Data Management: Why You Should Care and What You Should Know

June 19, 2013

Christine E. Chaisson
Data Coordinating Center, BUSPH
Clinical and Translational Sciences Institute (CTSI), BUMC

Boston University School of Public Health
Data Coordinating Center

Why Worry About the Data?

At the end of the study, after:

- Interviews
- Clinical assessments
- Lab tests
- Other data elements...

All you have is the data

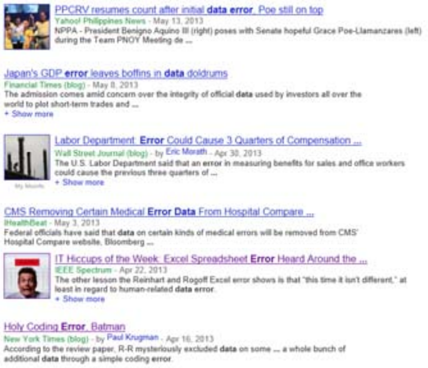
Things that can go wrong

- Data may not be saved or backed up
- Crucial data elements may be missing
- Data may be incorrect due to
  - Data collection errors
  - Data entry errors
- Data may be incorrectly identified
  - Cannot be merged
  - May be merged incorrectly
- Data files may be lost or corrupted

Real World Examples

- A few illustrations from the popular news sources

A Google Search



PPCRV resumes count after initial **data error**, Poe still on top
Yahoo! Philippines News - May 13, 2013
NPPA - President Benigno Aquino III (right) poses with Senate hopeful Grace Poe-Llamanzares (left) during the Team PNOY Meeting de ...

Japan's GDP **error** leaves boffins in **data** doldrums
Financial Times (blog) - May 8, 2013
The admission comes amid concern over the integrity of official **data** used by investors all over the world to plot short-term trades and ...
+ Show more

Labor Department: **Error** Could Cause 3 Quarters of Compensation ...
Wall Street Journal (blog) - by Eric Morath - Apr 30, 2013
The U.S. Labor Department said that an **error** in measuring benefits for sales and office workers could cause the previous three quarters of ...
+ Show more

CMS Removing Certain Medical **Error Data** From Hospital Compare ...
iHealthBeat - May 3, 2013
Federal officials have said that **data** on certain kinds of medical errors will be removed from CMS' Hospital Compare website, Bloomberg ...

IT Hiccups of the Week: Excel Spreadsheet **Error** Heard Around the ...
IEEE Spectrum - Apr 22, 2013
The other lesson the Reinhart and Rogoff Excel error shows is that "this time it isn't different," at least in regard to human-related **data error**.
+ Show more

Holy Coding **Error**, Batman
New York Times (blog) - by Paul Krugman - Apr 16, 2013
According to the review paper, R-R mysteriously excluded **data** on some ... a whole bunch of additional **data** through a simple coding **error**.

Data error gives Wisconsin justice big lead

- Waukesha County apologized for the error that 14,000 votes were not reported. The clerk explained that she had imported vote totals transmitted by the city of Brookfield but *had not saved the data.*

Data not Saved

philly.com

August 3, 2012

Forbes: Bad data hurt Haverford in college rankings

“Forbes' annual list is out, and Haverford plummeted from No. 7 to No. 27 - for no obvious reason. A College spokesman explained that the error was based on single figure:
A zero was incorrectly entered in database instead of 108 for the graduation rate of white women who enrolled in 2004.
…But no revision is planned, since the magazine and the online list has already been published.”

Data Entry Error

Oops: Excel Error Calls Into Question…

IEEE SPECTRUM

Posted 22 Apr 2013

- “Serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period” were recently identified ‘*This Time It’s Differen,t’ a* 2009 book by Harvard researchers
- The Authors admitted they forgot to include five rows in an Excel file resulting in exclusion of data from Australia, Austria, Belgium, Canada, and Denmark —a “coding error” which they said was “a significant lapse on our part.”

excluded key data

PharmaTimes ONLINE

May 6, 2012

Vertex stock slides over cystic fibrosis data mistake

“Shares in Vertex Pharmaceuticals have taken a hit after the company had to take the rather embarrassing step of correcting previously-announced interim mid-stage results of a combination cystic fibrosis treatment.
…the result of a misinterpretation [of the denominator of the treatment group] between the firm and its outside statistical ve

Data Mismanaged

The New York Times July 7, 2011

How Bright Promise in Cancer Testing Fell Apart

- Duke Cancer Center’s gene-based tests proved worthless, research behind them was discredited
- Statisticians from MD Anderson discovered errors such as columns moved over in a spread-sheet which Duke team “shrugged them off” as “clerical errors.”
- Four papers were retracted
- Duke shut down three trials
- Center leaders resigned or were removed.
- People died and relatives sued Duke

Data Entry/Management in Excel

Why Plan for Data Management?

- Standardized procedures
- Appropriate systems for
  - Data collection
  - Monitoring/auditing
  - Tracking
- Lead to higher quality data
- More reliable conclusions

With a Successful *Data Management System:*

Data will be:

- Complete
- Accurate
- Timely
- Answer the scientific questions

What you should consider when budgeting:

- Data entry if using paper forms
- Software (if applicable)
- Personnel (FTE) may include:
  - Data manager
  - SAS programmer
  - Web/database programmer
- Tasks include
  - Data cleaning and hecking
  - Reporting
  - Auditing

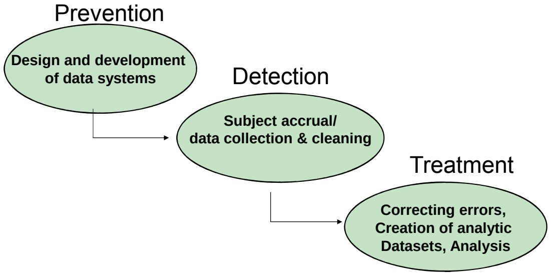
Goal: Convert Data into Electronic Format

- Get the data from report (CRFs) double-entered as soon as possible – don't wait until the end of the study!
- Consider direct electronic capture of data
- Decide what you're going to do before you start
- Make sure you have someone on your team that can handle the data tasks (preferably not a work study student that will change every semester)

Data Management 101:

- More than one approach to managing data.
- Consider the:
  - Environment
  - Available resources
- "Requirements analysis"
  - ***Planning*** prior to beginning the study
- Do what works for the study at hand

Data Management Stages Using a Public Health Model


Prevention
Design and development of data systems
Detection
Subject accrual/ data collection & cleaning
Treatment
Correcting errors, Creation of analytic Datasets, Analysis

Good Systems Prevent Errors

Begin with:

- Identification of tasks and timelines
- Written, standardized procedures
- Well designed data collection forms
- Staff training programs
- Documentation of systems
- Plans for monitoring data

Where to Begin?

- □ Budget appropriately
- □ Finalize protocol and analysis plan
- □ Determine type/frequency of data
  - Case report forms (CRFs)
  - Biologic samples, X-Rays
- □ Electronic Data Capture vs. paper
- □ Determine key people and roles
- □ Establish timelines (work backwards)

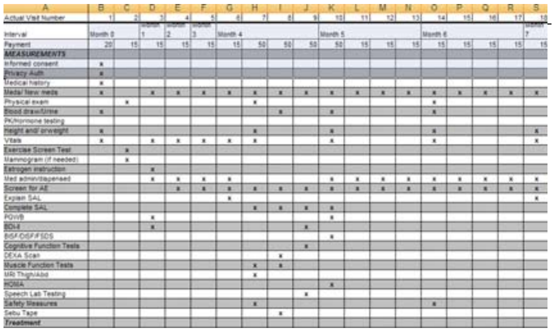
Example: Example Reverse Time-line

- □ June 2013: Study Enrollment will begin
- □ May 2013: Final staff training/Investigator meeting
- □ April 2013: Finalize systems/pilot/testing
- □ Jan-March 2013**:** Construct data systems
- □ December 2012: Finalize assessment
- □ December 2012: Final visit protocol
- □ 11/12: Other final decisions

Visit Protocol: Data by Time-point

- □ Determine Visit Schedule and type (e.g., semi-annual in-person, 3 and 9 month phone)
- □ Determine data collected at each visit
  - Questionnaires
  - Labs
  - Adverse Events
  - Other data elements?
- □ Consider data that may not be connected to a time-point (hospitalization, death)
- □ Finalize the windows around time-points and how will data be associated if applicable

Sample Visit Grid



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Actual Visit Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Interval	Month 0		Month 1	Month 2	Month 3	Month 4				Month 5				Month 6				Month 7
Payment	20	15	15	15	15	15	50	50	50	50	15	15	15	15	15	15	15	15
MEASUREMENTS																		
Informed consent	x																	
Privacy Auth	x																	
Medical history	x																	
Medal New meds	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Physical exam		x					x							x				
Blood draw/Urine	x							x		x				x				
PK/hormone testing																		
Height and/ or weight	x						x			x				x				x
Vitals	x		x	x	x	x	x			x				x				x
Exercise Screen Test		x																
Mammogram (if needed)		x																
Estrogen instruction			x															
Med admin/dispensed			x	x	x	x				x	x	x	x	x	x	x	x	x
Screen for AE				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Explain SAL						x												x
Complete SAL							x	x	x	x								
POWB			x							x								
BDI-II			x						x									
BSF/DSF/FSDS										x								
Cognitive Function Tests									x									
DEXA Scan								x										
Muscle Function Tests							x	x										
MRI Thigh/Abd							x											
HOMA										x								
Speech Lab Testing									x									
Safety Measures							x							x				
Sebu Tape								x										
Treatment																		

