

Data Management Solutions: Where is the Grass Greener?

Diane Levy, MS
diane.levy@columbia.edu
April 22, 2014

My Background

- Pre-Columbia
 - Community Health Education
 - Consulting
 - Training
- At Columbia (2001)
 - Data manager in Biostatistics (MSPH)
 - Created/taught P8180

Data Management: [Finally] Accepted as Necessary by Investigators

- Columbia has put security requirements into place
 - Investigators must now indicate which secure server they will be using
 - Must be certified by Columbia's IT department
 - https://secure.cumc.columbia.edu/cumcit/secure/security/scp_systems.html
- Funding agencies require a data management plan

Where is the grass greener?

- How do you decide which tools?
 - Clinical and Translational Science Award (CTSA)
 - Request a 1 hour consultation for data management
 - Best use: when writing a proposal
 - Alternatively: after receiving funding (but then you're limited by the funding you received and *maybe by the people you already hired*)
 - Must be faculty and have a UNI and an eRA commons ID
 - <http://www.mailman.columbia.edu/academic-departments/biostatistics/consultation-service>

What are the tools?

- What you hear:
 - Flat files
 - Relational databases
 - REDCap

Flat Structures: Generally Worst Choice

- Excel
- Statistical Software Packages
 - SAS
 - SPSS
 - R

The purpose of these packages is for data analysis.
Not for data collection and data management

Why?

- Excel: tries to be too helpful
 - Assumes all data to be dates or numbers
 - If your data contains text that resembles a date, Excel assumes date
 - If your data contains text that resemble a number, Excel assumes a number

Opening a text file of gene names in Excel

Text File

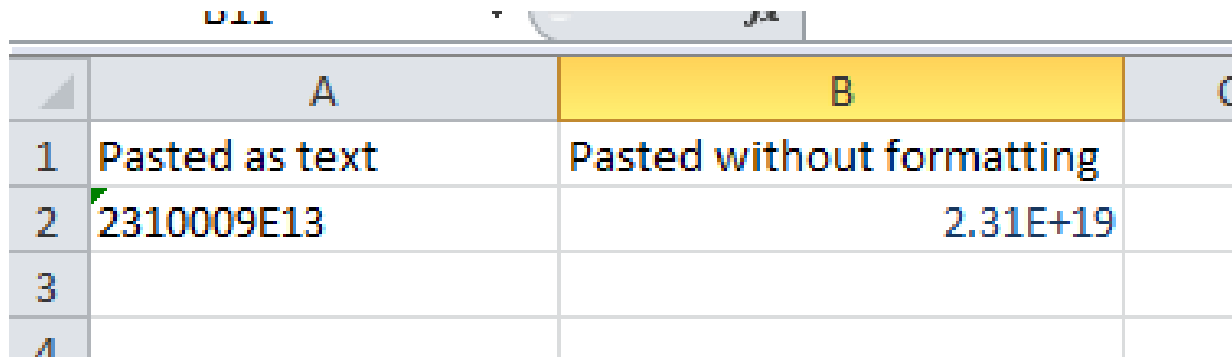
	A
GSTT1	GSTT1
CYP17A1	CYP17A1
XRCC6	XRCC6
LIG4	LIG4
NBN	NBN
MSH2	MSH2
GC	GC
APR1	1-Apr
APR2	2-Apr
APR4	4-Apr
APR5	5-Apr
DEC-1	1-Dec
DEC2	2-Dec
MAR1	1-Mar
MAR3	3-Mar
NOV1	1-Nov
NOV2	2-Nov

These are now dates

RIKEN Clone Identifiers

- In the form of: nnnnnnnnEnn (where n denotes a digit)
- identifiers are comprised of:
 - serial number of the plate that contains the library
 - information on plate status
 - address of the clone
- 2310009E13 would be converted irreversibly to the floating-point number "2.31E+19."

Example of Excel Assuming a Number



The image shows a screenshot of an Excel spreadsheet with three columns labeled A, B, and C. Column B is highlighted in yellow. Row 1 contains the text 'Pasted as text' in column A and 'Pasted without formatting' in column B. Row 2 contains the text '2310009E13' in column A and '2.31E+19' in column B. Rows 3 and 4 are empty.

	A	B	C
1	Pasted as text	Pasted without formatting	
2	2310009E13	2.31E+19	
3			
4			

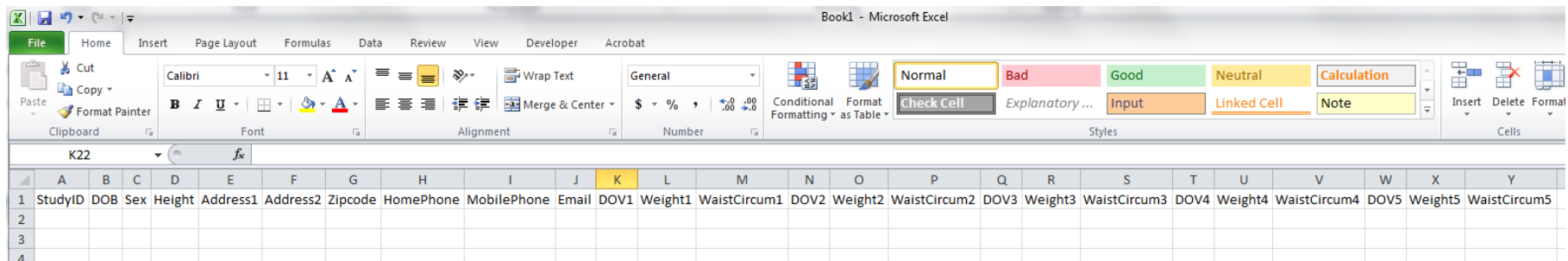
All Flat Files: Other Issues

- Structure of the data
 - Too wide
- VS
- Mismatched records

Example of a Study

- Weight Control Study
 - Initial Visit
 - Collection of demographic data
 - First weigh in
 - Subsequent:
 - Weekly weigh in visits for x number of months

Too Wide For Ease of Use



In 5 Visits: Up to column Y

Difficult to add a new variable

- What if you needed to add a comments column to capture notes for each visit?

