Matching methods for estimating causal effects using multiple control groups

Elizabeth A. Stuart

Mathematica Policy Research, Inc.

Donald B. Rubin

Harvard University

Outline

1. Overview: Matching with multiple control groups
   - Why? How?
2. Motivating example: SDDAP
   - Existing methods don’t help
3. Theoretical framework and results
4. Adjusting for difference between groups
5. Extensions: “pseudo-units,” matching on a binary covariate
6. Conclusions, Other topics for discussion

Multiple control groups

- Sometimes template of one treated and one control group not sufficient: Look at use of multiple control groups
  - Randomized clinical trial of drug that becomes commercially available so randomized controls no longer “controls” (e.g., Genzyme example)
  - Initial control group has small sample size (e.g., too expensive to enroll many control patients so want to use historical data to supplement original control group)
  - SDDAP
- Previous use of multiple control groups
  - Dempster (1983), Rosenbaum (1987)

School Dropout Demonstration Assistance Program (SDDAP)

- **Two approaches to reduce school dropout rate:**
  1. **Targeted: Intervene with individual “at-risk” students**
  2. **Restructuring: Change fundamental nature of entire schools**
- **Random assignment used for targeted programs**
- **Comparison design used for restructuring programs: Comparison schools chosen in same school districts**
- **Within each treated/comparison pair of schools, students often look significantly different on individual-level covariates**

Grand Rapids: Covariate Means

	Treated	Comparison	p-value of diff
Age	15.93	15.93	0.94
% Black	53	31	0.00
Mother's education	6.61	6.31	0.05
Father's education	6.59	6.39	0.25
% Ever skip school	52	73	0.00
School climate	-0.03	-0.05	0.67
Self-esteem score	0.32	0.20	0.01
Baseline reading score	50.09	52.88	0.02
Baseline math score	49.42	53.84	0.00
% Non-2 parent household	37	35	0.47
% Household receives TANF	9	6	0.14
% Average grade below C	17	23	0.02
% Discipline problems	40	41	0.86
% Is a parent	3	3	0.97
Sample size	434	428	

Standard approaches

- **Modeling**
  - **Regression models of outcome with indicator for treatment assignment:** $E(Y|T,X) = \alpha T + \beta X$
  - **Can lead to biased answers if not much overlap: Relies on extrapolation and (untestable) model assumptions**
    - **Cochran and Rubin 1973**
- **Matching**
  - **Would like to get one matched control for each student in a restructuring school**
  - **Hard to find enough good matches from the local comparison school**
  - **May be that no single school would provide sufficient matches**

“Pseudo-schools”

- Solution: Form comparison “pseudo-school” made up of students from multiple comparison schools
  - Irie (2001); Abadie and Gardeazabal (2003); MDRC (2003); Rubin, Stuart, and Zanutto (2004)
- Combination of “good” local matches and non-local matches
  - Local controls: Exact matches on area level variables (observed and unobserved)
  - Non-local controls: Closer matches on individual level covariates
- The best of both worlds?

