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:	Role in Meeting:
Ruta Brazauskas, PhD	Planning Committee
Haley Montsma, BBA	Planning Committee
Dan Eastwood, MS	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

4

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
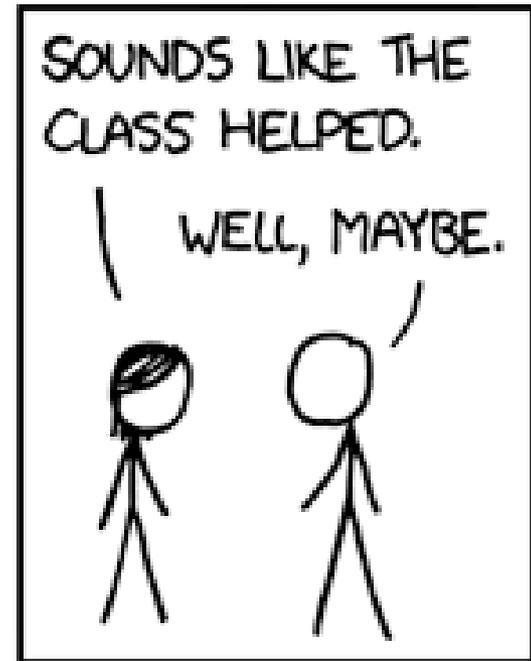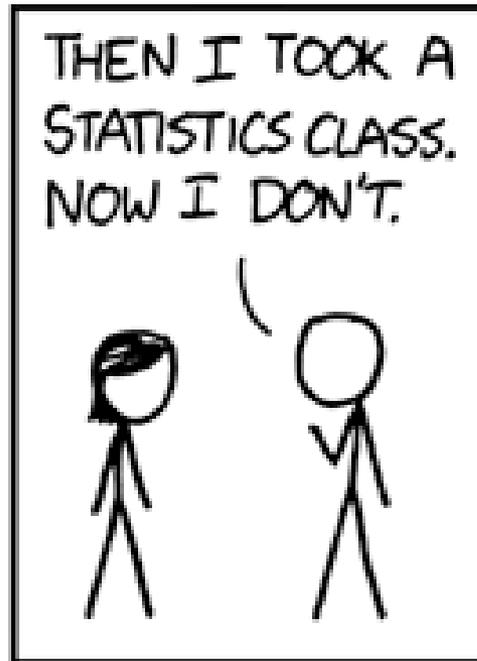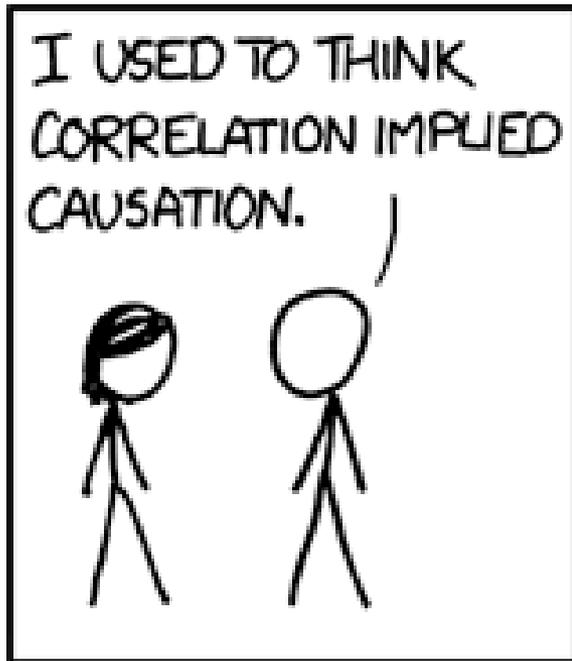
XKCD.com

Why Biostatistics Consulting?

- **Shared experience**

- Discuss your study

- Consider alternate views

- Formulate ideas into hypotheses

8

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 ..."

11

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

13

Questions for the Investigator

What makes a good hypothesis? (2)

- Statistical and clinical significance

- Related factors

- Preliminary data

14

It Never Hurts to Ask

Independent Samples	Sleep difficulty, Medication X?	Sleep difficulty, Medication Y?	Total
"No"	48	78	126
"Yes"	**136**	**106**	242
Total	**184**	**184**	368

Chi-Square test, p=0.0010,
Medication X 57.6%, Medication Y 73.9%

It Never Hurts to Ask

Paired Data	Med Y, No difficulties	Med Y, Sleep Difficulties	Total
Med X, No difficulties	34	44	78
Med X, Sleep Difficulties	14	92	106
Total	48	136	184

McNemar's Test, p=0.0001,
paired data odds ratio = 44/14 = 3.14

16

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

21

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

24

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

25

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

26

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

27

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?

