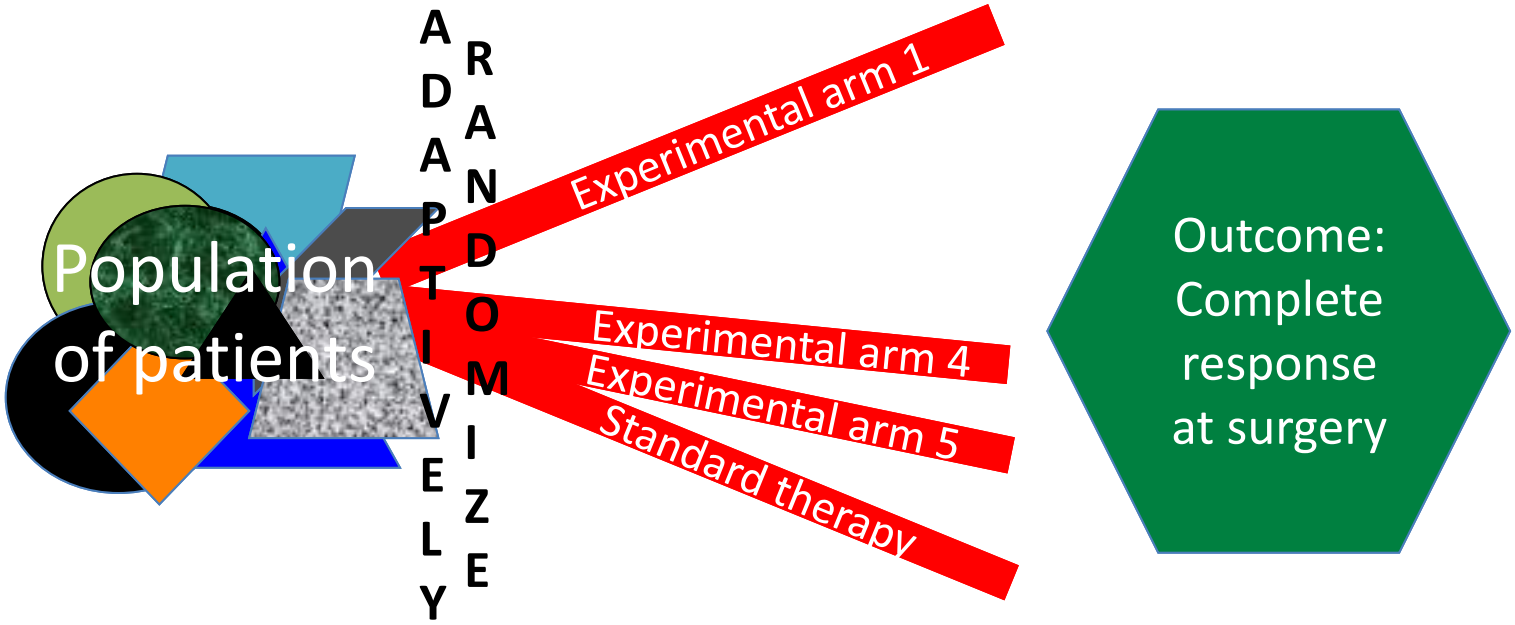
Arm 2 graduates to small focused Phase 3 trial

I-SPY2 TRIAL



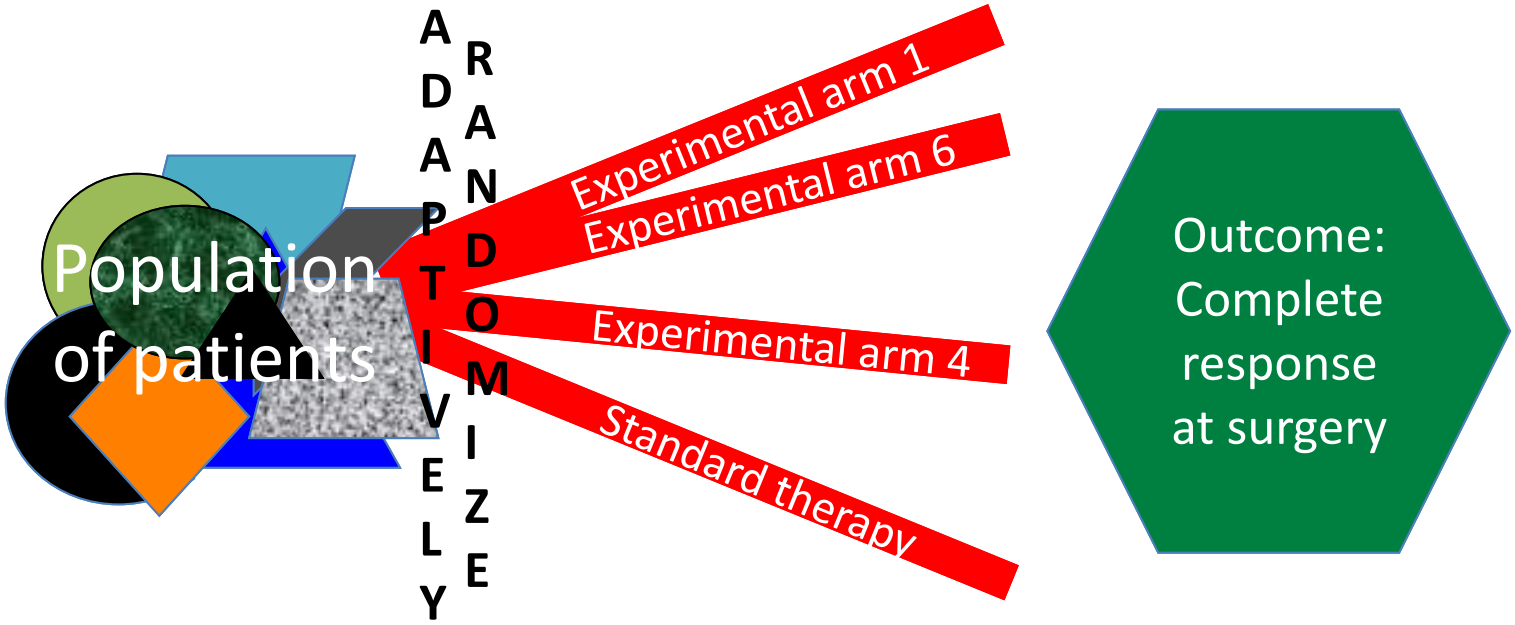
Arm 3 drops for futility

I-SPY2 TRIAL



Arm 5 graduates to small focused Phase 3 trial

I-SPY2 TRIAL



Arm 6 is added to the mix

Infrastructure Considerations for Biomarker-Based Trials

- Adaptive randomization is highly dependent on near instantaneous synchronized data
- Research IT
  - Significant IT infrastructure is required to support biology-based risk-stratified or adaptive designs
  - Expensive, but there may be some economies of scale by establishing a single center to coordinate
- Repurposing
  - Design facilitates repurposing data and supporting future CER

Summary

- Do not underestimate the 90/10 rule
- You never want to visit a biostatistician for the first time with an already collected set of data
  - Same thing here, plan out your data requirements and plan BEFORE you start your study
  - Multidisciplinary team:
    - Biostatistician/epidemiologist
    - Research IT / informatician
    - Data management personnel
    - Regulatory personnel
- Comprehensive and thoughtful database design is key

Summary

- Comprehensive and thoughtful database design is key:
  - Database content and documentation
  - Software
  - Hardware
- Consider capability as well as sustainability over the long-haul in how you develop your data management plan:
  - Future proof as much as possible
  - Stick to industry standards as much as possible
  - Consider future regulatory issues

Summary (continued)

- I think that informatics/research IT should be core competencies in clinical and translational research.
- Computational technologies for managing data are changing faster than technologies for analysis.
- Good data management → High quality data
- High quality data → Analytic quality