

Analytic Methods

Janet Wittes

Statistics Collaborative

Michael Proschan

NHLBI

Two rules

- Avoid bias: Janet
- Reduce variability: Mike

Avoiding bias

- Randomize
- Respect the randomization
- Use estimators that give the right answer

Bias and Consistency

- *Unbiased estimator* – on the average, we have the right number
- *Consistent estimator* – if our sample size is very large, we'll converge to the right number



Coin toss

- Unbiased: H/N





Coin toss

- Unbiased: H/N
- Not consistent: $\text{Min}[0.999, (H/N)+.001]$

No heads

0.001

All heads

0.999





Coin toss



- Unbiased: H/N
- Not consistent: $\text{Min}(0.999, (H/N)+.001)$
- Biased, but consistent: $(H+1)/(N+2)$

No heads

$$1/(N+2)$$

All heads

$$(N+1)/(N+2)$$

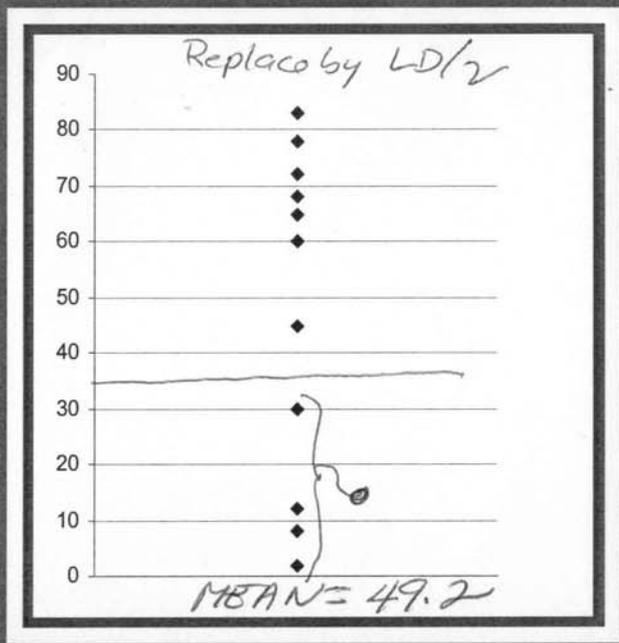
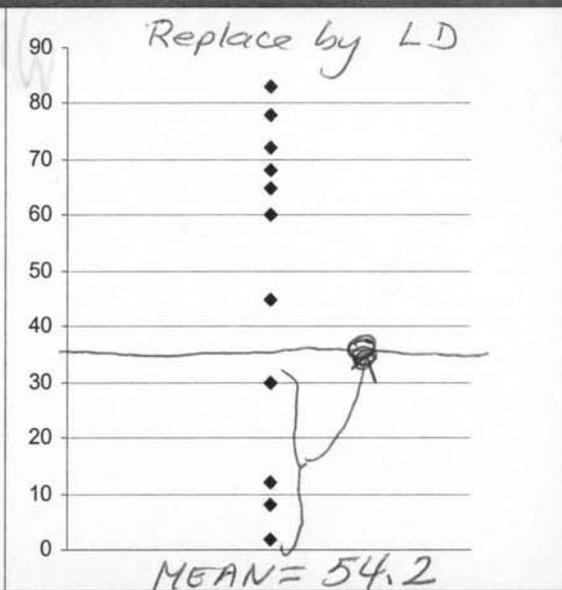
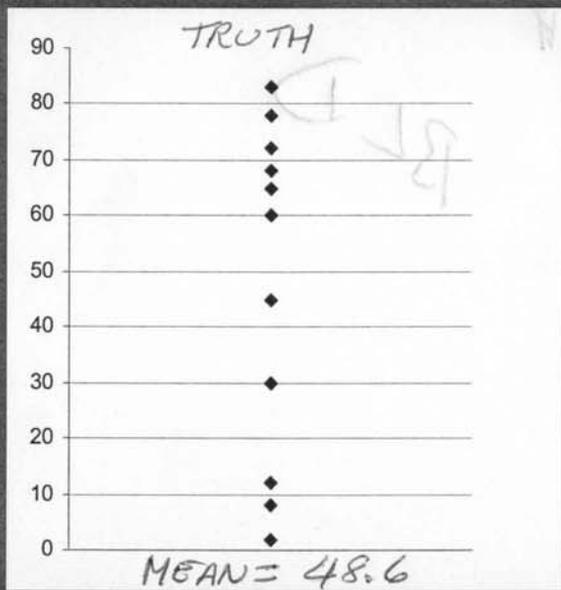
– It estimates $[Np+1]/(N+2) \neq p$

– Consistent: as $N \rightarrow \infty$, $H+1 \rightarrow H$ and $N+2 \rightarrow N$



Floor and ceiling effects

- Assay
 - Replace undetectable by BLD
 - Neither unbiased nor consistent



Floor and ceiling in psych tests

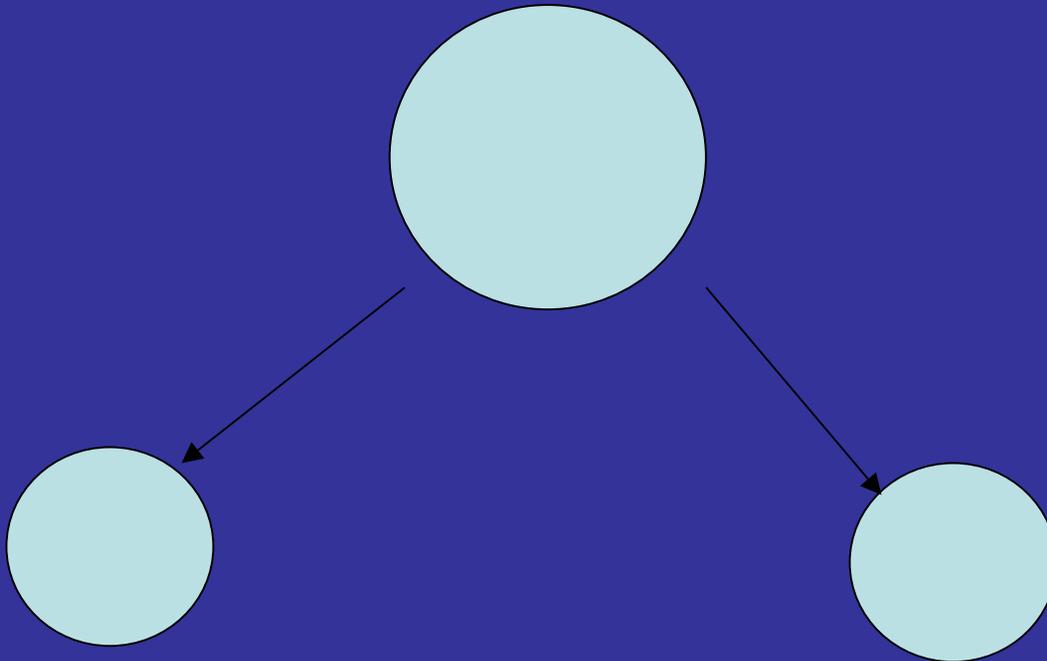
- OK if “true” parameter is test value
- Not OK if “true” parameter is latent value

Lesson #1 in bias

- Use standard statistical methods
- Avoid statistician-free *ad hoc* solutions
“This is how it’s done” is not an answer

