Introduction to Clinical Trials - Day 2

Session 3 - Methods of Randomization

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

Daniel L. Gillen
Department of Statistics
University of California, Irvine

Why randomization?

Consider the scientific objective

- ICH guidelines (www.ich.org) part E9 Statistical Principles

  “The most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be normal features of most controlled clinical trials intended to be included in a marketing application."

- Similar criteria are required in the CONSORT guidelines.

Bias

What is bias?

- Bias is a tendency of a statistical estimate to deviate in one direction from a“true value"
- What defines the “truth" is dictated by the scientific goal
- Randomization is the primary tool of a clinical trialist for reducing bias
- In order to illustrate the role in which bias arises in clinical studies and motivate the role of randomization, it is useful to review the components of a statistical model in order to:
  1. Develop a standard nomenclature
  2. Illustrate the goals and impact of randomization
- To this end, we can begin withe role of adjustment variables in statistical models

Example - FEV Data

Is there an association between smoking and lung function in children?

- Scientific justification
  - Longterm smoking is associated with lower lung function
  - Are similar effects observed in short term smoking in children?
- Causal pathway of interest
  - Interested in whether smoking will cause a decrease in lung function

Smoking ⟶ Lung function

Example - FEV Data

Study design

- Observational study
  - Measurements obtained on a sample of 654 healthy children
  - Children were sampled while being seen for a regular checkup
  - Data available on smoking, age, gender, and height
  - Predictor of interest: Self-reported smoking
  - Response: FEV (Forced Expository Volume)

FEV Data

SMOKERS

1.953 2.236 3.428 3.208 1.694 3.957 4.789 2.384 3.074 2.387 3.835 2.599 4.756 3.086 4.309 3.413 2.975 3.169 3.343 3.751
2.216 3 .078 3.186 3.297 2.304 3.102 2.677 3.297 3.498 2.759 2.953 3.785 2.276 4.637 3.038 3.120 3.339 3.152 3.104 4.045
4.763 3.069 4.506 3.519 3.688 2.679 2.198 3 .345 3.082 2.903 3.004 3.406 3.122 3.330 2.608 3.799 4.086 4.070 2.264 4.404
2.278 4.872 4.270 3.727 2.795

NONSMOKERS

1.708 1.724 1.720 1.558 1.895 2.336 1.919 1.415 1.987 1.942 1.602 1.735 2.193 2.118 2.258 1.932 1.472 1.878 2.352 2.604
1.400 1 .256 0.839 2.578 2.988 1.404 2.348 1.755 2.980 2.100 1.282 3.000 2.673 2.093 1.612 2.175 2.725 2.071 1.547 2.004
3.135 2.420 1.776 1.931 1.343 2.076 1.624 1 .344 1.650 2.732 2.017 2.797 3.556 1.703 1.634 2.570 3.016 2.419 1.569 1.698
2.123 2.481 1.481 1.577 1.940 1.747 2.069 1.631 1.536 2.560 1.962 2.531 2.715 2 .457 2.090 1.789 1.858 1.452 3.842 1.719
2.111 1.695 2.211 1.794 1.917 2.144 1.253 2.659 1.580 2.126 3.029 2.964 1.611 2.215 2.388 2.196 1.751 2.165 1.682 1 .523
1.292 1.649 2.588 0.796 2.574 1.979 2.354 1.718 1.742 1.603 2.639 1.829 2.084 2.220 1.473 2.341 1.698 1.196 1.872 2.219
2.420 1.827 1.461 1.338 2.090 1 .697 1.562 2.040 1.609 2.458 2.650 1.429 1.675 1.947 2.069 1.572 1.348 2.288 1.773 0.791
1.905 2.463 1.431 2.631 3.114 2.135 1.527 2.293 3.042 2.927 2.665 2 .301 2.460 2.592 1.750 1.759 1.536 2.259 2.048 2.571
2.046 1.780 1.552 1.953 2.893 1.713 2.851 1.624 2.631 1.819 1.658 2.158 1.789 3.004 2.503 1.933 2.091 2 .316 1.704 1.606
1.165 2.102 2.320 2.230 1.716 1.790 1.146 2.187 2.717 1.796 1.335 2.119 1.666 1.826 2.709 2.871 1.092 2.262 2.104 2.166
1.690 2.973 2.145 1 .971 2.095 1.697 2.455 1.920 2.164 2.130 2.993 2.529 1.726 2.442 1.102 2.056 1.808 2.305 1.969 1.556
1.072 2.042 1.512 1.423 3.681 1.991 1.897 1.370 1.338 2 .016 2.639 1.389 1.612 2.135 2.681 3.223 1.796 2.010 1.523 1.744
2.485 2.335 1.415 2.076 2.435 1.728 2.850 1.844 1.754 1.343 2.303 2.246 2.476 3.239 2.457 2 .382 1.640 1.589 2.056 2.226
1.886 2.833 1.715 2.631 2.550 1.912 1.877 1.935 1.539 2.803 2.923 2.358 2.094 1.855 1.535 2.135 1.930 2.182 1.359 2.002
1.699 2 .500 2.366 2.069 1.418 2.333 1.514 1.758 2.535 2.564 2.487 1.591 1.624 2.798 1.691 1.999 1.869 1.004 1.427 1.826
2.688 1.657 1.672 2.015 2.371 2.115 2.328 1 .495 2.884 2.328 3.381 2.170 3.470 3.058 1.811 2.524 2.642 3.741 4.336 4.842
4.550 2.841 3.166 3.816 2.561 3.654 2.481 2.665 3.203 3.549 3.222 3.111 3.490 3 .147 2.520 2.292 2.889 2.246 1.937 2.646
2.957 4.007 2.386 3.251 2.762 3.011 4.305 3.906 3.583 3.236 3.436 3.058 3.007 3.489 2.864 2.819 2.250 4.683 2.352 3 .108
3.994 4.393 2.592 3.193 2.346 3.515 2.754 2.720 2.463 2.633 3.048 3.111 3.745 2.094 3.183 3.977 3.354 3.411 3.171 3.887
2.646 2.504 3.587 3.845 2.971 2 .891 1.823 2.417 2.175 2.735 4.273 2.976 4.065 2.318 3.596 3.395 2.751 2.673 2.556 2.542
2.608 2.354 1.458 3.795 2.491 3.060 2.545 2.993 3.305 3.774 2.855 2 .988 2.498 3.169 2.887 2.704 3.515 3.425 2.287 2.434
2.365 2.696 2.868 2.813 3.255 4.593 4.111 1.916 1.858 3.350 2.901 2.241 4.225 3.223 5.224 4.073 4.080 2 .606 4.411 3.791
3.089 2.465 3.200 2.913 4.877 2.358 3.279 2.581 2.347 2.691 2.827 1.873 2.538 2.758 3.050 3.079 2.201 1.858 3.403 3.501
2.578 1.665 2.081 2 .974 4.073 4.448 3.984 2.250 2.752 3.680 2.862 3.023 3.681 3.255 3.692 2.356 4.591 3.082 3.258 2.216
3.247 4.324 2.362 2.563 3.206 3.585 4.720 3.331 5.083 2 .417 2.364 2.341 3.231 3.078 3.369 3.529 2.866 2.891 3.022 3.127
2.866 2.605 3.056 2.569 2.501 3.320 2.123 3.780 3.847 3.924 2.132 2.752 2.449 3.456 3.073 2 .688 3.329 4.271 3.530 2.928
2.689 2.332 2.934 3.110 2.894 2.435 2.838 3.035 4.831 2.812 2.714 3.086 3.519 4.232 2.770 3.341 3.090 2.531 2.822 2.935
2.568 2 .387 2.499 4.130 3.001 3.132 3.577 3.222 3.280 2.659 2.822 2.140 4.203 2.997 2.562 3.082 3.806 2.458 2.391 3.141
2.579 2.100 2.785 4.284 2.906 5.102 4.429 4 .279 4.500 2.635 3.082 3.387 5.793 3.985 4.220 4.724 3.731 3.500 3.674 5.633
3.645 2.887 3.960 4.299 2.981 4.504 5.638 2.853 3.211