And difficult to retrieve data when creating a dataset (TBD later!)

Using Multiple Sheets

	A	B	C	D	E	F	G	H	I	J
1	StudyID	DOB	Sex	Height	Address1	Address2	Zipcode	HomePhone	MobilePhone	Email
2	123									
3	345									
4	456									
5										

Demographic Sheet: One row per person

	A	B	C	D	E	F
1	StudyID	DOV	Weight	WaistCircum		
2	123	1/1/2000	192			
3	345	1/14/2000	256			
4	123	2/1/2000	189			
5	456	1/23/2000	199			
6	123	2/9/2000	188			
7						

Visit Sheet: One row per visit
(the combination of studyID and DOV must be unique)

Biggest Problem using Multiple Sheets

- Disconnected Records
 - Entering a visit with an incorrect Study ID

Limited Ability to Create User Friendly Interfaces

- Skip patterns are almost impossible to implement
- Validation is difficult to implement
 - Especially if validating one column based on info in another column
 - (e.g.: weight or height based on sex)
- Difficult to satisfy requirement of unique values
 - Especially if the combination of 2 columns makes a record unique
- Maintaining a regular structure requires discipline (no cutting and pasting over formats)

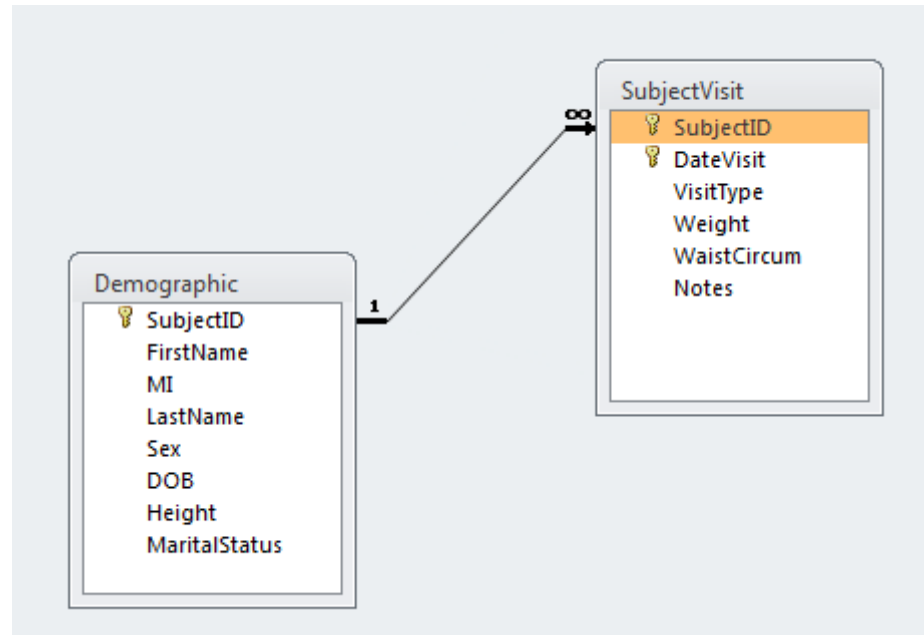
Relational Databases

- Access
- FileMaker
- FoxPro
- SQL Server
- MySql
- Oracle
- SIR

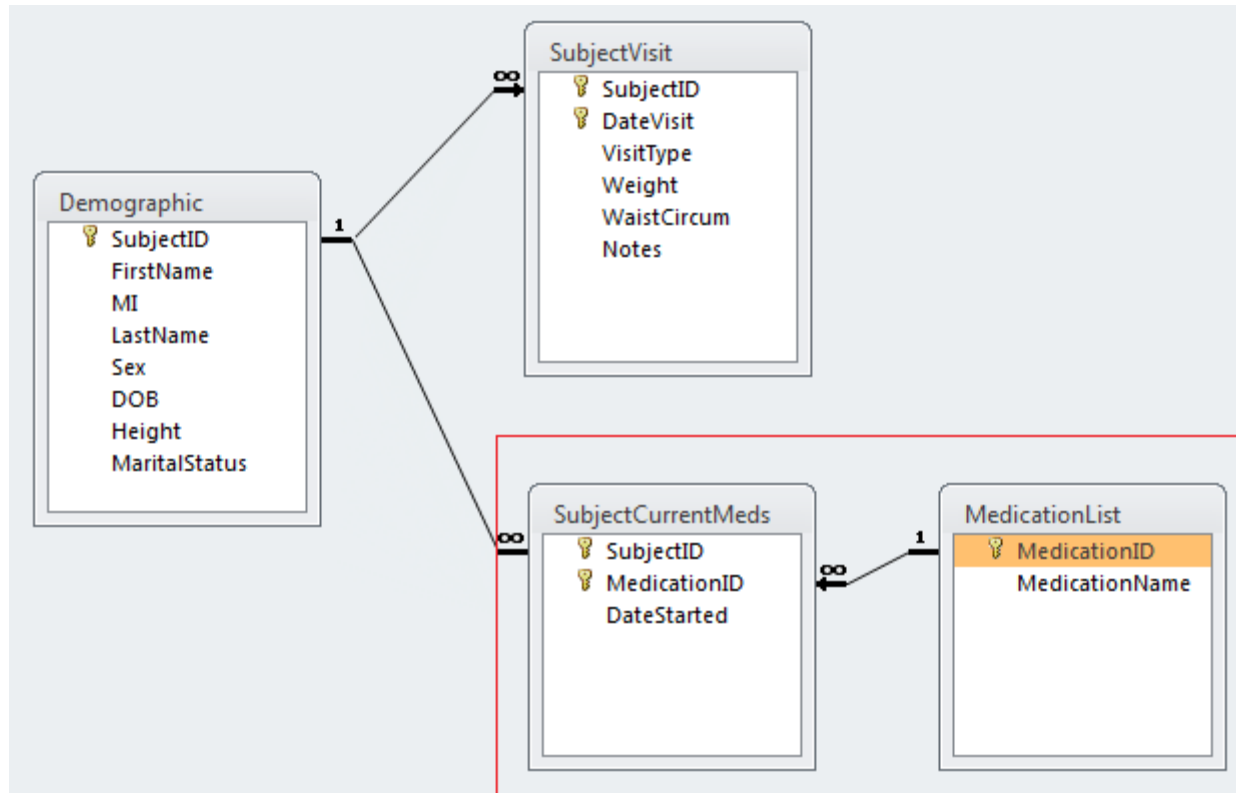
Can Accommodate Any Data

- Data are stored in consistent tables
 - Design is based on set theory so that:
 - Each table has a subject
 - Each record is a description of that subject
 - Each record in each table is unique
 - Additional tables can be added at any time
 - No redesign of database necessary

Example of Weigh-In Database



Add Current Medications (with lookup)



Relationships

- Represented by a line between 2 tables
- Programmed into the database
 - Automatically enters the StudyID into the visits table
(Access is automatic; otherwise programmable)
 - Automatically updates the StudyID in the visits table if user changes it in the demographic table
 - Prevents orphaned records
 - Can't add a visit for a person who isn't in the demographic table
