DATA MANAGEMENT

What you should know and why you should care

CREST Seminar
November 10, 2015

Christine Chaisson
Director, Data Coordinating Center
Assistant Research Professor, Biostatistics
Boston University School of Public Health

BU
Boston University School of Public Health
Data Coordinating Center

Things that can go wrong with data

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

Real World Examples

- A few illustrations of common data problems from the popular news sources

The Inquirer DAILY NEWS
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

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

Oops: Excel Error Calls Into Question…

IEEE SPECTRUM

Posted 22 Apr 2013

- A book by Harvard Researchers entitled *'This time it's different"* contained "…serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period."
- 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

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; Duke team “shrugged them off” as “clerical errors.”
- Four papers were retracted
- Duke shut down three cancer trials
- Center leaders resigned or were removed
- People died and their relatives sued Duke

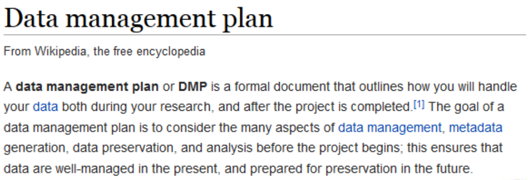
Data Entr agement in el

Goal: Convert Data into Electronic Format as Quickly as Possible

Data Management 101:

- No single “right” way to collect or manage data
- Consider:
  - Environment/location
  - Resources
  - Regulations
- Be sure to ***plan*** prior to study start
- Do what works for the study at hand

Where to start?



Data management plan

From Wikipedia, the free encyclopedia

A **data management plan** or **DMP** is a formal document that outlines how you will handle your data both during your research, and after the project is completed.[1] The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins; this ensures that data are well-managed in the present, and prepared for preservation in the future.

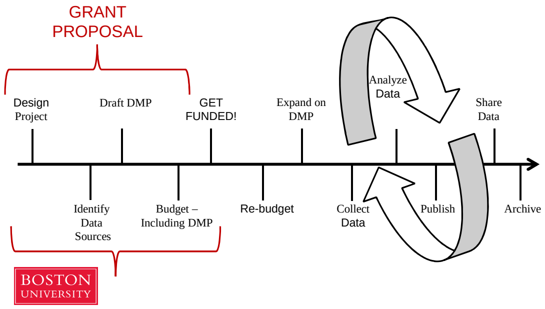
DMP Purpose: To help you manage and share your data; meet funder requirements. General elements include:

- Project or study description
- Documentation, organization, storage
- Access and sharing
- Archiving

DMP: Basic Elements

- Study design, data types and sources
- Storage format and location
- Naming conventions, documentation
- Software used for manipulation
- Project Staff – who has permission to what
- Identifiers (if applicable)
- Back ups, security
- Archiving
- Sharing

Timeline



Beginning: Identify Key Data Elements

- Review hypotheses
- What are primary, secondary outcomes?
- What covariates and confounders must be collected?
- What are the data sources?
  - Questionnaires
  - Labs, imaging
  - Medical record review
  - other external sources (e.g., lab results, medical records, death certificates)

Other Data Elements

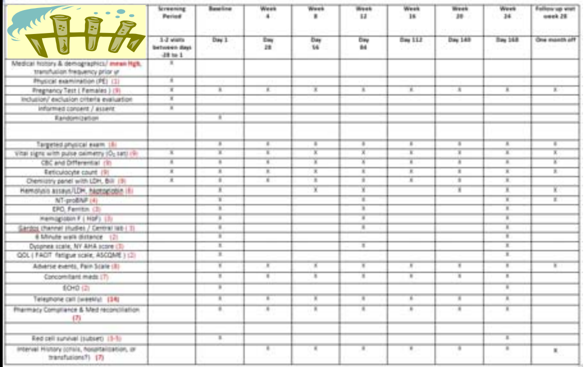
- Regulatory data:
  - IRB requirements
  - Safety (DSMB)
  - FDA (e.g., 21 CFR, part 11)
  - Other?
- Tracking/Study management data:
  - Tracking participants
  - Data elements by time-points
- Harmonization
  - NIH
  - Other

Visit Protocol: Data by Time-point

- Determine visit Schedule and data collected at each visit
  - Questionnaires
  - Labs
  - Other?

- Consider data not be connected to visits
  - Adverse events, serious adverse events
  - Hospitalization
  - Death
  - Medical records

Sample data/visit grid



	Screening Period	Baseline	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Follow up visit week 28
	1-2 visits between days -28 to 1	Day 1	Day 28	Day 56	Day 84	Day 112	Day 140	Day 168	One month off
Medical history & demographics/ mean Hgb, transfusion frequency prior yr	x								
Physical examination (PE) (1)	x								
Pregnancy Test (Females) (9)	x	x	x	x	x	x	x	x	x
Inclusion/ exclusion criteria evaluation	x								
Informed consent / assent	x								
Randomization		x							
Targeted physical exam (8)		x	x	x	x	x	x	x	x
Vital signs with pulse oximetry (O_2 sat) (9)	x	x	x	x	x	x	x	x	x
CBC and Differential (9)	x	x	x	x	x	x	x	x	x
Reticulocyte count (9)	x	x	x	x	x	x	x	x	x
Chemistry panel with LDH, Bili (9)	x	x	x	x	x	x	x	x	
Hemolysis assays/LDH, haptoglobin (6)		x		x	x		x	x	x
NT-proBNP (4)		x			x			x	x
EPO, Ferritin (3)		x			x			x	
Hemoglobin F (HbF) (3)		x			x			x	
Gardos channel studies / Central lab (3)		x			x			x	
6 Minute walk distance (2)		x						x	
Dyspnea scale, NY AHA score (3)		x			x			x	
QOL (FACIT fatigue scale, ASCQME) (2)		x						x	
Adverse events, Pain Scale (8)		x	x	x	x	x	x	x	x
Concomitant meds (7)		x	x	x	x	x	x	x	
ECHO (2)		x						x	
Telephone call (weekly) (24)		x	x	x	x	x	x	x	
Pharmacy Compliance & Med reconciliation (7)		x	x	x	x	x	x	x	
Red cell survival (subset) (3-5)		x						x	
Interval History (crisis, hospitalization, or transfusions?) (7)			x	x	x	x	x	x	x

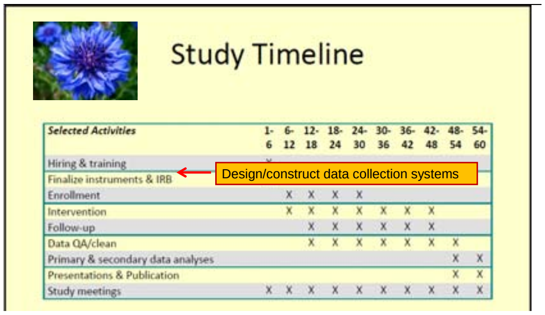
Timelines and Tasks

- Develop Protocol and Analytic Plan
- Create and pilot of forms/assessments
- Design/construct data systems
  - Data Collection/entry
  - Participant/Data Tracking
- Subject recruitment
- Data collection (baseline and follow-up)
- Data cleaning, auditing, and QA
- Preliminary 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