Example - FEV Data

Interpretation of smoking effect in unadjusted analysis

- Restrict sample to children 9 years and above (age of youngest smoker in sample)
- Consider log-transformation of FEV based upon past studies
  - Scientific focus on median FEV
  - Distribution of log-transformed FEV approximately symmetric

Example - FEV Data

Unadjusted association between smoking and FEV

- Consider an unadjusted comparison of FEV between smokers and non-smokers
  - Unadjusted Result: The median FEV of a smoker is estimated to be 10.8% higher than that of a non-smoker (95% CI: 1.04, 1.18). This difference is statistically significant $p = 0.002$.

Example - FEV Data

Adjustment for age

- Consider adjustment for age in a linear regression model
  - Age-adjusted result: The median FEV of a smokers is estimated to be 5.0% lower than that of non-smokers *similar in age* (95% CI: 0.90, 1.01). This difference is not statistically significant at the .05 level ($p = 0.093$).

Example - FEV Data

Adjustment for age and height

- After adjustment for age, height should have little association with smoking status but is still likely to have an association with FEV.
- Consider additional adjustment for height...
  - Age and height-adjusted result: The median FEV of smokers is estimated to be 5.2% lower than that of non-smokers *similar in age and height* (95% CI: 0.91, 0.99). This difference is statistically significant at the .05 level ($p = 0.011$).

Example - FEV Data

Comparison of age and age-height adjusted analyses

- Notice that there is little difference in estimated effect of smoking between age adjusted models with and without height
- Effect of height adjustment on precision
  - Lower Root MSE (.144 vs .209) in height adjusted model resulting in increased precision of estimate of smoking effect
  - Net effect: Much greater precision (SE 0.021 vs 0.031)

Example - FEV Data

Take-home message

- Our scientific question was not

  "Is there a difference between smokers' and nonsmokers' median FEV?"

- But rather

  "Do smokers have lower median FEV than otherwise comparable nonsmokers?"

Example - FEV Data

Take-home message

- This example highlights:
  1. How a scientific question should dictate a chosen statistical model
  2. The role of a *confounding* variable on association estimates
  3. The impact that adjustment has on the precision of association estimates
- These ideas provide the motivation for randomization, as well as the types and implementation of various randomization methods
- However, before going there, it is useful to define the statsitical role of variables and to generalize the observations that were made in the FEV example...

Statistical role of variables

Effect modifiers (interaction terms)

- Suppose that we are interested in modeling the association between an outcome variable Y and a predictor X
- Consider four broad categories of variables (this terminology is not universal)
- Effect modifiers (interaction variables)
  - An effect modifier (W) is a covariate for which the association between the predictor of interest (X) *and* the outcome of interest (Y) differs with each level of W

Statistical role of variables

Example: Effect modification

- Example: The association between gender and the risk of chd differs by systolic blood pressure

sbpgrp	Odds Ratio	chi2(1)	P>chi2	[95% Conf.	Interval]
1	0.394493	86.23	0.0000	0.32186	0.48351
2	0.429583	56.59	0.0000	0.34243	0.53892
3	0.597384	9.91	0.0016	0.43193	0.82621
4	0.741269	1.75	0.1858	0.47495	1.15693

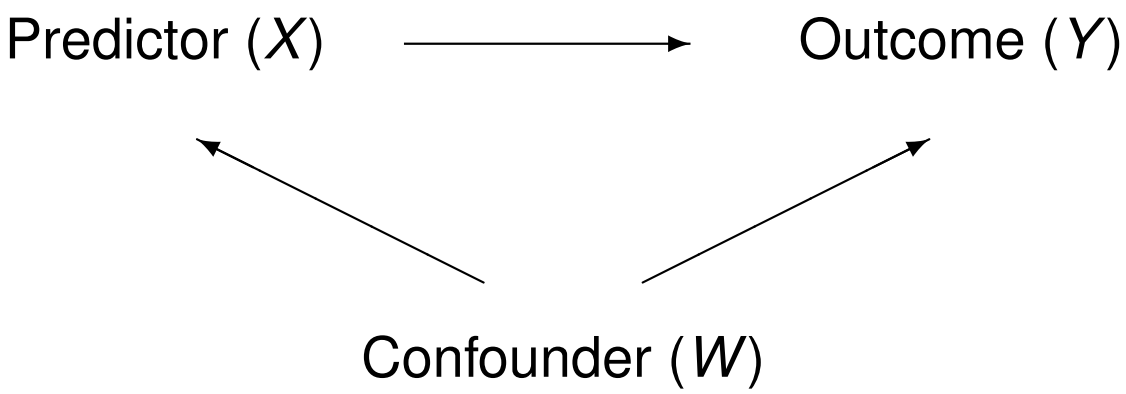
How do we deal with effect modifiers?

- When the scientific question involves effect modification, analyses must be within each stratum separately

Statistical role of variables

Confounders

- One definition: A **confounder** is a variable that is causally related to the predictor of interest (X) *and* the outcome of interest (Y).