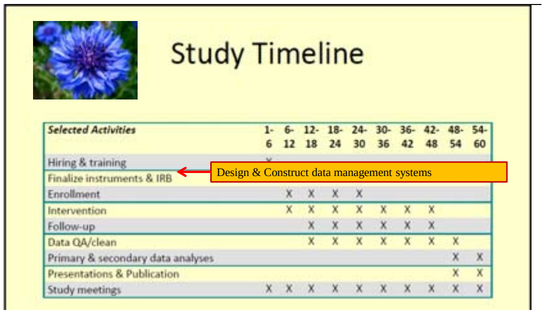
Timelines and Tasks

- □ Development of Protocol, analytic plan
- □ Creation & piloting of forms
- □ Design/construction of data entry and participant/data tracking systems
- □ Development of Manual of Operations
- □ Subject recruitment
- □ Data collection & follow-up
- □ Data cleaning, auditing, QA
- □ Analysis
- □ Manuscript preparation & submission

Create a Visual Timeline

- □ It doesn’t have to be fancy
- □ More detail is better but something simple is better than nothing
- □ Plan to review and revise it often

Simple overview Timeline

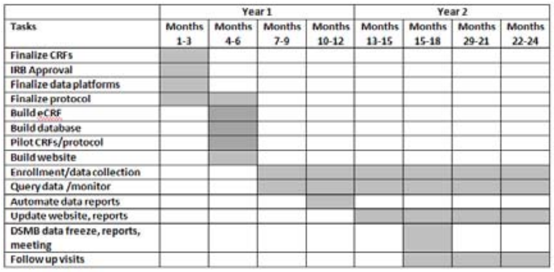


Study Timeline

Selected Activities	1-6	6-12	12-18	18-24	24-30	30-36	36-42	42-48	48-54	54-60
Hiring & training	X									
Finalize instruments & IRB										
Enrollment		X	X	X	X					
Intervention		X	X	X	X	X	X	X		
Follow-up			X	X	X	X	X	X		
Data QA/clean			X	X	X	X	X	X	X	
Primary & secondary data analyses									X	X
Presentations & Publication									X	X
Study meetings	X	X	X	X	X	X	X	X	X	X

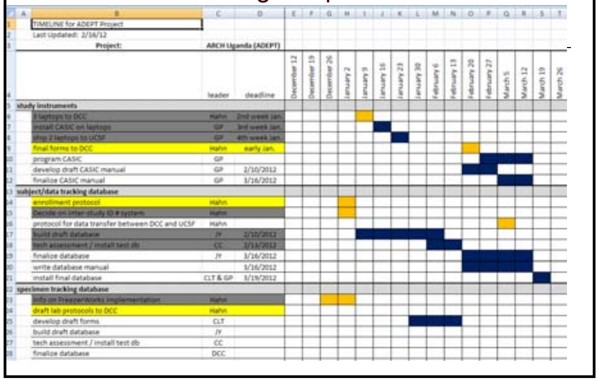
Design & Construct data management systems

Sample Task-based Gantt



	Year 1				Year 2			
Tasks	Months 1-3	Months 4-6	Months 7-9	Months 10-12	Months 13-15	Months 15-18	Months 29-21	Months 22-24
Finalize CRFs	▨							
IRB Approval	▨							
Finalize data platforms	▨							
Finalize protocol	▨	▨						
Build eCRF		▨						
Build database		▨						
Pilot CRFs/protocol		▨						
Build website		▨						
Enrollment/data collection			▨	▨	▨	▨	▨	▨
Query data /monitor			▨	▨	▨	▨	▨	▨
Automate data reports				▨				
Update website, reports					▨	▨	▨	▨
DSMB data freeze, reports, meeting						▨		
Follow up visits						▨	▨	▨

Multi-task Indicating Responsible Parties



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		TIMELINE for ADEPT Project																		
2		Last Updated: 2/16/12																		
3		Project:	ARCH Uganda (ADEPT)																	
4			leader	deadline	December 12	December 19	December 26	January 2	January 9	January 16	January 23	January 30	February 6	February 13	February 20	February 27	March 5	March 12	March 19	March 26
5	study instruments																			
6		3 laptops to DCC	Hahn	2nd week Jan.					🟧											
7		install CASIC on laptops	GP	3rd week Jan.						🟦										
8		ship 2 laptops to UCSF	GP	4th week Jan.							🟦									
9		final forms to DCC	Hahn	early Jan.											🟧					
10		program CASIC	GP													🟦	🟦	🟦		
11		develop draft CASIC manual	GP	2/10/2012											🟦	🟦				
12		finalize CASIC manual	GP	3/16/2012													🟦	🟦		
13	subject/data tracking database																			
14		enrollment protocol	Hahn					🟧												
15		Decide on inter-study ID # system	Hahn					🟧												
16		protocol for data transfer between DCC and UCSF	Hahn														🟧			
17		build draft database	JY	2/10/2012					🟦	🟦	🟦	🟦	🟦							
18		tech assessment / install test db	CC	3/13/2012									🟦	🟦						
19		finalize database	JY	3/16/2012											🟦	🟦	🟦	🟦		
20		write database manual		3/16/2012											🟦	🟦	🟦	🟦		
21		install final database	CLT & GP	3/19/2012															🟦	
22	specimen tracking database																			
23		info on FreezerWorks implementation	Hahn				🟧	🟧												
24		draft lab protocols to DCC	Hahn																	
25		develop draft forms	CLT									🟦	🟦	🟦						
26		build draft database	JY																	
27		tech assessment / install test db	CC																	
28		finalize database	DCC																	

Tools Of The Trade

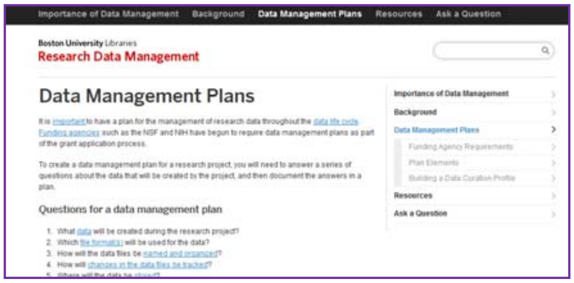
- □ Well designed forms
- □ Data management plan
- □ Tracking system / tracking database
- □ Data Capture System
- □ Database
- □ Study manuals
- □ Data query system

Data Management Plan*

* Required by many funders

- □ Outlines how data will be handled – should include descriptions of:
  - How data will be collected
  - How data will be processed (software, procedures)
  - How and where data will be stored (including software, formats, coding)
  - What QC the procedures will be and the schedule for carrying them out
  - Plan for long-term storage or archiving
  - Annotated forms

Information about DM Plans



Boston University Libraries
Research Data Management

Data Management Plans

It is important to have a plan for the management of research data throughout the data life cycle. Funding agencies such as the NSF and NIH have begun to require data management plans as part of the grant application process.

To create a data management plan for a research project, you will need to answer a series of questions about the data that will be created by the project, and then document the answers in a plan.

Questions for a data management plan

1. What data will be created during the research project?
2. Which file format(s) will be used for the data?
3. How will the data files be named and organized?
4. How will changes in the data files be tracked?
5. Where will the data be stored?

Importance of Data Management
Background
Data Management Plans
Funding Agency Requirements
Plan Elements
Building a Data Curation Profile
Resources
Ask a Question

Web Site User Manual

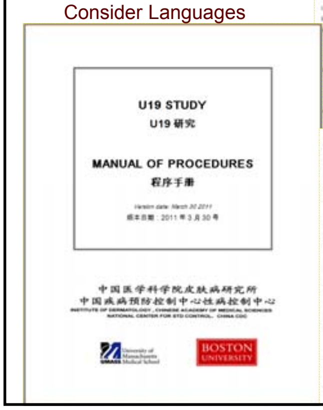
Version date: August 24, 2010

https://dccweb.bumc.bu.edu/Sizanani/

Sizanani
Login

TABLE OF CONTENTS

Consider Languages



U19 STUDY
U19 研究

MANUAL OF PROCEDURES
程序手册

Version date: March 30 2011
版本日期：2011 年 3 月 30 号

中国医学科学院皮肤病研究所
中国疾病预防控制中心性病控制中心
INSTITUTE OF DERMATOLOGY, CHINESE ACADEMY OF MEDICAL SCIENCES
NATIONAL CENTER FOR STD CONTROL, CHINA CDC

TABLE OF CONTENTS (目录)

I. INTRODUCTION 简介

The objective of the Manual of Procedures is to provide an easy-to-follow, step-by-step, comprehensive guide to assist you in data collection, data entry, and the maintenance of study subject data.