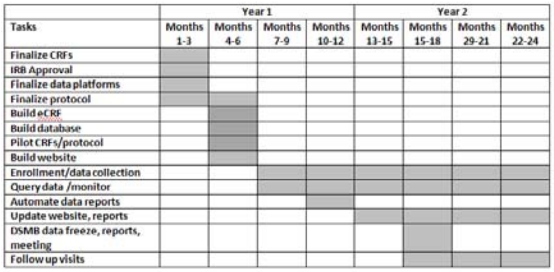


Study Timeline

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

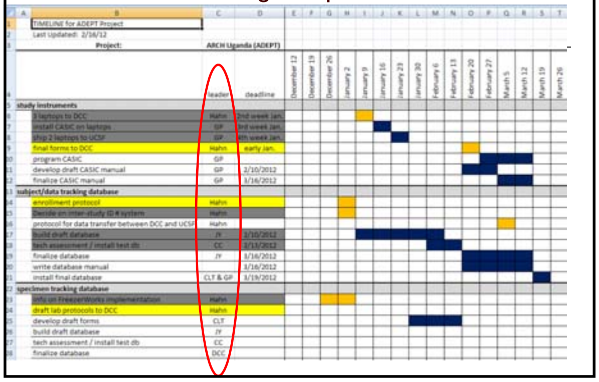
Design/construct data collection systems

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	3/10/2012					🟦	🟦	🟦	🟦	🟦							
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

- Analytic plan
- Detailed protocol
- Well designed data collection forms
- Tracking system
- Data capture/entry system
- Plan for data query (checking/cleaning)
- Manuals
- Data dictionaries

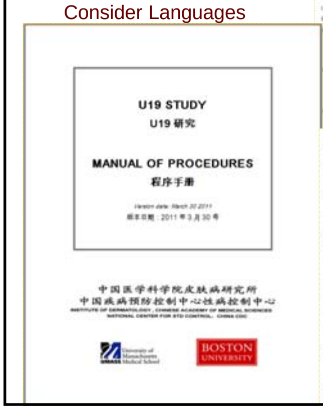
Web Site User Manual

Version date: August 24, 2010

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

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

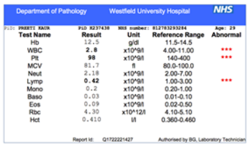
Alisa Biebes, Project Manager Phone: 617.638.[illegible] Email: [illegible]@bu.edu	Zhenyan Yang, Data Manager Phone: 617.638.5028 Email: [illegible]@bu.edu
[illegible] 电话：617.638.[illegible] 电子邮箱：[illegible]@bu.edu	[illegible] 数据管理 电话：617.638.5028 电子邮箱：[illegible]@bu.edu
John Lu, Database/Web Programmer Phone: 617.638.4241 Email: lu@bu.edu	Christine Chaisson, Project Director Phone: 617.638.5009 Email: chaisson@bu.edu
[illegible] 电话：617.638.4241 电子邮箱：lu@bu.edu	[illegible] 电话：617.638.5009 电子邮箱：chaisson@bu.edu

Sample Data Dictionary: SAS formatted dataset

Number	Variable	Label	Type	Format
1	adept_study_id	ADEPT Study ID	Num	8.
2	MissedAssessment	MissedAssessment	Num	11.
3	Interviewer	Interviewer	Char	$50.
4	PatientLanguage	PatientLanguage	Char	$20.
5	Completed	Completed	Num	11.
6	DtAssessStart	DtAssessStart	Num	DATETIME22.3
7	DtAssessEnd	DtAssessEnd	Num	DATETIME22.3
8	dateintv	What is today's date?	Num	DATETIME22.3
9	intv_initials	Please enter your (the research assistants) initials	Char	$1024.
10	lang	Language of interview	Num	LANGF.
11	sex	Gender	Num	SEXF.
12	weight	Participant's weight in Kg	Num	20.2
13	height	Participant's height in cm	Num	20.1
14	age	Age	Num	8.
15	DOBDay	DoB: Day	Num	8.
16	DOBMonth	DoB: Month	Num	8.
17	DOBYear	DoB: Year	Num	8.
18	tribenkole	Tribe: Nkole	Num	NOYES.
19	tribekiga	Tribe: Kiga	Num	NOYES.
20	tribehima	Tribe: Hima	Num	NOYES.
21	tribeganda	Tribe: Ganda	Num	NOYES.
22	tribetoro	Tribe: Toro	Num	NOYES.
23	tribekonjo	Tribe: Konjo	Num	NOYES.
24	tribeothe	Tribe: Other	Num	NOYES.
25	tribedk	Tribe: Don't Know	Num	NOYES.
26	tribena	Tribe: Not applicable	Num	NOYES.

Data Collection Forms

- Data will usually end up on some type of "form" whether it is interview, chart review, or imaging results
- Make sure you plan carefully and leave time as this can be a lengthy process



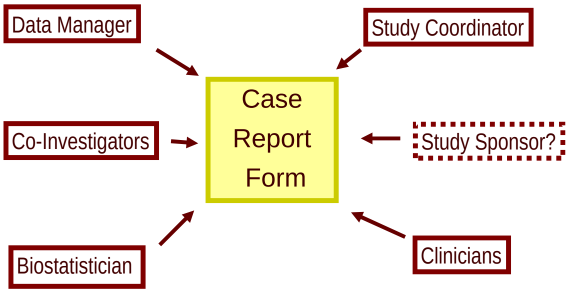
Department of Pathology Westfield University Hospital NHS

PID: PREETI KAUR PID H237436 NHS number: 612703293284 Age: 29

Test Name	Result	Unit	Reference Range	Abnormal
Hb	12.5	g/dl	11.5-14.5	
WBC	2.8	x10^9/l	4.00-11.00	***
Plt	98	x10^9/l	140-400	***
MCV	81.7	fl	80.0-100.0	
Neut	2.18	x10^9/l	2.00-7.00	
Lymp	0.42	x10^9/l	1.00-3.00	***
Mono	0.2	x10^9/l	0.20-1.00	
Baso	0.03	x10^9/l	0.01-0.10	
Eos	0.09	x10^9/l	0.02-0.50	
Rbc	4.30	x10^12/l	4.10-5.10	
Hct	0.410	l/l	0.360-0.460	

Report Id: Q1722221427 Authorised by BG, Laboratory Technician

Creating a Data Collection Form: Get input from all points of view



Data Manager → Case Report Form

Study Coordinator → Case Report Form

Co-Investigators → Case Report Form

Study Sponsor? → Case Report Form

Biostatistician → Case Report Form

Clinicians → Case Report Form

Why is this topic important?

- Sloppy forms indicate sloppy research
- CRF may not answer study questions
- Danger of collecting:
  - too much data
  - too little data
  - the wrong data
- Annoyed:
  - Participants
  - Study Coordinator
  - Data Analyst…

GET ALL THE INFORMATION YOU CAN, WE'LL THINK OF A USE FOR IT LATER.