Statistical role of variables

Example: Confounding

- Example: Age in the FEV example:
  - Older kids tend to smoke
  - Older kids tend to have larger lungs

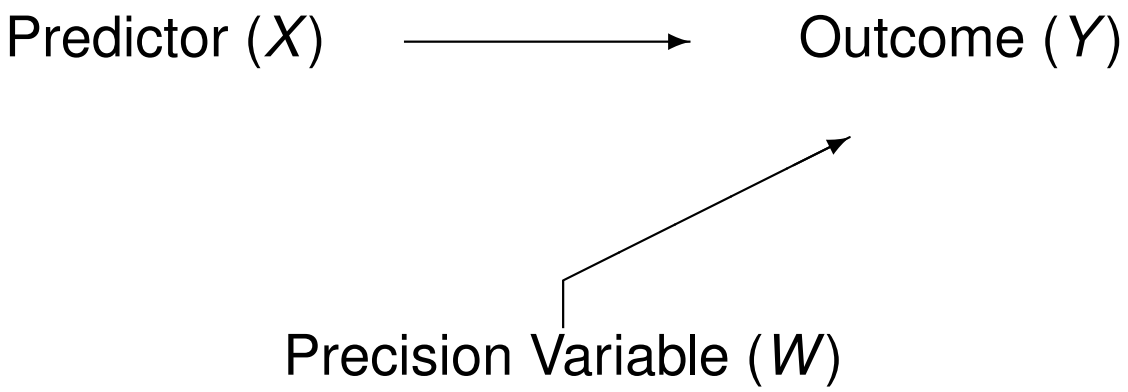
How do we deal with confounding?

- Adjust for the confounder

Statistical role of variables

Precision variables

- I define a **precision variable** as a covariate that is related to the outcome Y, but independent of the predictor of interest X.



Statistical role of variables

Example: Precision variable

- Example: Height (after adjustment for age) in the FEV example:
  - Conditional on age, little difference in prevalence of smoking by height
  - Conditional on age, taller kids tend to have larger lungs

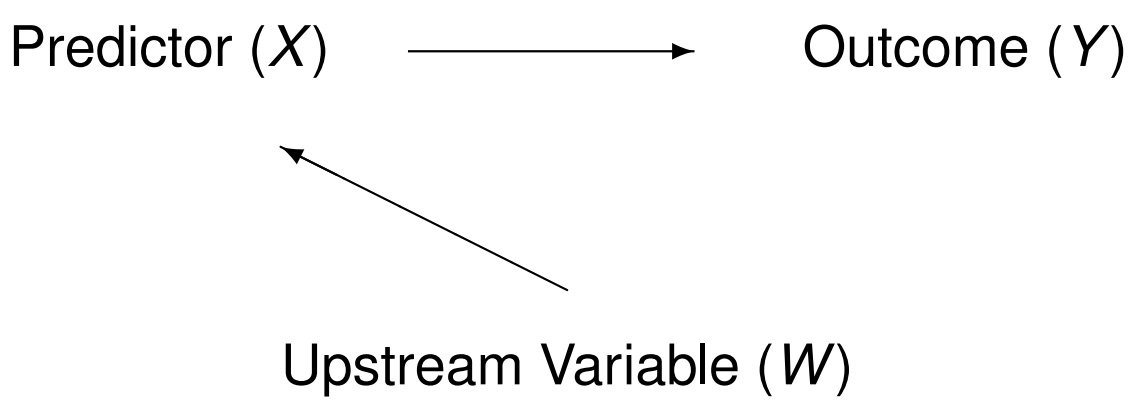
How do we deal with precision variables?

- Often a good idea to control for them
- For example, in a two sample comparison of means, we might control some variable in order to decrease the within group variability

Statistical role of variables

“Upstream" variables

- I define an **upstream variable** as a covariate that is independent of the outcome Y, but may or may not be related to the predictor of interest X.



- Generally a bad idea to adjust for “upstream" variables

Statistical role of variables

Why randomize?

- The fundamental statistical distinctions between unadjusted and adjusted regression models are central to the goals of randomization
- We thus want to be able to consider the relationships between
  - unadjusted and adjusted parameters, and
  - the standard errors of the two parameter estimates
- This is easily done in the context of linear regression and that will be the setting for our discussion
  - Results are less straightforward for non-linear models (eg. logistic regression or proportional hazards)
  - However, the general principles still apply

Statistical role of variables

Adjusted vs. unadjusted covariate effects

- Consider the following linear regression models:

1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
   - β_1 is the difference in the mean of Y for groups differing by 1-unit in X
2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i$
   - γ_1 is the difference in the mean of Y for groups differing by 1-unit in X, but agreeing in their value of W

Adjusted vs. unadjusted covariate effects

- Proposition 1: Let $\hat{\beta}_1$ denote the OLS estimate of β_1. Then under the adjusted model,

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{XW}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

where r_{XW}, $\mathrm{var}(X)$, and $\mathrm{var}(W)$ are the sample correlation between X and W, sample variance of X, and sample variance of W, respectively.

Statistical role of variables

The implication...

- $\hat{\beta}_1$ is biased (and inconsistent) for γ_1 unless at least one of the following hold

  1. $r_{XW} = 0$: X and W are uncorrelated (in the sample), OR
  2. $\gamma_2 = 0$: W is not related to Y

- In either case, $\hat{\beta}_1$ is unbiased (and consistent) for β_1

- Implication for confounders?

  - By definition, a confounder is related to the predictor of interest and the response

  - This implies that if W is a confounder, then both conditions above fail

  - Hence the parameter from the reduced model is biased for the adjusted estimate

Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

- Consider the following linear regression models:

1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i$

Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

- Proposition 2:
  1. For the unadjusted model,

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2_{Y|X}}{n\text{var}(X)}$$

  2. For the adjusted model,

$$\text{Var}[\hat{\gamma}_1] = \frac{\sigma^2_{Y|X,W}}{n\text{var}(X)(1 - r^2_{XW})}$$

  where $\sigma^2_{Y|X,W} = \sigma^2_{Y|X} - \gamma_2^2 \text{var}(W|X)$

- Hence, if $\gamma_2 \neq 0$ then $\sigma^2_{Y|X,W} < \sigma^2_{Y|X}$

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 1: $r_{XW} = 0$ (X and W uncorrelated) and $\gamma_2 = 0$ (W and Y unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] = \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Lose 1 degree of freedom for hypothesis tests and CIs if adjusting for W

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 2: $r_{XW} \neq 0$ (X and W correlated) and $\gamma_2 = 0$ (W and Y unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] < \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Mathematically estimating the same quantity but *lose* precision when adjusting for W (nuisance variable)

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 3: $r_{XW} = 0$ (X and W uncorrelated) and $\gamma_2 \neq 0$ (W and Y related)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] > \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Mathematically estimating the same quantity but *gain* precision when adjusting for W (precision variable)

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 4: $r_{XW} \neq 0$ (X and W correlated) and $\gamma_2 \neq 0$ (W and Y related)
  - From Proposition 1, $\hat{\beta}_1$ biased for γ_1
  - From Proposition 2, no definitive statement about the variances
  - Conclusion: W is a confounder and decision to adjust should be based on what you are trying to estimate.