Theoretical basis

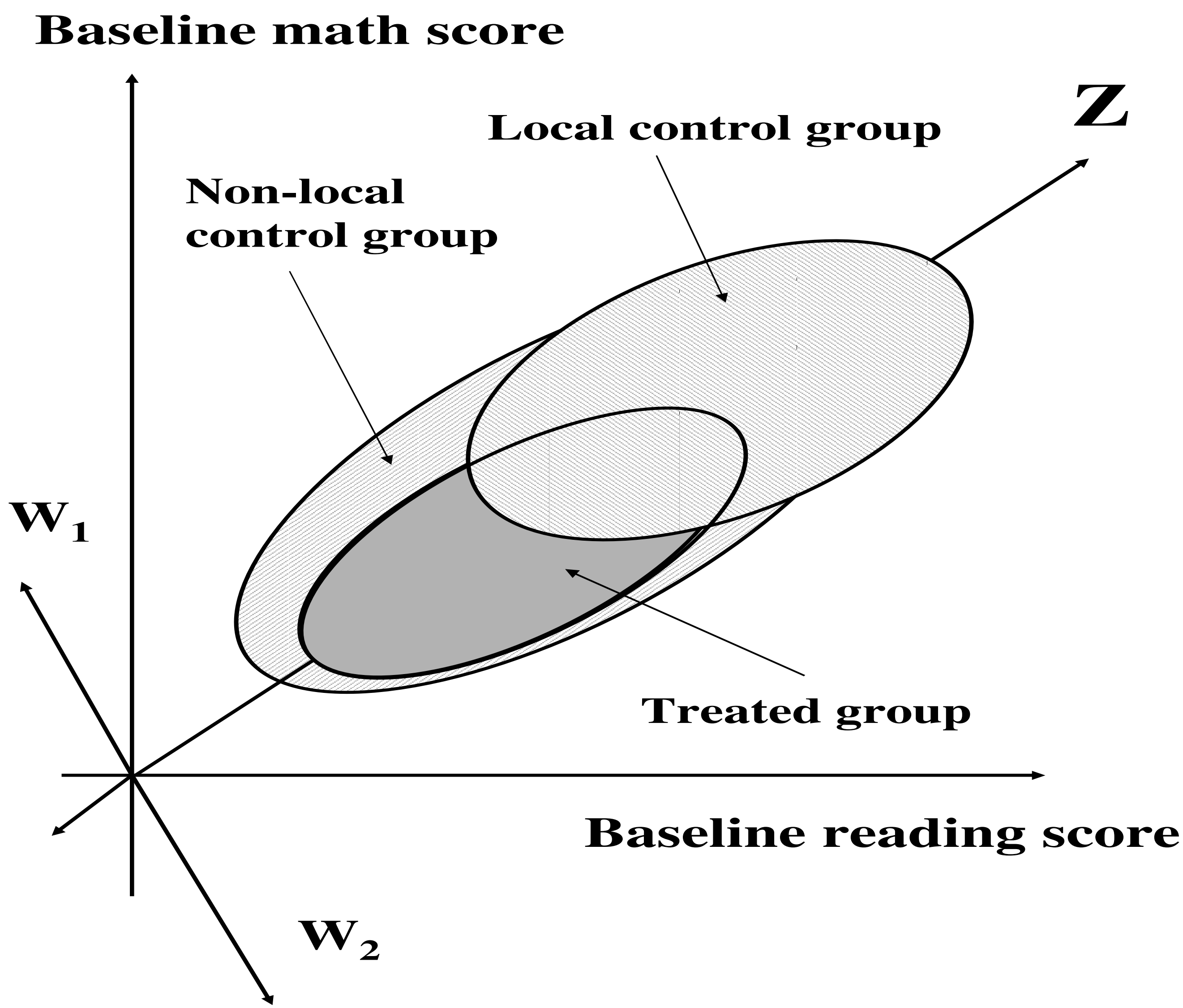
- How do we know the matching isn't increasing bias?
- Context of one treated group, multiple control groups
- Rubin and Thomas (1992a,b; 1996) provide assurance in situation with one treated and one control group
- Extend those results to affinely invariant matching methods with mixtures of ellipsoidally symmetric distributions
  - e.g., Propensity score or discriminant matching with covariates normally distributed within each group
- Consider each of the control groups to be a mixture component

Effect of matching

- Many results based on symmetry
- Consider effect on Y, an arbitrary linear combination of the covariates X (e.g., outcome of interest)
- Best linear discriminant (Z): linear combination of X with largest difference in population means
- Y can be written as sum of components along (Z) and orthogonal to (W) standardized discriminant:

  $$Y = \rho Z + \sqrt{1-\rho^2}W$$

- Compare means (variances, covariances) in matched versus random subsamples

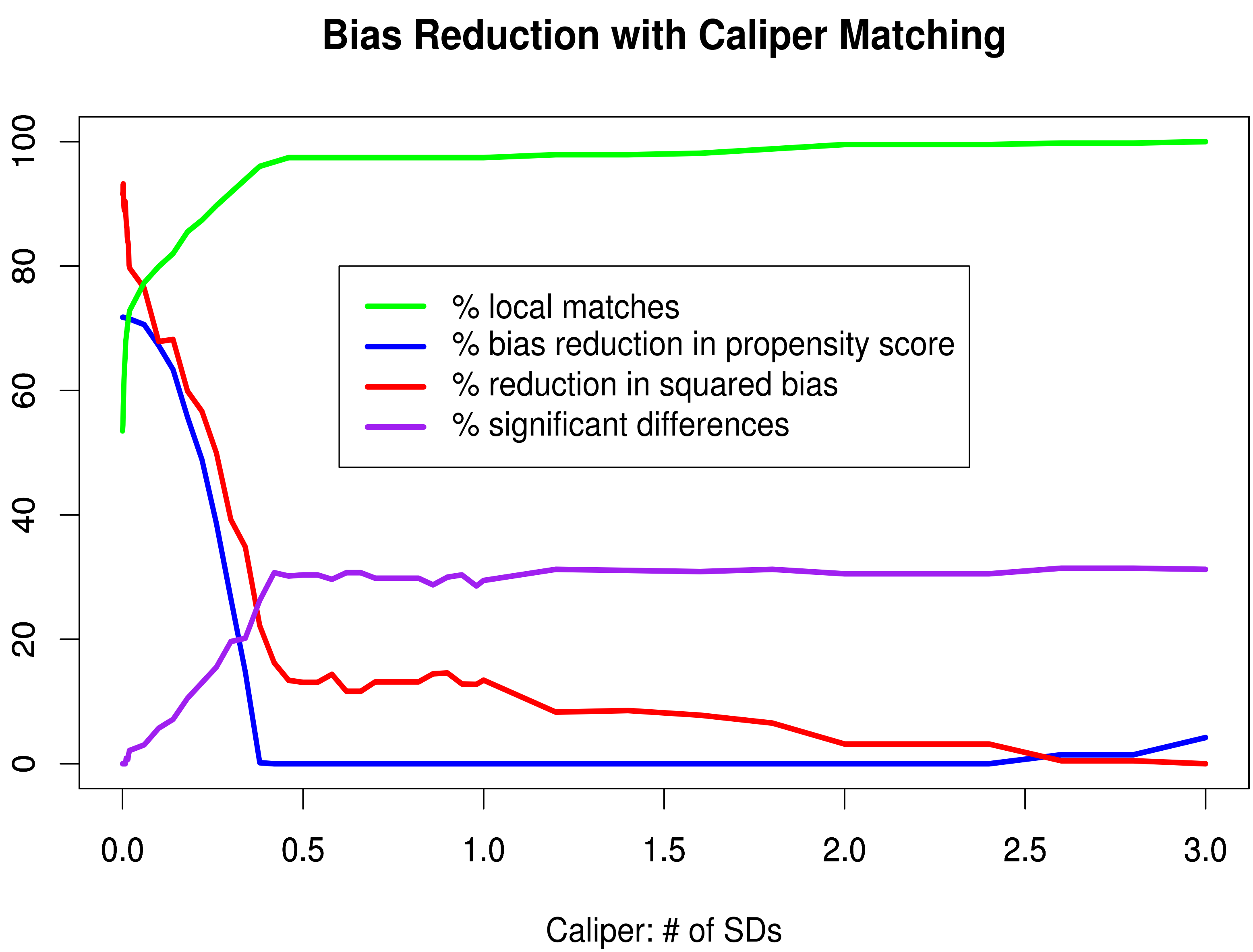

Baseline math score
Z
Local control group
Non-local
control group
W1
Treated group
Baseline reading score
W2

Summary of theoretical results

1. **Method is EPBR (Equal Percent Bias Reducing)**
   - **Reduces bias in every direction by the same amount**
   - **Can't increase bias in the outcome!**
2. **Theoretical results generate analytic approximations regarding the effects of matching**
   - **Before matching, can estimate maximum possible bias reduction**
   - **Variance benefits if match on estimated rather than true discriminant**
3. **Most Rubin and Thomas (*Annals of Statistics*, *Biometrika*) results generalize well**

How can we do this in practice? “Extended” matching

- **Algorithm to select matches in the SDDAP example:**
  1. **Set caliper, e.g., 0.25 propensity score standard deviations**
  2. **For each treated student, choose the closest local match if there are local control students within the caliper. If no local control student within the caliper, take the closest non-local match.**
- **Explicitly chooses non-local matches only for treated students who have no good local matches**
  - **Use non-local matches only as much as necessary**
- **Related to “partial” matching of any binary covariate**