[illegible]

II. CONTACT INFORMATION 联系方式

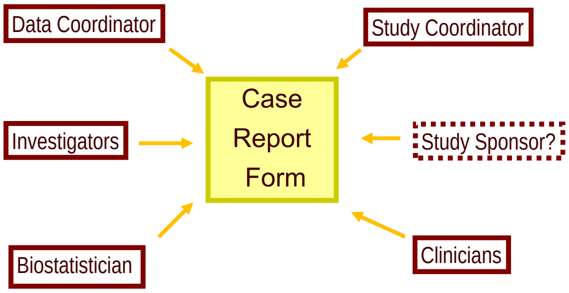
For questions regarding the website or study data, please contact any of the following people:

[illegible]

Data Coordinating Center, Boston, MA USA

[illegible]

Creating a Data Collection Form

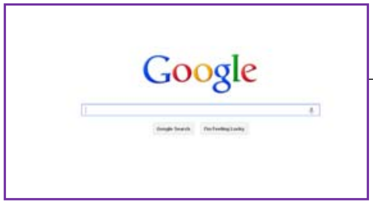


Why is this topic important?

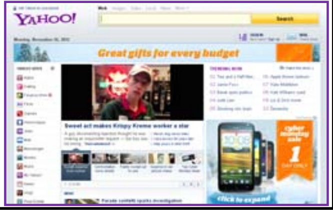
- ❑ Sloppy forms indicate sloppy research
- ❑ CRF doesn't answer study questions
- ❑ Danger of collecting:
  - ❑ too much data
  - ❑ too little data
  - ❑ the wrong data
- ❑ Annoyed:
  - ❑ Study Coordinator
  - ❑ Participants
  - ❑ Analyst…

Successful Form: Consider ALL Functions

- ❑ Data Collection - who is completing form?
  - ❑ Study Staff (Coordinator, Clinician)
  - ❑ Participant
  - ❑ Clinician
- ❑ Data entry - who is entering data?
  - ❑ Study staff
  - ❑ Students
  - ❑ Outsourced
- ❑ Data management/cleaning
- ❑ Data analysis

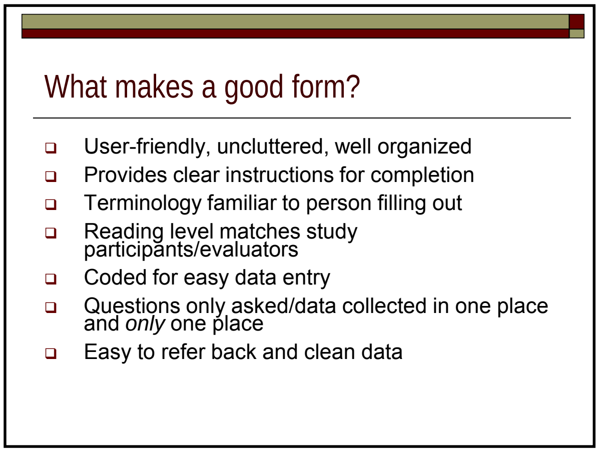


Think Google



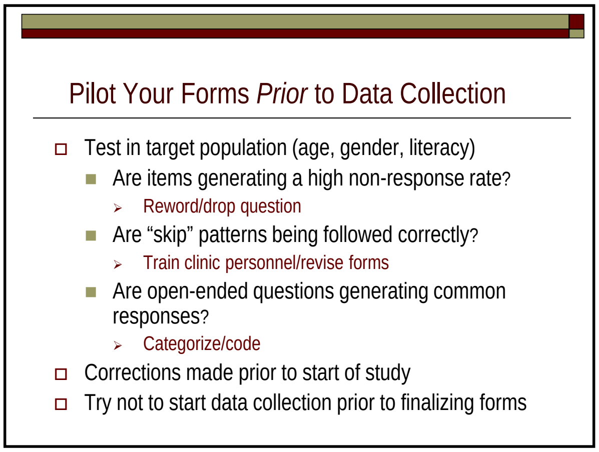
YAHOO!
Search
Great gifts for every budget
Sweet act makes Krispy Kreme worker a star
click to expand

Not Yahoo



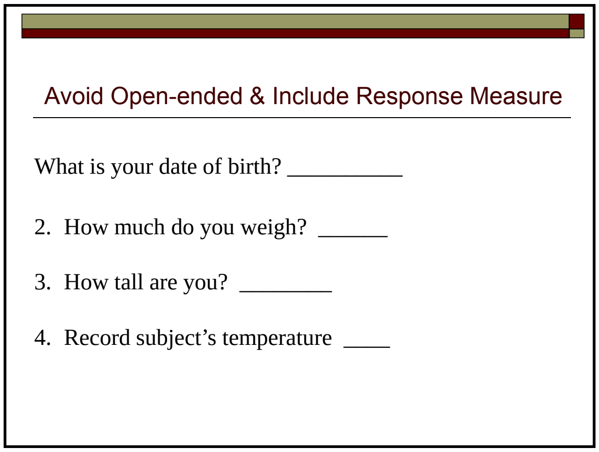
What makes a good form?

- ❑ User-friendly, uncluttered, well organized
- ❑ Provides clear instructions for completion
- ❑ Terminology familiar to person filling out
- ❑ Reading level matches study participants/evaluators
- ❑ Coded for easy data entry
- ❑ Questions only asked/data collected in one place and *only* one place
- ❑ Easy to refer back and clean data



Pilot Your Forms *Prior* to Data Collection

- □ Test in target population (age, gender, literacy)
  - Are items generating a high non-response rate?
    - ➢ Reword/drop question
  - Are “skip” patterns being followed correctly?
    - ➢ Train clinic personnel/revise forms
  - Are open-ended questions generating common responses?
    - ➢ Categorize/code
- □ Corrections made prior to start of study
- □ Try not to start data collection prior to finalizing forms



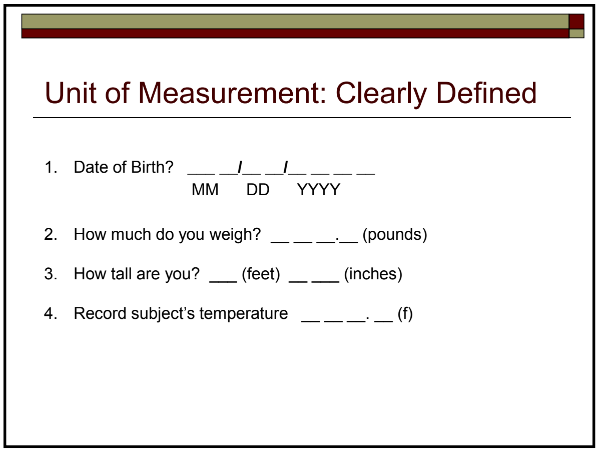
Avoid Open-ended & Include Response Measure

What is your date of birth? __________

2. How much do you weigh? ______

3. How tall are you? ________

4. Record subject’s temperature ____



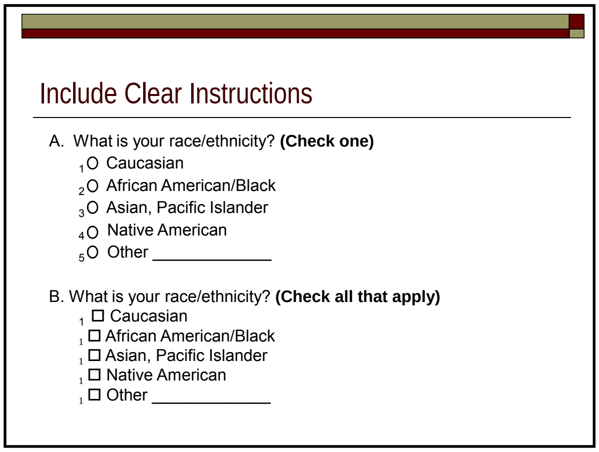
Unit of Measurement: Clearly Defined

1. Date of Birth? ___ __/__ __/__ __ __ __
   MM DD YYYY

2. How much do you weigh? __ __ __.__ (pounds)

3. How tall are you? ___ (feet) __ ___ (inches)

4. Record subject’s temperature __ __ __. __ (f)



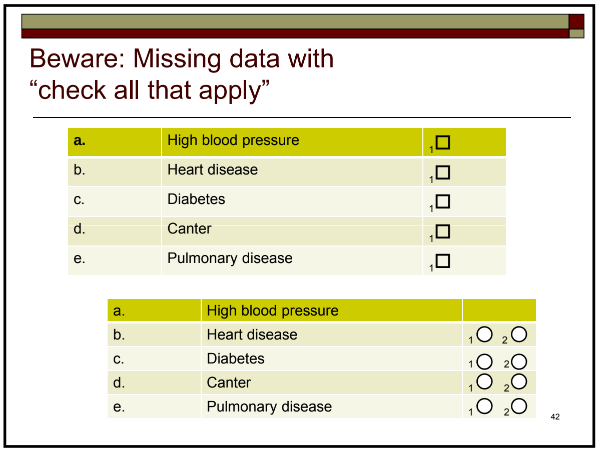
Include Clear Instructions

A. What is your race/ethnicity? **(Check one)**

$_1$○ Caucasian
$_2$○ African American/Black
$_3$○ Asian, Pacific Islander
$_4$○ Native American
$_5$○ Other ____________