Lesson #2: First we randomize---

-



Randomize

- Unrestricted

AABABBBBAABBAAAABBABAB

- Stratified

– M: AABABABBBAAB

– F: BBBBAAABABABAA

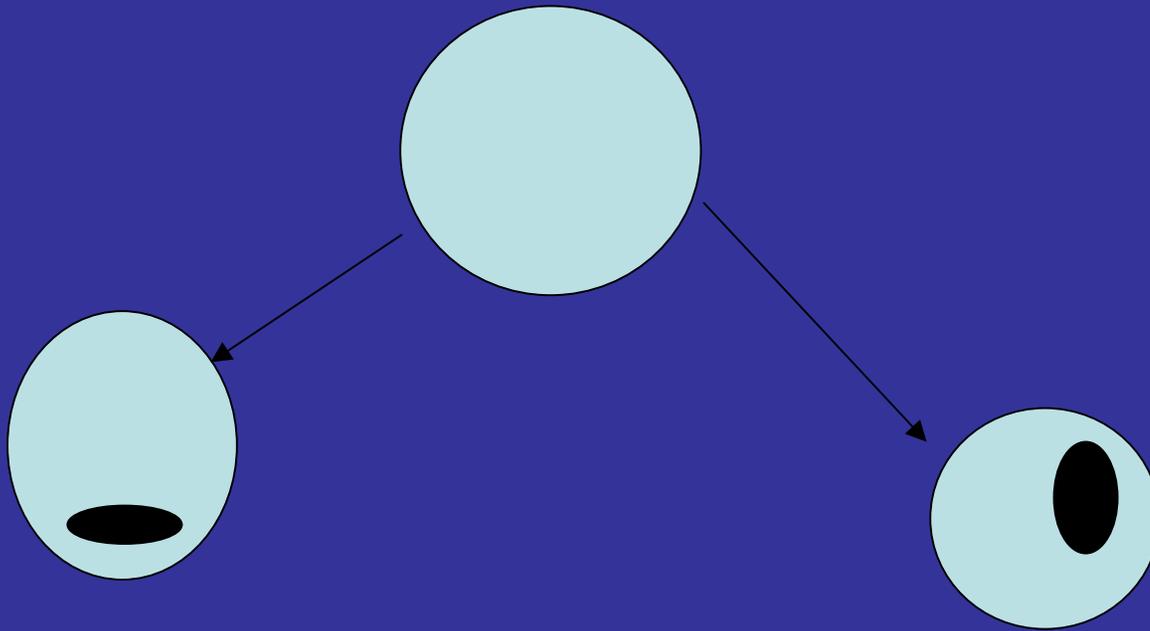
- Blocked

– |AABBAB|ABAABB|ABBABA|

Respect the randomization

- Once randomized, always analyzed

What if we exclude some people?

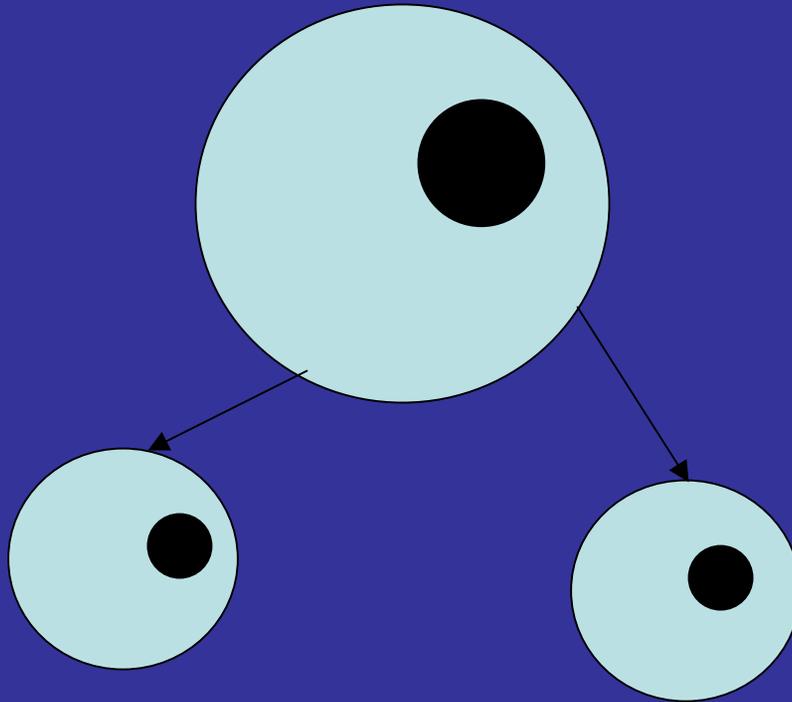


Ways to exclude

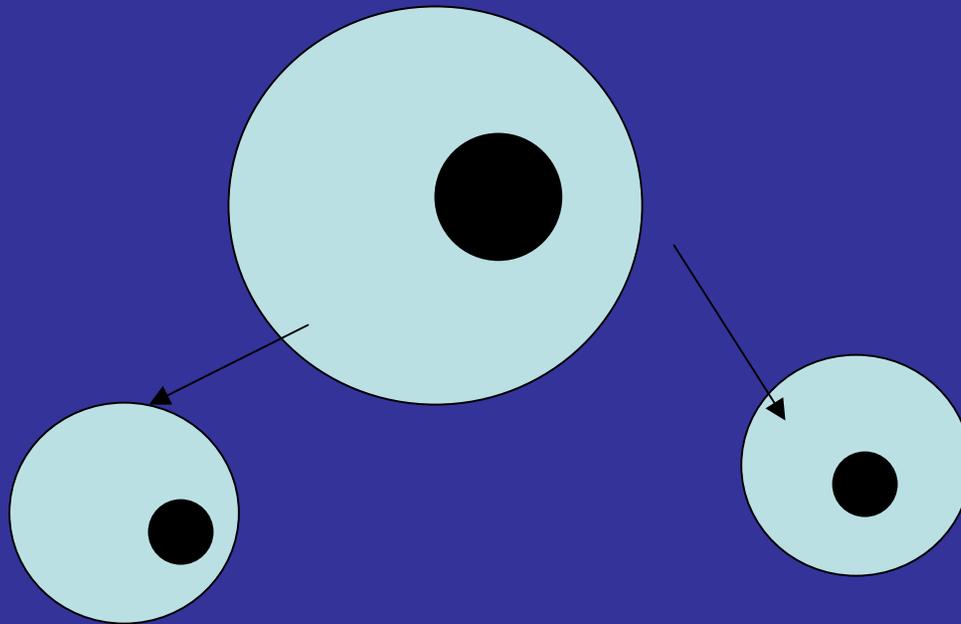
- Analyze “as treated” BIASED
- Stop coming to visits BIASED
- Loss to follow-up BIASED
- Presumptive treatment NOT BIASED

Why is baseline exclusion OK?

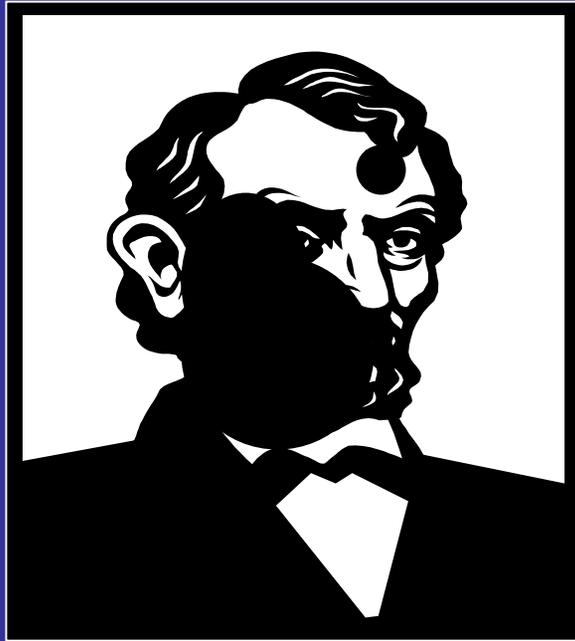
Remove, then randomize is same as...