Statistical role of variables

Why do we care?

- The above results provide the fundamental motivation for

  1. The use and types of randomization (balance of confounders)

  2. The consideration of analytic methods under various types of randomization

Methods of Randomization

Cause and Effect

- Necessary conditions for establishing cause and effect of a treatment

  1. The treatment should precede the effect

     - Beware protopathic signs (eg. Marijuana and risk of MI within 3 hours)

  2. When comparing groups differing in their treatment, the groups should be comparable in every other way (at baseline) (see previous discussion on confounding)

Methods of Randomization

Cause and Effect

Randomization is the major way in which cause and effect is established

- Ensures comparability of populations
  - Each treatment group drawn from same population
  - Differences in other prognostic factors will only differ by random sampling
    - Provides balance on the total effect of all other prognostic factors
    - May not provide balance on each individual factor
- Note: Sequential allocation of patients is not randomization
  - Possible time trends in recruitment, treatments, etc.

Nonadaptive Randomization

General statements on randomization

- Randomization is our friend...
  - If we randomize, we do not (on average) need to worry about differences between the treatment groups with respect to factors present at time of randomization
    - Any difference in outcomes can be attributed to treatment
    - However, recognize that treatment can lead to differential use of other ancillary treatments
- But like all friends, we must treat it with respect.
  - We must analyze our data in groups defined at the time of randomization
    - Discarding or missing data on randomized subjects may lead to bias (It certainly leads to diminished scientific credibility)

Nonadaptive Randomization

Impact on data analysis

In presence of randomized treatment assignment

- Intent to treat analysis (ITT)
  - Based on randomization
- Confounding not an issue (on average)
  - P value measures probability of observed effects occurring due only to randomization imbalance
- Gain precision if
  - Control important predictors, or
  - Adjust for stratification variables
- Subgroup analyses
  - If effect modification is concern
  - Pre-specification

Nonadaptive Randomization

Randomization strategies

- Complete randomization (CRD)
- Blocked randomization
  - Ensure balance after every k patients
  - Ensure closer adherence to randomization ratio
  - Undisclosed block sizes to prevent bias
- Stratified randomization
  - Separately within strata defined by strong risk factors
    - Lessens chance of randomization imbalance
  - Need to consider how many variables can be used
- Dynamic randomization
  - Adaptive randomization to achieve best balance on marginal distribution of covariates

Nonadaptive Randomization

Complete randomization

- The simplest form of randomization is independent randomization of each individual
- With each accrued subject a (possibly biased) coin is tossed to determine which arm
  - Probability of treatment arm = $r/(r+1)$
  - Independence of successive randomizations
- Possible issues with complete randomization include
  - Bias,
  - Face validity, and
  - Precision

Nonadaptive Randomization

Complete randomization

- On average (across repeated experiments)
  - No correlation between treatment variable and other covariates
  - Individual type I errors come from samples in which other covariates are imbalanced

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{xw}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

Nonadaptive Randomization

Complete randomization

- Typical to consider face validity of randomization in a “Table 1"

	Methotrexate Arm		Placebo Arm	
	n	Mean (SD; Min – Max)	n	Mean (SD; Min – Max)
Age (yrs)	132	50.4 (8.5; 32 - 69)	133	52.2 (8.5; 26 - 67)
Female	132	92.4%	133	92.5%
Pruritus score	116	7.7 (3.8; 4 - 16)	124	6.9 (3.8; 4 - 20)
Splenomegaly	131	8.4%	133	10.5%
Telangiectasia	132	4.6%	133	11.3%
Edema	132	6.1%	133	3.0%
Alkaline phosphatase	132	242.6 (145.9; 53 - 933)	133	245.0 (187.6; 66 - 1130)
ALT	131	54.5 (41.7; 12 - 202)	132	50.6 (41.4; 12 - 311)
Total bilirubin	132	0.7 (0.4; 0.1 - 2.7)	133	0.7 (0.4; 0.1 - 2.4)
Albumin	132	4.0 (0.3; 3.1 - 6.0)	133	4.0 (0.3; 3.0 - 4.8)
Prothrombin time INR	124	1.0 (0.1; 0.7 - 1.3)	132	1.0 (0.1; 0.7 - 1.3)
Mayo score	128	3.8 (0.8; 1.6 - 6.3)	133	3.9 (0.8; 1.6 - 6.1)

Nonadaptive Randomization

Complete randomization

- Consider differences in baseline stoke severity in a multi-center randomized clinical trial comparing tissue plasminogen activator (tPA) for the treatment of acute ischemic stroke
  - Percentage of patients (N = 320) in the 91 to 180-minute subgroups with a specific baseline National Institutes of Health Stroke Scale (NIHSS) score (Marler et al., *Neurology*, 2000)

Baseline NIHSS score	*tPA-treated patients, % (n = 153)*	*Patients given placebo, % (n = 167)*
0-5	19.0	4.2
6-10	24.2	27.5
11-15	17.0	21.0
16-20	21.6	19.8
>20	18.3	27.5
tPA = tissue plasminogen activator		

"The marked imbalance in baseline stroke severity in the 91 to 180-minute groups of the NINDS trial suggests that the NINDS trial lacks internal validity." -Mann, *West J. of Med* (2002)

Nonadaptive Randomization

Complete randomization

- Table 1: Potential for imbalance in covariates
  - Depends on number of covariates and correlations among them
  - Probability of at least one “significant" imbalance

Number	Worst	Correlation			
Displayed	Case	0.00	0.30	0.50	0.75
1	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081
3	.150	.143	.137	.126	.104
5	.250	.226	.208	.184	.138
10	.500	.401	.353	.284	.193
20	1.000	.642	.540	.420	.258
50	1.000	.923	.806	.624	.353

Nonadaptive Randomization

Complete randomization

- Of course, statistical significance is not the issue
- The real concern is “conditional confounding"
  - How does unadjusted estimate compare to adjusted estimate?
  - Product of sample correlation between X (treatment) and W (potential confounder) and adjusted association between Y (outcome) and W

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{XW}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

Nonadaptive Randomization

Complete randomization

- Spurious results due to covariate imbalance
  - Unconditionally: Unbiased so no problem
    - CONSORT Item 15 : “Although proper random assignment prevents selection bias, it does not guarantee that groups are equivalent at baseline. Any differences in baseline characteristics are, however, the result of chance rather than bias."
  - Conditional on obtained randomization:
    - IF covariates are strongly predictive of outcome, then covariate imbalance increases type I error
    - But need to consider that combined effect of other measured and unmeasured covariates may provide balance
- Ultimately, however, we need to have credible results
  - We do not always get to choose what others believe

