Getting Help for your Biostatistics Questions & Database Basics

Dan Eastwood, MS, Biostatistician
Medical College of Wisconsin, Division of Biostatistics

Friday, October 5, 2012
12:00-1:00 pm
The Medical College of Wisconsin is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

The Medical College of Wisconsin designates this Live activity for a maximum of 1.0 AMA PRA Category 1 Credit(s)™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Hours of Participation for Allied Health Professionals
The Medical College of Wisconsin designates this activity for up to 1.0 hours of participation for continuing education for allied health professionals.
Financial Disclosure

• In accordance with the ACCME® standard for Commercial Support Number 6, all in control of content disclosed any relevant financial relationships. The following in control of content had no relevant financial relationships to disclose.

Name: Ruta Brazauskas, PhD
Role in Meeting: Planning Committee

Name: Haley Montsma, BBA
Role in Meeting: Planning Committee

Name: Dan Eastwood, MS
Role in Meeting: Speaker
Learning Objectives

• Discover the capabilities and resources available within the Biostatistics Consulting Service

• Transition your research idea into a testable hypothesis

• Effectively organize research data
Evaluation Forms

Your opinion matters!
Help us plan future meetings, by completing and submitting your evaluation forms.

Thank you.
Getting Help for your Biostatistics Questions

Dan Eastwood, MS
Understanding Statistics

I used to think correlation implied causation.

Then I took a statistics class. Now I don't.

Sounds like the class helped. Well, maybe.

XKCD.com
Why Biostatistics Consulting?

• Shared experience
• Discuss your study
• Consider alternate views
• Formulate ideas into hypotheses
Why Biostatistics Consulting?

Why might you need help?
• “I’ve got this research idea about ...”
• ...
• ...
• ...
• “The reviewers asked me to ...”
Why Biostatistics Consulting?

Why might you need help?
• “I’ve got this research idea about ...”
• “How large should my sample be?”
• ...
• ...
• “A significant result! What does it mean?”
• “The reviewers asked me to ...”
Why Biostatistics Consulting?

Why might you need help?
• “I’ve got this research idea about ...”
• “How large should my sample be?”
• “I need help organizing my data.”
• “How do I perform a Chi-square test?”
• “A significant result! What does it mean?”
• “The reviewers asked me to ...”
Questions for the Investigator

• What is your hypothesis?
• What is the experimental design?
• What type of data are available?
• What is the plan for analysis?
Questions for the Investigator

What makes a good hypothesis? (1)

• A simple sentence, from which the null and alternate hypothesis should be clear
• The alternate hypothesis should be reasonable (power and effect size)
• Difference, equivalence, agreement
• No hypothesis- a descriptive study
Questions for the Investigator

What makes a good hypothesis? (2)

• Statistical and clinical significance
• Related factors
• Preliminary data
## It Never Hurts to Ask

<table>
<thead>
<tr>
<th>Independent Samples</th>
<th>Sleep difficulty, Medication X?</th>
<th>Sleep difficulty, Medication Y?</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>“No”</td>
<td>48</td>
<td>78</td>
<td>126</td>
</tr>
<tr>
<td>“Yes”</td>
<td>136</td>
<td>106</td>
<td>242</td>
</tr>
<tr>
<td>Total</td>
<td>184</td>
<td>184</td>
<td>368</td>
</tr>
</tbody>
</table>

Chi-Square test, $p=0.0010$, Medication X 57.6%, Medication Y 73.9%
## It Never Hurts to Ask

<table>
<thead>
<tr>
<th>Paired Data</th>
<th>Med Y, No difficulties</th>
<th>Med Y, Sleep Difficulties</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Med X, No difficulties</td>
<td>34</td>
<td>44</td>
<td>78</td>
</tr>
<tr>
<td>Med X, Sleep Difficulties</td>
<td>14</td>
<td>92</td>
<td>106</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>136</td>
<td>184</td>
</tr>
</tbody>
</table>

McNemar’s Test, \( p=0.0001 \),
paired data odds ratio = \( \frac{44}{14} = 3.14 \)
Questions for the Investigator

What data resources are available?

• Understand your data
• Clinical data and public databases
• Data management

• Good data = good research
The Answers

What is the plan for analysis?

• Best methods for the available data
• Best data for the available methods
• Potential for other analyses
Biostatistics Consulting Service

How to find us:
• Schedule a meeting.
• Just “Drop-In”.
• Special Sessions.