...randomize, then randomly remove



M_ss_ng D_t_a





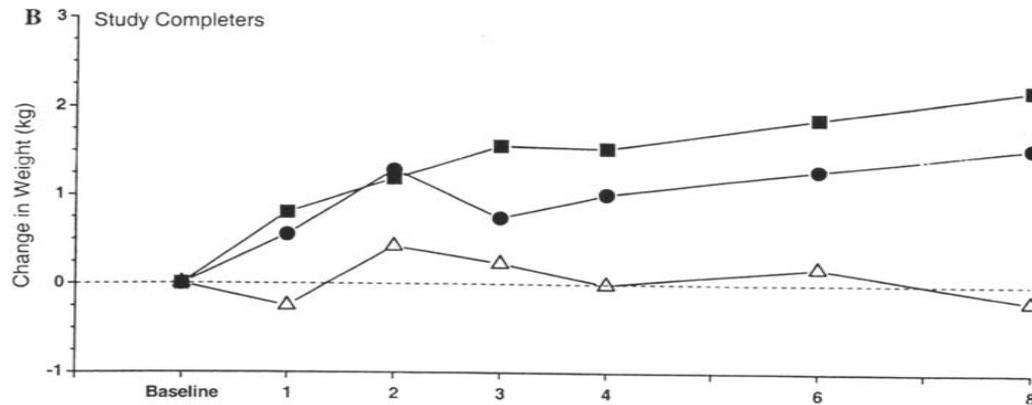
Defining ITT

“When I use a word,” Humpty Dumpty said, in a rather scornful tone, “it means just what I choose it to mean—neither more nor less.”

“The question is,” said Alice, “whether you can make words mean so many different things.”

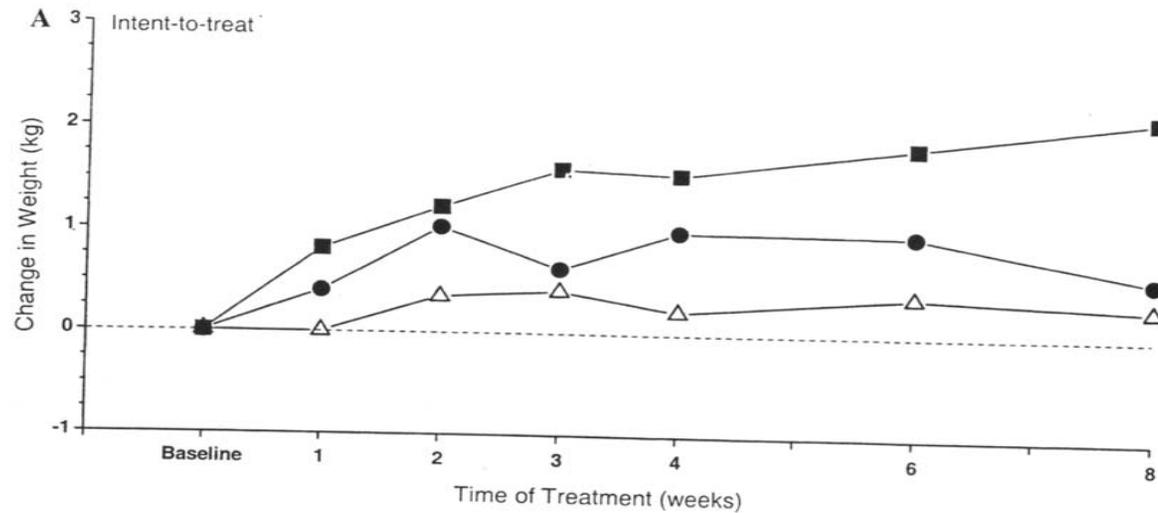
“The question is,” said Humpty Dumpty, “which is to be master—that’s all.”

Study Completers



▲ = placebo ■ = low ● = high

“Intent-to-treat”

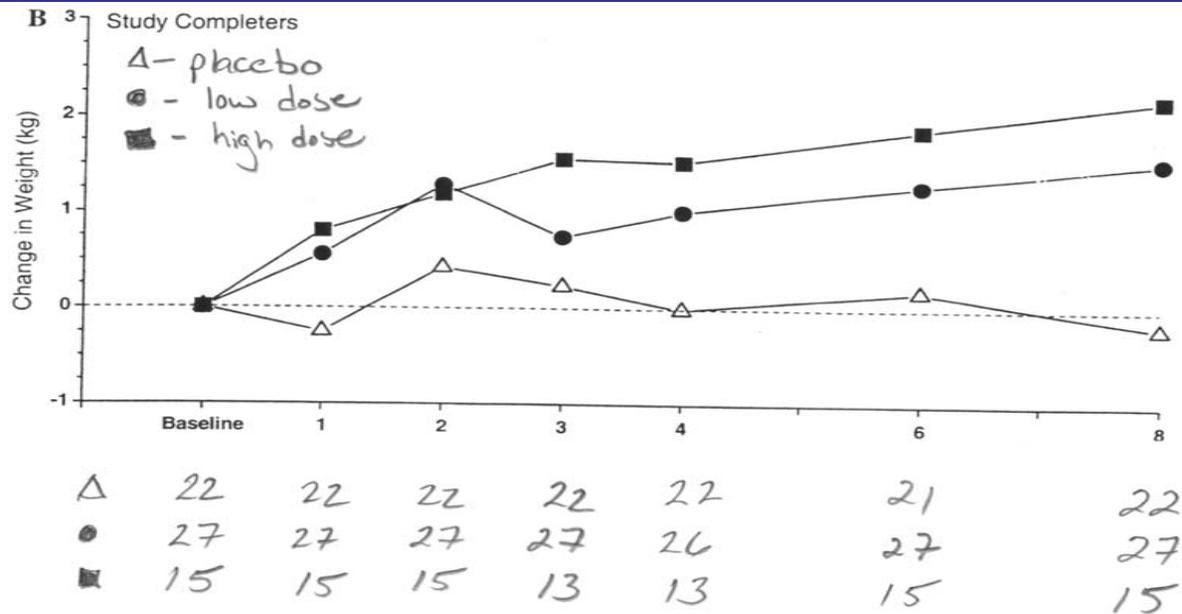


▲ = placebo

■ = low

● = high

Completers

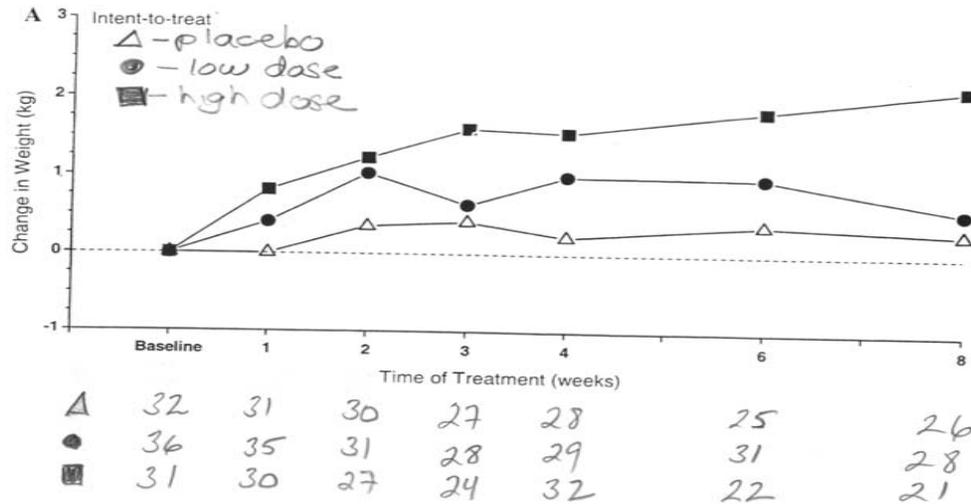


▲ = placebo

■ = low

● = high

Intent-to-treat'