08172011 PostOp Outcomes- sen

F18 = 20.1698630136986

	Match for elderly pt	Sex	Ethnicity 1= White 2= Black 3= Hispanic 4= Asian	IBD 1=CD 2=UC	BMI (kg/m2)	Disease duration at time of surgery (yrs) raw value	Age at Surgery	Charlson Comorbidity Index 0=0 1=1 2=2 3=3 or more	Co-morbidity: Cardiac (HTN, arrhythmia, coronary artery disease CHF, MI)
2	1	M	3	1	31.53	9.0684932	55	2	Yes
3	2	F	2	1	21.48	8.1452055	50	0	No
4	2	F	1	1	23.866	32.230137	56	0	No
5	3	M	1	1	20.08	25.060274	51	0	No
6	3	M	1	1	24.73	0.3150685	50	0	No
7	4	M	1	1	27.66	14.221918	62	1	Yes
8	5	M	1	1	19.55	2.6931507	61	0	Yes
9	6	F	1	1	23.01	17.339726	54	2	No
10	6	F	1	1	18.12	1.7315068	50	0	No
11	7	F	2	1	30.78	4.2164384	61	1	Yes

Spreadsheet vs. Database

- Spreadsheets have few or no rules

- Databases have strict rules

- Rules make spreadsheets more like a database

32

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

33

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

34

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
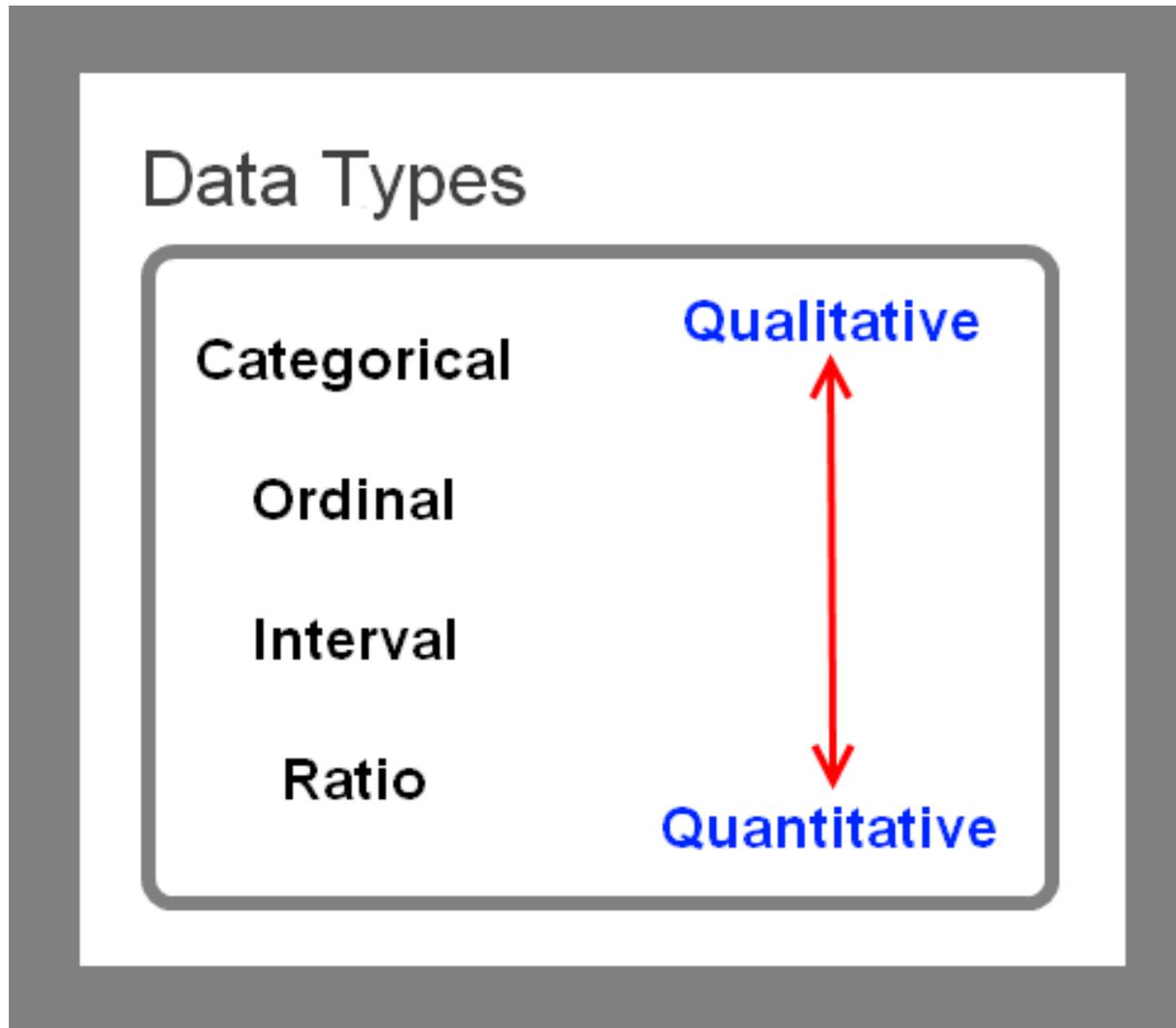
Computational Qualities of Data