• Ask Us.
Consulting Services Faculty

• John Klein, PhD, Professor & Director of Division of Biostatistics, Fellow ASA
  • Survival analysis, cancer statistics, gait analysis

• Aniko Szabo, PhD, Associate Professor & BCS Director
  • Cancer statistics, genetics, clinical trials

• Sergey Tarima, PhD, Assistant Professor
  • Missing data problems, health service research

• Tao Wang, PhD, Associate Professor
  • Statistical genetics

• Jessica Pruszynski, PhD, Assistant Professor
  • Logistic regression
Consulting Services Staff

- Dan Eastwood, MS, BCS Manager
  - Gait analysis, health policy, general biostatistics
- Alexis Visotcky, MS, Biostatistician
  - VA databases, REDCap
- Qun (Katelyn) Xiang, MS, Biostatistician
  - Large databases, pediatric data
- Shi (Heather) Zhao, MS, Biostatistician
- Haley Montsma, BBA, Administrator
What should you bring to a meeting?

- Ideas
- Protocol?
- Example of your data
- Electronic copy of your data?
- “The boss”
Services

Data entry (fee service).

Help with:

• Design
• Analysis
• Grant Preparation
• Reading Papers
• Reports
• Graphics
• Assistance with Public Databases
• Advice on Methods
Biostatistics Consulting Service

- We are now supported by the Medical College’s *Clinical and Translational Science Institute* (CTSI)
- Biostatistics key function
- Monthly Lecture Series (more stats!):
  www.mcw.edu/biostatistics/LectureSeries.htm
- **DATUM** newsletter:
  www.mcw.edu/biostatistics/datum.htm
Biostatistics Consulting Service

CTSI services available to faculty, staff, and students working on Clinical and Translational Science Research at:

• MCW
• VA Medical Center
• Blood Center
• UW-Milwaukee
• Marquette
• Milwaukee School of Engineering
Free Drop-in Consulting

• Medical College of Wisconsin:
  Tuesdays and Thursdays
  Time: 1:00 PM—3:00 PM
  Building: Health Research Center
  Room: H2400 Biostatistics

• MCW Cancer Center
  Wednesdays 10:00 AM—12:00 PM
  Fridays 1:00 PM—3:00 PM
  Building: MCW Clinical Cancer Center
  Room: Clinical Trials Support Room
  CLCC: 3236 (Enter through C3233)

• Froedtert Pavilion:
  Mondays & Wednesdays
  Time: 1:00 PM—3:00 PM
  Building: Froedtert Pavilion
  Room: L772A- TRU Offices (Lower Level)

• Clement J. Zablocki VA Medical Center:
  1st & 3rd Monday of the month
  Time: 9:00 AM—11:00 AM
  Building: 111, 5th Floor B-wing
  Room: 5423

• Marquette University:
  Every Tuesday
  Time: 8:30 AM—10:30 AM
  Building: School of Nursing—Clark Hall
  Room: Office of Research and Scholarship: 112D
  Contact: Jessica Pruszynski, PhD to make an appointment
  Please note: Priority given to MU Nursing and Dental School personnel
Contact

- Haley Montsma
  - (414) 955-7439
  - hmontsma@mcw.edu
  - consult@mcw.edu

- Dan Eastwood, MS
  - (414) 955-4855
  - eastwood@mcw.edu
  - consult@mcw.edu

www.mcw.edu/biostatsconsult.htm
Database Basics

Dan Eastwood, MS
What is a Database?

• An organized collection of data

• Accessible in a computer

• Accessible in various ways
  • sortable
  • searchable
  • indexed
What is a Database?

- Well organized data enables good research
- Complex studies require careful organization
- Simple studies benefit from good organization
Is a spreadsheet a database?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match for elderly pt</td>
<td>Sex</td>
<td>Ethnicity 1= White 2= Black 3= Hispanic 4= Asian</td>
<td>IBD 1=CD 2=UC</td>
<td>BMI (kg/m²)</td>
<td>Disease duration at time of surgery (yrs) raw value</td>
<td>Age at Surgery</td>
<td>Charlson Comorbidity Index 0=0 1=1 2=2 3=3 or more</td>
<td>Co-morbidity: Cardiac (HTN, arrhythmia, coronary artery disease CHF, MI)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>M</td>
<td>3</td>
<td>1</td>
<td>31.53</td>
<td>9.0684932</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>F</td>
<td>2</td>
<td>1</td>
<td>21.48</td>
<td>8.1452055</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>23.866</td>
<td>32.230137</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>20.08</td>
<td>25.060274</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>24.73</td>
<td>0.3150685</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>27.66</td>
<td>14.221918</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>19.55</td>
<td>2.6931507</td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>23.01</td>
<td>17.339726</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>18.12</td>
<td>1.7315068</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>F</td>
<td>2</td>
<td>1</td>
<td>30.78</td>
<td>4.2164384</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>F</td>
<td>2</td>
<td>1</td>
<td>30.78</td>
<td>4.2164384</td>
<td>61</td>
<td>1</td>
</tr>
</tbody>
</table>
Spreadsheet vs. Database