▲ = placebo

■ = low

● = high

Daily Mean Number of Urinary Incontinence Episodes

		Placebo	Behavior	Drug
BL	n	100	100	100
	mean	5.0 ± 3.2	4.8 ± 2.9	4.7 ± 2.9
	median	4	4	4
Proportion missing final value				
		10%	12%	18%
Change to end of study				
	LOCF	-2.1	-2.9	-3.2
	Observed	-2.6	-2.8	-3.0

Data collected

Completion of study/Withdrawal from study	
Is the subject being withdrawn before the planned end of study? N <input type="checkbox"/> No Y <input type="checkbox"/> Yes	
<i>If NO, please sign and date below and fill out only section "Completion of study".</i>	
<i>If YES, please fill out the whole page.</i>	
Withdrawal from study	
Date of withdrawal: <input type="text"/> / <input type="text"/> / <input type="text"/> 2 0 0 <input type="text"/>	
(day) (month) (year)	
Reason for premature withdrawal from study (check only one)	
A <input type="checkbox"/> the subject no longer meets the criteria to remain in the study	
B <input type="checkbox"/> new adverse event or worsening of an existing adverse event	
<i>Please supply an "Adverse event report" form as appropriate.</i>	
C <input type="checkbox"/> lack of efficacy	
D <input type="checkbox"/> poor compliance with treatment	
E <input type="checkbox"/> the subject did not wish to continue in the study	
F <input type="checkbox"/> the subject is lost to follow-up	
G <input type="checkbox"/> administrative reasons	
H <input type="checkbox"/> protocol violation	
I <input type="checkbox"/> the subject died	
<i>Please fax a "Serious adverse event report" and complete "In case of death" form.</i>	
L <input type="checkbox"/> at the discretion of the investigator	
J <input type="checkbox"/> other reason (please specify) _____	

Choices

- Don't have any missing values
- Use what you have
- Redefine your endpoint
- Use slope
- Impute
 - If so, how?

Avoid missing values

- Important to get follow-up measures
 - Cessation of program not excuse for failing to measure last observation

Use what you have

Does not respect the randomization

Redefine your endpoint

- Ventilator failure in acute lung injury
 - Number of days ALIVE and not on ventilator
- Alcoholism
 - Number of days of **known** abstinence
 - Missing data = heavy drinking

Redefine...

- AIDS
 - Success=Known increase in weight ≥ 1 kg
- Incontinence
 - Success= Known number of episodes < 3 /wk

Slope

- Assume that slope extends beyond last measure
 - Even after death?

Impute

- Idea: assign number to the last value
- Choices
 - LOCF
 - Windows
 - Worst case
 - Worst reasonable case
 - Multiple imputation

What does it mean to impute?

- Reason for missing
 - Moved
 - Died
 - Adverse event
 - Quit

Classify your missing

- Completely at random
- At random
- Related to treatment

“Sensitivity” Analysis

- Do the conclusions vary depending on the method of analysis you use?

AIDS p-values vs. placebo

	Low dose	High dose
Completers	0.006	0.04
Still on original	0.0007	0.10
LOCF	0.012	0.40
Dawson/Lagakos	0.012	0.53
WRC	0.12	0.78
Multiple impute	0.045	0.31

Incontinence example

- How sure are we that
 - The drug works?
 - The behavioral intervention works?

Conclusion

- Think of how to minimize bias
 - Analytically
 - By sensibly dealing with missing
- Think of how much missing data you will have
 - Design study to minimize missing data
 - In analysis, check robustness of your analyses

Variability

Mike Proschan

Outline

- Variance reduction strategy: Focus on largest sources
- When patient is biggest source: Trying to reduce variability in a parallel arm trial
 - End of study measurement
 - Analysis of change
 - Adjusting for baseline covariates
- When group is biggest source
- Surviving survival

General principle: Focus on largest sources of variability

Example: Variability in BP trial



One way to counteract variability: Average

- Patient-patient variability largest—need to average over many patients
- Day-day variability not trivial—should average over several days for each patient
- Minute-minute variability much smaller—don't need to average over several measurements minutes apart

Another solution is to reduce variability

- With parallel-arm trial with continuous outcome (cholesterol, BP) have 3 choices:
 - Use end of study value
 - Use change from baseline
 - Adjust for baseline value using ANCOVA
- What's best?

Analyzing changes

- X =baseline value, Y =end of study value
- Assume variance of baseline and end of study measurements are equal
- $\text{var}(X)=\text{var}(Y)=\sigma^2$
- $\text{var}(Y-X)=2\sigma^2(1-\rho)$, where ρ is the correlation between X and Y
- $\text{var}(Y-X)<\text{var}(Y)$ when $\rho>1/2$

Analyzing changes

- In short duration trials, $\rho > 1/2$ is likely. E.g., for BP, $\rho \approx .90$ for 6 week study
- In long term study, X and Y may almost be uncorrelated ($\rho \approx 0$)
- The shorter the duration, the more appealing a change score is

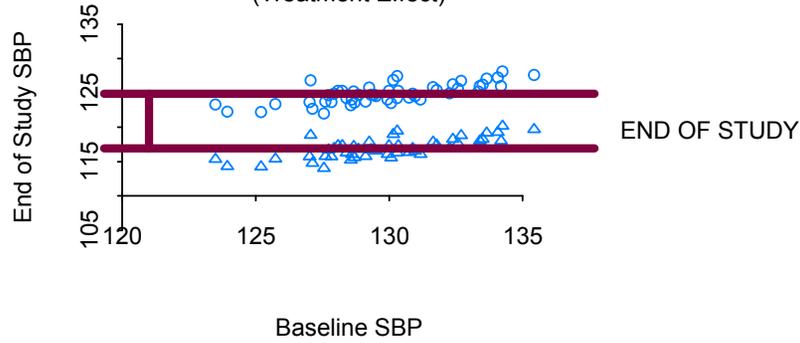
ANCOVA

- An alternative: Assume ANCOVA model $Y = \beta_0 + \beta_1 x + \beta_2 z$, where $z = \text{treatment variable}$
- Slope in control ($z=0$) and treatment ($z=1$) arms are same
- Intercepts differ: β_0 in control and $\beta_0 + \beta_2$ in treatment
- β_2 is treatment effect ($\beta_0 + \beta_2 - \beta_0$)

