

Modeling/Adjustments for Covariates

Michael Proschan, Ph.D.
Mathematical Statistician
National Heart, Lung, and Blood
Institute

Outline

- Continuous outcome trials: ANCOVA
 - Reduction of variability
 - Correction for baseline imbalances
 - Assumptions
 - Choice of covariates
- Dichotomous outcome trials: Logistic regression
 - What's wrong with linear regression
 - How to fix it: Log odds

Continuous Outcome Trials

Two reasons to adjust for covariates:

- 1) Reduce variability
- 2) Correct baseline imbalances

To illustrate 1), consider trial comparing new blood pressure reduction diet to standard, adjusting for baseline blood pressure

First Reason for ANCOVA: Reduce Variability

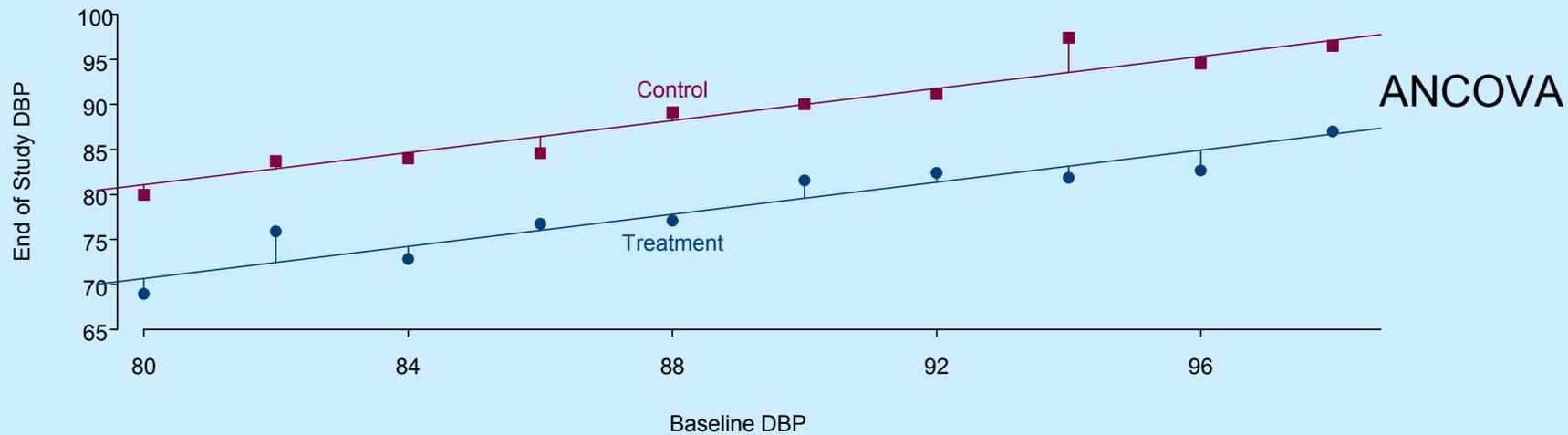
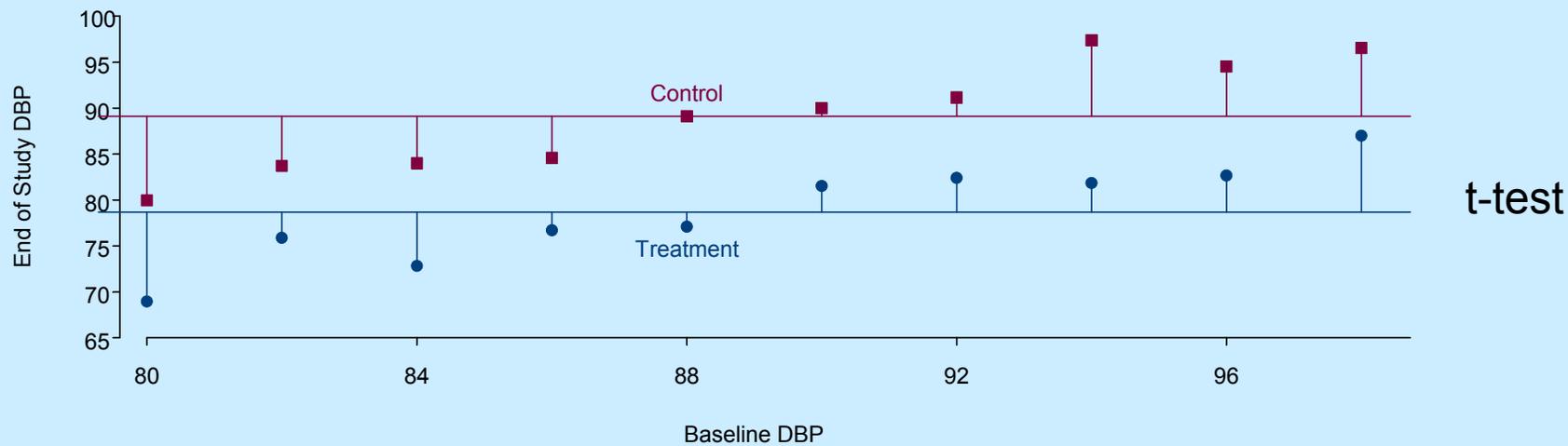
Primary outcome-Diastolic blood pressure (DBP)