• Spreadsheets have few or no rules

• Databases have strict rules

• Rules make spreadsheets more like a database
Spreadsheet vs. Database

• Use a database program to enforce rules

• Additional capability of databases

• A simple database can be viewed in spreadsheet or “table” form
Spreadsheet vs. Database

- Spreadsheets are prone to copy/paste, partial sorting, and other entry errors
  - Errors may be uncorrectable
  - Errors may be undetectable
- Changes to databases are generally reversible
  - Queries display data in different ways
  - Revert to original
  - Errors are more easily detected
What goes into a database?

- Type of data, formatting
- The “bad” list
- Factors and variables
- Sample units or observations?
- Multiple tables and linked tables
Statistical Qualities of Data

Data Types

- Categorical
  - Ordinal
  - Interval
  - Ratio
- Qualitative
- Quantitative
Computational Qualities of Data

Data Types, Part 2

Character (text)

Numeric (numbers)

Dates

(Missing or Censored)
The List of Bad Things

- More than one value in a single cell
- Mixed character and numbers data
- Merged cells
- Color coding
- UPPER and lower case text are different
- Confused coding or formats
- “Prettifying” is generally unhelpful
- Identifying information (try to minimize)
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>READ</td>
<td></td>
<td></td>
<td></td>
<td>Location of stenosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 = none/minimal narrowing</td>
<td></td>
<td></td>
<td></td>
<td>2 = moderate but obstructing &lt;50% of the lumen</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 = significant ≥50 - 70% but no severe narrowing</td>
<td></td>
<td></td>
<td></td>
<td>4 = severe narrowing ≥70% to total occlusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QUALITY</td>
<td></td>
<td></td>
<td></td>
<td>1 = poor image quality, uninterpretable image w/ severe artifacts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 = adequate image quality, mild to moderate artifact</td>
<td></td>
<td></td>
<td></td>
<td>3 = good image quality w/ no artifact</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td>1 = midstent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 = prox marker</td>
<td></td>
<td></td>
<td></td>
<td>3 = distal marker</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4 = diffuse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>43</td>
<td>Non-Hispanic</td>
<td>Angie</td>
<td>7/23/2002 MRA Head</td>
<td></td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>White</td>
<td>Angie</td>
<td>3/18/2003 CTA Head</td>
<td></td>
<td>1</td>
<td>n/a</td>
<td>3</td>
<td>1</td>
<td>n/a</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>Non-Hispanic</td>
<td>Angie</td>
<td>3/18/2003 CTA Head</td>
<td></td>
<td>3</td>
<td>1,2</td>
<td>3</td>
<td>3</td>
<td>1,2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>72</td>
<td>Non-Hispanic</td>
<td>Angie</td>
<td>3/19/2003 CTA Head</td>
<td></td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>White</td>
<td>Angie</td>
<td>3/19/2003 CTA Head &amp; Neck</td>
<td></td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>77</td>
<td>Non-Hispanic</td>
<td>Angie</td>
<td>4/3/2002 CTA Head &amp; Neck</td>
<td></td>
<td>1</td>
<td>n/a</td>
<td>3</td>
<td>1</td>
<td>n/a</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>47</td>
<td>Non-Hispanic</td>
<td>Angie</td>
<td>2/19/2003 CTA Head</td>
<td></td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>Black</td>
<td>Angie</td>
<td>4/23/2004 CTA Head</td>
<td></td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>index1</td>
<td>PID</td>
<td>Sex</td>
<td>Age</td>
<td>Race</td>
<td>Angio</td>
<td>Angio_date</td>
<td>Read Angio 1</td>
<td>Location Angio 1</td>
<td>QUALITY angio 1</td>
<td>Read Angio 2</td>
<td>Location Angio 2</td>
<td>QUALITY Angio 2</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>---------------</td>
<td>-------</td>
<td>------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>F</td>
<td>43.00</td>
<td>white non-hispanic</td>
<td>Angio</td>
<td>07/23/02</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>F</td>
<td>43.00</td>
<td>white non-hispanic</td>
<td>Angio</td>
<td>03/18/03</td>
<td>3</td>
<td>n/a</td>
<td>3</td>
<td>1</td>
<td>n/a</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>F</td>
<td>43.00</td>
<td>white non-hispanic</td>
<td>Angio</td>
<td>03/18/03</td>
<td>3</td>
<td>1,2</td>
<td>3</td>
<td>3</td>
<td>1,2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>F</td>
<td>72.00</td>
<td>white non-hispanic</td>
<td>Angio</td>
<td>03/19/03</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>F</td>
<td>72.00</td>
<td>white non-hispanic</td>
<td>Angio</td>
<td>03/19/03</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
Variables