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
  - ❑ Outsourced
- ❑ Data management/cleaning
- ❑ Auditing
- ❑ Data analysis

When designing forms…

Think Google

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
- ❑ Unambiguous questions
- ❑ 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 (e.g., age, literacy)
  - Are items left blank?
    - Reword/drop question
  - Are "skip" patterns followed correctly?
    - Train clinic personnel/revise or simplify forms
  - Are open-ended questions generating common responses?
    - Categorize/code
- □ Make corrections prior to start of study
- □ Do not start data collection until forms are final

Avoid Open-ended & Include Response Measure

What is your date of birth? ____

How much do you weigh? ___

How tall are you? ________

Record subject's
temperature ______

Date of Birth? __/__/____
MM DD YYYY

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

How tall are you? ___ /__
(feet)/(inches)

Record subject's
oral temperature __ __. __ (f)

Include Clear Instructions

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

$_1$O Caucasian
$_2$O African American/Black
$_3$O Asian, Pacific Islander
$_4$O Native American
$_5$O 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 of “check all that apply”

	COMORBIDITIES	
b.	Heart disease	1□
c.	Diabetes	1□
d.	Hypertension	1□
e.	Pulmonary disease	1□

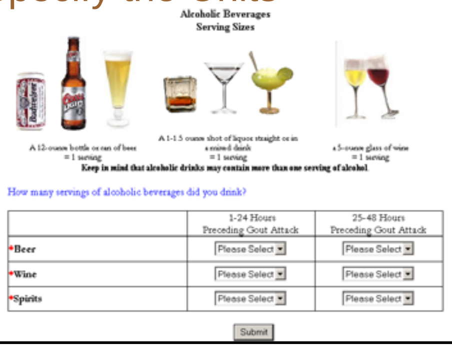
a.	COMORBIDITIES	
b.	Heart disease	1○ 2○
c.	Diabetes	1○ 2○
d.	Hypertension	1○ 2○
e.	Pulmonary disease	1○ 2○

37

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 ▾

Categorize Anticipated Responses

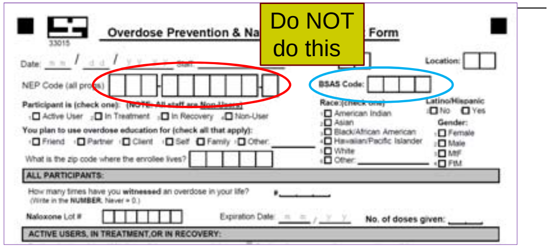
□1 USA
□2 Guatemala
□3 Mexico
□4 Dominican Republic
□5 Other

□1 USA
□2 Guatemala
□3 Mexico
□4 Dominican Republic
□6 El Salvador
□5 Other

ID Assignment

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

Example of an ID that is not unique



Do NOT do this

33015 Overdose Prevention & Na Form

Date: m m / d d / y y y y Staff: Location:

NEP Code (all pro s)

BSAS Code:

Participant is (check one): (NOTE: All staff are Non-Users)
1□ Active User 2□ In Treatment 3□ In Recovery 4□ Non-User

You plan to use overdose education for (check all that apply):
1□ Friend 1□ Partner 1□ Client 1□ Self □ Family 1□ Other: ______

What is the zip code where the enrollee lives?

Race:(check one)
1□ American Indian
2□ Asian
3□ Black/African American
4□ Hawaiian/Pacific Islander
5□ White
6□ Other: ______

Latino/Hispanic
0□ No □ Yes

Gender:
1□ Female
2□ Male
3□ MtF
4□ FtM

ALL PARTICIPANTS:

How many times have you witnessed an overdose in your life? #____
(Write in the NUMBER. Never = 0.)

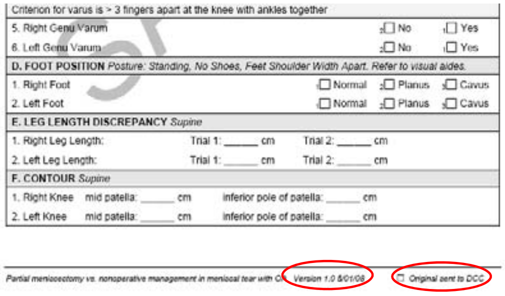
Naloxone Lot # Expiration Date: m m / y y No. of doses given: ____

ACTIVE USERS, IN TREATMENT,OR IN RECOVERY:

Codes and Abbreviations:

FtM	Female to Male transgender
MtF	Male to Female transgender
NEP Code	First three letters of mother's first name+ date of birth (mm/dd/yy) Ex: GER053077)
BSAS Code	First & third letters of first and last name Ex: Joseph "Joe" Francis Blow=JSBO

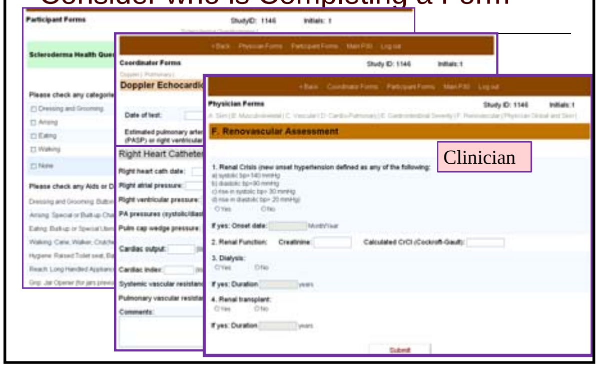
Don't Underestimate Need for Version# /Date



Criterion for varus is > 3 fingers apart at the knee with ankles together			
5. Right Genu Varum		₀□ No	₁□ Yes
6. Left Genu Varum		₀□ 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 8/01/08 □ *Original sent to DCC*

Consider who is Completing a Form



Participant Forms StudyID: 1146 Initials: 1

Scleroderma Health Ques

Please check any categorie

□ Dressing and Grooming

□ Arising

□ Eating

□ Walking

□ None

Please check any Aids or D

< Back Physician Forms Participant Forms Main Fill Log out

Coordinator Forms Study ID: 1146 Initials: 1

Doppler Echocardio

Date of test:

Estimated pulmonary arter (PASP) or right ventricular

Right Heart Catheter

Right heart cath date:

Right atrial pressure:

Right ventricular pressure:

PA pressures (systolic/dias

Pulm cap wedge pressure:

Cardiac output:

Cardiac index:

Systemic vascular resistanc

Pulmonary vascular resista

Comments:

< Back Coordinator Forms Participant Forms Main Fill Log out

Physician Forms Study ID: 1146 Initials: 1

F. Renovascular Assessment

Clinician

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

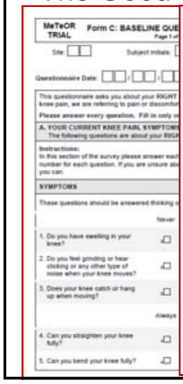
If yes: Duration years

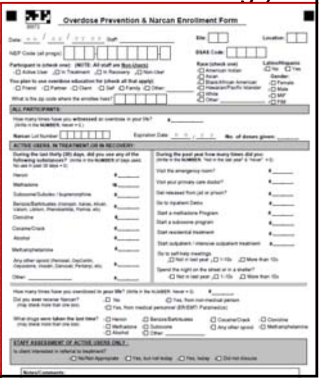
Submit

In Summary, when designing questions:

- Avoid ambiguous questions and open ended responses
- Include clear instructions
- Be sure form complexity matches collector (self, study coordinator, clinician)
- Collect data elements in correct format ("continuous" or "categorical")
- Make categories mutually exclusive
- Pilot your forms in the target population

The Good The Bad And The Ugly





Paper or Electronic?