Bias Reduction with Caliper Matching
% local matches
% bias reduction in propensity score
% reduction in squared bias
% significant differences
0
20
40
60
80
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Caliper: # of SDs

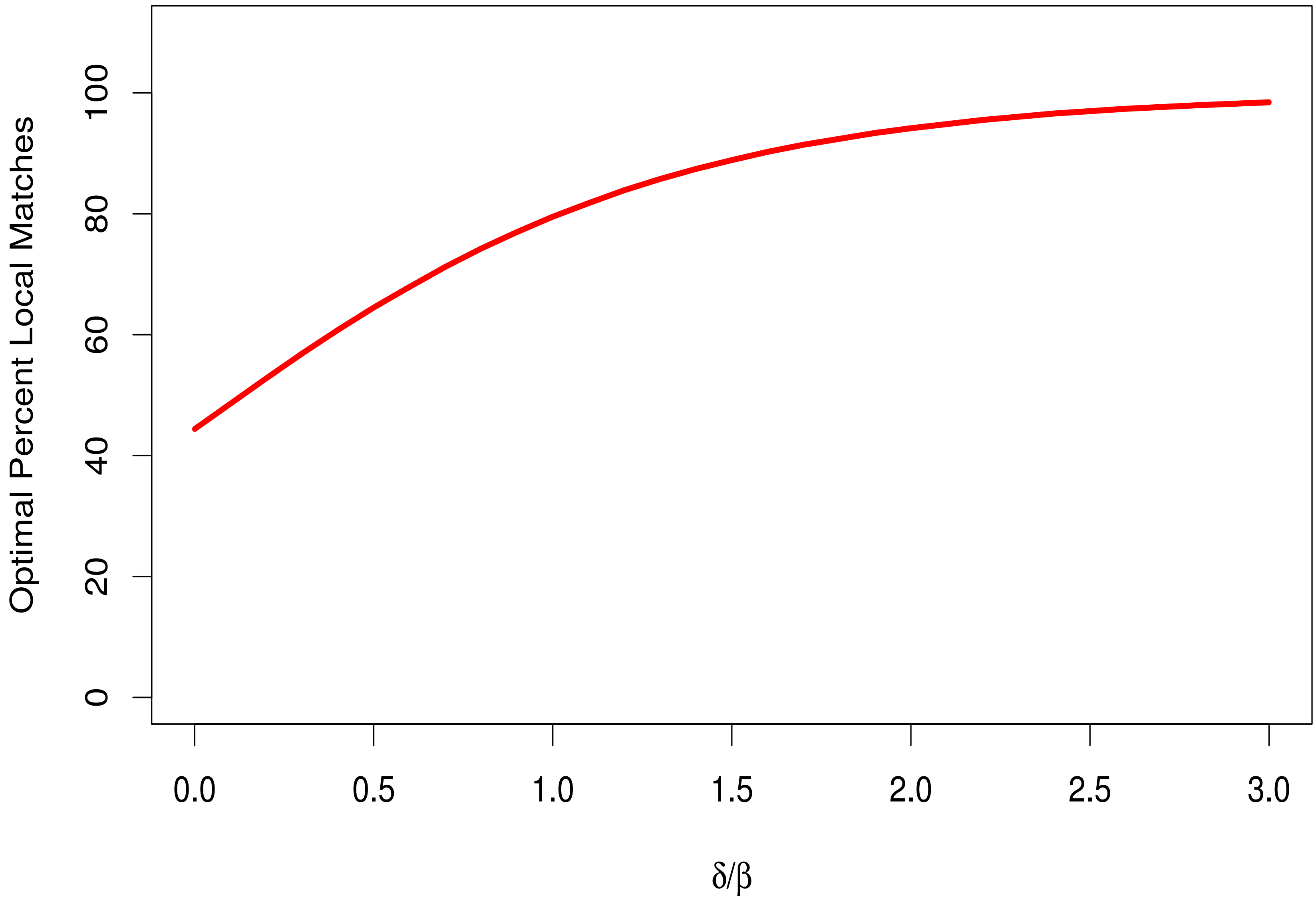
Calculating optimal % local matches

- **Getting non-local matches reduces bias in X, but what about bias from including matches from a different area?**
- **Assume normality, $E(Y|X,D) = \beta X + \delta D$**