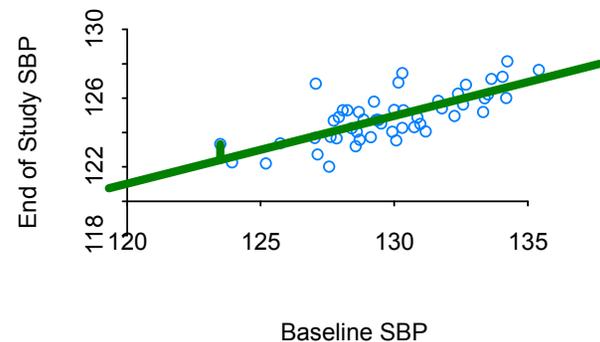
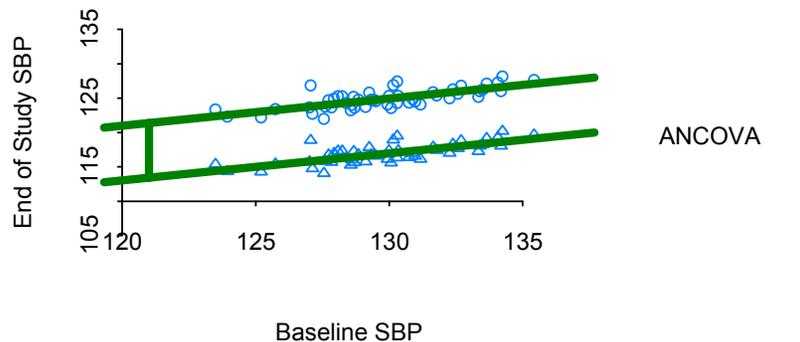
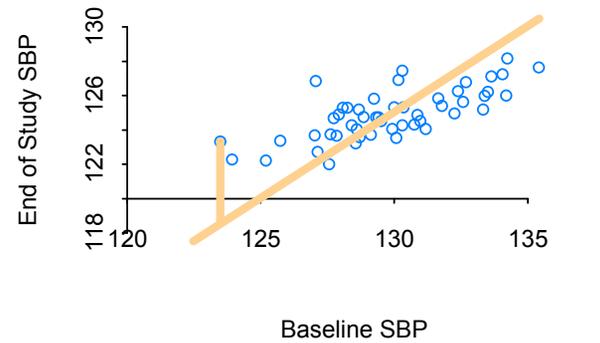
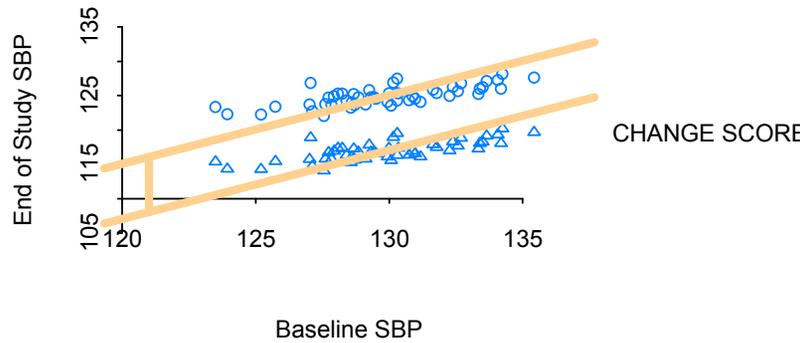
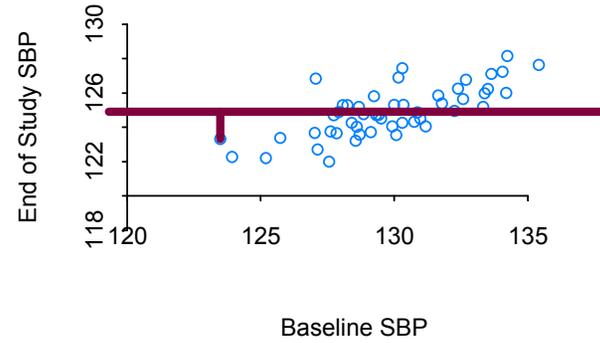
Comparison of methods

- Treatment effect estimates:
 - End of study approach: $Y_T - Y_C$
 - Change score approach: $Y_T - Y_C - (X_T - X_C)$
 - ANCOVA: $Y_T - Y_C - \beta_1(X_T - X_C)$
- Change scores and ANCOVA adjust for baseline imbalances in different ways

Between-Arm Variability (Treatment Effect)



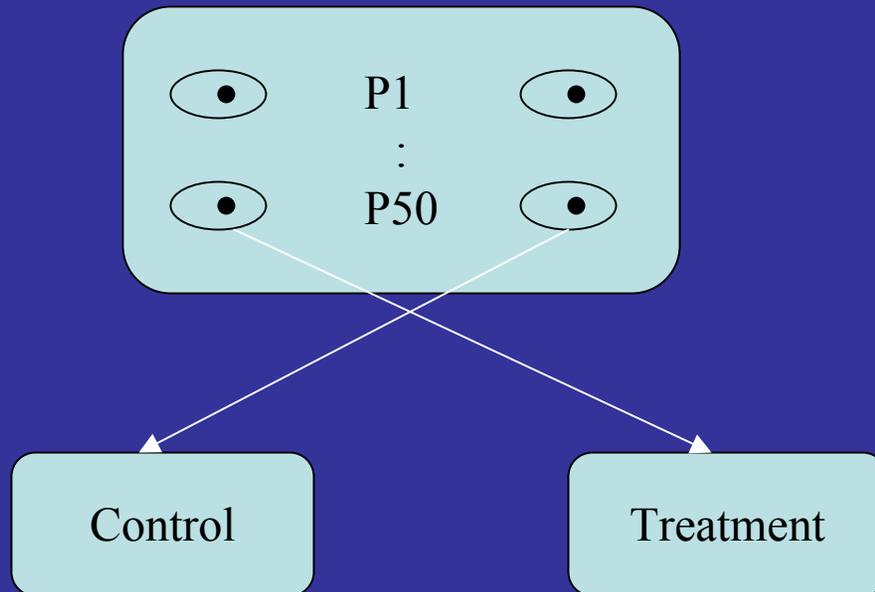
Within-Arm Variability



Proper use of ANCOVA

- ANCOVA
 - Reduces variability even if arms balanced
 - Adjusts treatment effect if arms unbalanced
- Pick limited number of **BASELINE** variables (no post-randomization variables) most related to outcome
- Specify covariates in protocol

Another way to reduce variability: Give patient T & C



Not appropriate for many behavioral trials because of carryover

Group-randomized trials

- Sometimes treatment is applied at group level
 - REACT (Rapid Early Action for Coronary Treatment): Media campaign to call 911 when having chest pain. Randomized within pairs
 - PAD (Public Access Defibrillation): Should we put defibrillators in public places and let lay-people use them?