Paper Forms / Manual Entry

Advantages

- The old "standard"
- Shorter start-up time (Word/PDF)
- Relatively easy to train staff
- Hardcopy document to refer back to
- Can be done anywhere

Disadvantages

- Costs: data entry, storage and shipping
- Longer time from collection to database
- Errors in data collection (missing, out of range, skips)

Electronic Data Capture

Advantages

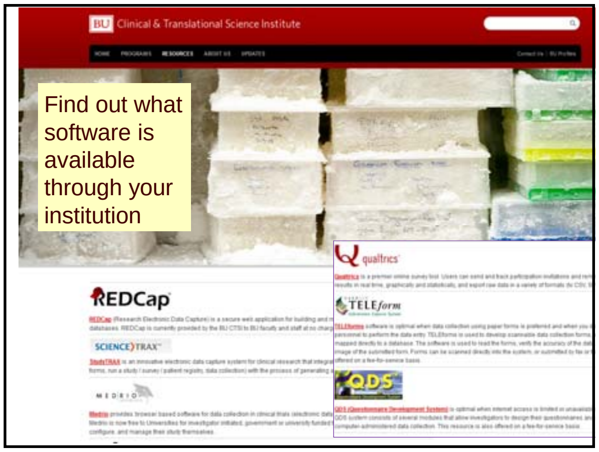
- Cleaner data at entry (required fields, skips, ranges)
- Can use data in real time (or close to it)
- No extra data entry costs
- Data can inform next visit even for short follow up

Disadvantages

- Programming time and costs
- Increased hardware and software costs
- Infrastructure concerns (software versions, internet connection, back-up equipment)
- Data security

A Word About "Canned" Software

- Many "canned" sofware packages available
- No single best choice
- Cost can vary widely
- Database structures vary
- Do your homework to make sure what you get will work for your project



Find out what software is available through your institution

REDCap (Research Electronic Data Capture) is a secure web application for building and m
databases. REDCap is currently provided by the BU CTSI to BU faculty and staff at no charg

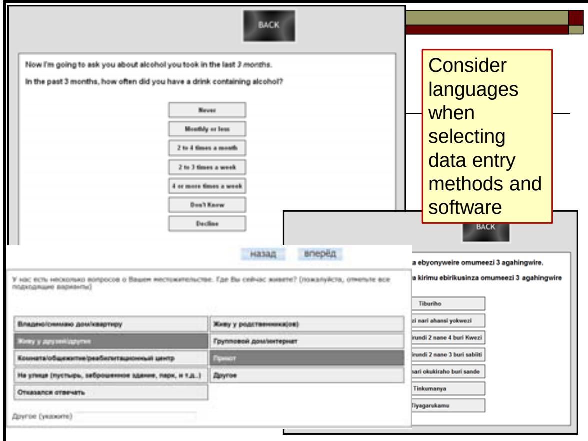
StudyTRAX is an innovative electronic data capture system for clinical research that integra
forms, run a study / survey / patient registry, data collection) with the process of generating

Medrio provides browser based software for data collection in clinical trials (electronic data
Medrio is now free to Universities for investigator initiated, government or university funded
configure, and manage their study themselves.

Qualtrics is a premier online survey tool. Users can send and track participation invitations and rem
results in real time, graphically and statistically, and export raw data in a variety of formats (to CSV, S

TELEforms software is optimal when data collection using paper forms is preferred and when you d
personnel to perform the data entry. TELEforms is used to develop scannable data collection forms,
mapped directly to a database. The software is used to read the forms, verify the accuracy of the dat
image of the submitted form. Forms can be scanned directly into the system, or submitted by fax or t
offered on a fee-for-service basis.

QDS (Questionnaire Development System) is optimal when internet access is limited or unavailab
QDS system consists of several modules that allow investigators to design their questionnaires an
computer-administered data collection. This resource is also offered on a fee-for-service basis.



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?

У нас есть несколько вопросов о Вашем местожительстве. Где Вы сейчас живете? (пожалуйста, отметьте все подходящие варианты)

Владею/снимаю дом/квартиру	Живу у родственника(ов)
Живу у друзей/других	Групповой дом/интернат
Комната/общежитие/реабилитационный центр	Приют
На улице (пустырь, заброшенное здание, парк, и т.д.)	Другое
Отказался отвечать	

Другое (укажите)

a ebyonyweire omumeezi 3 agahingwire.

a kirimu ebirikusinza omumeezi 3 agahingwire

Consider languages when selecting data entry methods and software

What to use…?

- To determine what software is best suited for your project see:
  - What is available to you?
  - What is the cost (can you afford it)?
  - What has the features you need (e.g., language)

Once data are collected…

- Get your data into a useful format
- No "right" format – use what works for you
  - SQL database
  - SAS datasets
  - SPSS
  - Excel (be careful!)

Crosswalk for Personal Identifiers

- □ Do not store any identifier unless you have a good reason for it
- □ Do not store identifiers in same files with study data. Identifiers should be kept separate!
- □ Create "crosswalk" files of identifiers and store them someplace secure.

Personal Identifiers

First Name	Last Name	Study ID	Screen ID	Phone #
Joseph	Blow	1234	50001	555-131-1111

Identifiable Data

Crosswalk

Subject ID	MRN	SSN
1234	64322	******

Study Data

ID	Visit	Var 1	Var 2	Var 3
1234	1	5	2	3

HIPAA Identifiers

1. Names
2. Addresses other than state, and first three digits of the zip code
3. All elements of date other than year, and all specific ages over 89 years
4. Telephone numbers
5. fax numbers
6. Email addresses
7. Social Security numbers
8. Medical Record numbers
9. Health plan beneficiary numbers

HIPAA Identifiers (cont)

10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers
13. Device identifiers and serial numbers
14. Web universal resource locators (URLs; web site addresses)
15. Internet protocol (IP) addresses
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code