 - Can't delete a person from the demographic table if visits exist

Strong Data Type Enforcement

- Decimals
- Integers
- Dates/times
- Text
- Memo

Enables User Friendly Interfaces

- Access: Has user interface programming ability built in
- MySQL, SQL Server, Oracle (and others)
 - Allow web interfaces to be built/programmed

SQL Query Language

- Universal querying language
- English based structure
- **Allows easy querying of well designed and structured relational databases**
 - Datasets for analysis can be compiled with minimal effort

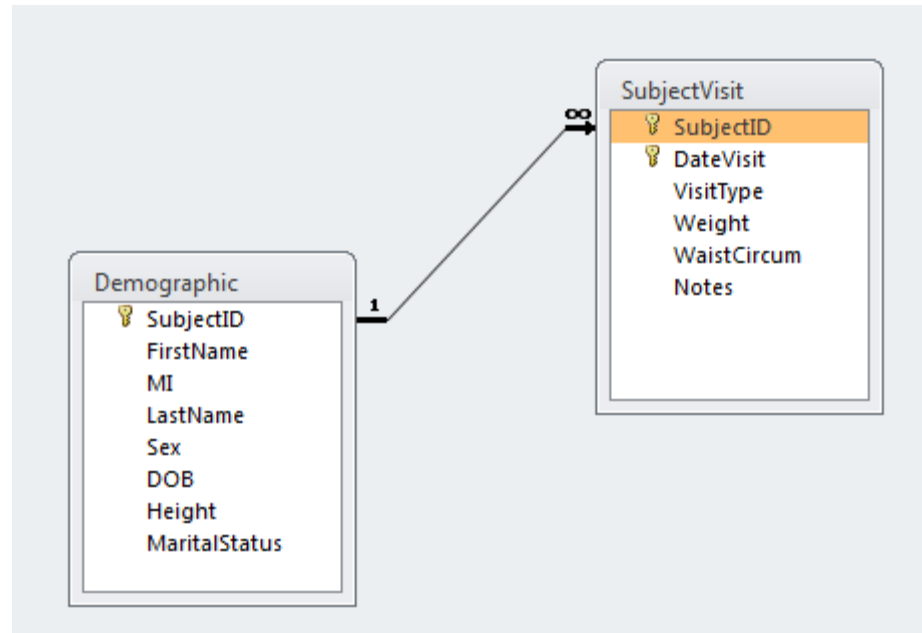
Simplicity of SQL

StudyID	DOB	Sex	Height	Address1	Address2	Zipcode	HomePhone	MobilePhone	Email
123	4/12/1982	1	73	123 Main St	Apt 34	11234			
345	6/18/1979	0	64	56 Long Ave	Suite 44	10255			
456	9/23/1985	1	69	99 North St		33345			

To find all the men born on or after January 1, 1982:

Select * from tblDemographic where Sex=1 and DOB >= '1/1/1982'

Relational Database Data Storage



Want to retrieve all subjectIDs of people who had visits in 2013

Peek at the Subject Visit Data

Unique Combination

SubjectID	DateVisit	VisitType	Weight	WaistCircum	Notes
1	7/4/2009	1	123		
1	11/5/2009	2	124		
1	3/3/2010	3	122		
1	7/8/2010	4	123		
1	11/9/2010	5	120		
2	6/26/2009	1	178		
2	10/28/2009	2	177		
2	2/28/2010	3	175		
2	6/4/2010	4	172		
3	1/11/2009	2	195		
3	5/20/2009	3	195		
3	9/3/2009	1	199		
3	9/1/2010	4	196		
4	4/27/2009	1	189		
4	8/29/2009	2	190		
4	12/26/2009	3	188		
5	5/5/2009	1	134		
5	9/2/2009	2	132		
5	1/6/2010	3	130		
5	5/5/2010	4	131		
8	10/31/2009	1	112		

```
SELECT distinct subjectID FROM
tblSubjectVisit where
DateVisit>= '1/1/2013' and
DateVisit<='12/31/2013'
```

**Query never changes
even when more people
or more visits for a
person are added**

Poorly Designed Data: Repeated Columns

Difficult to apply SQL to repeated column design used with flat files or incorrect table design

StudyID	DOB	Sex	Height	Address1	Address2	Zipcode	HomePhone	MobilePhone	Email	DOV1	Weight1	WaistCircum1	DOV2	Weight2	WaistCircum2	DOV3	Weight3	WaistCircum3	DOV4	Weight4	WaistCircum4	DOV5	Weight5	WaistCircum5