Intraclass correlation

- Key issue: Observations within group are correlated. Intra-class correlation ICC=correlation between two observations within group.
- Equivalently, $ICC = \sigma^2_B / (\sigma^2_B + \sigma^2_W)$, where σ^2_B and σ^2_W are between- and within-group variance
 - $ICC \approx 0$ means within-group variance dwarfs between-group variance. Want large number people/group
 - $ICC \approx 1$ means between-group variance dwarfs within-group variance: Want large # groups

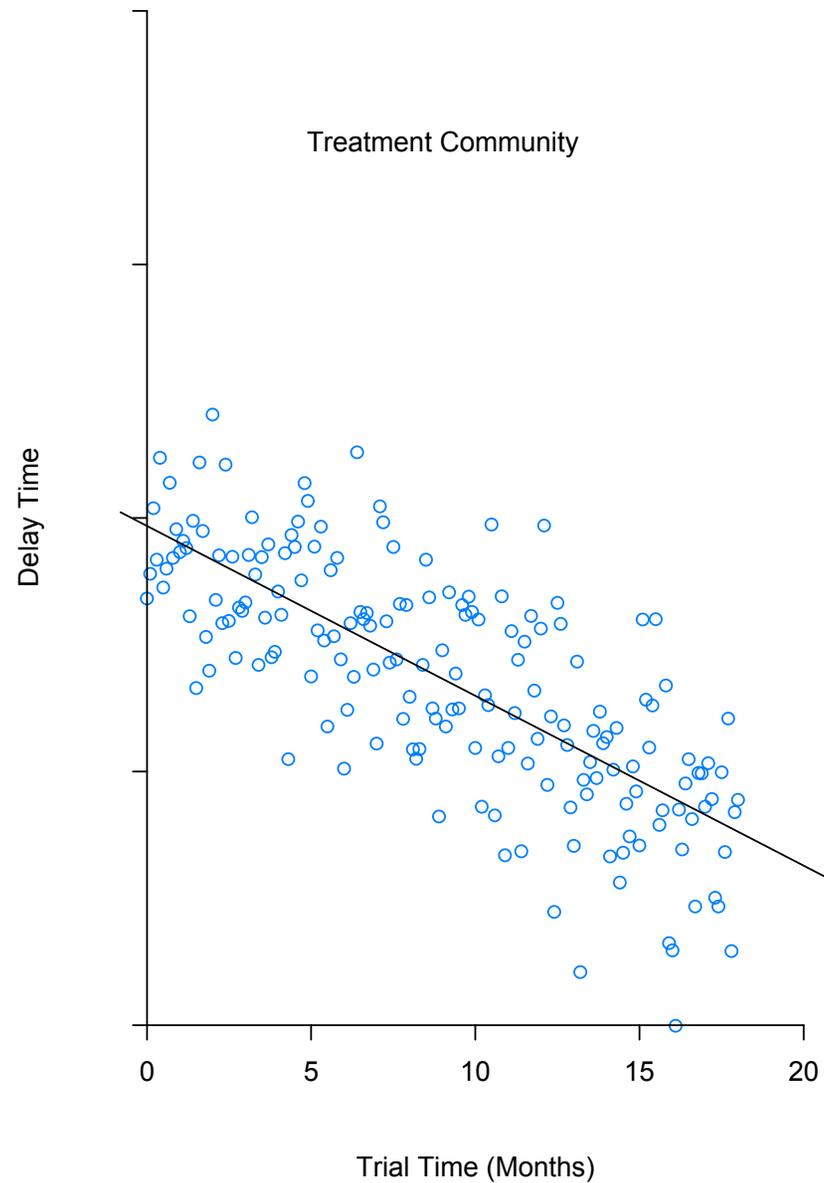
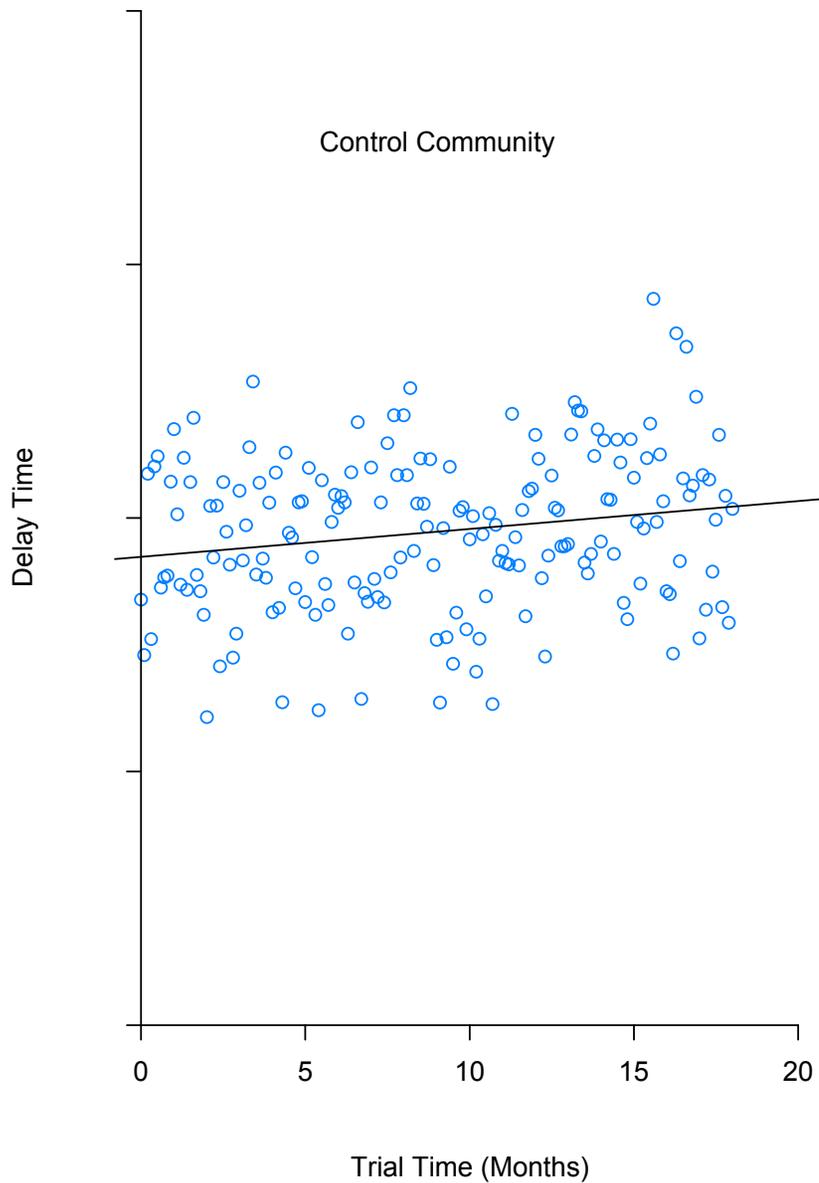
Never use < 8 groups/arm

- Problem: σ^2_B estimated very poorly. Don't really know if ICC large or small. Need fairly large # communities even if millions of people/group
- E.g., if REACT randomized only 2 communities, the U.S. and Russia, can't know whether differences are attributable to treatment or differences between countries

Useful Analysis Method

- Useful analysis method-Use a summary measure for each group, & analyze by usual methods
- E.g., REACT summarized delay times over calendar time in each community with slope. Then did paired t-test on slopes. Chose not to weight by # observations per community.