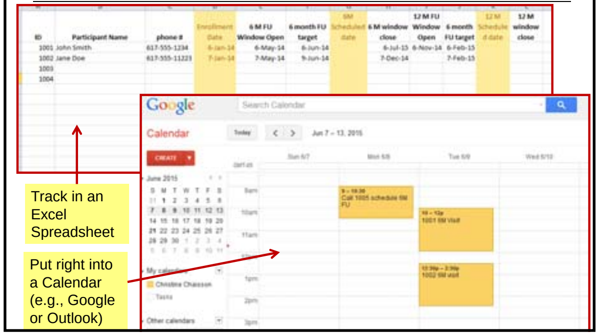
Participant Tracking

Tracking the Participants

You need a system to track participants

- □ Tracking for Study Management:
  - Screened, Eligible, Enrolled
  - Monitor and report progress
- □ Tracking tools for study staff:
  - Schedule/reminders follow up visits
  - Collection of all data points at each visit
- □ Small study may use Outlook or Excel; large study may need a tracking system

Simple Participant Tracking Tools

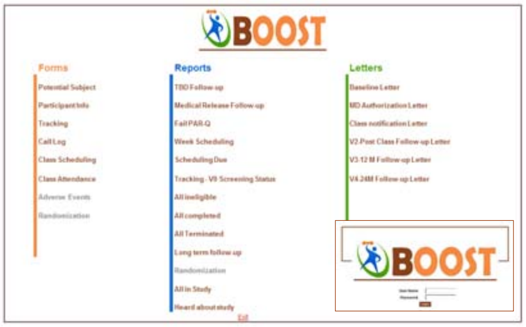


ID	Participant Name	phone #	Enrollment Date	6 M FU Window Open	6 month FU target	6M Scheduled date	6 M window close	12 M FU Window Open	6 month FU target	12 M Schedule d date	12 M window close
1001	John Smith	617-555-1234	6-Jan-14	6-May-14	6-Jun-14		6-Jul-13	6-Nov-14	6-Feb-15		
1002	Jane Doe	617-555-11223	7-Jan-14	7-May-14	9-Jun-14		7-Dec-14		7-Feb-15		
1003											
1004											

Track in an Excel Spreadsheet

Put right into a Calendar (e.g., Google or Outlook)

More Complex Tracking Tool



Forms
- Potential Subject
- Participant Info
- Tracking
- Call Log
- Class Scheduling
- Class Attendance
- Adverse Events
- Randomization

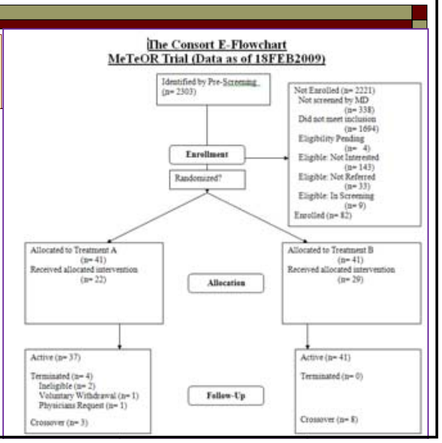
Reports
- TBD Follow up
- Medical Release Follow up
- Fail PAR-Q
- Week Scheduling
- Scheduling Due
- Tracking - V8 Screening Status
- All ineligible
- All completed
- All Terminated
- Long term follow up
- Randomization
- All in Study
- Heard about study

Letters
- Baseline Letter
- MD Authorization Letter
- Class notification Letter
- V2-Post Class Follow-up Letter
- V3-12 M Follow-up Letter
- V4 24M Follow-up Letter

Reports

- For the study team to view at regular meetings or online as needed

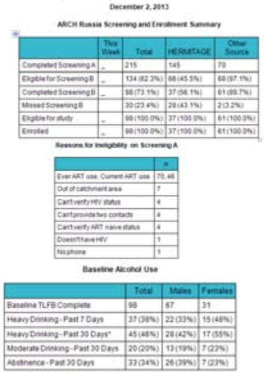
Consort Diagram



The Consort E-Flowchart
MeTeOR Trial (Data as of 18FEB2009)

Identified by Pre-Screening (n= 2303)
Not Enrolled (n= 2221)
Not screened by MD (n= 338)
Did not meet inclusion (n= 1694)
Eligibility Pending (n= 4)
Eligible: Not Interested (n= 143)
Eligible: Not Referred (n= 33)
Eligible: In Screening (n= 9)
Enrolled (n= 82)
Enrollment
Randomized?
Allocated to Treatment A (n= 41)
Received allocated intervention (n= 22)
Allocation
Allocated to Treatment B (n= 41)
Received allocated intervention (n= 29)
Active (n= 37)
Terminated (n= 4)
Ineligible (n= 2)
Voluntary Withdrawal (n= 1)
Physicians Request (n= 1)
Crossover (n= 3)
Follow-Up
Active (n= 41)
Terminated (n= 0)
Crossover (n= 8)

Enrollment



December 2, 2013

ARCH Russia Screening and Enrollment Summary

	This Week	Total	HERMITAGE	Other Source
Completed Screening A	_	215	145	70
Eligible for Screening B	_	134 (62.3%)	66 (45.5%)	68 (97.1%)
Completed Screening B	_	98 (73.1%)	37 (56.1%)	61 (89.7%)
Missed Screening B		30 (23.4%)	28 (43.1%)	2 (3.2%)
Eligible for study	_	98 (100.0%)	37 (100.0%)	61 (100.0%)
Enrolled	_	98 (100.0%)	37 (100.0%)	61 (100.0%)

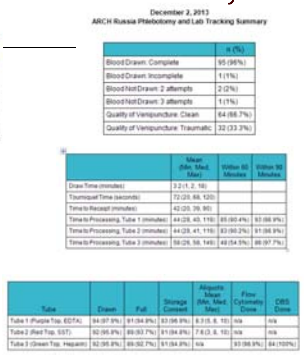
Reasons for Ineligibility on Screening A

	n
Ever ART use, Current ART use	70, 46
Out of catchment area	7
Can't verify HIV status	4
Can't provide two contacts	4
Can't verify ART naive status	4
Doesn't have HIV	1
No phone	1

Baseline Alcohol Use

	Total	Males	Females
Baseline TLFB Complete	98	67	31
Heavy Drinking - Past 7 Days	37 (38%)	22 (33%)	15 (48%)
Heavy Drinking - Past 30 Days*	45 (46%)	28 (42%)	17 (55%)
Moderate Drinking - Past 30 Days	20 (20%)	13 (19%)	7 (23%)
Abstinence - Past 30 Days	33 (34%)	26 (39%)	7 (23%)

Phlebotomy



December 2, 2013
ARCH Russia Phlebotomy and Lab Tracking Summary

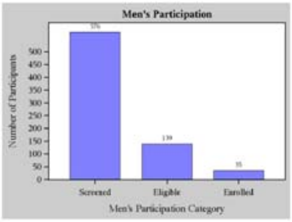
	n (%)
Blood Drawn: Complete	95 (96%)
Blood Drawn: Incomplete	1 (1%)
Blood Not Drawn: 2 attempts	2 (2%)
Blood Not Drawn: 3 attempts	1 (1%)
Quality of Venipuncture: Clean	64 (66.7%)
Quality of Venipuncture: Traumatic	32 (33.3%)