B. What is your race/ethnicity? **(Check all that apply)**

$_1$ □ Caucasian
$_1$ □ African American/Black
$_1$ □ Asian, Pacific Islander
$_1$ □ Native American
$_1$ □ Other ____________



Beware: Missing data with “check all that apply”

a.	High blood pressure	$_1$□
b.	Heart disease	$_1$□
c.	Diabetes	$_1$□
d.	Canter	$_1$□
e.	Pulmonary disease	$_1$□

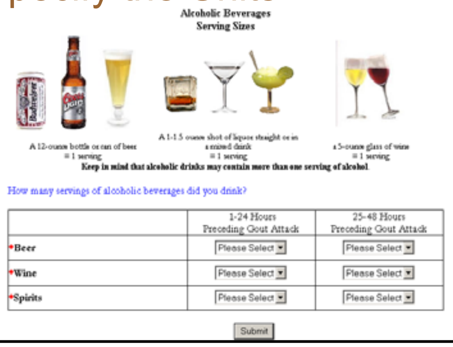
a.	High blood pressure	
b.	Heart disease	$_1$○ $_2$○
c.	Diabetes	$_1$○ $_2$○
d.	Canter	$_1$○ $_2$○
e.	Pulmonary disease	$_1$○ $_2$○

42

Account For Missing Data

CBC	Unit	Value	
1. Hemoglobin	g/dl	__ __.__	□ Not Done
2. Hematocrit	%	__ __.__	□ Not Done
3. RBC	M/mm^3	__ __.__	□ Not Done

Specify the Units



Alcoholic Beverages
Serving Sizes

A 12-ounce bottle or can of beer
= 1 serving

A 1-1.5 ounce shot of liquor straight or in
a mixed drink
= 1 serving

a 5-ounce glass of wine
= 1 serving

Keep in mind that alcoholic drinks may contain more than one serving of alcohol.

How many servings of alcoholic beverages did you drink?

	1-24 Hours Preceding Gout Attack	25-48 Hours Preceding Gout Attack
***Beer**	Please Select ▾	Please Select ▾
***Wine**	Please Select ▾	Please Select ▾
***Spirits**	Please Select ▾	Please Select ▾

Submit

ID Assignment

- Should appear on every form (preferably page)
  - Links paper form with specific record in database
  - Multiple forms, “merge key” in database
- Must be UNIQUE for each subject
- May be a simple number 1001
- May be multi-part: 102101
  - 1 = Site
  - 02 = Language
  - 101= ID

Inclusion of “Other (specify)”

- May cut down on items left blank
- Position “Other” last in list of possible responses to ensure all responses considered first
- Continue to monitor “Other” to ensure a common response category was not overlooked

Anticipated Responses Categorized

1. In what country were you born (check one)?

□$_1$ USA

□$_2$ Guatemala

□$_3$ Mexico

□$_4$ Dominican Republic

□$_5$ Other

Updated Due to Overwhelming Response

1. In what country were you born (check one)?

□$_1$ USA

□$_2$ Guatemala

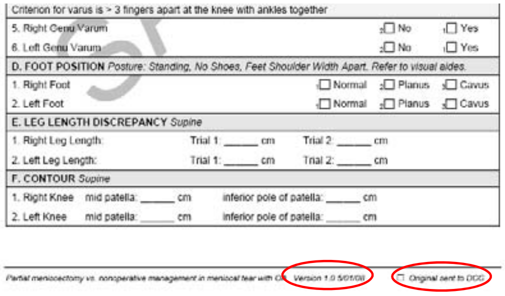
□$_3$ Mexico

□$_4$ Dominican Republic

→ □$_6$ El Salvador

□$_5$ Other

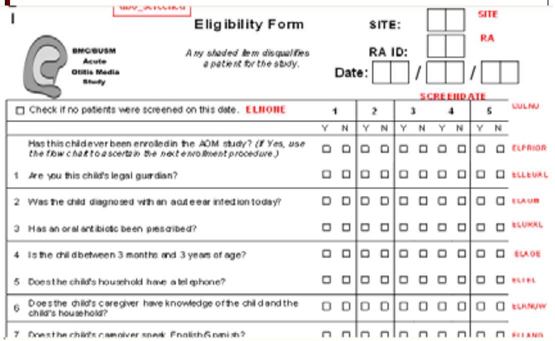
Don’t Underestimate Need for Version# /Date



Criterion for varus is > 3 fingers apart at the knee with ankles together		
5. Right Genu Varum	2☐ No	1☐ Yes
6. Left Genu Varum	2☐ No	1☐ Yes
D. FOOT POSITION *Posture: Standing, No Shoes, Feet Shoulder Width Apart. Refer to visual aides.*		
1. Right Foot	1☐ Normal 2☐ Planus	3☐ Cavus
2. Left Foot	1☐ Normal 2☐ Planus	3☐ Cavus
E. LEG LENGTH DISCREPANCY *Supine*		
1. Right Leg Length:	Trial 1: ______ cm	Trial 2: ______ cm
2. Left Leg Length:	Trial 1: ______ cm	Trial 2: ______ cm
F. CONTOUR *Supine*		
1. Right Knee mid patella: ______ cm	inferior pole of patella: ______ cm	
2. Left Knee mid patella: ______ cm	inferior pole of patella: ______ cm	

Partial meniscectomy vs. nonoperative management in meniscal tear with OA *Version 1.0 5/01/08* ☐ *Original sent to DCC*

Annotate your forms



Eligibility Form

BMC/BUSM Acute Otitis Media Study

Any shaded item disqualifies a patient for the study.

SITE: ☐☐ SITE

RA ID: ☐☐ RA

Date: ☐☐ / ☐☐ / ☐☐

SCREENDATE

	1		2		3		4		5		COLNO
☐ Check if no patients were screened on this date. ELNONE	Y	N	Y	N	Y	N	Y	N	Y	N	
Has this child ever been enrolled in the AOM study? *(If Yes, use the flow chart to ascertain the next enrollment procedure.)*	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELPRIOR
1 Are you this child's legal guardian?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELLEGAL
2 Was the child diagnosed with an acute ear infection today?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELAOM
3 Has an oral antibiotic been prescribed?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELORAL
4 Is the child between 3 months and 3 years of age?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELAGE
5 Does the child's household have a telephone?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELTEL
6 Does the child's caregiver have knowledge of the child and the child's household?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELKNOW
7 Does the child's caregiver speak English/Spanish?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ELLANG

In Summary, when designing questions:

- Avoid open ended responses
- Determine to whether question should be collected as “continuous” or “categorical”
- Consider all possible responses
- Make categories mutually exclusive
- Allow for unanticipated responses
- Put ID on every form/page
- Pilot your forms in the target population

Once you know what to collect…

- Decide how it will be collected
  - Paper
  - Electronic
  - Both
- If electronic, how?
- Who will:
  - Enter data
  - Handle data

Data Collection: Paper Or Paperless?

METHOD	Scan/ FAX	CAPI* (laptop /tablet)	Web	Kiosk/ Touch Screen	Hand Held/ Smart-phone	ACASI† (Audio)
PAPER	☑	☑	☑			
PAPERLESS		☑	☑	☑	☑	☑

*Computer Assisted Personal Interview
† Audio Computer Assisted Self Interview

Paper Forms / Manual Entry

Advantages

- The “standard”
- Shorter start-up time
- Relatively easy to train staff
- Hardcopy document to refer back to
- Can be done anywhere

Disadvantages

- Longer time to inclusion in database
- Errors in data collection (missing, out of range, skips)
- Data entry/shipping costly for large studies

Electronic Data Capture

Advantages

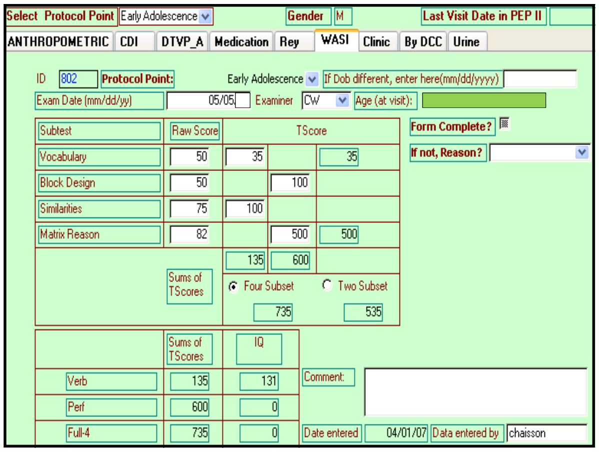
- Cleaner data at entry (required fields, skips, ranges)
- More efficient for larger studies
- Electronic data in real time (or close to it)
- Eliminate data entry/shipping costs
- Data can inform next visit even for short follow up