- Short yet meaningful name
  - top row of spreadsheet
  - longer description or label

- Create a “key” to formatted values
  - ex: 1=yes, 2=no   1=treatment, 0=control
  - usually on a different sheet
Factors & Variables

• A factor is a complete description of one contributing element in the analysis
• A variable is a representation of a factor, or part of a factor, as used in the analysis
• A factor may be described by several variables
Factors & Variables

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>level</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
Sample Units & Observations

• Depends on Study Design

• Usually one row of data per sample unit
  • ie: one row per patient
  • “wide” layout
  • side-to-side scrolling problems

• Sometimes one row per observation
  • “long” layout
  • wasted space with demographics
Wide Layout Form

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Sex</th>
<th>Age</th>
<th>Group</th>
<th>SBP1</th>
<th>DBP1</th>
<th>SBP2</th>
<th>DBP2</th>
<th>SBP3</th>
<th>DBP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>56</td>
<td>Treatment</td>
<td>136</td>
<td>82</td>
<td>130</td>
<td>84</td>
<td>148</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>65</td>
<td>Placebo</td>
<td>138</td>
<td>95</td>
<td>22</td>
<td>88</td>
<td>120</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>76</td>
<td>Treatment</td>
<td>124</td>
<td>88</td>
<td>130</td>
<td>88</td>
<td>136</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>77</td>
<td>Treatment</td>
<td>120</td>
<td>84</td>
<td>140</td>
<td>78</td>
<td>122</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>54</td>
<td>Placebo</td>
<td>126</td>
<td>86</td>
<td>124</td>
<td><strong>80</strong></td>
<td>134</td>
<td>82</td>
</tr>
<tr>
<td>Study ID</td>
<td>Visit</td>
<td>Sex</td>
<td>Age</td>
<td>Group</td>
<td>SBP</td>
<td>DBP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>M</td>
<td>56</td>
<td>T</td>
<td>136</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>M</td>
<td>56</td>
<td>T</td>
<td>130</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>M</td>
<td>56</td>
<td>T</td>
<td>148</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>F</td>
<td>65</td>
<td>P</td>
<td>138</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>F</td>
<td>65</td>
<td>P</td>
<td>22</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>F</td>
<td>65</td>
<td>P</td>
<td>120</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>M</td>
<td>76</td>
<td>T</td>
<td>124</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>M</td>
<td>76</td>
<td>T</td>
<td>130</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>M</td>
<td>76</td>
<td>T</td>
<td>136</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>M</td>
<td>77</td>
<td>T</td>
<td>120</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>M</td>
<td>77</td>
<td>T</td>
<td>140</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>M</td>
<td>77</td>
<td>T</td>
<td>122</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>F</td>
<td>54</td>
<td>P</td>
<td>86</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>F</td>
<td>54</td>
<td>P</td>
<td>80</td>
<td>134</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>F</td>
<td>54</td>
<td>P</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiple Tables -

Demographics and Clinical Data with Linking index variable

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Sex</th>
<th>Age</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>56</td>
<td>Treatment</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>65</td>
<td>Placebo</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>76</td>
<td>Treatment</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>77</td>
<td>Treatment</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>54</td>
<td>Placebo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Visit</th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>136</td>
<td>82</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>130</td>
<td>84</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>148</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>138</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>22</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>120</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>124</td>
<td>88</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>130</td>
<td>88</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>136</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>120</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>140</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>122</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>86</td>
<td>124</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>80</td>
<td>134</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>
Database Programs

• Microsoft Access
  • Everybody has it
  • Nobody uses it
Database Programs

- REDCap
  - web based
  - secure server
  - survey package (no more Survey Monkey)
  - 267 institutional partners
  - 20K+ studies, 30K+ end users
  - project-redcap.org

- Contact for more information:
  - Mark Oium, moium@mcw.edu, 805-2051
Concluding Remarks

• Have a plan for your data
• You can “pilot” a database at the same time you gather pilot data for a study
• Good data leads to good research
Questions?