	Mean (Min, Med, Max)	Within 60 Minutes	Within 90 Minutes
Draw Time (minutes)	3.2 (1, 2, 18)		
Tourniquet Time (seconds)	72 (20, 68, 120)		
Time to Receipt (minutes)	42 (20, 39, 95)		
Time to Processing, Tube 1 (minutes)	44 (28, 40, 118)	85 (90.4%)	93 (98.9%)
Time to Processing, Tube 2 (minutes)	44 (28, 41, 116)	83 (90.2%)	91 (98.9%)
Time to Processing, Tube 3 (minutes)	58 (26, 58, 149)	48 (54.5%)	86 (97.7%)

Tube	Drawn	Full	Storage Consent	Aliquots Mean (Min, Med, Max)	Flow Cytometry Done	DBS Done
Tube 1 (Purple Top, EDTA)	94 (97.9%)	91 (94.8%)	93 (96.9%)	6.3 (5, 6, 10)	n/a	n/a
Tube 2 (Red Top, SST)	92 (95.8%)	90 (93.7%)	91 (94.8%)	7.6 (3, 8, 10)	n/a	n/a
Tube 3 (Green Top, Heparin)	92 (95.8%)	89 (92.7%)	91 (94.8%)	n/a	93 (96.9%)	84 (100%)

Reports: Visual as well as Tabular

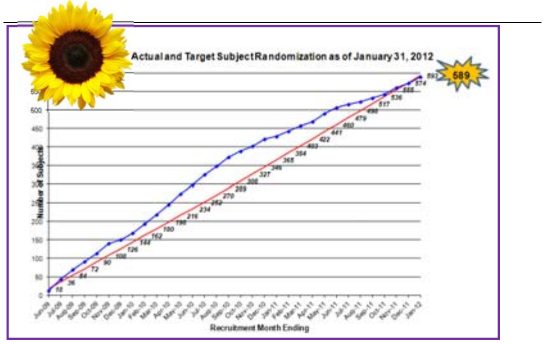
U19 Study: Enrollment Reports as of 10/15/2012



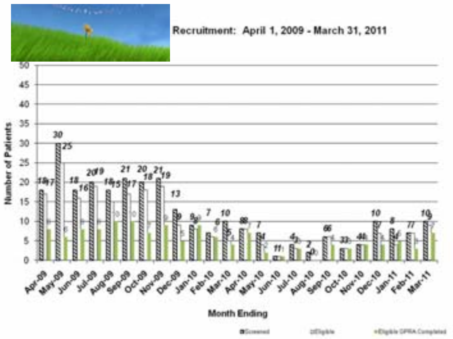
Men's Participation

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

Actual vs. Targeted Enrollment



Screened, Eligible and Enrolled



Participation Summary: Total

Beginning: 5/21/2009 Ending: 4/5/2011

	Number Pending (1)	Number Due (2)	Number Complete (3)	Number Incomplete (4)	Number Inactive (5)	Number Out of Study (6)	Total	Min	Max
Baseline	-----	-----	460	-----	0	0	460	-----	-----
Six Week	8	15	434	1	1	1	460	95.0%	99.3%
Six Month	4	94	355	2	3	0	458	78.2%	98.9%

Participation Summary: ASSIST Score ≥ 4

Beginning: 5/21/2009 Ending: 4/5/2011

	Number Pending (1)	Number Due (2)	Number Complete (3)	Number Incomplete (4)	Number Inactive (5)	Number Out of Study (6)	Total	Min	Max
Baseline	-----	-----	416	-----	0	0	416	-----	-----
Six Week	7	14	393	1	0	1	416	96.1%	99.5%
Six Month	0	84	326	2	3	0	415	78.6%	98.8%

Participation Summary: ASSIST Score < 4

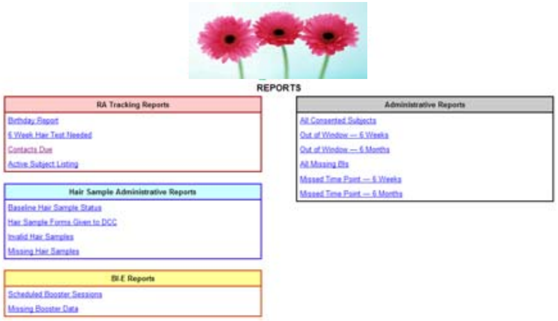
Beginning: 5/21/2009 Ending: 4/5/2011

	Number Pending (1)	Number Due (2)	Number Complete (3)	Number Incomplete (4)	Number Inactive (5)	Number Out of Study (6)	Total	Min	Max
Baseline	-----	-----	44	-----	0	0	44	-----	-----
Six Week	1	1	41	0	1	0	44	95.3%	97.7%
Six Month	4	10	29	0	0	0	43	74.4%	100.0%

(1) Active Subjects who have not yet completed the current interview(not due)
(2) Active Subjects who are due for the current interview
(3) Active Subjects who have completed the current interview
(4) Active Subjects who have missed the current interview(fallen out of their window)
(5) Inactive Subjects
(6) Out of Study Subjects or Deceased Subjects

Reports

- For the study staff to help manage tasks and know what needs to be done, when



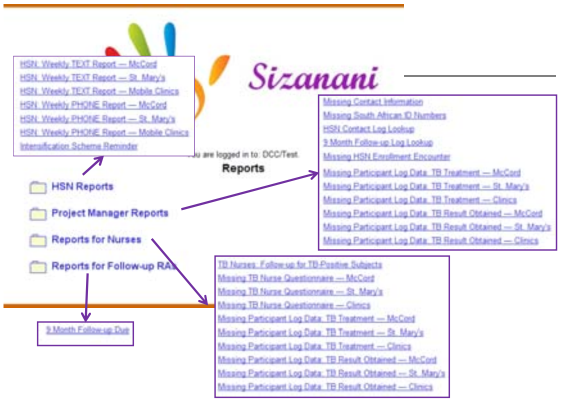
REPORTS

RA Tracking Reports
- Birthday Report
- 6 Week Hair Test Needed
- Contacts Due
- Active Subject Listing

Hair Sample Administrative Reports
- Baseline Hair Sample Status
- Hair Sample Forms Given to DCC
- Invalid Hair Samples
- Missing Hair Samples

BI-E Reports
- Scheduled Booster Sessions
- Missing Booster Data

Administrative Reports
- All Consented Subjects
- Out of Window — 6 Weeks
- Out of Window — 6 Months
- All Missing BIs
- Missed Time Point — 6 Weeks
- Missed Time Point — 6 Months



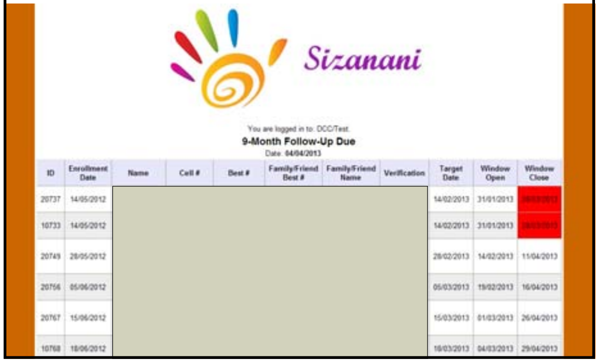
Sizanani

You are logged in to: DCC/Test.