Disadvantages

- May entail increased upfront programming costs
- Additional training of staff
- Increased equipment costs
- Infrastructure (software versions, internet connection, back-up equipment)
- Data security

Types of electronic data capture

- Local device data capture (data saved on local machine/tablet)
- Web-based data collection (data saved on central server)
  - Desktop computer
  - Laptop/Tablet/PC
- Handheld devices (HAPI, smart-phones)
- Faxed/scanned data forms system
- Others



Select Protocol Point: Early Adolescence | Gender: M | Last Visit Date in PEP II:

ANTHROPOMETRIC | CDI | DTVP_A | Medication | Rey | WASI | Clinic | By DCC | Urine

ID: 802 | Protocol Point: Early Adolescence | If Dob different, enter here(mm/dd/yyyy):

Exam Date (mm/dd/yy): 05/05 | Examiner: CW | Age (at visit):

Subtest	Raw Score	TScore		
Vocabulary	50	35		35
Block Design	50		100	
Similarities	75	100		
Matrix Reason	82		500	500
		135	600	
	Sums of TScores	● Four Subset 735		○ Two Subset 535

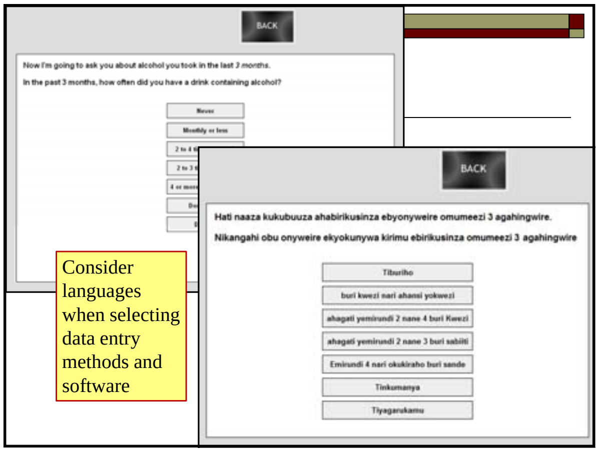
Form Complete? ▨

If not, Reason?

	Sums of TScores	IQ
Verb	135	131
Perf	600	0
Full-4	735	0

Comment:

Date entered: 04/01/07 | Data entered by: chaisson



Now I'm going to ask you about alcohol you took in the last *3 months*.

In the past 3 months, how often did you have a drink containing alcohol?

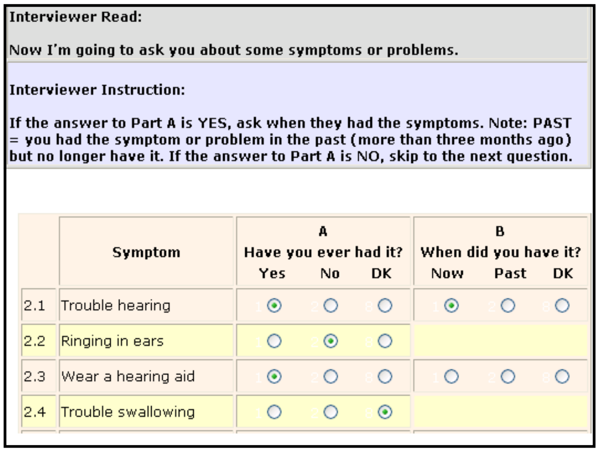
Never
Monthly or less
2 to 4 ti
2 to 3 ti
4 or more
Do
D
BACK

Hati naaza kukubuuza ahabirikusinza ebyonyweire omumeezi 3 agahingwire.

Nikangahi obu onyweire ekyokunywa kirimu ebirikusinza omumeezi 3 agahingwire

Tiburiho
buri kwezi nari ahansi yokwezi
ahagati yemirundi 2 nane 4 buri Kwezi
ahagati yemirundi 2 nane 3 buri sabiiti
Emirundi 4 nari okukiraho buri sande
Tinkumanya
Tiyagarukamu

Consider languages when selecting data entry methods and software



Interviewer Read:

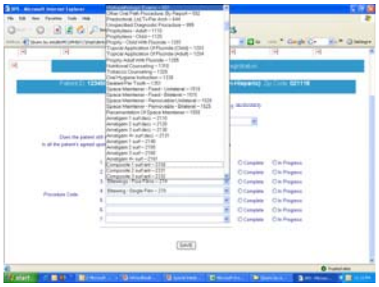
Now I'm going to ask you about some symptoms or problems.

Interviewer Instruction:

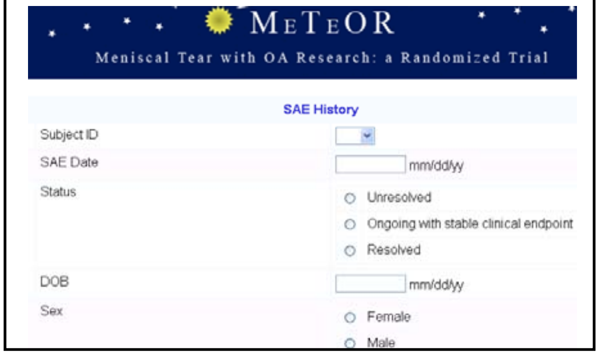
If the answer to Part A is YES, ask when they had the symptoms. Note: PAST = you had the symptom or problem in the past (more than three months ago) but no longer have it. If the answer to Part A is NO, skip to the next question.

	Symptom	A Have you ever had it? Yes	 No	 DK	B When did you have it? Now	 Past	 DK
2.1	Trouble hearing	●	○	○	●	○	○
2.2	Ringing in ears	○	●	○			
2.3	Wear a hearing aid	●	○	○	○	○	○
2.4	Trouble swallowing	○	○	●			

Dropdown Menu for Codes



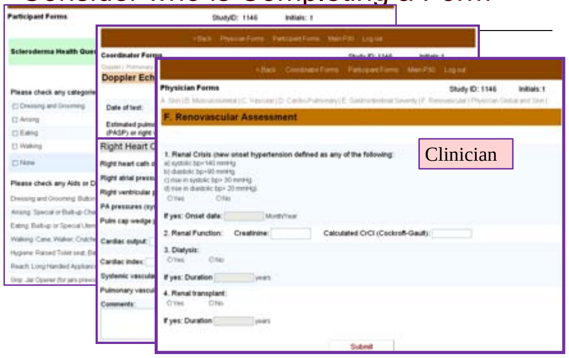
Timely Submission of Urgent Data



SAE History

Subject ID	
SAE Date	mm/dd/yy
Status	○ Unresolved ○ Ongoing with stable clinical endpoint ○ Resolved
DOB	mm/dd/yy
Sex	○ Female ○ Male

Consider who is Completing a Form



Participant Forms StudyID: 1146 Initials: 1

Scleroderma Health Que

Please check any categorie
☐ Dressing and Grooming
☐ Arising
☐ Eating
☐ Walking
☐ None

Please check any Aids or D

Coordinator Forms

Doppler Ech

Date of test:

Estimated pulmo (PASP) or right v

Right Heart C

Right heart cath d

Right atrial pressu

Right ventricular p

PA pressures (sy

Pulm cap wedge p

Cardiac output:

Cardiac index:

Systemic vascula

Pulmonary vascu

Comments:

Physician Forms Study ID: 1146 Initials: 1

F. Renovascular Assessment

1. Renal Crisis (new onset hypertension defined as any of the following:
a) systolic bp>140 mmHg
b) diastolic bp>90 mmHg
c) rise in systolic bp> 30 mmHg
d) rise in diastolic bp> 20 mmHg)
○ Yes ○ No

If yes: Onset date: Month/Year

2. Renal Function: Creatinine: Calculated CrCl (Cockroft-Gault):

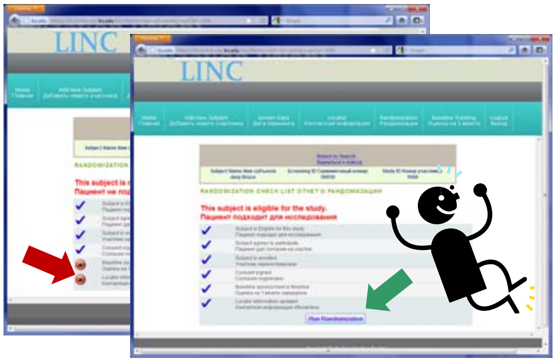
3. Dialysis:
○ Yes ○ No

If yes: Duration years

4. Renal transplant:
○ Yes ○ No

If yes: Duration years

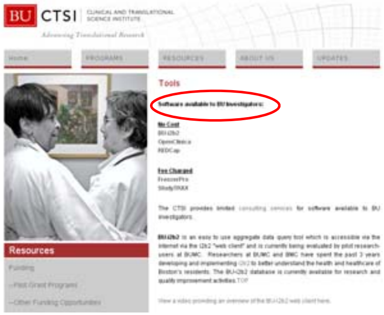
Advantage of Web-based Randomization



RANDOMIZATION CHECK LIST ОТЧЕТ О РАНДОМИЗАЦИИ

This subject is eligible for the study.
Пациент подходит для исследования

Find out what may be available to you


BU CTSI
CLINICAL AND TRANSLATIONAL SCIENCE INSTITUTE
Advancing Translational Research
HOME
PROGRAMS
RESOURCES
ABOUT US
UPDATES
Resources
Funding

Tools

Software available to BU Investigators:

No Cost
BU i2b2
OpenClinica
REDCap

Fee Charged
FreezerPro
StudyTRAX

The CTSI provides limited consulting services for software available to BU investigators.