  $D = 0$ **in local area (treated and local control group)**

  $D = 1$ **for non-local control group**
- **Can estimate optimal number of local matches (m)**
- **Solution depends on how close the local groups are and on relative importance of individual level covariates (X) and area indicator (D) in terms of outcome**
- **Large $\delta \rightarrow$ more local matches, Small $\delta \rightarrow$ fewer local matches**

SDDAP: Optimal Percent Local Matches



Adjusting for control group differences

- **Would like to adjust for unobserved difference between local and non-local groups in outcome analysis**
- **Compare well-matched subsets of the two control groups**
  - **Model $Y(0) \sim N(\beta X + \delta D, \sigma^2)$ to obtain plausible values of δ, δ_***
- **Create (multiple) data sets with matched controls**
  - **Make non-local matches look like they could have been local**
  - **Use $Y(0)$ for local matches**
  - **Use $Y(0) - \delta_*$ for non-local matches**
  - **Do this multiple times, use MI combining rules**

Matching adjustment vs. OLS

- **Simulations run to assess performance of OLS and matching adjustment procedure**
  - **True model:** $E(Y|X,D) = e^{aX} + (\delta_0 + \delta_1 X)D$
  - **Three groups, extended matching with caliper = 0.1 sd's**
- **Quantity of interest: Average treatment effect (ate)**
- **Estimated treatment effects using each method ($\widehat{ate}_m$):**
  - **OLS: Coefficient on treatment indicator in regression using local groups only**
  - **Matching: Difference in means of outcome in treated and matched control groups**

Simulation Details

- **Varying: δ_1, a, covariate means, covariate variances, sample sizes**

- **Examine bias: Residual variance assumed to be 0**
  - **Rubin 1979, Rubin and Thomas 2000**
  - **Compare % reduction in integrated squared bias (ISB)**
  - **For method m, $\mathrm{ISB_m} = (\widehat{ate}_m - ate)^2$**
  - **Without loss of generality, assume $ate = 0$**

Results

- $\delta_1 = 0$ **(no-interaction): Matching adjustment gives unbiased estimate of treatment effect**

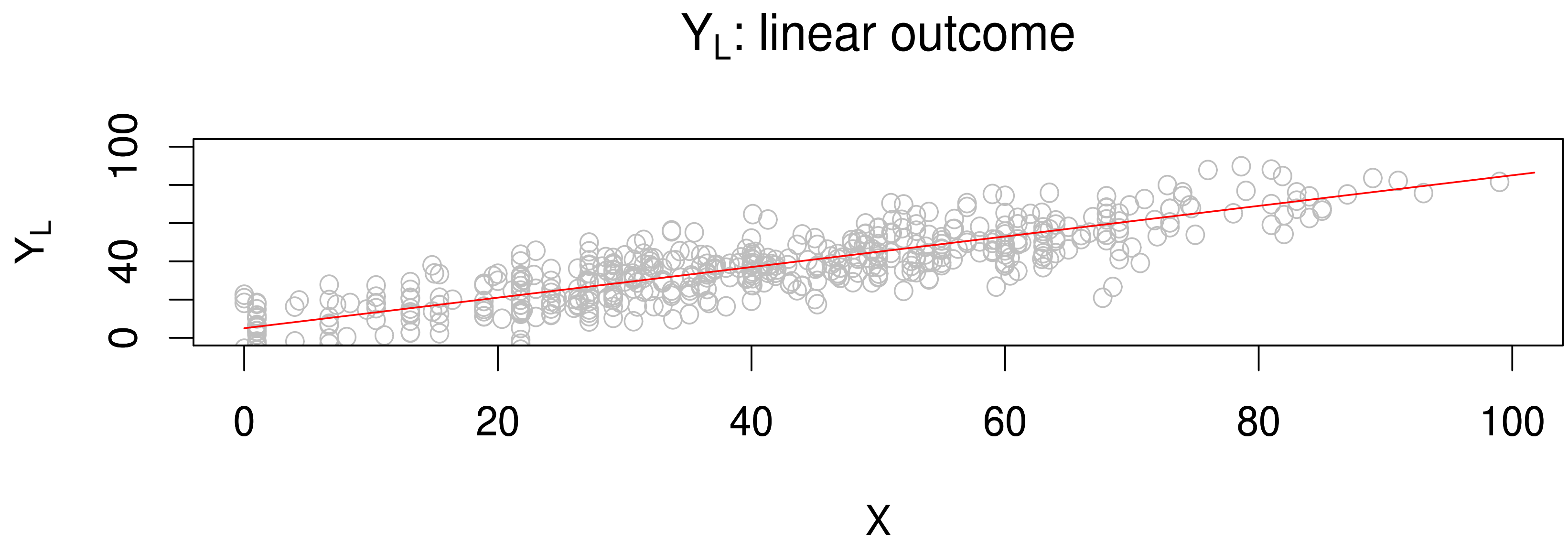
  δ_0 **perfectly estimated in group of well-matched controls**