Nonadaptive Randomization

Precision

- Impact of completely randomized design on precision of inference
  - Impact of imbalance in sample sizes
    - The number accrued to each arm is random
  - Impact of imbalance in covariates
    - “One statistician’s mean is another statistician’s variance"

Nonadaptive Randomization

Randomization ratio

- Most efficient
  - When test statistics involve a sum, choose ratio equal to ratio of standard deviations
- Most ethical for patients on study
  - Assign more patients to best treatment
    - Many sponsors / patients presume new treatment
    - (Adaptive randomization: Play the winner)
- Most ethical for general patient population
  - Whatever is most efficient (generally not adaptive)
- Other goals
  - Attaining sufficient patients exposed to new treatment
  - Maintaining DSMB blind

Nonadaptive Randomization

Randomization ratio : Optimal r (fixed n)

- Suppose we are constrained by maximal sample size $n = n_1 + n_2$
- Smallest standard error when

$$r = \frac{n_1}{n_2} = \frac{s_1}{s_2}$$

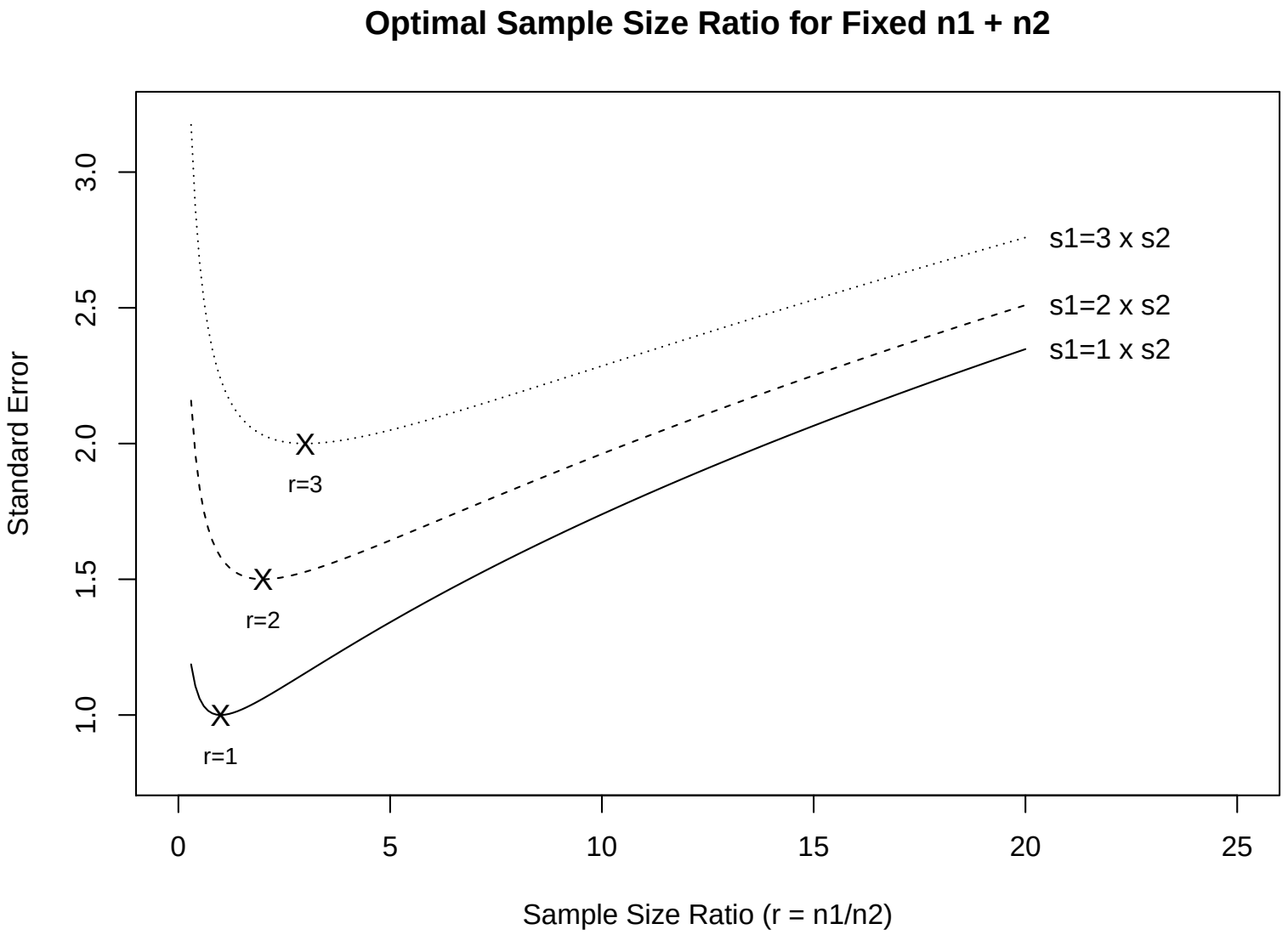
where s_i is the standard deviation of response in group i,

$i = 1, 2$

- When we are unconstrained by maximal sample size we still hit a point of diminishing returns
  - Often quoted: $r = 5$
  - Really depends on ratio of standard deviations...