A Word About “Canned” Software

- There are many types of “canned” or commercially available software available
- No single “best” choice
- Cost can vary widely
- Database structure can vary
- Do your homework to make sure what you get will work for your project

Consider the Database Structure

- Relational database:
  - each *form* is a record/row
  - Can be queried in real-time
  - Better choice for *data management*
- “Long Skinny” file:
  - each *variable* is its own short record/row,
  - difficult or impossible to query in real-time
  - Good choice for *data capture*

Example of “Relational database”

ID	VisDate	Sex	Race
1001	09/09/2011	F	AA
1002	09/08/2011	M	W
1003	09/08/2011	F	AA
…			

“One” row per subject

“Many” rows per subject

Linked by ID

ID	Lesion	LesionSize	Unit
1001	1	6.2	mm
1001	2	4.6	mm
1002	1	2.9	mm
…			

Best structure for real-time querying of data. Can be structured this way in most database packages including: Oracle, SQL, MS Access

Example of “Long Skinny”

ID	Date	VarName	VarType	Value
1001	09/09/2012	ID	Numeric	1001
1001	09/09/2012	VisDate	Date	09092011
1001	09/09/2012	Sex	Numeric	2
1001	09/09/2012	Race	text	AA
1001	09/09/2012	Lesion_mm	numeric	6
…				
1002	09/08/2012	ID	numeric	10021
Etc.				

One row per variable

Many “canned” web software packages use this structure including:
•Survey Monkey, REDCap, StudyTrax

REDCap “Long Skinny”

PI ID	Study ID	Partic. ID	Date	VarName	VarType	Value
12345	44556	1001	09/09/12	ID	Numeric	1001
12345	44556	1001	09/09/12	VisDate	Date	09092011
12345	44556	1001	09/09/12	Sex	Numeric	2
12345	44556	1001	09/09/12	Lesion_1	numeric	6
1234					numeric	10021
…						
12345	77987	201	1/1/13	Site	alpha	bmc
12345	77987	201	1/1/13			
…						
78723	11112	2211	1/15/13	ID	numeric	2211
I34543	22312	FE12	2/2/13	ID	alpha	FE12
…						

Includes Investigator and Study code

One huge table with multiple investigators and studies

Consider the Database Structure

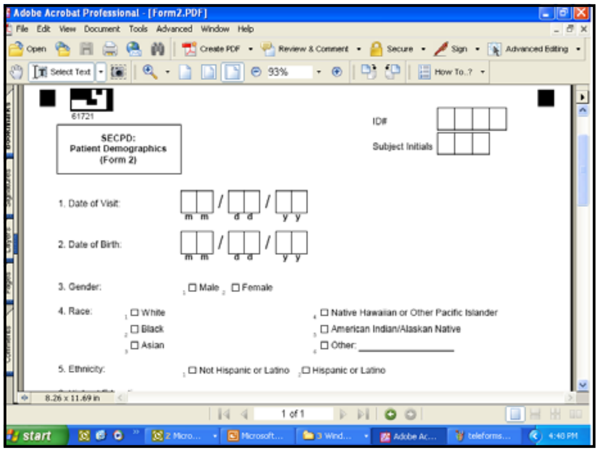
- □ For straight electronic data capture underlying structure may not matter
- □ If you want an “intelligent” e-form with sophisticated checking or custom error and warning messages, database structure does matter

Paper/Electronic Hybrid Systems: Optical Character Recognition software (scan/fax)

- □ Data collected on paper “TELEForm”
- □ Form scanned/uploaded or faxed to processing center
- □ Software “reads” forms and enters data into a database
- □ Questionable characters are set aside for manual review
- □ “Verifier” may be customized for each form
  - ■ Different level can be set for various fields
  - ■ 100% for key fields or hand written fields

Optical Scanning/Faxes

- □ Advantages
  - ■ Don’t need constant internet access
  - ■ Easy to train clinical staff
  - ■ *Relatively* inexpensive
  - ■ Shorter time between data collection and inclusion in database
- □ Disadvantages
  - ■ Software costs, skills
  - ■ Not practical for text or hand written data
  - ■ More sensitive to quality of forms



61721

SECPD:
Patient Demographics
(Form 2)

ID# □□□□

Subject Initials □□□

1. Date of Visit: □□ / □□ / □□ (m m / d d / y y)

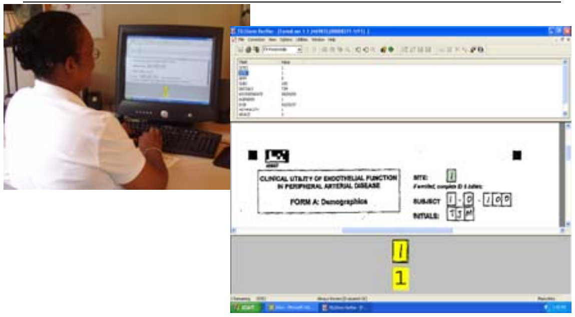
2. Date of Birth: □□ / □□ / □□ (m m / d d / y y)

3. Gender: 1 □ Male 2 □ Female

4. Race: 1 □ White 2 □ Black 3 □ Asian 4 □ Native Hawaiian or Other Pacific Islander 5 □ American Indian/Alaskan Native 6 □ Other: ______________

5. Ethnicity: 1 □ Not Hispanic or Latino 2 □ Hispanic or Latino

TELEForm® Verifying Software



CLINICAL UTILITY OF ENDOTHELIAL FUNCTION IN PERIPHERAL ARTERIAL DISEASE

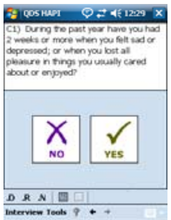
FORM A: Demographics

SITE:

SUBJECT

INITIALS:

Technology Changes: A word of Caution on Handheld Devices



C1) During the past year have you had 2 weeks or more when you felt sad or depressed; or when you lost all pleasure in things you usually cared about or enjoyed?

A Caution About New Technology

Too late, becoming obsolete

Too soon, not ready for prime time

A Word of Caution on Smart-phones

- Encryption can be difficult (or impossible)
- Small screens make it difficult to view some question types
- Navigating around questionnaire (going back) is challenging
- Battery life is short (need to recharge frequently)
- Target for theft

Beginning A Study: Quality Control Systems

Public Health Model:

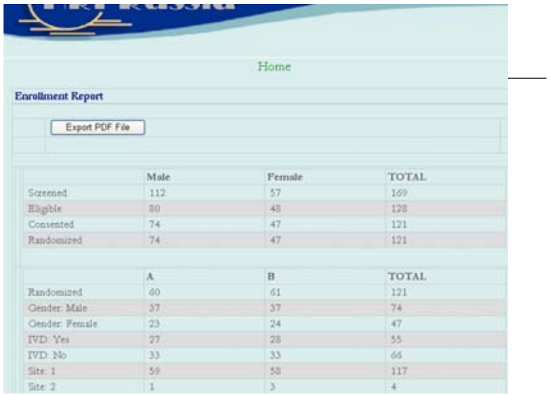
- Prevention
- → Detection
- Treatment

Reports

- For the study team to manage the project
  - Screened, eligible enrolled
  - Key demographics (by randomization group)
  - Follow up rates
- For the study staff to help them manage components (e.g., call lists, follow up visit schedules)
- For Data managers to identify data problems

Tracking: Reports

- Run regular reports of information collected in tracking database
  - Subjects
  - Forms
  - Follow ups
  - Key variables (when applicable)
  - Missing:
    - Visits
    - Forms
    - Data elements

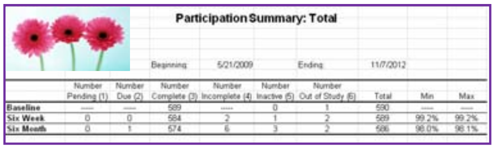


Enrollment Report

	Male	Female	TOTAL
Screened	112	57	169
Eligible	80	48	128
Consented	74	47	121
Randomized	74	47	121

	A	B	TOTAL
Randomized	60	61	121
Gender: Male	37	37	74
Gender: Female	23	24	47
IVD: Yes	27	28	55
IVD: No	33	33	66
Site: 1	59	58	117
Site: 2	1	3	4

Sample Follow Up Report



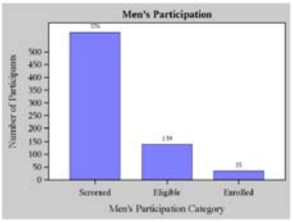
Participation Summary: Total

Beginning: 5/21/2009 Ending: 11/7/2012

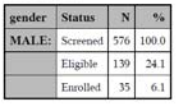
	Number Pending (1)	Number Due (2)	Number Complete (3)	Number Incomplete (4)	Number Inactive (5)	Number Out of Study (6)	Total	Min	Max
Baseline	-----	-----	589	-----	0	1	590	-----	-----
Six Week	0	0	584	2	1	2	589	99.2%	99.2%
Six Month	0	1	574	6	3	2	586	98.0%	98.1%

Reports: Visual as well as Tabular

U19 Study: Enrollment Reports as of 10/15/2012

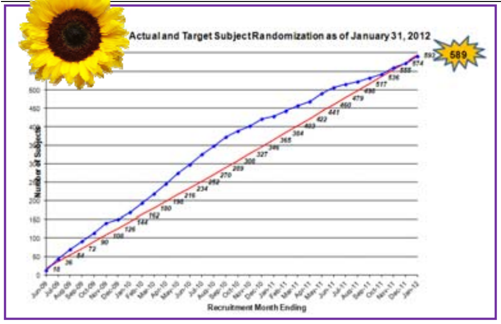


Men's Participation



gender	Status	N	%
MALE:	Screened	576	100.0
	Eligible	139	24.1
	Enrolled	35	6.1

Actual vs. Targeted Enrollment



Tracking the Data

- Identify what data have been collected
  - For each Subject
  - For each Visit
    - Questionnaires
    - Exams, images
    - Labs results, specimen shipping
  - Other data elements

Tracking the data (continued)

Record:

- What data will not be collected
- What data have been received
- What data have been entered

Create:

- Schedules
- Reports

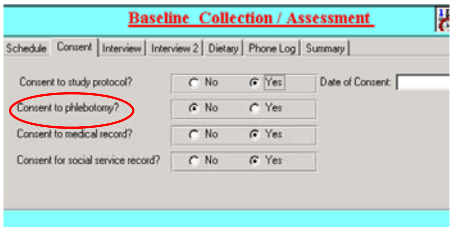
Form Shipment and Receipt

MeTeOR

Meniscal Tear with OA Research: a Randomized Trial

Site	Date	FedEx #	Count	Received
BRIGHAM AND WOMEN'S HOSPITAL	7/24/2008		20	7/25/2008
BRIGHAM AND WOMEN'S HOSPITAL	8/7/2008		14	8/8/2008
BRIGHAM AND WOMEN'S HOSPITAL	8/20/2008	7970-6044-0399	9	8/22/2008
BRIGHAM AND WOMEN'S HOSPITAL	9/5/2008	7970-7899-8378	11	9/9/2008
BRIGHAM AND WOMEN'S HOSPITAL	9/26/2008	7911-4804-6459	24	9/30/2008
BRIGHAM AND WOMEN'S HOSPITAL	10/23/2008	7994- 000-9 46	23	10/27/2008
BRIGHAM AND WOMEN'S HOSPITAL	10/23/2008	7921-3246-2475	16	10/27/2008
BRIGHAM AND WOMEN'S HOSPITAL	11/13/2008	7961-1142-9175	27	11/14/2008

Track what will NOT be done



Baseline Collection / Assessment

Consent to study protocol? ○ No ● Yes Date of Consent:

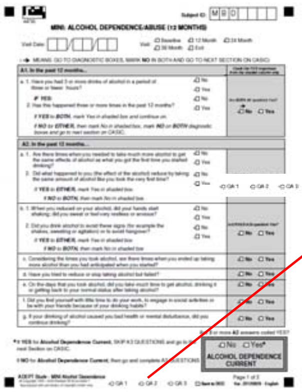
Consent to phlebotomy? ● No ○ Yes

Consent to medical record? ○ No ● Yes

Consent for social service record? ○ No ● Yes

Real Time Data Cleaning

- Database updated as soon as form “Submitted”
- Reports and queries can be run in real time (as opposed to “freezing” the database and running reports at specified intervals)
- If data irregularities found, can notify sites and correct immediately
- Can identify missing data while there may be an opportunity to collect it



MINI: ALCOHOL DEPENDENCE/ABUSE (12 MONTHS)

Subject ID: M B D

A1. In the past 12 months...

A2. In the past 12 months...

ALCOHOL DEPENDENCE CURRENT

Data should be reviewed for accuracy *prior* to entry whenever possible

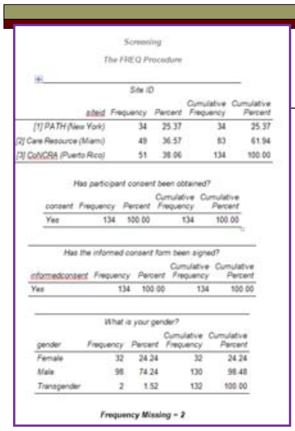
○ QA 1 ○ QA 2 ○ QA 3

Look at the Data Early and Often

- You cannot fix a problem if you don’t know it exists
- Get data into electronic format ASAP so it can be more easily reviewed
- Monitor the first few and participants
- Ongoing audit percentage of forms
- Pay extra attention to key variables

Do simple checks

- Frequency (count) and distribution (range) of each and every variable
- Do crosstabs of variables where appropriate
- What is missing?
- What is out of range?
- What contradicts (e.g., pregnant males)
- Are there systematic problems?



Screening

The FREQ Procedure

Site ID

siteid	Frequency	Percent	Cumulative Frequency	Cumulative Percent
[1] PATH (New York)	34	25.37	34	25.37
[2] Care Resource (Miami)	49	36.57	83	61.94
[3] CoNCRA (Puerto Rico)	51	38.06	134	100.00

Has participant consent been obtained?

consent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	134	100.00	134	100.00

Has the informed consent form been signed?

informedconsent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	134	100.00	134	100.00

What is your gender?

gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	32	24.24	32	24.24
Male	98	74.24	130	98.48
Transgender	2	1.52	132	100.00

Frequency Missing = 2

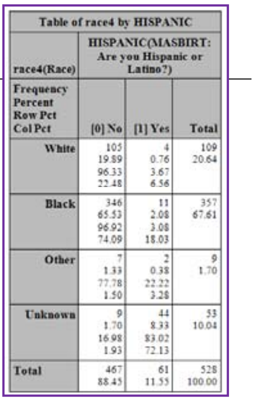
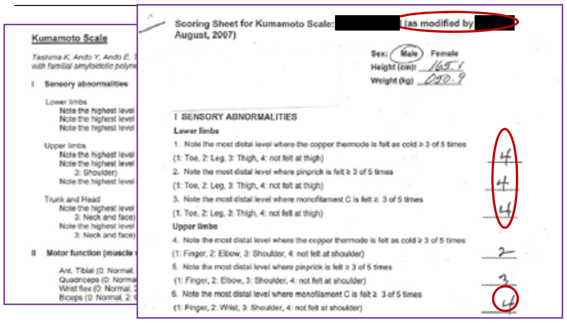


Table of race4 by HISPANIC

race4(Race)	HISPANIC(MASBIRT: Are you Hispanic or Latino?)		
Frequency Percent Row Pct Col Pct	[0] No	[1] Yes	Total
White	105 19.89 96.33 22.48	4 0.76 3.67 6.56	109 20.64
Black	346 65.53 96.92 74.09	11 2.08 3.08 18.03	357 67.61
Other	7 1.33 77.78 1.50	2 0.38 22.22 3.28	9 1.70
Unknown	9 1.70 16.98 1.93	44 8.33 83.02 72.13	53 10.04
Total	467 88.45	61 11.55	528 100.00

This is why you check…



Kumamoto Scale

Tashima K, Ando Y, Ando E, T
with familial amyloidotic polyne

I Sensory abnormalities

Lower limbs
Note the highest level
Note the highest level
Note the highest level

Upper limbs
Note the highest level
Note the highest level
3: Shoulder)
Note the highest level

Trunk and Head
Note the highest level
3: Neck and face)
Note the highest level
3: Neck and face)

II Motor function (muscle

Ant. Tibial (0: Normal,
Quadriceps (0: Norma
Wrist flex (0: Normal, 2
Biceps (0: Normal, 2: (

Scoring Sheet for Kumamoto Scale: [redacted] (as modified by [redacted]
August, 2007)

Sex: [✓ Male] Female
Height (cm): 165.1
Weight (kg): 050.9

I SENSORY ABNORMALITIES

Lower limbs

1. Note the most distal level where the copper thermode is felt as cold ≥ 3 of 5 times (1: Toe, 2: Leg, 3: Thigh, 4: not felt at thigh) 4
2. Note the most distal level where pinprick is felt ≥ 3 of 5 times (1: Toe, 2: Leg, 3: Thigh, 4: not felt at thigh) 4
3. Note the most distal level where monofilament C is felt ≥ 3 of 5 times (1: Toe, 2: Leg, 3: Thigh, 4: not felt at thigh) 4