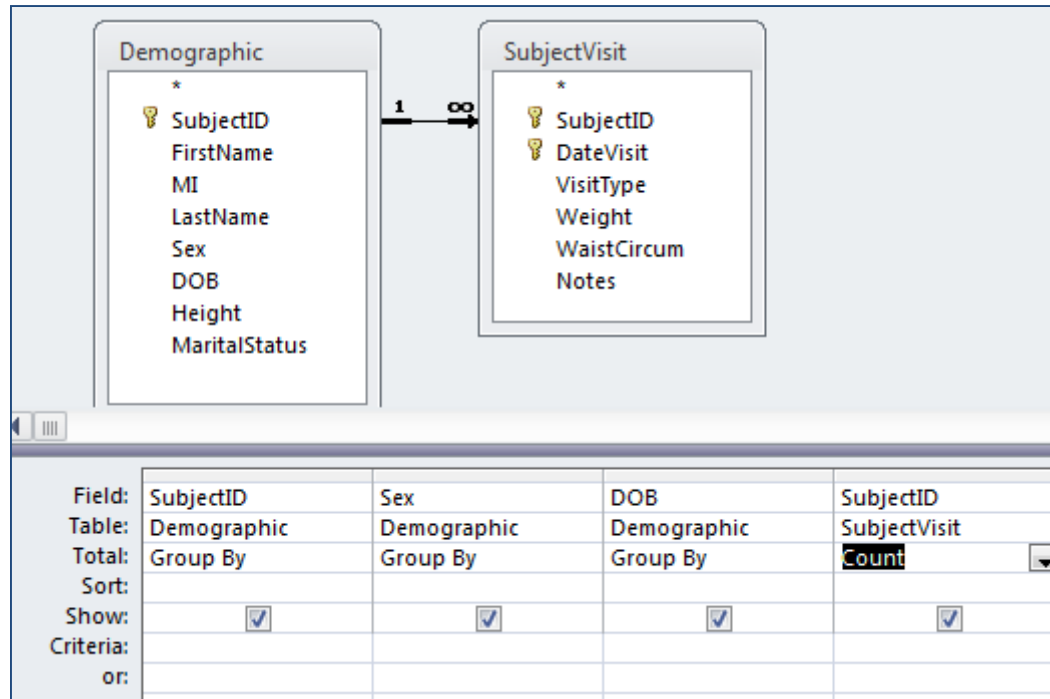
List the studyID of each participant who had a weigh in visit between 1/1/2013 and 12/31/2013:

```
Select StudyID from tblDemographic
where (DOV1 >= 1/1/2013 and DOV1 <= 12/31/2013)
or (DOV2 >= 1/1/2013 and DOV2 <= 12/31/2013)
or (DOV3 >= 1/1/2013 and DOV3 <= 12/31/2013)
or (DOV4 >= 1/1/2013 and DOV4 <= 12/31/2013)
or (DOV5 >= 1/1/2013 and DOV5 <= 12/31/2013)
```

Query will Grow

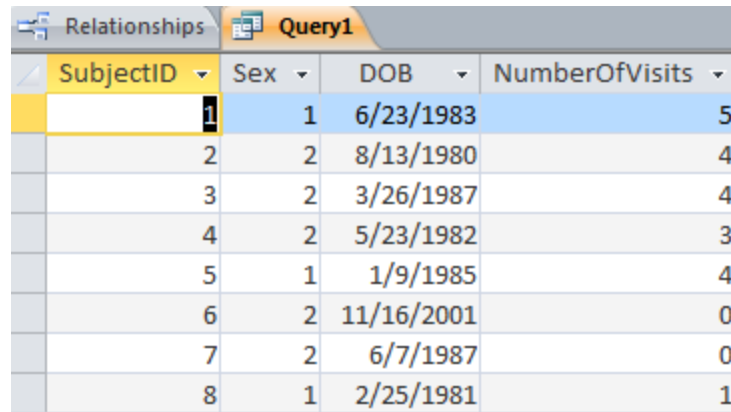
- The more visits a person can have (maximum number of visits for the study), the longer that query will have to be

Query Builder Tools



```
SELECT Demographic.SubjectID, Sex, DOB, Count(SubjectID) AS NumberOfVisits
FROM LEFT JOIN SubjectVisit ON Demographic.SubjectID = SubjectVisit.SubjectID
GROUP BY Demographic.SubjectID, Sex, DOB
```

Produces this Dataset



SubjectID	Sex	DOB	NumberOfVisits
1	1	6/23/1983	5
2	2	8/13/1980	4
3	2	3/26/1987	4
4	2	5/23/1982	3
5	1	1/9/1985	4
6	2	11/16/2001	0
7	2	6/7/1987	0
8	1	2/25/1981	1

Query results are a combination of demographic data and aggregated visit data

Can be exported to:

- Text
- Excel
- SAS

Access

- Access must be installed on an encrypted computer (*for use by 1 user at a time only*)
- or -
- Can be installed on a secure server

Access: Pros

- Has front end built in
- Easy to implement relational structure
 - Fast development

Concerns about using Access

- It is a file (1 file) and can be copied and moved
- Although it allows for password encryption, it can be skipped in design or broken
- Doesn't allow for different levels of security
- Can be used incorrectly
- Data can get too large and it can become corrupted

Other Options: Industrial Databases

- MySQL
- SQL Server
- Oracle

Industrial Databases: Pros

- Powerful
- Sophisticated
- Flexible
- Complex
- Enhanced security
 - Different security levels for different users
- Can be programmed onto the web
 - Easy collaboration (e.g.: multi-center studies)

Industrial Databases: Concerns

- Expensive
- Require server installation
- Requires working with a data programmer
- Longer development time to create data entry screens (usually web based)

All Databases: Development Steps

- Start with data structure first
 - Tables are designed to accommodate all data needed for the project
- Data entry screens follow
 - Duplicate of investigators' data collection tools
 - Data are entered into data entry screens and parsed (behind the scenes) into the tables
- These steps ensure ease of data going in cleanly and being retrieved easily

Sample Data Entry Screen: Weigh In Study

Demographic Data

SubjectID DOB Marital Status

Sex Height

Weigh In Visits

SubjectID	DateVisit	Weight
1	7/4/2009	123
1	11/5/2009	124
1	3/3/2010	122
1	7/8/2010	123
1	11/9/2010	120
1		

Normally
Hidden

1. Infinite number of visits
2. All automatically assigned correct subject ID

Sample of An Access Screen: List of Participants

EDC ID#	Version:	Subject Initials	Interviewer	Interview Date
ED001	24 to 36 mos	CN	JRC	3/9/2012
ED001	36 to 48 mos	CN	JRC	3/6/2013
ED002	24 to 36 mos	EP	JRC	2/28/2013
ED002	36 to 48 mos			
ED003	36 to 48 mos			
ED004	24 to 36 mos	YNF	JRC	4/5/2012
ED004	36 to 48 mos	RD	YNF	3/21/2013
ED005	36 to 48 mos	JE	YNF	3/27/2013
ED006	36 to 48 mos			
ED007	36 to 48 mos			
ED008	36 to 48 mos			
ED009	36 to 48 mos			
ED010	24 to 36 mos	BRG	JRC	4/21/2012
ED010	36 to 48 mos			
ED011	36 to 48 mos			

Exit Edit Add Interviewers View/Update Issues

Record: 2 of 384 No Filter Search

Preloaded StudyIDs

Page 1: Duplicate of Paper Questionnaire

Interview Page 1

Date: 03/06/2013 EDC_ID: ED001 Subject Initials: CN Interviewer's Initials: JRC Enter Issue

Hello, my name is [redacted]. I want to start by thanking you for your help with this survey. All of your answers to these questions are completely confidential. If you feel uncomfortable with any questions, you may choose not to answer them. We appreciate you being honest in all your answers. We are updating information from previous interviews, so you may be familiar with several items on the questionnaire. Before we begin, I would like you to have your baby's vaccination record or "blue card" with you, as I will be asking for some of this information later.

Do you have any questions before we begin?

Confirm the interview is with the mother; if not identify the relationship of the person with whom you are interviewing. This person should be the primary caregiver.

Biological Father
Adoptive Mother
Adoptive Father
Grandmother
Legal Guardian
Other Adult
Refused
Don't know
N/A

Specify Other Relation: [redacted]

Which form is being administered? English Spanish

1. What language is spoken at home? (Select Yes for as many as apply)

English: Not checked as spoken at home Yes spoken at home

Spanish: Not checked as spoken at home Yes spoken at home

Chinese: Not checked as spoken at home Yes spoken at home

Russian: Not checked as spoken at home Yes spoken at home

French: Not checked as spoken at home Yes spoken at home

German: Not checked as spoken at home Yes spoken at home

Other: Not checked as spoken at home Yes spoken at home Other Language: [redacted]

2. How many years of school have you completed? 13 -1 = refused; -2 = Don't know; -3 = N/A

3. Are you currently in school?: Yes No Refused

<<Previous Next>> Notes: [redacted]
Go To Page Close

Page 7: Data are Saved into Separate Tables

Interview Page 7

EDC_ID: ED181 Enter Issue

C. Early Intervention

33. Since the birth of your child, has he/she received any special services such as Early Intervention, speech therapy, physical therapy, mental health services, or behavioral therapy or other special services?

Yes
No
Refused
Don't know
N/A

34. If yes, did your child receive any of the following services?

E Service	Yes/No	Specify (Other)
▶ Speech therapy	Yes	
*		

<<Previous Next>> Notes:

Go To Page Close

→ Saved in demographic table

1. Saved in Early Intervention table
2. Data entry is only allowed if previous question's answer is YES

Example of Web Based Screen

https://www.sac-cu.org/AsCohort/DataEntry/NHChildPage4.aspx

HEALTH EFFECTS OF ARSENIC
Programmed By The Statistical Analysis Center

[Home](#) [GoTo Questionnaire](#) [Download Only](#) [Admin ONLY](#) [Change Password](#)

NH Well & Demographic Questionnaire Page 4 (Child ID = 62)

13. Have you ever noticed that your water causes your bathtub, toilet or sinks to turn black?

Answered yes but no fixtures on questionnaire

14. When the child is home, does he/she drink the tap water from this well? (Include all drinks mixed with water such as tea, hot chocolate, kool-aid, etc.)

15. When your child is home, does he/she have any other source of drinking water?

	Source	Description
Edit Delete	2	under sink filter (lead micro organism)

Answered yes but no sources on questionnaire

16. If the child has an alternative source of drinking water, how often does the child use the alternative source?

All the time (100%)
Often (75%)
Occasionally (50%)
Rarely (25%)
Never

Question #13

Other sources:
saved to another
table

Summary

- Table design is first in process
- Data entry screens (with endless validation) follow
- Data are stored in various tables in an organized manner (relational structure)
 - Users are oblivious
- Data set building (retrieval) using SQL is easy
- Datasets themselves are stored
- When using server based database technology, the entire database cannot be copied to a flashdrive or other computer

[Home](#)[My Projects](#)[+ Create New Project](#)[🎓 Training Resources](#)

Welcome to REDCap!

REDCap is a secure, web-based application for building and managing online surveys and databases. Using REDCap's stream-lined process for rapidly developing projects, you may create and design projects using 1) the online method from your web browser using the Online Designer; and/or 2) the offline method by constructing a 'data dictionary' template file in Microsoft Excel, which can be later uploaded into REDCap. Both surveys and databases (or a mixture of the two) can be built using these methods.

REDCap provides automated export procedures for seamless data downloads to Excel and common statistical packages (SPSS, SAS, Stata, R), as well as a built-in project calendar, a scheduling module, ad hoc reporting tools, and advanced features, such as branching logic, file uploading, and calculated fields.

