Data Management Solutions: Where is the Grass Greener?
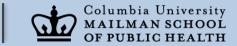
Diane Levy, MS
diane.levy@columbia.edu
April 22, 2014

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

My Background

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
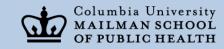OF PUBLIC HEALTH

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
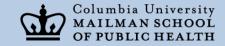   - http://www.mailman.columbia.edu/academic-departments/biostatistics/consultation-service

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

What are the tools?

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
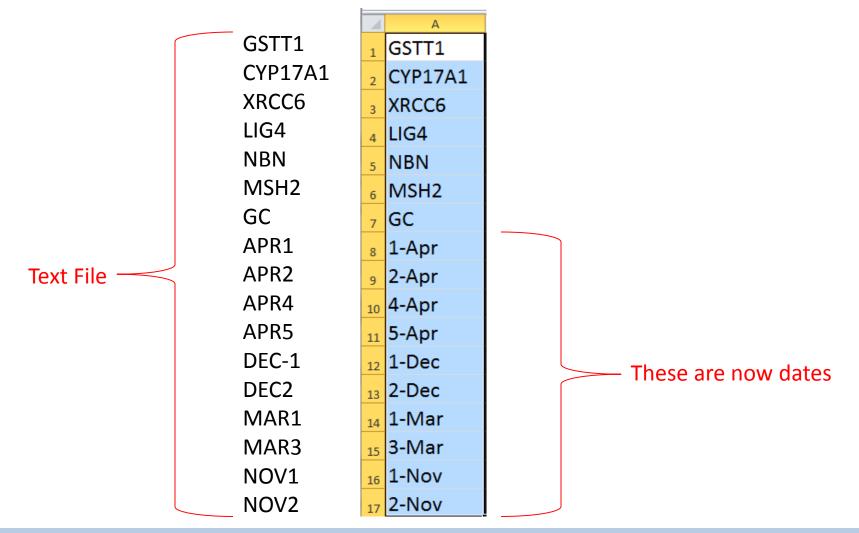MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Opening a text file of gene names in Excel

Text File
- GSTT1
- CYP17A1
- XRCC6
- LIG4
- NBN
- MSH2
- GC
- APR1
- APR2
- APR4
- APR5
- DEC-1
- DEC2
- MAR1
- MAR3
- NOV1
- NOV2

	A
1	GSTT1
2	CYP17A1
3	XRCC6
4	LIG4
5	NBN
6	MSH2
7	GC
8	1-Apr
9	2-Apr
10	4-Apr
11	5-Apr
12	1-Dec
13	2-Dec
14	1-Mar
15	3-Mar
16	1-Nov
17	2-Nov

These are now dates

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

RIKEN Clone Identifiers

- In the form of: nnnnnnnEnn (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."
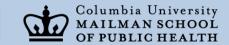
THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Example of Excel Assuming a Number

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

All Flat Files: Other Issues

- Structure of the data
  - Too wide

    VS

  - Mismatched records

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Too Wide For Ease of Use

In 5 Visits: Up to column Y

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Difficult to add a new variable

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

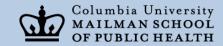**And difficult to retrieve data when creating a dataset (TBD later!)**

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University MAILMAN SCHOOL OF PUBLIC HEALTH

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

G21

	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)

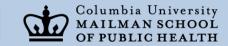THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Biggest Problem using Multiple Sheets

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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)

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Relational Databases

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

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
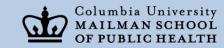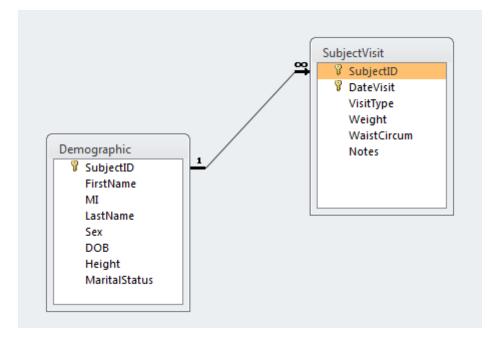MAILMAN SCHOOL
OF PUBLIC HEALTH