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)

38

						READ 1= none/minimal narrowing 2= moderate but obstructing <50% of the lumen 3= significant ≥50 - 70% but no severe narrowing 4= severe narrowing ≥70% to total occlusion		LOCATION 1= midstent 2 =prox marker 3 = distal marker 4 =diffuse		QUALITY 1=poor image quality, uninterpretable image w/ severe artifacts 2=adequate image quality, mild to moderate artifact 3=good image quality w/ no artifact	
	Sex	Age	Race		Angio	ANGIO read #1	Location of stenosis	QUALITY	ANGIO read #2	Location of stenosis	QUALITY
1	F	43	white non-hispanic	Angio	7/23/2002 MRA Head	3	4	3	3	1	3
				Angio	3/18/2003 CTA Head	1	n/a	3	1	n/a	3
				Angio	3/18/2003 CTA Head	3	1,2	3	3	1,2	3
2	F	72	white non-hispanic	Angio	3/19/2003 CTA Head	4	1	3	4	4	3
				Angio	3/19/2003 CTA Head & Neck	4	1	3	4	4	3
3	M	77	white non-hispanic	Angio	4/3/2002 CTA Head & Neck	1	n/a	3	1	n/a	3
4	M	47	black non-hispanic	Angio	2/19/2003 CTA Head	2	1	3	3	1	3
				Angio	4/23/2004 CTA Head	4	4	3	3	3	3

Better now?

index1	PID	Sex	Age	Race	Angio	Angio_date	Read Angio 1	Location Angio 1	QUALITY angio 1	Read Angio 2	Location Angio 2	QUALITY Angio 2
1	1	F	43.00	white non-hispan	Angio	07/23/02	3	4	3	3	1	3
2	1	F	43.00	white non-hispanic	Angio	03/18/03	3	n/a	3	1	n/a	3
3	1	F	43.00	white non-hispanic	Angio	03/18/03	3	1,2	3	3	1,2	3
4	2	F	72.00	white non-	Angio	03/19/03	4	1	3	4	4	3

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

41

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

level	Factor ABC	A	B	C
1	None	N	N	N
2	A only	Y	N	N
3	B only	N	Y	N
4	C only	N	N	Y
5	A and B	Y	Y	N
6	A and C	Y	N	Y
7	B and C	N	Y	Y
8	A, B, and C	Y	Y	Y

43

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

Study ID	Sex	Age	Group	SBP1	DBP1	SBP2	DBP2	SBP3	DBP3
1	M	56	Treatment	136	82	130	84	148	82
2	F	65	Placebo	138	95	22	88	120	76
3	M	76	Treatment	124	88	130	88	136	80
4	M	77	Treatment	120	84	140	78	122	84
5	F	54	Placebo	126	86	124	80	134	82

Long Layout Form

Study ID	Visit	Sex	Age	Group	SBP	DBP
1	1	M	56	T	136	82
1	2	M	56	T	130	84
1	3	M	56	T	148	82
2	1	F	65	P	138	95
2	2	F	65	P	22	88
2	3	F	65	P	120	76
3	1	M	76	T	124	88
3	2	M	76	T	130	88
3	3	M	76	T	136	80
4	1	M	77	T	120	84
4	2	M	77	T	140	78
4	3	M	77	T	122	84
5	1	F	54	P	86	124
5	2	F	54	P	80	134
5	3	F	54	P	82	

Multiple Tables -

Demographics and Clinical Data
with Linking index variable

Study ID	Sex	Age	Group
1	M	56	Treatment
2	F	65	Placebo
3	M	76	Treatment
4	M	77	Treatment
5	F	54	Placebo

Study ID	Visit	SBP	DBP
1	1	136	82
1	2	130	84
1	3	148	82
2	1	138	95
2	2	22	88
2	3	120	76
3	1	124	88
3	2	130	88
3	3	136	80
4	1	120	84
4	2	140	78
4	3	122	84
5	1	86	124
5	2	80	134
5	3	82	

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

49

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?