REDCap: What is it?

- Developed by Vanderbilt University's Informatics Department
 - Funded by the NIH
- “Free” download for **researchers**
 - Not to be used for non-research purposes
 - Never in a for-profit environment
 - Including clinical practices

More about “Free”

- Must be installed on a secure server
 - Deemed secure by CUIT
- Servers must be maintained by network administrators
 - Not programmers
- Must be upgraded several times a year
- A REDCap administrator must be the one to:
 - Create new projects
 - Create new IDs
 - Approve significant changes to a project

CTSA and REDCap

- CTSA supports the use of REDCap at CUMC
 - Biostatistics is funded to maintain it
 - Anyone eligible for CTSA support can request:
 - a consultation (required if REDCap is to be used)
 - a new project
 - necessary IDs
- Users of our REDCap installation must cite:
 - the CTSA in their publications
 - REDCap in their publications

Programming REDCap

- Quick learning curve
- Training videos
- Programming concentrates on the user interface
 - Online designer
 - Spreadsheet data dictionary
 - User learns how to do this after using data designer and downloading into spreadsheet

Programming REDCap

- REDCap starts with the user interface
- It builds a flat table structure into MySQL
 - The data are not stored in a relational structure
 - Downloaded data look much the same as a large repeated column spreadsheet
 - Creation of datasets (using aggregation) is more difficult
 - Can't easily run SQL against the data to create a dataset for analysis

REDCap: Creating Datasets

- Investigators export to a specified format (e.g.: Excel, SAS, SPSS)
 - Dataset creation occurs in that software package
 - That snapshot of the data cannot be saved in REDCap
 - Finding the dataset that was used for a specific publication is harder
 - Recreation of the dataset in REDCap may be impossible
 - **Important:** the data in REDCap itself are constantly evolving!
(Not static)

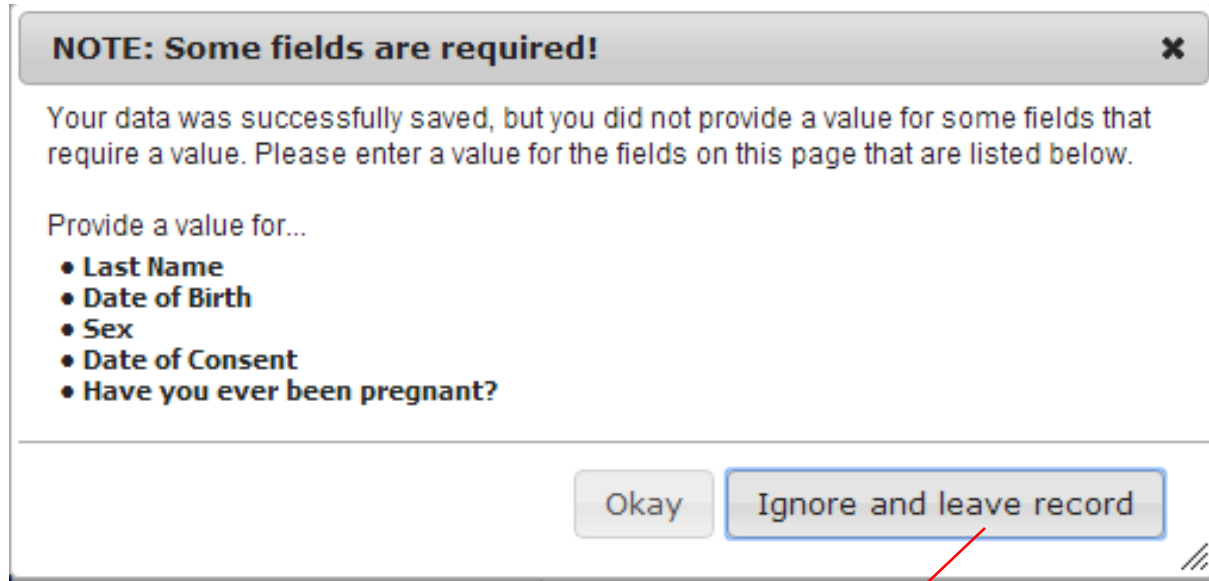
REDCap: Building User Interface

- Can create screens that look similar to questionnaires
 - Vertical listing of questions only
 - Not possible to have 2 questions on 1 line
- Use of the following allow for consistent and (somewhat*) clean data:
 - Coding (drop down choices)
 - Data types (only dates allowed in date variables)
 - Skip patterns

*Validation is limited

- Validation rules cannot be based on values entered in previous fields
 - i.e.: weights or heights based on sex
- **Required and range checking are only suggestions**
 - System will warn user if data are being left out or entered out of range
 - But will allow the data to be entered and record to be saved
- Data cleaning will be necessary

Warning Message



User can save the record without entering required variables

Difficult to Program Repetition

- Medications
- Allergies
- Information about children
- Adverse events
- Anything unscheduled

Example of Capturing Medications

Are you taking medications? Yes No

Form Status

Complete?

Lock this record for this form?
If locked, no user will be able to edit this record on this form until someone with Lock/Unlock privileges unlocks it. Lock

Medications

How many times have you been pregnant?
Min=1; Max=5 (for adolescent girls)

Are you taking medications? Yes No

How many medications are you taking?
Enter a number from 1 to 3

Form Status

Complete? Incomplete ▾

Maximum of 3 medications

Data Entry for Multiple Medications

Are you taking medications?	<input checked="" type="radio"/> Yes <input type="radio"/> No
How many medications are you taking?	<input type="text" value="2"/> Enter a number from 1 to 3
Name of first medication	<input type="text"/>
Dose of first medication	<input type="text"/> Enter the dose including units and # times/day
Reason for taking first medication	<input type="text"/>
Name of 2nd medication	<input type="text"/>
Dose of 2nd medication	<input type="text"/> Enter the dose including units and # times/day
Reason for taking 2nd medication	<input type="text"/>