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
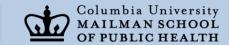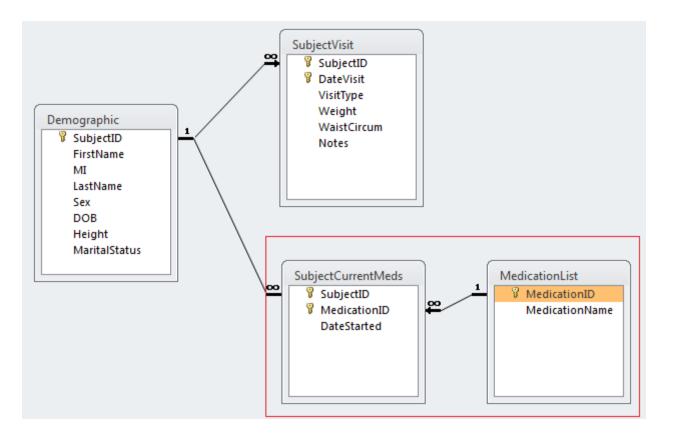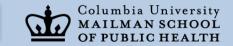   - No redesign of database necessary

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Example of Weigh-In Database

Add Current Medications (with lookup)

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Strong Data Type Enforcement

- Decimals

- Integers

- Dates/times

- Text

- Memo

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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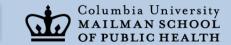
Select * from tblDemographic where Sex=1 and DOB >= '1/1/1982'

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Relational Database Data Storage

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

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University MAILMAN SCHOOL OF PUBLIC HEALTH

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)

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Access: Pros

- Has front end built in

- Easy to implement relational structure
  - Fast development

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Other Options: Industrial Databases

- MySQL
- SQL Server
- Oracle

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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)

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Sample Data Entry Screen: Weigh In Study

Normally Hidden

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Sample of An Access Screen:
List of Participants

Preloaded StudyIDs

Page 1: Duplicate of Paper Questionnaire

Page 7: Data are Saved into Separate Tables

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Example of Web Based Screen

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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.

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Programming REDCap

- REDCap <u>starts</u> 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)

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

*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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Warning Message

NOTE: Some fields are required! ✕

Your data was successfully saved, but you did not provide a value for some fields that require a value. Please enter a value for the fields on this page that are listed below.

Provide a value for...

- **Last Name**
- **Date of Birth**
- **Sex**
- **Date of Consent**
- **Have you ever been pregnant?**

Okay | Ignore and leave record

User can save the record without entering required variables

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Difficult to Program Repetition

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Example of Capturing Medications

Medications

Maximum of 3 medications

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH
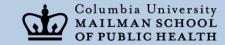
Data Entry for Multiple Medications

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 DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

The Medications: Behind the scenes (table storage)

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Scheduled Events

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

REDCap Version: Weigh In Project

REDCap Version: Weigh In Project

REDCap Version: Weigh In Project

Dashboard

Represents 3 programmed forms:
- Demographic Information
- Weigh In Information
- Blood test results

This project scheduled:
- 1 demographic record
- 3 visits
- 2 blood draws

THE DEPARTMENT OF **BIOSTATISTICS**

Columbia University MAILMAN SCHOOL OF PUBLIC HEALTH

Export: Requesting Specific Variables

Form: demographicInfo

Record ID *(record_id)*	☑
DateOfBirth *(dateofbirth)*	☑
Gender *(gender)*	☑
Do you have a dog? *(dog)*	☐
weight at registration: *(regweight)*	☑
Form Status	
Complete? *(demographicinfo_complete)*	☐

Form: WeighIn

Date of weigh in: *(dateweight)*	☑
Age at weigh in: *(weight_age)*	☐
weight *(weight)*	☑
Pounds Lost: *(lblost)*	☐
Waist Circum (inches): *(waistcircum)*	☑
Form Status	
Complete? *(weighin_complete)*	☐

Form: bloodwork

date of blood draw *(blooddate)*	☐
Arsenic *(arsenic)*	☐
Lead *(lead)*	☐
Selenium *(selenium)*	☐
Form Status	
Complete? *(bloodwork_complete)*	☐

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Downloaded Data

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

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

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

THE DEPARTMENT OF
BIOSTATISTICS

Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH