

Engaging Statistical Collaboration in Research at UMC

William D. Johnson, Ph.D.

Professor, Division of Biostatistics
Department of Preventive Medicine

Division of Biostatistics
Room N408
(601) 984-1935

- Edward F. Meydrech, Ph.D. Professor and Chair
- William D. Johnson, Ph.D. Professor
- Warren L. May, Ph.D. Professor
- Jake Olivier, Ph.D. Assistant Professor

Missions of the Division of Biostatistics

- **Teaching:** Medical students and graduate students
- **Collaborative Research:** Provide statistical advice and collaboration to researchers throughout The University of Mississippi Medical Center
- **Research on Statistical Methods:** Conduct research on statistical theory and methods for definitive and reliable research conclusions

Collaborative Skills

- Expertise in design of animal experiments, clinical trials, sample surveys, case-control studies and other observational investigations
- Power/sample size calculations
- Randomization
- Preparation of statistical components of grant applications

Collaborative Skills (Cont'd)

- Expert statistical analysis ranging from the routine to cutting-edge techniques
- Knowledge of appropriate computer software to use for analyzing data
- Ability to design appropriate tables and graphics to effectively summarize results
- Experienced at writing statistical methods and results sections for publications

Who is Eligible?

The Division of Biostatistics offers collaboration to all UMC investigators whose research is connected to their UMC mission

What is the Charge?

- There is no monetary charge for statistical collaboration but long term relationships are enhanced by obtaining input from and including funding for biostatisticians in grant applications
- We do expect to be included as co-authors in presentations and publications where we contribute statistical advice or analyses

How to Initiate Collaboration

- Call Ms. Linda Dunn at (601) 984-1935 and ask for an appointment
- Call or e-mail one of the biostatisticians and ask for an appointment (Groupwise)
- Access departmental or individual information via UMC webpage

When should contact be initiated?

- The best time to engage a biostatistician is very early in the research planning stage
- Alternatively, as early as possible prior to deadlines for meetings, etc.

The Initial Meeting

- Copy of protocol if available
- Copy of some closely related publications
- Research objectives—primary and secondary
- Identify primary study unit (e.g. patient, family, animal)
- Study design
 - Clinical trial or other experimental study
 - Retrospective study based on charts
 - Individually Matched Case-Control study

The Initial Meeting (Cont'd)

- Inclusion/exclusion criteria
- Randomization schedule if applicable
- Blinding/Masking?
- Multi-center Study?
- List of variables relevant to analysis
 - Primary outcomes (e.g., glucose level)
 - Concomitants (e.g., age, height, weight)

The Initial Meeting (Cont'd)

- Time points when interventions occur or variables are measured
- Research hypotheses or goals for analysis
- Provide a clean data file (e.g., EXCEL)
- Progress to date
- Time constraints / Deadlines
- Expected presentations, publications, etc
- Authorship of presentations / publications

Sample Size / Power

- Express research hypothesis as a statistical (null) hypothesis
- Determine whether alternative hypothesis is one or two directional
- Specify level of significance (e.g., 0.05)
- Specify estimate of standard deviation of variables to be analyzed
- Specify estimates of group means (effect size)**
- Specify desired power (e.g., 80%) **
- Determine sample size per group **

Patient	Trtgrp	BLage	Htin	Wtlb	HisMI	ChBL	Ch3m	Ch4m
1	c	52	73	216	yes	367	300	278
2	a	40	68	178	no	273	291	283
3	b	49	70	185	no	273	234	245
4	a	64	65	173	no	314	309	325
5	c	37	67	191	no	243	223	227
6	b	48	71	175	no	269	208	221
7	b	51	69	144	no	221	169	155
8	c	62	68	156	no	272	190	183
9	a	46	74	240	yes	276	269	284
10	a	33	65	139	no	251	233	218
11	c	69	72	274	yes	306	182	201
12	b	70	69	155	no	239	167	191
13	b	34	66	151	no	248	172	159
14	a	55	71	185	no	407	316	292
15	c	43	70	210	no	206	196	199
16	a	47	73	227	no	259	216	148

Randomization

- Two Treatments: A and B
- 24 subjects, 12 per treatment
- Restrict randomization to balance after every 4 patients so 2 receive A and 2 receive B

Randomization

Six possible sequences of 4:

AABB

ABAB

ABBA

BAAB

BABA

BBAA

01-B	02-A	03-B	04-A
05-B	06-A	07-A	08-B
09-A	10-B	11-A	12-B
13-A	14-A	15-B	16-B
17-A	18-B	19-B	20-A
21-B	22-B	23-A	24-A

Calculating Power

Experiment: Compare efficacy of A to B

20 subjects: 10 individually matched pairs

Each pair: Observe whether $A > B$ or $B > A$

Null Hypothesis: Probability $A = B = 0.5$

Alternative hypothesis: Probability not = 0.5

Probability $A > B = 0.5$ $n = 10$ matched pairs

Trials with $A > B = y$	Probability of y
0	0.0010
1	0.0098
2	0.0439
...	...
8	0.0439
9	0.0098
10	0.0010

Probability of outcomes that are not expected if $A = B$

$$\begin{aligned} &\text{Probability of observing } y = 0, 1, 9 \text{ or } 10 \\ &= 0.0010 + 0.0098 + 0.0098 + 0.0010 \\ &= 0.0216 \end{aligned}$$

We reject the null hypothesis

$A = B$ if $y = 0, 1, 9$ or 10

Probability $A > B = 0.75$ $n = 10$ matched pairs

Trials with $A > B = y$	Probability of y
0	0.0000
1	0.0000
2	0.0004
...	...
8	0.0563
9	0.1877
10	0.2816

Power if Probability $A > B = 0.75$

We reject the null hypothesis $A = B$
if $y = 0, 1, 9$ or 10

$$\begin{aligned} &\text{Probability of observing } y = 0, 1, 9 \text{ or } 10 \\ &= 0.0000 + 0.0000 + 0.1877 + 0.0563 \\ &= 0.2440 \end{aligned}$$

$$\text{Power} = 0.2440$$

Increase sample size to $n = 20$

Decide to reject if $y =$

$0, 1, 2, 3, 4, 5, 15, 16, 17, 18, 19$ or 20

Probability rejecting if $A = B = 0.0414$

Power if probability $A > B = 0.75 = 0.4966$

Note: Power increases as n increases.

Simplified Example: Estimating Sample Size

Clinical Trial: Investigate efficacy of treatment A versus treatment B for glucose control

Primary outcome variable: A1C at 6 months

Null Hypothesis: Mean A1C the same for two treatment patient “Populations”

Alternative: Means not the same

Simplified Example (Cont'd)

Parameter estimates based on a small sample of treated diabetic patients treated with A: Mean = 8.4 Standard Dev = 2.3

Investigator wants p-values equal to or less than 0.05 to reject the null hypothesis

Power at least 80% to reject if mean for B is 1 or more units different from 8.4

Simplified Example (Cont'd)

$$\text{Effect size} = |8.4 - 9.4|/2.3 = 0.435$$

Required sample size = 85 in each of the two treatment groups

Results of Clinical Trial

n = 85 patients per group

Mean A1C for Treatment A patients = 8.25

Mean A1C for Treatment B patients = 7.42

Pooled estimate of SD = 2.34

P-value < 0.0001

Results (Cont'd)

Simultaneous 95% Confidence intervals:

-- for treatment A:

8.25 +/- 0.40 or 7.85 – 8.65

-- for treatment B:

7.42 +/- 0.40 or 7.02 – 7.82

Conclusion: Mean A1C for Treatment B patients is significantly less than the mean for Treatment A patients.

AD Hyperactivity Disorder

Status	Placebo		Ritalin
Normal	50		67
	45		60
	55		58
	52		65
Hyperactive	70		51
	72		57
	68		48
	75		55

AD Hyperactivity (Cont'd)

Status	Placebo		Ritalin
Normal	50.5		62.5
Hyperactive	71.3		52.8

SV	df	MS	F	P-value
Act	1	121.00	8.00	0.0152
Trt	1	42.25	2.79	0.1205
ActxTrt	1	930.25	61.50	<0.0001
Error	12	15.13		
C. Tot	15	1275.00		

Comments

Significant interaction changes approach to analysis

Re-analysis in terms of 4 groups

SV	df	MS	F	P-value
Groups	3	364.50	24.10	<0.0001
Error	12	15.13		
C. Tot	15			

Six Comparisons(Adjusted)

<u>Comparison</u>	<u>p-value</u>
Norm-Placebo vs Norm-Ritalin (50.5 vs 62.5)	0.0055
Norm-Placebo vs Hyper-Placebo (50.5 vs 71.3)	<0.0001
Norm-Placebo vs Hyper-Ritalin (50.5 vs 52.8)	1.0000

Comparisons (Cont'd)

Norm-Ritalin vs Hyper-Placebo (62.5 vs 71.3)	0.0474
Norm-Ritalin vs Hyper-Ritalin (62.5 vs 52.8)	0.0242
Hyper-Placebo vs Hyper-Ritalin (71.3 vs 52.8)	0.0001

Status	Placebo		Ritalin
Normal	50.5		62.5
Hyperactive	71.3		52.8