Upper limbs

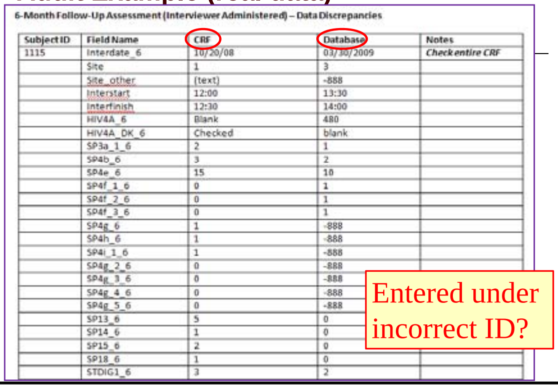
4. Note the most distal level where the copper thermode is felt as cold ≥ 3 of 5 times (1: Finger, 2: Elbow, 3: Shoulder, 4: not felt at shoulder) 2
5. Note the most distal level where pinprick is felt ≥ 3 of 5 times (1: Finger, 2: Elbow, 3: Shoulder, 4: not felt at shoulder) 3
6. Note the most distal level where monofilament C is felt ≥ 3 of 5 times (1: Finger, 2: Wrist, 3: Shoulder, 4: not felt at shoulder) 4

Perform Systematic Data Audits

- Data forms and source documents are compared with database on X % of forms
- Set an “acceptable” error rate. For example:
  - 0.5% overall
  - 0.1% for key fields)

- If audit yields a larger error rate, you must check and correct the database

Audit Example (real data)



6-Month Follow-Up Assessment (Interviewer Administered) – Data Discrepancies

Subject ID	Field Name	CRF	Database	Notes
1115	Interdate_6	10/20/08	03/30/2009	Check entire CRF
	Site	1	3	
	Site_other	(text)	-888	
	Interstart	12:00	13:30	
	Interfinish	12:30	14:00	
	HIV4A_6	Blank	480	
	HIV4A_DK_6	Checked	blank	
	SP3a_1_6	2	1	
	SP4b_6	3	2	
	SP4e_6	15	10	
	SP4f_1_6	0	1	
	SP4f_2_6	0	1	
	SP4f_3_6	0	1	
	SP4g_6	1	-888	
	SP4h_6	1	-888	
	SP4i_1_6	1	-888	
	SP4g_2_6	0	-888	
	SP4g_3_6	0	-888	
	SP4g_4_6	0	-888	
	SP4g_5_6	0	-888	
	SP13_6	5	0	
	SP14_6	1	0	
	SP15_6	2	0	
	SP18_6	1	0	
	STDIG1_6	3	2	

Entered under incorrect ID?

Subject Baseline 12 Month F/U

ID 1034

Baseline:

ID	1034
Sex	M
Age	28
Drinks	No/ Нет
IVDU	No/ Нет

12 Month F/U:

ID	1034
Sex	M
Age	28
Drinks	No/Нет
IVDU	No/ Нет

ID 1043

Baseline:

ID	1043
Sex	M
Age	28
Drinks	Yes/Да
IVDU	Yes/Да

12 Month F/U:

ID	1034
Sex	M
Age	28
Drinks	Yes/Да
IVDU	Yes/Да

Pay Extra Attention To Key Data

Be sure to pay particular attention to key data points where applicable.

- Query all SAE's ?
- Query all entries of crucial variables (e.g., study outcome)
- Extra attention to problematic variables (e.g., time-line-follow-back)

Data Cleaning: Essentials For Success

- Clean data in stages:
  - "Freeze" dataset for interim analysis (DSMB)
  - Subsequent cleaning of the data will be from that date forward
- The programmer must be familiar with the CRF
- The investigator or someone who really "knows" the study and the data must be involved in setting cleaning parameters and making decisions on what is invalid

Document, Document, Document!

Once you have identified errors in the data, be sure to document:

- All instances of errors
- All edits and corrections of the data
- History of manipulations, modifications, corrections to files/variables
- Location, type of media storage
- Archival procedures

Take Home Message

- Budget appropriately
- Be careful and be accurate
- Double and triple check the data
- Bring problems to the attention of study staff or PI right away
- Learn from your/other's mistakes
- If you do things right it's less work and you are more likely to discover the truth at the end

Data Security - General

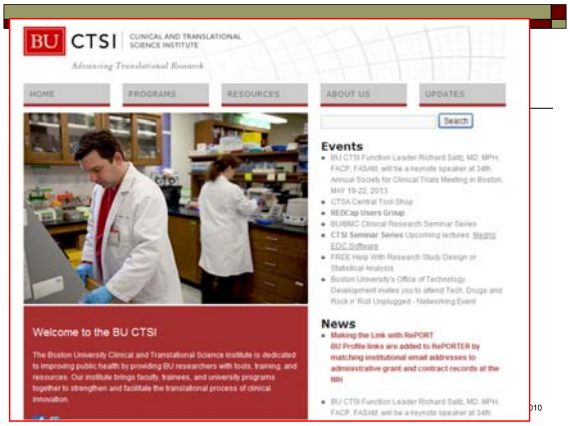
- Keep paper records should be kept in locked cabinets and/or offices
- Store identifiers like names and addresses separate from clinical data
- Keep particularly sensitive data apart from other identifiers (e.g., SSN) – in a separate file, by ID
- Do not collect sensitive data unless you *really* need it

Data Security - Hardware

- □ Password protect all computers
- □ Set to automatically timeout if inactive
- □ Encrypt laptops, flash-drives and other storage devices when possible
- □ Do not put identifiable data on portable media (e.g., CDs, flash-drives) unless password protected, preferably encrypted

Data Security - Electronic Data

- □ Make sure web and database servers are behind firewalls
- □ Encrypt all data transmissions from data collection point to servers (e.g., SSL)
- □ If sensitive fields must be collected, (e.g., SSN) encrypt them
- □ System users should have own logins and be instructed not to share usernames and passwords



BU CTSI CLINICAL AND TRANSLATIONAL SCIENCE INSTITUTE

Advancing Translational Research

Events

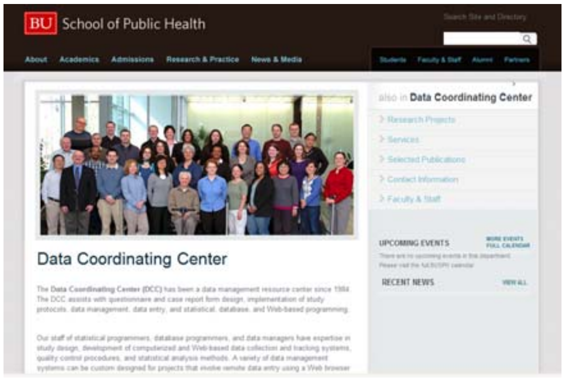
- BU CTSI Function Leader Richard Saitz, MD, MPH, FACP, FASAM, will be a keynote speaker at 34th Annual Society for Clinical Trials Meeting in Boston. MAY 19-22, 2013
- CTSA Central Tool Shop
- REDCap Users Group
- BU/BMC Clinical Research Seminar Series
- CTSI Seminar Series Upcoming lectures: Medrio EDC Software
- FREE Help With Research Study Design or Statistical Analysis
- Boston University's Office of Technology Development invites you to attend Tech, Drugs and Rock n' Roll Unplugged - Networking Event

News

- Making the Link with RePORT
  BU Profile links are added to RePORTER by matching institutional email addresses to administrative grant and contract records at the NIH
- BU CTSI Function Leader Richard Saitz, MD, MPH, FACP, FASAM, will be a keynote speaker at 34th

Welcome to the BU CTSI

The Boston University Clinical and Translational Science Institute is dedicated to improving public health by providing BU researchers with tools, training, and resources. Our institute brings faculty, trainees, and university programs together to strengthen and facilitate the translational process of clinical innovation.



BU School of Public Health

also in Data Coordinating Center

- Research Projects
- Services
- Selected Publications
- Contact Information
- Faculty & Staff

UPCOMING EVENTS MORE EVENTS FULL CALENDAR

There are no upcoming events in this department. Please visit the full BUSPH calendar

RECENT NEWS VIEW ALL

Data Coordinating Center

The Data Coordinating Center (DCC) has been a data management resource center since 1984. The DCC assists with questionnaire and case report form design, implementation of study protocols, data management, data entry, and statistical, database, and Web-based programming.

Our staff of statistical programmers, database programmers, and data managers have expertise in study design, development of computerized and Web-based data collection and tracking systems, quality control procedures, and statistical analysis methods. A variety of data management systems can be custom designed for projects that involve remote data entry using a Web browser

For more information…

- □ Contact the CTSI
  - See the CTSI website: http://ctsi.bu.edu/
  - Attend a CTSI drop-in session
  - Send an email: ctsi@bu.edu
- □ Contact the DCC
  - See the website: http://sph.bu.edu/DCC
  - Send an email: chaisson@bu.edu