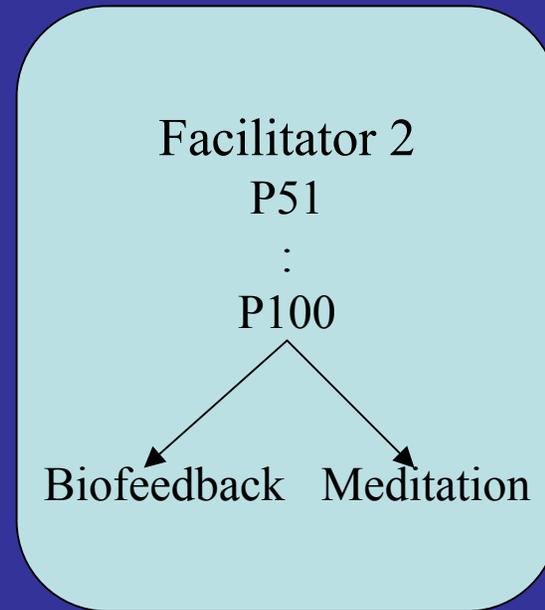
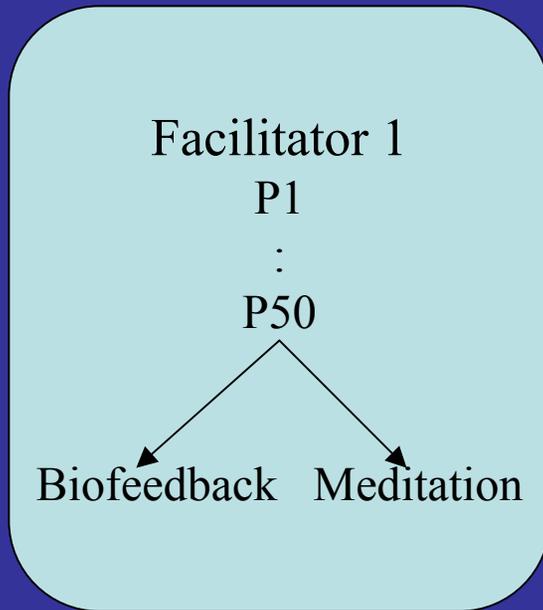
REACT



Another source of variability in group-randomized trials: Facilitator

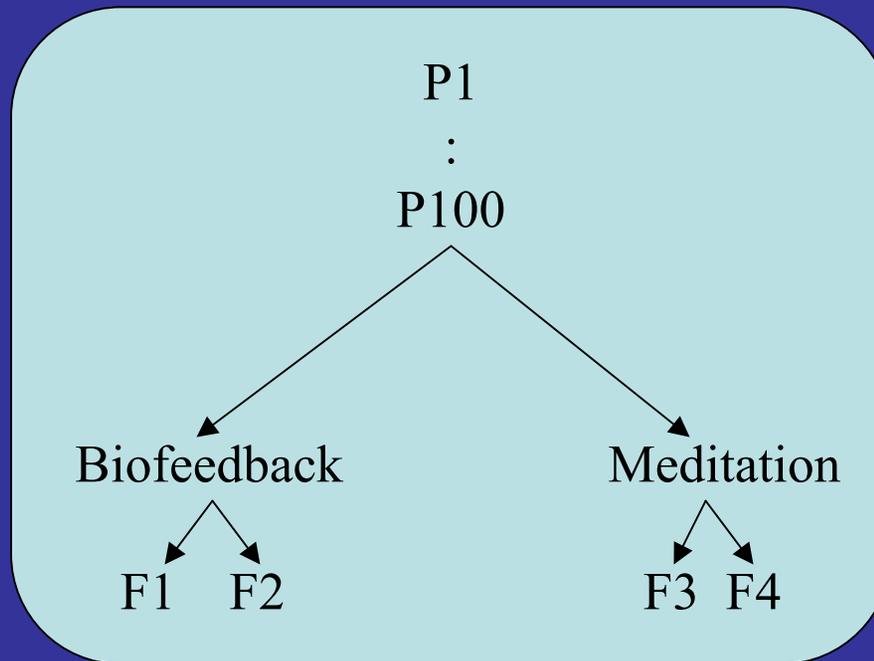
- In behavioral trials, may have group sessions with biofeedback, meditation, CBT, etc
- Facilitator may have big effect
- Could average over large number of facilitators, but practical?

Eliminate facilitator effect by having facilitators do both treatments?



Problem: Want best advocates of treatments

More common design confounds treatment & facilitator

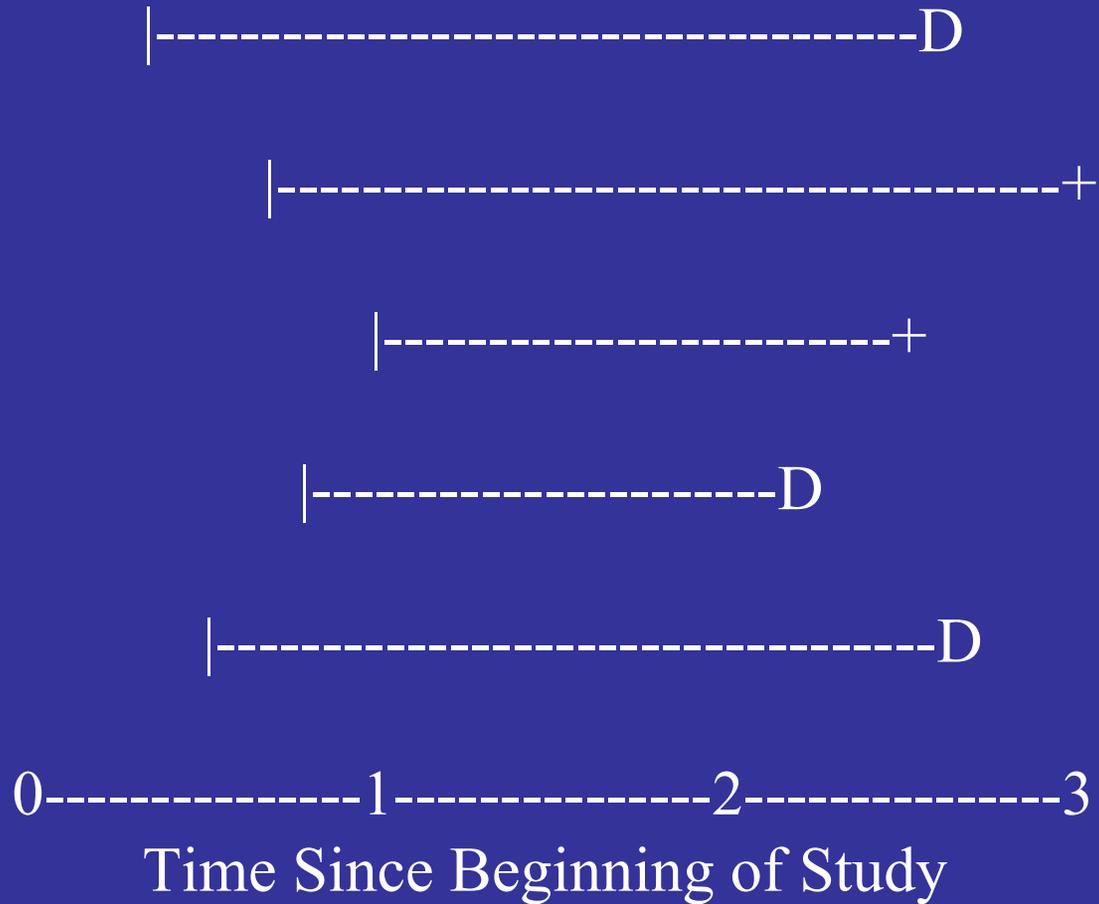


Solution: Look at facilitator effect
within each intervention & pray it is
small

Survival Analysis

- In some behavioral trials, outcome is time to event like death or MI
- Problem: Some patients lost to followup, some die from other causes, & some don't have event by end of study (censoring)
- Know only that time to event is at least so long