- $a = 0$ **(linear response): OLS gives unbiased estimate of treatment effect**

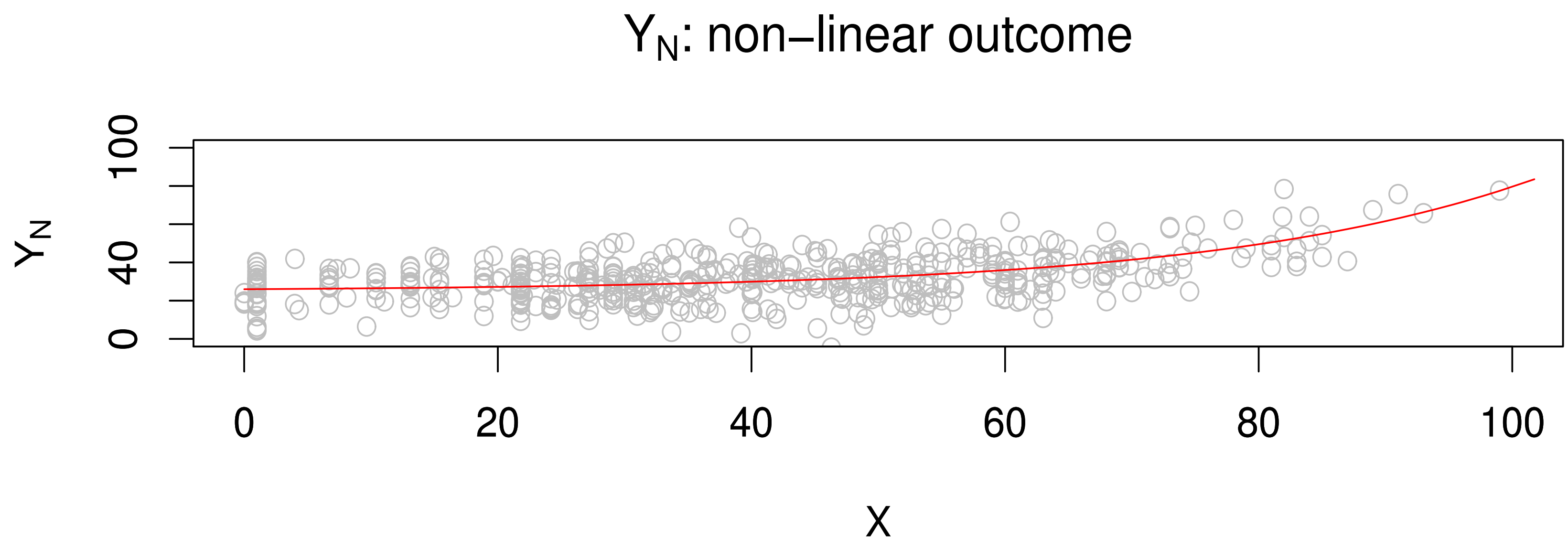
- **Situations in between:**
  - **OLS appears to be more sensitive to assumption of linearity than matching is to no-interaction assumption**
  - **OLS works slightly better than matching when** $a = 0, \pm 0.5$**, but much worse when** $a = \pm 1$