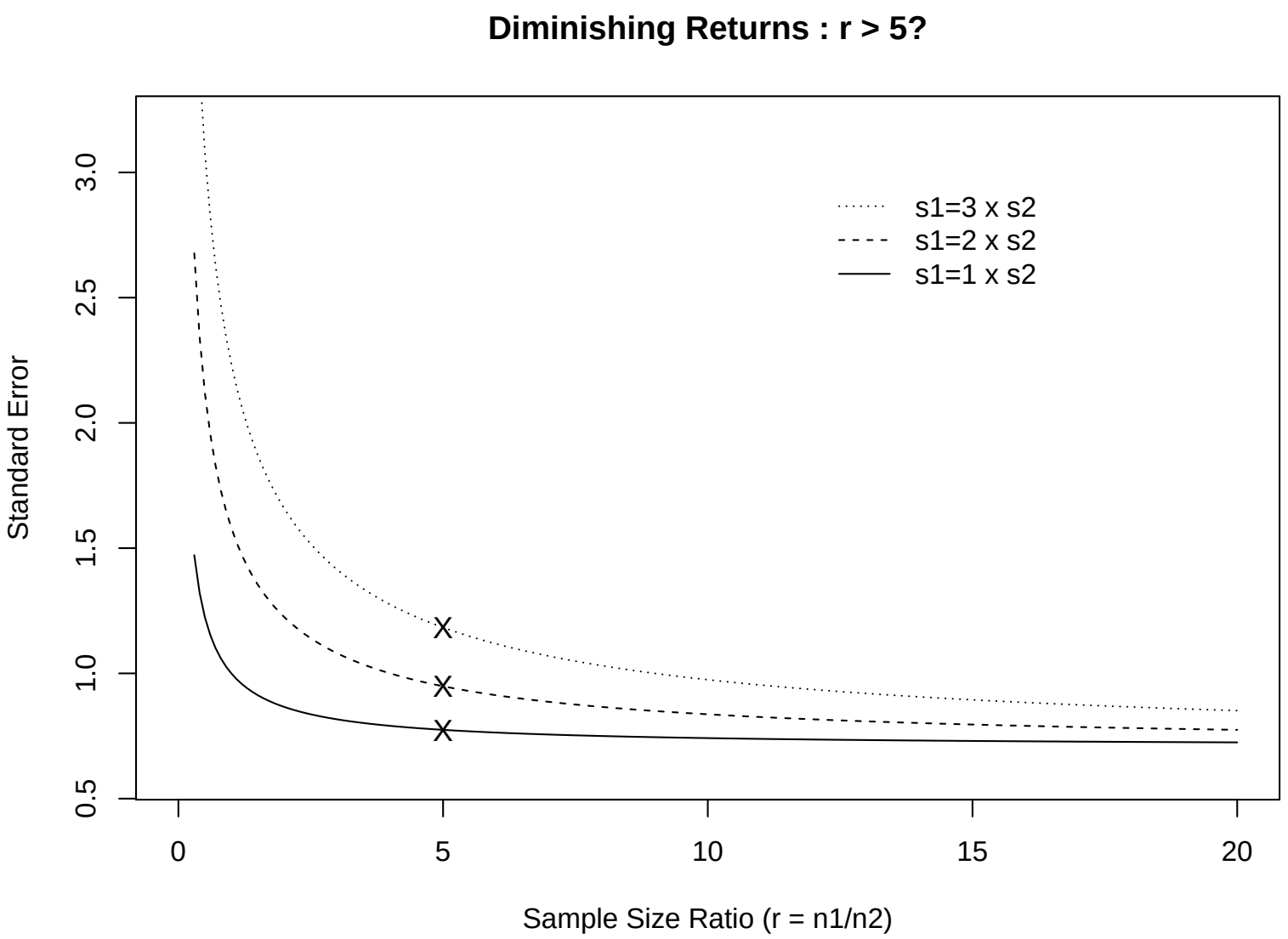
Nonadaptive Randomization

Randomization ratio : Optimal *r* (fixed *n*)



Nonadaptive Randomization

Randomization ratio : Diminishing returns



Nonadaptive Randomization

Complete randomization

- It is possible, in smaller studies, that a completely randomized design with high randomization ratio may not randomize at least two subjects to each arm
- Consider the probability that a CRD may not randomize at least two subjects to each arm as a function of the total trial size and randomization ratio

N	r= 1	r= 2	r= 3	r= 5	r=10
20	0.0000	0.0033	0.0243	0.1304	0.4459
50	0.0000	0.0000	0.0000	0.0012	0.0511
100	0.0000	0.0000	0.0000	0.0000	0.0008
200	0.0000	0.0000	0.0000	0.0000	0.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000

Efficiency loss from imbalance

- Covariates may be imbalanced across arms
  - Variability across replicated experiments increased if important predictor not controlled
  - Recall

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2_{Y|X}}{n\text{var}(X)}$$

with

$$\sigma^2_{Y|X} = \gamma_2^2 \text{var}(W|X) + \sigma^2_{Y|X,W}$$

Nonadaptive Randomization

How to improve performance?

- If we adjust for important covariates, we will often gain precision
  - Face validity in Table 1 if readers recognize that adjustment accounts for any observed imbalance
- Caveats:
  - If covariate imbalance by arm, model misspecification can be an issue regarding conditional bias
  - If covariate imbalance by arm, lack of effect can be an issue regarding variance inflation
  - If adjustment not TOTALLY prespecified, “intent to cheat" analysis can be an issue
    - Loss of precision from imperfect model should not be too much of an issue in most situations

Nonadaptive Randomization

Issues with complete randomization

- Imbalance across arms in sample sizes
  - Not much of an issue with large sample sizes
  - Could be problematic with sequential sampling
    - Interim analyses of data early in the study
- Imbalance across arms in time trends
  - Outcome may be associated with time of accrual
- *Blocking* is sometimes used to ensure
  - Proper ratio of sample sizes across groups, and
  - Balance across arms over time

Nonadaptive Randomization

Mechanisms leading to time trends

- Patients accrued early may differ from those accrued later, because
  - Backlog of eligible patients
  - Startup of new clinical sites
  - Pressure to increase accrual
  - Secular trends in beliefs about intervention
    - (Made much worse if any interim results leak out)
  - Secular trends in diagnostic tools used for eligibility
  - Secular trends in ancillary treatments

Nonadaptive Randomization

Mechanisms leading to time trends

- Within every sequence of k patients, the ratio of treatment to control is exactly $r : 1$
  - Within each “block" ordering of treatments is random
- Important caveats:
  - Investigators must not know block size
    - Otherwise, decisions to enroll patients might be affected by knowledge of next assignment
  - Hence, often use “concealed blocks of varying sizes" (often termed a “random block design")