Reports

- HSN Reports
  - HSN: Weekly TEXT Report — McCord
  - HSN: Weekly TEXT Report — St. Mary's
  - HSN: Weekly TEXT Report — Mobile Clinics
  - HSN: Weekly PHONE Report — McCord
  - HSN: Weekly PHONE Report — St. Mary's
  - HSN: Weekly PHONE Report — Mobile Clinics
  - Intensification Scheme Reminder
- Project Manager Reports
  - Missing Contact Information
  - Missing South African ID Numbers
  - HSN Contact Log Lookup
  - 9 Month Follow-up Log Lookup
  - Missing HSN Enrollment Encounter
  - Missing Participant Log Data: TB Treatment — McCord
  - Missing Participant Log Data: TB Treatment — St. Mary's
  - Missing Participant Log Data: TB Treatment — Clinics
  - Missing Participant Log Data: TB Result Obtained — McCord
  - Missing Participant Log Data: TB Result Obtained — St. Mary's
  - Missing Participant Log Data: TB Result Obtained — Clinics
- Reports for Nurses
  - TB Nurses: Follow-up for TB-Positive Subjects
  - Missing TB Nurse Questionnaire — McCord
  - Missing TB Nurse Questionnaire — St. Mary's
  - Missing TB Nurse Questionnaire — Clinics
  - Missing Participant Log Data: TB Treatment — McCord
  - Missing Participant Log Data: TB Treatment — St. Mary's
  - Missing Participant Log Data: TB Treatment — Clinics
  - Missing Participant Log Data: TB Result Obtained — McCord
  - Missing Participant Log Data: TB Result Obtained — St. Mary's
  - Missing Participant Log Data: TB Result Obtained — Clinics
- Reports for Follow-up RAs
  - 9 Month Follow-up Due

Sample Tracking Report: Follow UP



You are logged in to: DCC/Test

9-Month Follow-Up Due

Date: 04/04/2013

ID	Enrollment Date	Name	Cell #	Best #	Family/Friend Best #	Family/Friend Name	Verification	Target Date	Window Open	Window Close
20737	14/05/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14/02/2013	31/01/2013	[illegible]
10733	14/05/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14/02/2013	31/01/2013	[illegible]
20749	28/05/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	28/02/2013	14/02/2013	11/04/2013
20756	05/06/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	05/03/2013	19/02/2013	16/04/2013
20767	15/06/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	15/03/2013	01/03/2013	26/04/2013
10768	18/06/2012	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18/03/2013	04/03/2013	29/04/2013

Tracking the Data Elements

- Identify what data have been collected
  - For each Subject at each Visit:
    - Questionnaires
    - Imaging, labs results
    - Other external data
- Missing data: can you still get it?
- Data entry: 1st entry/ 2nd entry/ reconciled
- Data cleaning/ QA / auditing
- "Clean" frozen datasets

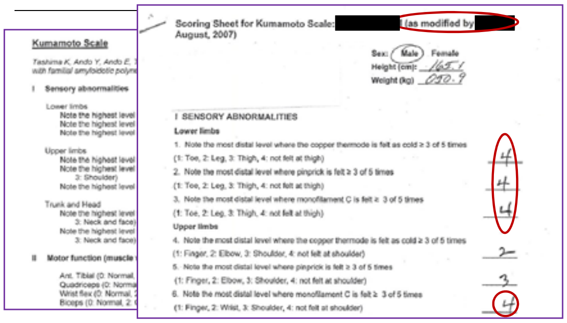
Look at the Data Early and Often

- 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 every data point for the first few 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 systemic problems?

This is why you check…



Kumamoto Scale

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

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

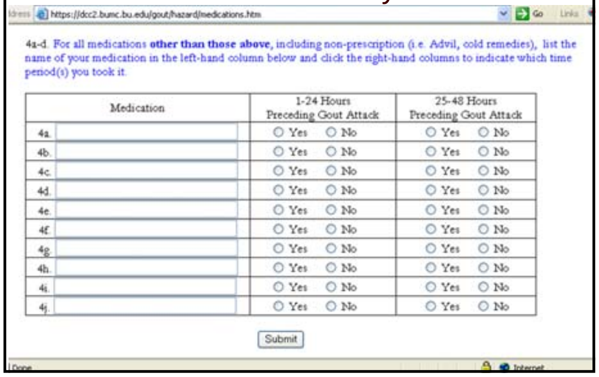
Lower limbs

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

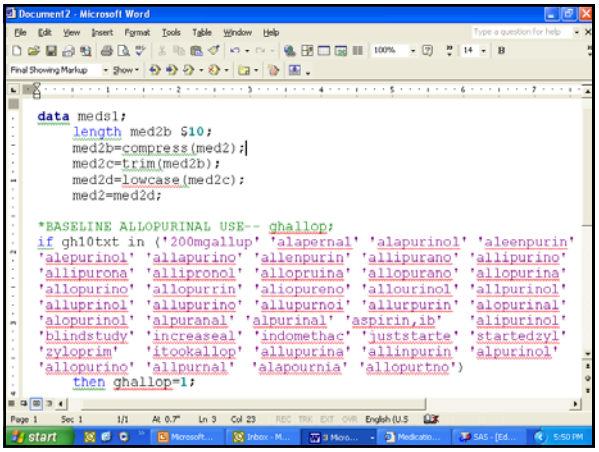
Medications are never easy



4a-d. For all medications **other than those above**, including non-prescription (i.e. Advil, cold remedies), list the name of your medication in the left-hand column below and click the right-hand columns to indicate which time period(s) you took it.

	Medication	1-24 Hours Preceding Gout Attack	25-48 Hours Preceding Gout Attack
4a.		○ Yes ○ No	○ Yes ○ No
4b.		○ Yes ○ No	○ Yes ○ No
4c.		○ Yes ○ No	○ Yes ○ No
4d.		○ Yes ○ No	○ Yes ○ No
4e.		○ Yes ○ No	○ Yes ○ No
4f.		○ Yes ○ No	○ Yes ○ No
4g.		○ Yes ○ No	○ Yes ○ No
4h.		○ Yes ○ No	○ Yes ○ No
4i.		○ Yes ○ No	○ Yes ○ No
4j.		○ Yes ○ No	○ Yes ○ No