SDDAP simulation

- **Assess methods in SDDAP situation**
  - **Actual covariate values, sample sizes**
- $X =$ **baseline reading test score**
- $Y =$ **simulated outcome reading test score**
- **All variables on 0-100 scale**
- **Two models examined:**
  1. $E(Y_L|X,D) = 5 + 0.8X + (1 + \delta_1 X)D$
  2. $E(Y_N|X,D) = 25 + e^{0.04X} + (1 + \delta_1 X)D$



Y_L: linear outcome
Y_L
0
40
100
0
20
40
60
80
100
X



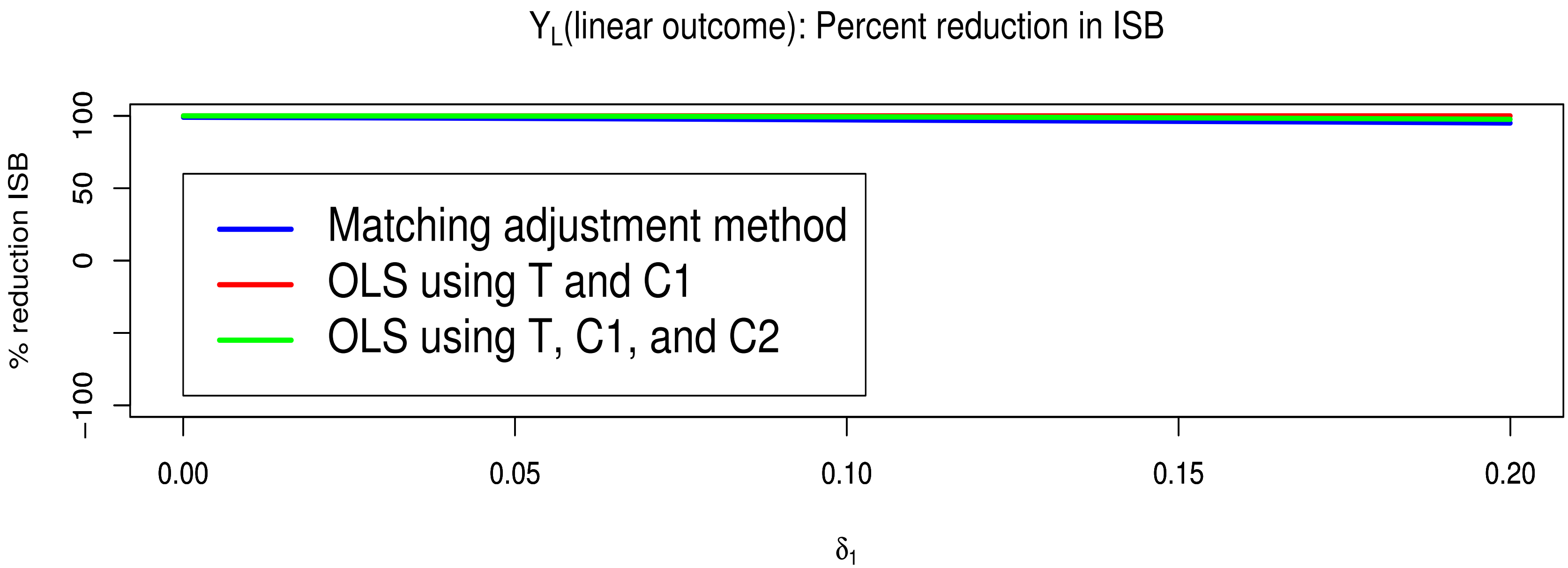
Y_N: non−linear outcome
Y_N
0
40
100
0
20
40
60
80
100
X