Differential Followup



Example: 2-year study with 11 patients:

.8, 1+, 1.2, 1.3, 1.3, 1.4, 1.8, 2+, 2+, 2+, 2+

$$P(\text{survive } .8 \text{ years given survive } 0) = 10/11$$

$$P(\text{survive } 1 \text{ year given survive } .8) = 10/10 = 1$$

$$P(\text{survive } 1.2 \text{ years given survive } 1) = 8/9$$

$$P(\text{survive } 1.3 \text{ years given survive } 1.2) = 6/8$$

$$P(\text{survive } 1.4 \text{ years given survive } 1.3) = 5/6$$

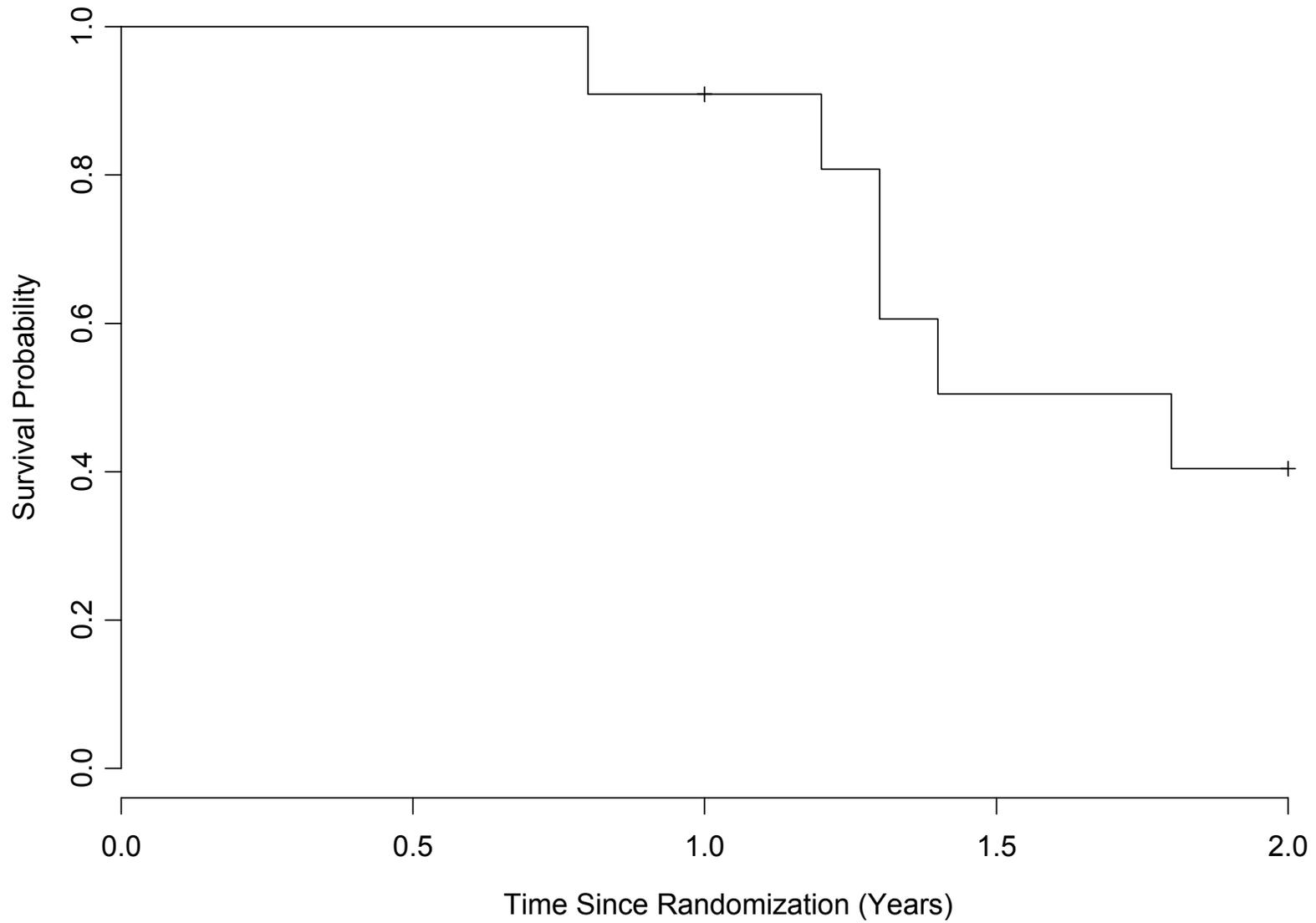
$$P(\text{survive } 1.8 \text{ years given survive } 1.4) = 4/5$$

$$P(\text{survive } 2 \text{ years given survive } 1.8) = 4/4 = 1$$

<u>t</u>	<u>P(survive t alive at t-)</u>	<u>S(t)=P(survive t)</u>
0	1	11/11=1
.8	10/11	10/11=.909
1	10/10	10/11=.909
1.2	8/9	80/99=.808
1.3	6/8	.606
1.4	5/6	.505
1.8	4/5	.404
2	4/4	.404

Called *Kaplan-Meier* estimate of survival curve

Note: Only needed to compute at death times



Censoring

- Survival methods require noninformative censoring
- Examples of noninformative censoring in trial of MI:
 - Patient moved away because of new job
 - Patient got run over by train
 - The trial ended (administrative censoring)
- Examples of informative censoring:
 - Patient had heart transplant
 - Patient quit because treatment wasn't working

Comparing two survival curves: The logrank test

Small example: Mortality trial with only 4 /arm :

<u>Control</u>	<u>Treatment</u>
.5, .75+, 2.0, 2.0+	1.5, 2.0+, 2.0+, 2.0+

Put data all together in order from smallest to largest:

.5 .75+ **1.5** 2.0 2.0+ **2.0+** **2.0+** **2.0+**

Example (continued)

								Death	P(T)	
	.5	.75+	1.5	2.0	2.0+	2.0+	2.0+	2.0+	1st	4/8
	.5	.75+	1.5	2.0	2.0+	2.0+	2.0+	2.0+	2nd	4/6
	.5	.75+	1.5	2.0	2.0+	2.0+	2.0+	2.0+	3rd	3/5

Example (continued)

Death	P(death was in treatment arm)	# deaths in treatment arm
1 st	$4/8=.500$	0
2 nd	$4/6=.667$	1
3 rd	$3/5=.600$	0
Total	1.767 (E)	1 (O)

$$(O-E)=\text{observed-expected}=1-1.767=-.767$$

To see whether statistically significant,
must know V, the variance of observed-expected

Can show that under null hypothesis,

$$V=(4/8)(1-4/8)+(4/6)(1-4/6)+(3/5)(1-3/5)=.712$$

$$\text{Std deviation}=(.712)^{1/2}=.844$$

$$Z=-.767/.844=-.909$$

If had more deaths, could refer Z to standard normal
distribution to find p-value: Two-tailed p-value= $2P(Z<-.909)=.36$

Not valid here because small number of events

Summary

- Focus on biggest variance sources—block (patient, community, facilitator ...)
 - Average over large number of blocks
 - Reduce block-block variability by adjusting for covariates
 - Eliminate block-block variability: Compute T-C in each block if possible

Conclusions (continued)

- When outcome is time to event, use survival methods
 - Kaplan-Meier curve to plot survival
 - Logrank test to compare two survival curves
- Survival methods
 - Allow differential followup
 - Assume noninformative censoring