Submit



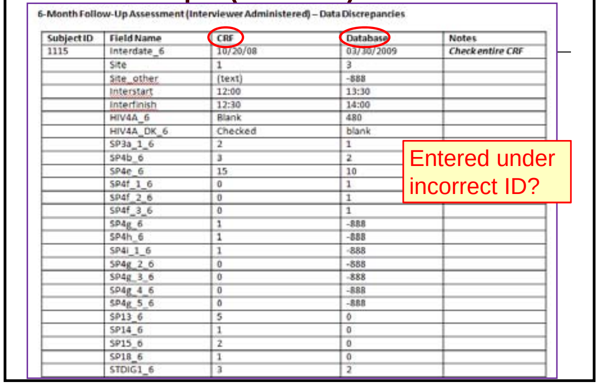
Document2 - Microsoft Word
File Edit View Insert Format Tools Table Window Help
Type a question for help
Final Showing Markup
Show
Page 1 Sec 1 1/1 At 0.7" Ln 3 Col 23 REC TRK EXT OVR English (U.S
start
Microsoft...
Inbox - M...
3 Micro...
Medicatio...
SAS - [Ed...
5:50 PM

```
data meds1;
     length med2b $10;
     med2b=compress(med2);
     med2c=trim(med2b);
     med2d=lowcase(med2c);
     med2=med2d;

*BASELINE ALLOPURINAL USE-- ghallop;
if gh10txt in ('200mgallup' 'alapernal' 'alapurinol' 'aleenpurin'
'alepurinol' 'allapurino' 'allenpurin' 'allipurano' 'allipurino'
'allipurona' 'allipronol' 'allopruina' 'allopurano' 'allopurina'
'allopurino' 'allopurrin' 'aliopureno' 'allourinol' 'allpurinol'
'alluprinol' 'allupurino' 'allupurnoi' 'allurpurin' 'alopurinal'
'alopurinol' 'alpuranal' 'alpurinal' 'aspirin,ib'  'alipurinol'
'blindstudy' 'increaseal' 'indomethac' 'juststarte' 'startedzyl'
'zyloprim'   'itookallop' 'allupurina' 'allinpurin' 'alpurinol'
'allopurino' 'allpurnal'  'alapournia' 'allopurtno')
     then ghallop=1;
```

Perform Systematic Data Audits

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

Pay Extra Attention To Key Data

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

- Query all entries of critical variables (e.g., primary outcome)
- Extra attention to problematic variables (e.g., time-line-follow-back)
- Query all Serious Adverse Events ?

Derived Variables

Many analyses require creation of a derived variable from multiple data points

- Be especially careful in creating derived variables
- Include all relevant data elements
- Don't forget to account for missing data
- Be sure to look at frequencies and cross-tabs of derived variables prior to including in models

Creating a Derived Variable

Q1. Does child smoke Q2. Do household members smoke? Q3. Do caretakers smoke? New Var: Smoke_Exp	Q1. Unprotected sex primary partner? Q2. Unprotected sex with casual partner? Q3. Share needles? New Variable: HIV_Exp

Sample SAS Code for Derived Variable

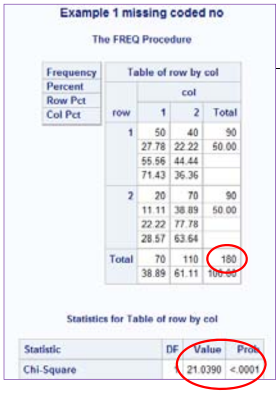
```
if (q1=1) or (q2=1) or (q3=1) then any_exp=1; else
    any_exp=2;

   proc freq;
     tables any_exp*site;
     run;
```

Any_Exp	Yes	No	Total
Site 1	50	40	90
Site 2	20	70	90
Total	70	110	180

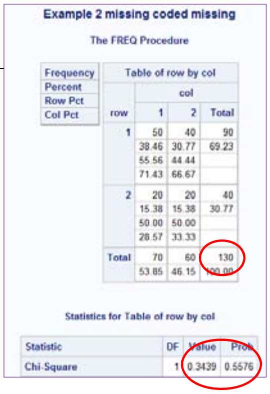
Sample SAS Code for Derived Variable

```
/* Corrected code to account for missing */
    q1=1) or (q2=1) or (q3=1) then any_expM=1; else
        if (q1=0) and (q2=0) and (q3=0) then any_expM=.; else
            any_expM=2;

   proc freq;
     tables any_expM*site;
     run;
```

Any_ExpM	Yes	No	Missing	Total
Site 1	50	40		90
Site 2	20	20	50	90
Total	70	60	50	180



```
Example 1 missing coded no

The FREQ Procedure

Frequency   Table of row by col
Percent
Row Pct              col
Col Pct     row      1      2   Total
              1     50     40      90
                 27.78  22.22   50.00
                 55.56  44.44
                 71.43  36.36
              2     20     70      90
                 11.11  38.89   50.00
                 22.22  77.78
                 28.57  63.64
          Total     70    110     180
                 38.89  61.11  100.00

Statistics for Table of row by col

Statistic     DF    Value    Prob
Chi-Square     1  21.0390  <.0001
```

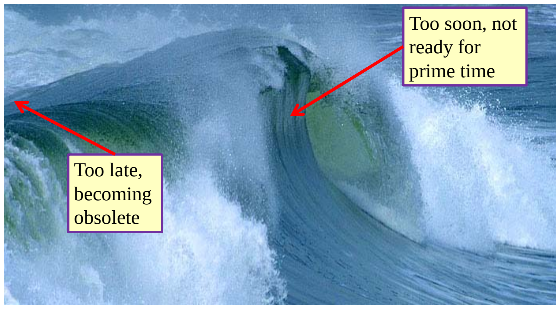


```
Example 2 missing coded missing

The FREQ Procedure

Frequency   Table of row by col
Percent
Row Pct              col
Col Pct     row      1      2   Total
              1     50     40      90
                 38.46  30.77   69.23
                 55.56  44.44
                 71.43  66.67
              2     20     20      40
                 15.38  15.38   30.77
                 50.00  50.00
                 28.57  33.33
          Total     70     60     130
                 53.85  46.15  100.00

Statistics for Table of row by col

Statistic     DF    Value    Prob
Chi-Square     1   0.3439  0.5576
```

What's up with the missing values?

- □ Go back and look at forms:
  - Is there an explanation?
  - Is the missing data differential?
- □ What are the implications?
  - Example: There are 2 sites and all the forms with missing values came from a single site
- □ Did you find this problem early enough to correct it?
- □ This is why you check "early and often"

A Caution About New Technology



Too soon, not ready for prime time

Too late, becoming obsolete

Smartphones

Benefits of smartphones

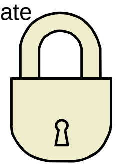
- Electronic data capture
- Secure if you use to connect to website with encryption
- SMS (text messaging) in addition to linking to web-form
- Easy to carry
- Everyone wants one
- “Sexy” so funders like the idea

Smartphone challenges

- Can be Expensive (hardware, data plans)
- Cannot encryption text messages
- One question per screen
- Small screens make view some question types difficult (e.g., grids)
- Navigating around questionnaire (going back) is challenging
- Battery life is short
- Attractive to thieves and easily stolen

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

Take Home Message

- Your team should include someone who understands data issues
- Budget for data management
- Planning ahead results in fewer revisions
- Check your data early and often
- If you do things correctly from the beginning:
  - It is less work
  - It is less expensive
  - You are more likely to discover the truth at the end

Questions?