- ANCOVA assumes end of study DBP, Y , linearly related to baseline DBP, X :

$$Y = \alpha_C + \beta X \quad (\text{control})$$

$$Y = \alpha_T + \beta X \quad (\text{treatment})$$

- Slopes assumed equal, intercepts may differ



Within-arm variability smaller with ANCOVA than with t-test

Second Reason For ANCOVA: Correct Baseline Imbalance

What if treatment patients had lower baseline DBP than controls?

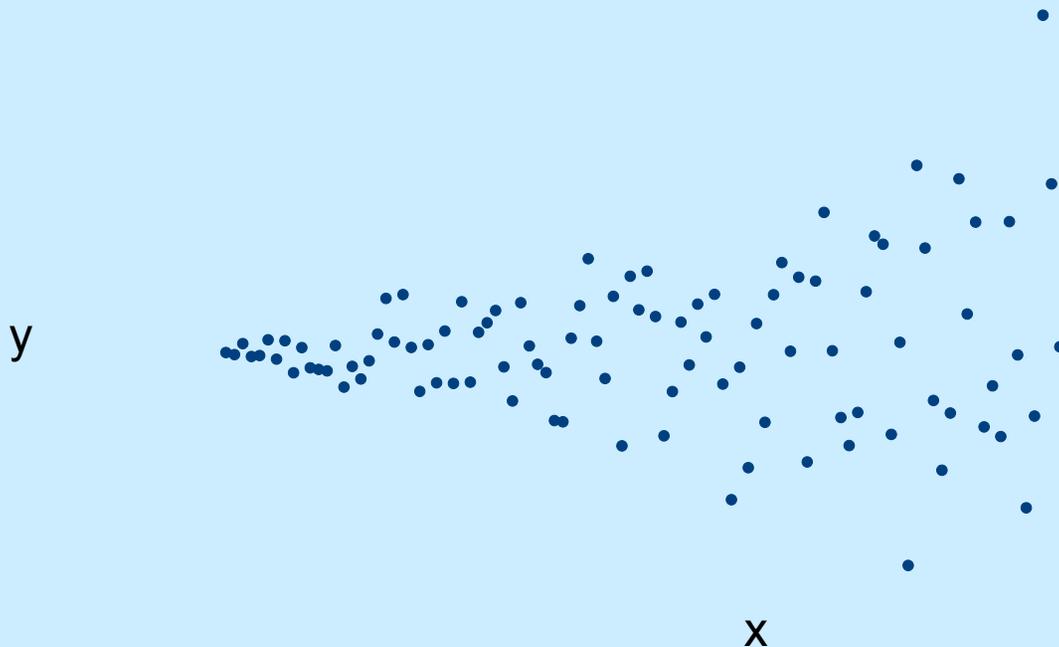
- One way to correct is to use baseline to end of study change: $Y_T - X_T - (Y_C - X_C)$
- ANCOVA is similar: $Y_T - \beta X_T - (Y_C - \beta X_C)$
- Assumes linear relationship with equal slope (β), but doesn't assume $\beta=1$

- Classical ANCOVA assumptions
 - 1) Errors are independent and normal
 - 2) Error variability constant
 - 3) Linear relationship between outcome and covariate
 - 4) Equal slopes in two arms
 - 5) Covariate measured without error

1) Errors independent and normal: P-values and confidence intervals approximately valid without assuming normal errors, but independence crucial

Can't have multiple Y values on same patient! Use mixed model

2) **Error variability constant:** Can be a problem but often ignorable unless very severe



3) Linear relationship between X and Y:

Usually a reasonable approximation—
don't overfit with higher order polynomials

4) Equal slopes in both arms*:

Assumes treatment effect does not
depend on baseline value

Holds under null hypothesis, so ANCOVA
p-values are valid

5) Covariate measured without error:

Covariate measured with error causes biased estimates of slopes and intercepts (regression dilution bias), but treatment effect estimate remains unbiased
(*Biometrics* 1987; **43**, 895-901)

- ANCOVA can also be used with binary/categorical covariates
- With single binary covariate, ANCOVA equivalent to stratified t-test
- Can have multiple covariates, some categorical & others continuous