Alternative strategy : Urn Model

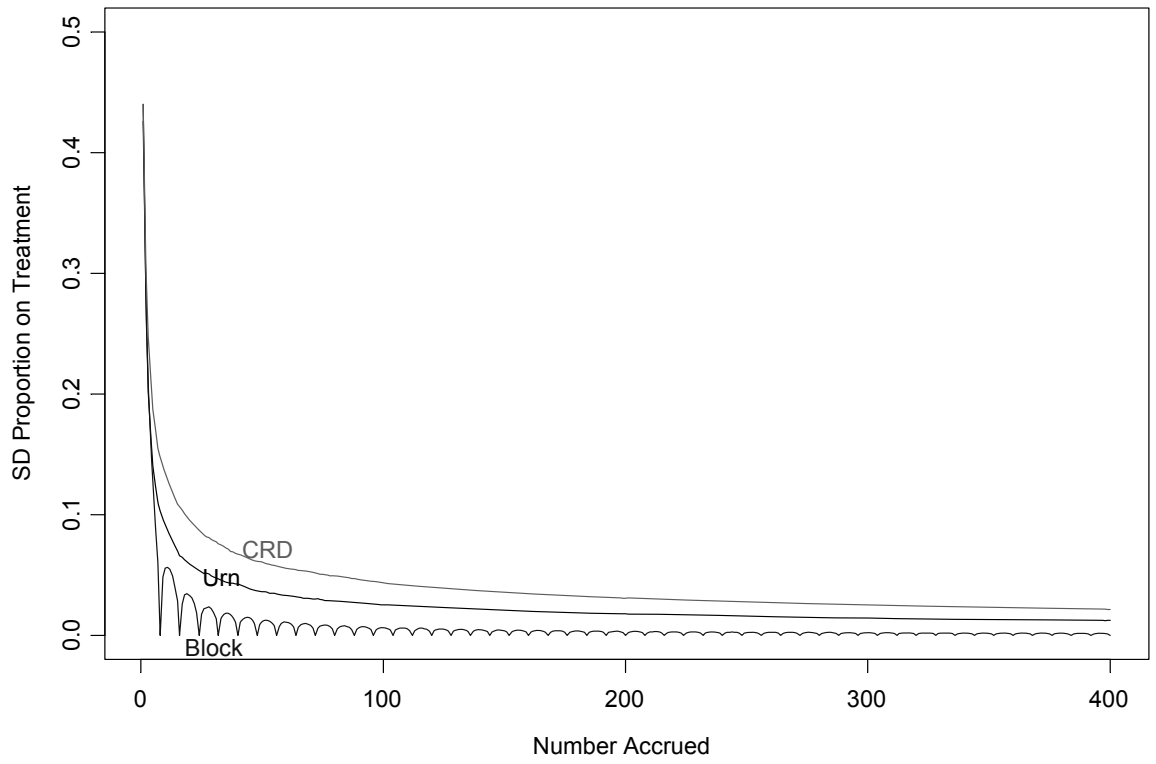
1. Begin with k white balls and $r \times k$ black balls in an urn
2. Upon accrual of a patient draw a ball from urn
   - White $\rightarrow$ control; black $\rightarrow$ treatment
   - After every white ball withdrawn, return 1 white ball and $r \times m$ black balls
   - After every r-th black ball withdrawn, return r black balls and m white balls

- Such a strategy tends to behave like small blocks early and complete randomization later, depending on k and m

Nonadaptive Randomization

Comparison of blocking strategies

- SD proportion on treatment for 3:1 randomization
  - Urn ($k = 1, m = 1$) vs Blocking (size = 8) vs CRD



Statistical inference after blocking

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity largely unchanged
    - We rarely report accrual patterns over time
  - Precision slightly improved due to achieving closer to desired randomization ratio
  - Precision could be improved if adjust for blocks as a random effect in analysis
    - This is rarely done, except in re-randomization test

Nonadaptive Randomization

Issues with complete randomization

- Imbalance across arms in covariate distribution
  - Loss of face validity
  - Conditional bias
  - Not much of an issue with large sample sizes
  - Could be problematic with sequential sampling
    - Interim analyses of data early in the study
  - Could be problematic with subgroup analyses
    - Possibility of very inefficient randomization ratio in small subgroups
- *Stratified randomization* is often used to ensure proper ratio of sample sizes across subgroups defined by important covariates

Stratified randomization

- Strata are defined based on values of important covariates
  - E.g., sex, age, disease severity, clinical site
- Within each stratum defined by a unique combination of stratification variables, CRD or blocked randomization
- Important caveats:
  - Number of strata is exponential in number of stratification variables
    - E.g., 4 two level stratification variables $\Rightarrow$ 16 strata

Nonadaptive Randomization

Statistical inference

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity improved for most important variables
  - Precision improved due to achieving closer to desired randomization ratio
  - Precision could be further improved if adjust for stratification variables in analysis
    - This should be done! (Without adjustment for strata, may even lose power for some alternatives)
    - Requires pre-specification of analysis model to avoid “intent to cheat" analysis

Additional advantages of stratified randomization

- Additional advantages of stratification
  - Balance within clinical center
    - Especially if quality control issues
  - Balance for interim analyse
  - Balance for subgroup analyses
- Also, stratified randomization does not preclude the use of blocking
  - Common to combine the two...blocking within strata

Adaptive Randomization

Issues with stratified randomization

- The need to stratify on all combinations of variables
  - Good news:
    - Balances on interactions as well as main effects
  - Bad news:
    - Effect of interactions might be quite small
    - Really only need to adjust on “counterfactual" outcome based on linear combination of all covariates
- Stratified randomizations has drawbacks in the presence of sparse data
- Because of this, some authors have described dynamic randomization processes that will allow balancing on more covariates

Adaptive Randomization

Dynamic randomization

- Subjects are assigned to the treatment arm that will achieve best balance
  - "Minimization": minimize the difference between the distribution of covariate effects between arms
    - Define a "distance" between arms for covariate vectors
    - Probability of assignment depends upon arm that would provide smallest difference
- Two arms are "distant" based on one of:
  - Randomization ratio very different from $r : 1$ in some stratum
  - Summary measure of distribution of $(W_{i1}, \ldots, W_{ip})$ differs
    - Mean, median, variance, ...
  - Distribution of $(W_{i1}, \ldots, W_{ip})$ differs
  - Contribution of covariates to the outcome differs

Conditional confounding

- Consider unadjusted and adjusted (linear) models for an outcome Y:

  1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
  2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \vec{W}_i^T \vec{\delta}$

  or in matrix notation

  1. Unadjusted model: $\mathrm{E}[\vec{Y}] = \mathbf{X}\vec{\beta}$
  2. Adjusted model: $\mathrm{E}[\vec{Y}] = \mathbf{X}\vec{\gamma} + \mathbf{W}\vec{\delta}$

Adaptive Randomization

Conditional confounding

- Then it can be shown that

$$\mathrm{E}[\widehat{\vec{\beta}}] = \vec{\gamma} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\vec{\delta}$$

- This implies that

$$\beta_1 = \gamma_1 + \sum_{j=1}^{p} \left(\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot}\right) \delta_j$$

with

$$\bar{W}_{kj\cdot} = \frac{1}{n_k} \sum_{i=1}^{n} W_{ij} 1_{[X_i=k]}$$

Adaptive Randomization

This provides reasonable ways to define distance metrics

- Based on contribution to confounding :

$$d(\vec{X}, \mathbf{W}) = \left| \sum_{j=1}^{p} \left(\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot} \right) \delta_j \right|$$

- Weighted distance between standardized means :

$$d(\vec{X}, \mathbf{W}) = \sum_{j=1}^{p} c_j \left| \frac{\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot}}{SD(W_j} \right|^{\lambda}$$