#1

#2

Problem

- If more than 3 medications, the application would have to be modified in order to accept another medication

Very different from the relational model where a grid would be available to enter as many medications as needed

The Medications: Behind the scenes (table storage)

A	B	C	D	E	F	G	H	I	J	K	L	M
participantid	redcap_event_name	meds	nummeds	med1	dose1	medreason1	med2	dose2	medreason2	med3	dose3	medreason3
1	interview_arm_1	1	2	Motrin	200 mg every 4-6 hours	abdominal pain	Lipitor	20 mg	high cholesterol			
1	visit_1_arm_1											
1	visit_2_arm_1											
1	visit_3_arm_1											
2	interview_arm_1	1	2	amoxicillin	250 mg every 8 hours	sinus infection	Sudafed	30 mg every 6 hours.	congestion			
2	visit_1_arm_1											
2	visit_2_arm_1											
3	interview_arm_1	1	2									

Several Issues:

- The more medications, the more columns
 - If demographics were included, sheet would be extremely wide
 - Saves columns even when no data (medication 3)
 - Very difficult to run SQL against this to aggregate (create dataset)
- There is more information here than requested from REDCap
 - Because study consists of scheduled visits, REDCap exports data with lines saved for those visits

Scheduled Events

- Predictable
 - Therefore can be programmed
 - A tool can be used to create a dashboard

REDCap Version: Weigh In Project

weightcontrol

demographicInfo

Share this instrument

VIDEO: Basic data entry (16 min)

Download PDF of - select PDF download option -

Editing existing Record ID 15

Event Name: Register

Record ID 15
(To rename this record, modify the value immediately below.)

Record ID

DateOfBirth Today M-D-Y

Gender

Do you have a dog?

weight at registration:

Form Status

Complete?

Lock this record for this form?

If locked, no user will be able to edit this record on this form until someone with Lock/Unlock privileges unlocks it.

Lock

Save Record


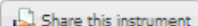
Save and Continue


-- Cancel --


Delete Record

REDCap Version: Weigh In Project

weightcontrol


 **WeighIn** 

 Download PDF of

 Editing existing Record ID 1

Event Name: **First Weigh In**

Record ID: 1


Date of weigh in:  Today M-D-Y

weight:

Waist Circum (inches):

Form Status


Complete?


Lock this record for this form?  Lock


If locked, no user will be able to edit this record on this form until someone with Lock/Unlock privileges unlocks it.


REDCap Version: Weigh In Project


weightcontrol

→  **bloodwork**

 Share this instrument


 [VIDEO: Basic data entry \(16 min\)](#)

 Download PDF of

 Editing existing Record ID 1

Event Name: **First Blood Draw**

Record ID 1

date of blood draw  Today M-D-Y

Arsenic


Lead

Selenium

Form Status

Complete?

Lock this record for this form?
If locked, no user will be able to edit this record on this form until someone with Lock/Unlock privileges unlocks it.




 Lock

Dashboard









































































Record Status Dashboard (all records)

Displayed below is a table listing all existing records/responses and their status for every data collection instrument (and if longitudinal, for every event). You may click any of the colored buttons in the table to open a new tab/window in your browser to view that record on that particular data collection instrument. Please note that if your form-level user privileges are restricted for certain data collection instruments, you will only be able to view those instruments, and if you belong to a Data Access Group, you will only be able to view records that belong to your group.

Legend for status icons:

-  Incomplete
-  Unverified
-  Complete

Displaying record "1" through "21" of 12 records

Record ID	demographicInfo Register	WeighIn First Weigh In	WeighIn Second Weigh In	WeighIn Third Weigh In	bloodwork First Blood Draw	bloodwork Second Blood Draw
1						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

Represents 3 programmed forms:

- Demographic Information
- Weigh In Information
- Blood test results

This project scheduled:

- 1 demographic record
- 3 visits
- 2 blood draws

Export: Requesting Specific Variables

Form: demographicinfo	
Record ID (<i>record_id</i>)	<input checked="" type="checkbox"/>
DateOfBirth (<i>dateofbirth</i>)	<input checked="" type="checkbox"/>
Gender (<i>gender</i>)	<input checked="" type="checkbox"/>
Do you have a dog? (<i>dog</i>)	<input type="checkbox"/>
weight at registration: (<i>regweight</i>)	<input checked="" type="checkbox"/>
Form Status	
Complete? (<i>demographicinfo_complete</i>)	<input type="checkbox"/>
Form: WeighIn	
Date of weigh in: (<i>dateweight</i>)	<input checked="" type="checkbox"/>
Age at weigh in: (<i>weight_age</i>)	<input type="checkbox"/>
weight (<i>weight</i>)	<input checked="" type="checkbox"/>
Pounds Lost: (<i>lblast</i>)	<input type="checkbox"/>
Waist Circum (inches): (<i>waistcircum</i>)	<input checked="" type="checkbox"/>
Form Status	
Complete? (<i>weighin_complete</i>)	<input type="checkbox"/>
Form: bloodwork	
date of blood draw (<i>blooddate</i>)	<input type="checkbox"/>
Arsenic (<i>arsenic</i>)	<input type="checkbox"/>
Lead (<i>lead</i>)	<input type="checkbox"/>
Selenium (<i>selenium</i>)	<input type="checkbox"/>
Form Status	
Complete? (<i>bloodwork_complete</i>)	<input type="checkbox"/>

Downloaded Data

A	B	C	D	E	F
record_id	redcap_event_name	dateofbirth	regweight	dateweight	weight
1	register_arm_1	5/12/1980	195	3/1/2013	189
1	first_weigh_in_arm_1			3/1/2013	185
1	first_blood_draw_arm_1				
1	second_weigh_in_arm_1			4/1/2013	175
1	second_blood_draw_arm_1				
1	third_weigh_in_arm_1			5/1/2013	162
11	register_arm_1	5/13/1986			
11	first_weigh_in_arm_1			5/22/2013	153
11	second_weigh_in_arm_1			5/24/2013	123
12	register_arm_1	1/7/2013			
12	first_weigh_in_arm_1			5/1/2013	2
13	register_arm_1	2/11/1999			
13	first_weigh_in_arm_1			5/1/2013	12
14	register_arm_1	1/7/2013	14		
14	first_weigh_in_arm_1			5/5/2013	12
15	register_arm_1				
16	register_arm_1				
17	register_arm_1				
18	register_arm_1				
19	register_arm_1				
20	register_arm_1				
21	register_arm_1	1/15/1982			
21	first_weigh_in_arm_1			2/20/2013	200
21	first_blood_draw_arm_1				
21	second_weigh_in_arm_1			3/22/2013	195
21	second_blood_draw_arm_1				
21	third_weigh_in_arm_1			5/1/2013	175

Demographic Data

Weigh In Data

Space saved for blood data

Request of All Variables

A	B	C	D	E	F	G	H	I	J	K	L
record_id	redcap_event_name	dateofbirth	gender	regweight	dateweight	weight	waistcircum	blooddate	arsenic	lead	selenium
1	register_arm_1	5/12/1980	1	195	3/1/2013	189		3/6/2013	25	13	
1	first_weigh_in_arm_1				3/1/2013	185	50				
1	first_blood_draw_arm_1							3/6/2013	33	44	31
1	second_weigh_in_arm_1				4/1/2013	175	51				
1	second_blood_draw_arm_1							4/2/2013	32	41	12
1	third_weigh_in_arm_1				5/1/2013	162	49				
11	register_arm_1	5/13/1986	0					1/1/2012	0.1	0.2	
11	first_weigh_in_arm_1				5/22/2013	153	54				
11	second_weigh_in_arm_1				5/24/2013	123	22				
12	register_arm_1	1/7/2013	0					1/2/2012	0.1	0.2	
12	first_weigh_in_arm_1				5/1/2013	2	7				
13	register_arm_1	2/11/1999	1					1/3/2012	0.1	0.2	
13	first_weigh_in_arm_1				5/1/2013	12	12				
14	register_arm_1	1/7/2013	1	14				1/4/2012	0.1	0.2	
14	first_weigh_in_arm_1				5/5/2013	12	2				
15	register_arm_1							1/5/2012	0.1	0.2	
16	register_arm_1							1/6/2012	0.1	0.2	
17	register_arm_1							1/7/2012	0.1	0.2	
18	register_arm_1							1/8/2012	0.1	0.2	
19	register_arm_1							1/9/2012	0.1	0.2	
20	register_arm_1							1/10/2012	0.1	0.2	
21	register_arm_1	1/15/1982									
21	first_weigh_in_arm_1				2/20/2013	200					
21	first_blood_draw_arm_1							3/19/2013	22	11	
21	second_weigh_in_arm_1				3/22/2013	195					
21	second_blood_draw_arm_1							4/10/2013	12	3	
21	third_weigh_in_arm_1				5/1/2013	175					

REDCap:Positives

- Excellent for surveys and simple questionnaires
- Set up to be multi-user
 - On the web
- Economical
- Secure
 - Handles multilevel security
- Good reporting tools
- Multiple formats available for downloading

Negatives

- Can't accommodate unscheduled events
 - Adverse events
 - Unscheduled visits
 - Complex protocols
 - If there's a death, fill out form x
- Exported data require manipulation for analysis
 - Exported datasets cannot be stored in the system for later retrieval
- More advanced programming requirements necessitate programming expertise
 - Example: lookup drop down boxes

	Text Files ^a	Spreadsheets ^a	Statistical Software Data Files ^a	REDCap ^a	File-based Databases ^a	Server-based Databases ^a	Clinical Trials Management Software ^a
Number of simultaneous users ^a	1 ^a	1 ^a	1 ^a	No limit ^a	<20 ^a	No limit ^a	No limit ^a
Differential access by user/site ^a	— ^a	— ^a	— ^a	Yes ^a	No ^a	Yes ^a	Yes ^a
Web access ^a	No ^a	No ^a	No ^a	Yes ^a	No ^a	Yes ^a	Yes ^a
Other remote access (app, ftp, etc.) ^a	No ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	No ^a
Programming expertise required ^a	No ^a	No ^a	Often ^a	Often ^a	Often ^a	Yes ^a	Often ^a
Calendar/scheduling ^a	No ^a	No ^a	No ^a	Built-in ^a	Custom ^a	Custom ^a	Built-in ^a
Project management ^a	No ^a	No ^a	No ^a	Custom ^a	Custom ^a	Custom ^a	Built-in ^a
Tracks modifications to data (audit trail) ^a	No ^a	No ^a	No ^a	Built-in ^a	Custom ^a	Custom ^a	Built-in ^a
Data stored natively in tables ^a	No ^a	No ^a	Yes ^a	No ^a	Yes ^a	Yes ^a	Yes ^a
Enforces datatypes ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Enforces value lists ^a	No ^a	Yes ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Enforces value ranges ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a
Enforces complex validation ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	No ^a
Parent-child/one-to-many structures ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a
Intra-record logic checks ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Intra-/inter-table logic checks ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	No ^a
Simple data querying ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Complex user data querying ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	No ^a
Unexpected events (adverse events, unscheduled visits, etc.) ^a	No ^a	No ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a
Exports to multiple formats ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Imports from multiple formats ^a	No ^a	No ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a	Yes ^a
Security ^a	0 ^a	+ ^a	++ ^a	++++ ^a	++++ ^a	++++ ^a	++++ ^a
Ease of data entry ^a	0 ^a	++ ^a	+ ^a	+++ ^a	++++ ^a	++++ ^a	+++ ^a



Thank You!

- Please email if you would like a copy of the slides and/or a summary article written by my colleague Richard Buchsbaum
- Diane.Levy@columbia.edu