- With multiple covariates, collinearity can cause problems
- E.g., suppose end of study weight, Y (lbs), related to baseline height, X_1 (in), by
$$Y = -200 + 6X_1 + \text{error}$$
- If stupidly include $X_2 = \text{baseline height (ft)}$, can't tell, for example, whether
$$Y = -200 + 6X_1 + 0X_2 + \text{error}$$
 or
$$Y = -200 + 0X_1 + (6 \times 12)X_2 + \text{error}$$

- Result: can't estimate slopes but still get unbiased estimate of treatment effect!
- Only collinearity **with treatment variable** causes problems, & randomization protects you! (Treatment & **baseline** covariates independent)
- Underscores admonition: **Adjust only for baseline variables** (and not too many)

Summary of Continuous Outcome Trials

- ANCOVA is good because it:
 - Reduces variability
 - Corrects baseline imbalances
- It can be used with continuous and/or categorical covariates
- It is especially important in small trials

Summary of Continuous Outcome Trials (continued)

- Adjust only for baseline covariates
- Select limited number of prognostic covariates
- Specify covariates a priori in protocol
- Don't select covariates after the fact based on treatment imbalance:

Remember, including important prognostic covariates reduces variance even if balanced

- Good guidelines for covariate adjustment: Committee for Proprietary Medicinal Products (CPMP) “Points To Consider On Adjustment For Baseline Covariates”
<http://www.emea.eu.int/pdfs/human/ewp/286399en.pdf>

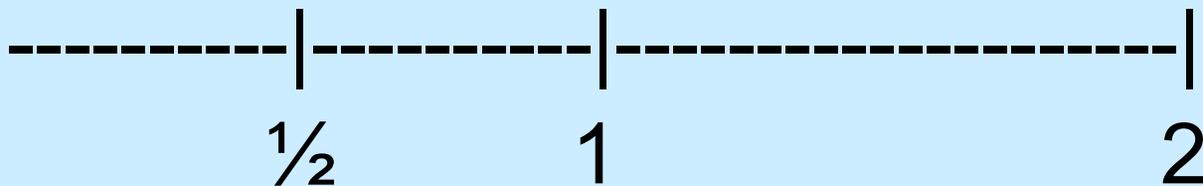
Binary Outcome Trials

- Sometimes trials use binary outcomes
 - In trial of alcoholics, outcome may be relapse within 12 months (yes/no)
 - In the Dietary Approaches to Stop Hypertension Trial (DASH) (*New England Journal of Medicine* 1997; **336**, 1117-1124), a secondary outcome was hypertensive during followup (yes/no)

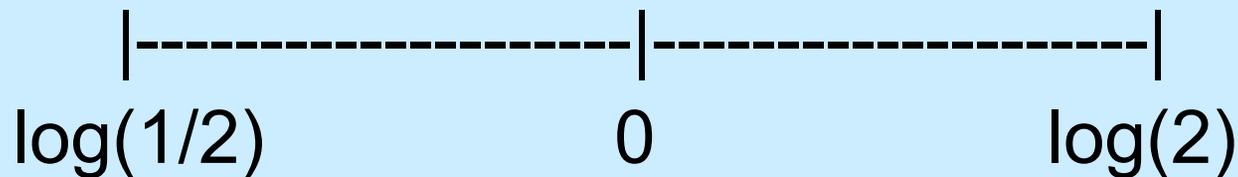
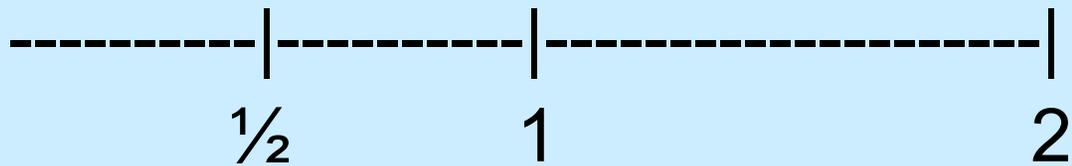
- Want to compare probabilities of events in treatment and control, adjusting for baseline covariates
- First consider one covariate
- By analogy with ANCOVA, could assume
$$p = \alpha_C + \beta x \quad (\text{Control})$$
$$p = \alpha_T + \beta x \quad (\text{Treatment})$$
- Problem: p is a probability; it can't be greater than 1, but $\alpha + \beta x$ can

- Solution: Use odds instead of probability: If $p=1/3$, then odds=1:2=1/2
- Odds= $p/(1-p)$
- No upper limit on odds (if p close to 1, odds huge)
- Try model
Odds= $\alpha_C + \beta x$ (Control)
Odds= $\alpha_T + \beta x$ (Treatment)
- Problem: Odds can't be negative, but $\alpha + \beta x$ can