Simulation results

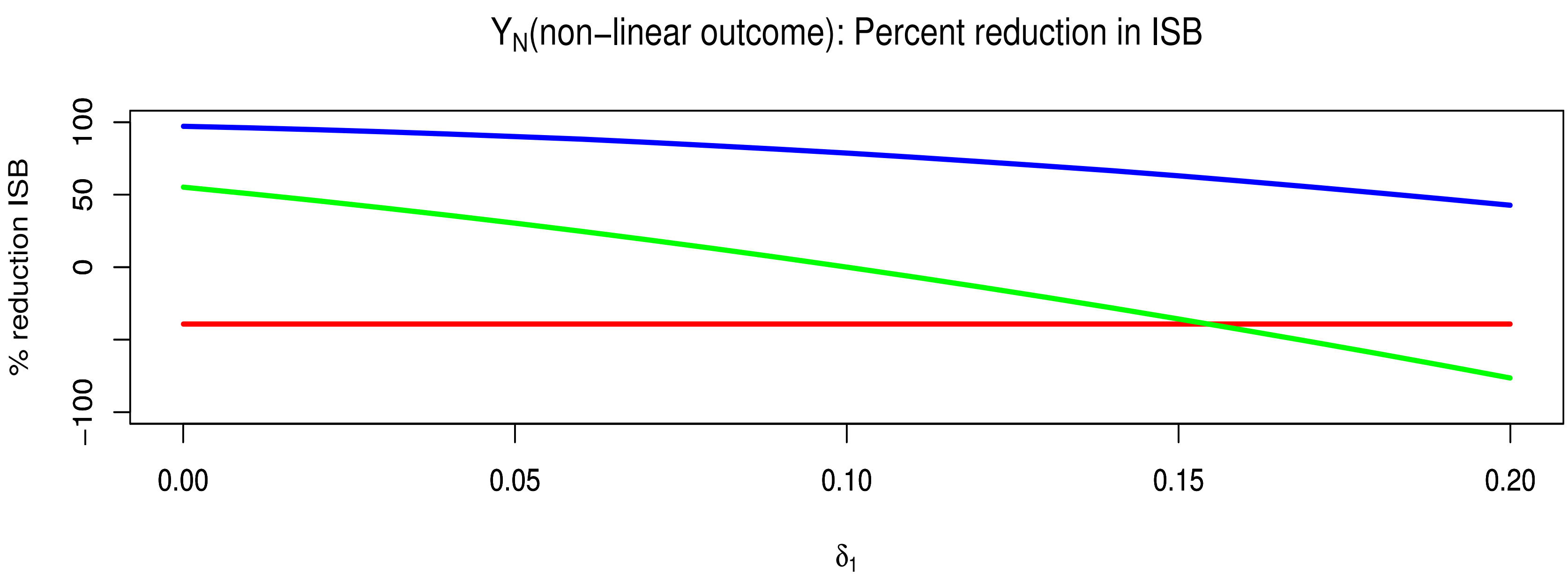
- **Linear outcome: OLS performs better, but matching also performs well**
- **Non-linear outcome: Matching adjustment performs much better when outcome even moderately non-linear**
  - **OLS increases ISB!**

% ISB Reduction

	$Y_L(X)$	$Y_N(X)$
OLS	**100**	**-39**
Matching	**95-99**	**43-97**



Y_L(linear outcome): Percent reduction in ISB
% reduction ISB
100
50
0
−100
Matching adjustment method
OLS using T and C1
OLS using T, C1, and C2
0.00
0.05
0.10
0.15
0.20
δ_1



Y_N(non−linear outcome): Percent reduction in ISB
% reduction ISB
100
50
0
−100
0.00
0.05
0.10
0.15
0.20
δ_1

Extensions and Future Work

- Related to idea of forming "pseudo-units"
- Matching on an "important" binary covariate
- Use in design: cost trade-offs
- Extend tests for sensitivity to an unobserved covariate to this setting
  - e.g., Rosenbaum and Rubin 1983
- Examine matching methods further
  - e.g., variations on caliper/extended matching?

Conclusions

- Extension of Rubin and Thomas results apply to this situation and help explain why those results held approximately even when assumptions not satisfied (Rubin and Thomas 1992b, 1996)
- Explicit discussion of the use of multiple control groups
- Matching adjustment procedure effective method for estimating causal effects
  - Performs much better than OLS when outcome variable is even moderately non-linear
- Provides a theoretical framework and practical guidance for matching with multiple control groups

Other topics...

- Assessing multivariate balance
- Treatment at one level, data at lower level (counties within states)
  - Do these "pseudo-unit" ideas work?
- Matching with weighted data
- Propensity scores with missing data
- Combining observational studies and randomized experiments
  - Get benefits of both?