- Weighted imbalance in n across strata $\Omega_1, \ldots, \Omega_s$:

$$d(\vec{X}, \mathbf{W}) = \sum_{s=1}^{S} c_s \left| \sum_{i=1}^{n} 1_{[X_i=1]} 1_{[\vec{W}_i \in \Omega_s]} - \sum_{i=1}^{n} 1_{[X_i=0]} 1_{[\vec{W}_i \in \Omega_s]} \right|^{\lambda}$$

Adaptive Randomization

Conditional confounding

- Spurious associations will be minimized if means of important predictors are balanced across treatment arms
  - The greater the value of δ_j the more important it is for the means of the j-th covariate to be equal
    - (Presumes linear model reasonable approximation)
  - We could use estimates of the of δ_j's to define the distance between the arms (or just balance means)
- Balancing group sizes across covariates will tend to have means balanced by randomization
  - Group sizes within strata may matter for subgroup analyses

Adaptive Randomization

Dynamic randomization

- Subjects are assigned to the treatment arm that will achieve best balance
  - When i-th patient accrued, compute a randomization probability, π_i, where

$$\Delta_i = d(\vec{X}, \mathbf{W}|X_i = 1) - d(\vec{X}, \mathbf{W}|X_i = 0)$$

and

$$\pi_i = \Pr[X_i = 1] = f(\Delta_i),$$

with

  - $0 \leq \pi_i \leq 1$
  - $f(\Delta_i)$ monotonically decreasing in π_i
  - (generally seek to avoid $\pi_i = 0$ and $\pi_i = 1$)

Adaptive Randomization

Inference : Population model

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity improved for most important variables
  - Precision improved due to achieving closer to desired randomization ratio
  - Precision could be further improved if adjust for stratification variables in analysis for population model
    - This should be done
    - Requires pre-specification of analysis model to avoid “intent to cheat" analysis

Adaptive Randomization

Advantages and disadvantages

- Advantages:
  - Typically improved face validity
  - Can handle an arbitrary number of covariates
    - Depending on distance metric
- Disadvantages:
  - Logistically more involved
  - Decreased credibility if too deterministic
    - Approaches sequential allocation
  - Some analytic strategies more complex (permutation tests for strong null)
  - Does not necessarily facilitate subgroup analyses
    - Unless distance metric chosen carefully

Adaptive Randomization

Ethics

- Clinical trials are experiments in human volunteers
  - Individual ethics:
    - Patients on trial: Avoid continued administration of inferior treatment
    - Patients not yet on trial: Avoid starting inferior treatment
  - Group ethics:
    - Facilitate rapid adoption of new beneficial treatments
    - Avoid prolonging study of ineffective treatments
- Some authors have described dynamic randomization processes that attempt to minimize exposure of patients to harmful treatments

Adaptive Randomization

Proposed solutions

- Most commonly used
  - Sequential sampling
    - Interim analyses of data
    - Terminate trials when credible decisions can be made
- Also proposed
  - Response adaptive randomization
    - Change randomization probabilities as evidence accumulates that one treatment might be best
    - “Play the winner"

Adaptive Randomization

Play the winner : Urn model

1. Begin with k white balls and k black balls in an urn
2. Upon accrual of a patient draw a ball from urn
   - White $\rightarrow$ control; black $\rightarrow$ treatment
3. Observe outcome
   - If outcome is good, return $m + 1$ balls of same color as withdrawn
   - If outcome is bad, return 1 ball of same color as withdrawn and m balls of opposite color

Adaptive Randomization

Bayesian methods

- An explicit Bayesian approach to dynamic randomization bases the randomization ratio on the current posterior probability that one treatment is superior
  - Ultimately, that posterior probability is based on the number of good outcomes on each treatment (in conjunction with a probability model for the response and a prior distribution)
- Advantage of using Bayesian posterior probability
  - Can easily handle continuous outcomes
  - Can easily handle continuous randomization probabilities

Adaptive Randomization

Analytic issues

- Treatment of successive patients is not independent of previous patients treatment and results
  - Possible bias in accrual of future patients
- Conditionally biased estimates of treatment effect in arm with lower sample sizes
  - Bad early results tend to preclude regression to mean

Response-Adaptive Randomization (Example)

ECMO: Extracorporeal Membrane Oxygenation in neo-natal respiratory failure

- Persistent pulmonary hypertension results in right to left shunt through the foramen ovale or ductus arteriosus causing hypoxemia. ECMO is used to maintain life until the condition resolves.
- Trial 1 (Play the winner absolutely): *Pediatrics* (1985) 76:479-487
  - First subject was randomized to conventional medical therapy (CMT); the infant died.
  - Second subject given ECMO; infant lived.
  - Next 8 subjects given ECMO; all lived.
  - Result:

    100% mortality with CMT
    0% with ECMO
    RR = 0.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

- Trial 2 (Play the winner with higher probability): *Pediatrics* (1989) 84(6):957-63
  - Randomize until the 4th CMT death, then treat remainder with best approach.
  - 19 babies in first phase (4/10 die with CMT; 0/9 die with ECMO).
  - 20 babies on ECMO in second phase (1 death).
  - Result:
    40% (4/10) mortality with CMT;
    3% (1/29) with ECMO;
    RR = 0.086.

- Trial 3 (conventional RCT): *Pediatrics* (1998) 101(4):E1
  - Randomize 185 infants (92 to CMT, 93 to ECMO)
  - Result:
    59% (54/92) mortality with CMT;
    32% (30/93) with ECMO;
    RR = 0.55.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

Implications of the ECMO example:

- ECMO looked better with response-adaptive randomization.
- Response-adaptive designs were not accepted as adequate justification for ECMO.
- Inadequate study designs can delay introduction of beneficial treatments or prolong use of inferior treatments.

"In fact, in the ECMO trial, the patient who failed on treatment B had the most extreme values on no fewer than four important covariates (Paneth & Wallenstein, 1985), and was clearly the sickest. In effect, the trial provides no information whatsoever regarding the treatment comparison. "

-Begg (1990)

Response-Adaptive Randomization (Example)

- The ECMO experience has tempered enthusiasm for randomized PTW
- This being said, there may be times were response-adaptive randomization will work, but
  - There needs to be a clear dilemma re individual ethics
  - There will tend to be decreased group ethics
  - It takes a lot of planning in order to obtain results that will be sufficiently credible

Methods: Logistics of Randomization

- Where to perform randomization:
  - Central randomization:
    - Phone calls to the coordinating center.
    - Sequences can be determined at the start of the study (except with adaptive randomization).
  - Distributed randomization: Computer programs, envelopes, or lists at pharmacies.
- Important principles:
  - Strong quality assurance must be in place to ensure proper randomization.
  - Ensure adequate concealment/blinding.
  - Provide for emergency unblinding.
  - Exact randomization scheme must be known for analysis.