- Another problem: Odds is on wrong scale; it shouldn't matter whether use odds of having event or odds of not having event
- E.g., odds of event is $1/2$, odds of no event is $2/1=2$; both should be equidistant from 1 (note: Odds of 1 means 50-50 chance)



- Solution to both problems: Use $\log(\text{odds})$
 - Can be any negative or positive number
 - Puts odds of event on same footing as odds of no event:



- $\text{Log}(\text{odds}) = \alpha_C + \beta x$ (Control)

- $\text{Log}(\text{odds}) = \alpha_T + \beta x$ (Treatment)

Called logistic regression

- As with ANCOVA, can have several binary and/or continuous covariates

- Though model is in terms of $\log(\text{odds})$, we transform treatment effect to odds ratio (O.R.):

$$\text{O.R.} = (\text{Odds of event})_T / (\text{Odds of event})_C$$

- Same covariate selection principles apply as with ANCOVA
- As with ANCOVA, logistic regression with prognostic variables increases power
- As with ANCOVA, we assume slopes are same in treatment and control*

Summary of Binary Outcome Trials

- Logistic regression is the binary outcome analog of ANCOVA
- It gives treatment effect adjusted for baseline imbalances in prognostic factors
- Like ANCOVA, it can be used with continuous and/or categorical covariates
- Like ANCOVA, it is especially important in small trials

Summary of Binary Outcome Trials (Continued)

- Principles are same as with ANCOVA
 - Include only **baseline** covariates
 - Select limited number of prognostic variables
 - Specify covariates a priori in protocol

Subgroups

- Our models have assumed treatment effect does not depend on baseline value
- Not always true

Dietary Approaches to Stop Hypertension (DASH), covariate=BL hypertension status

<u>Baseline Status</u>	<u>Combination-Control DBP</u>
Not hypertensive	-3.5 (-5.3,-1.6)
Hypertensive	-11.4 (-15.9, -6.9)

- Seems like strong evidence of difference, but is it?
- Peto (1995), section 52.4, *Treatment of Cancer*, 3rd edition, Price and Sikora, editors, Chapman and Hall, London:

With two equally-sized subgroups, if overall $Z=2$, roughly 1/3 chance that one group will have $p<.05$ and other group $p>.50$.

- With unequal size subgroups, worse

- With more than 2 groups, problem is multiplied
- ISIS-2 Trial (*The Lancet* **2**, 1988; 349-360)
Effect of aspirin on mortality

Gemini or Libra: 9% increase (NS)

All other signs: 28% decrease ($p < .000001$)

A Statistical Test of Subgroup Differences

- Modify model to include treatment by baseline hypertension interaction

$$Y = \alpha_C + \beta_C X \quad (\text{control})$$

$$Y = \alpha_T + \beta_T X \quad (\text{treatment})$$

$X=0$ for nonhypertensives, 1 for hypertensives

- Treatment effects:

$$\alpha_T - \alpha_C \quad \text{nonhypertensives}$$

$$\alpha_T - \alpha_C + (\beta_T - \beta_C) \quad \text{hypertensives}$$

Test whether $\beta_T = \beta_C$ (no interaction)

- Two counterbalancing points:
 - Statistical test has low power, so might miss true interactions (i.e., might have false negatives)
 - With many subgroups, bound to get false positives by chance alone
- What is worse: To miss a true subgroup difference or to claim a subgroup difference when none exists?

- Two types of interactions:
 - Quantitative (treatment is effective in each subgroup, but more in one than in the other)
 - Qualitative interaction (treatment helps one subgroup and hurts or has no effect on another)
- Quantitative interactions not uncommon, but qualitative interactions rare

- Be careful about claiming qualitative interactions
- Horwitz, Singer, Makuch, & Viscoli (1996)
Journal of Clinical Epidemiology **49**, 395-400

Divided trial into 21 “dominant” centers (mortality rate higher on placebo than treatment) and 10 “divergent” centers (mortality rate higher on treatment than placebo)

- Tested whether dominant centers differed from divergent centers as if subgroups had been pre-specified & got $p < .05$!
- Senn and Harrell (1997) Journal of Clinical Epidemiology 50, 749-751:
“Hindsight is so much more precise than foresight and but for its unfortunate habit of arriving too late, it would surely be used for prediction all the time”

“...we are prepared to predict that a very similar analysis applied to any multi-center trial whatsoever, in any condition with any treatment, will always be significant at the 5% level provided only that the number of centers is at least equal to 8.”

“...better a harmless drudge than an aimless dredge”

A Safer Strategy

- If think a priori may be a qualitative interaction, power study within each subgroup & do separate tests
- If not, power based on overall test; do interaction test and if not significant, go no further
- Whether include interaction or not, still get good estimate of overall effect