Lecture 25

1. Comparing treatments with an observational study
2. Comparing baseline values between treatment groups
3. Example from LaLonde (1986)
4. Propensity scores
5. Matching
6. Treatment effect estimated from sample matched on propensity scores

Comparing treatments with an observational study

Comparison of treatments aims to compare *effects* of treatments on something.

Experiment: researcher *assigns* treatment to subject

Researcher makes a change and observes the effect. If subjects were alike except for treatment (by randomization), difference in effect was caused by treatments.

Observational study: subjects choose their own treatment

Subjects may be different, and difference relates to both choice of treatment and outcome.

Subject differences may cause part or all of treatment difference.

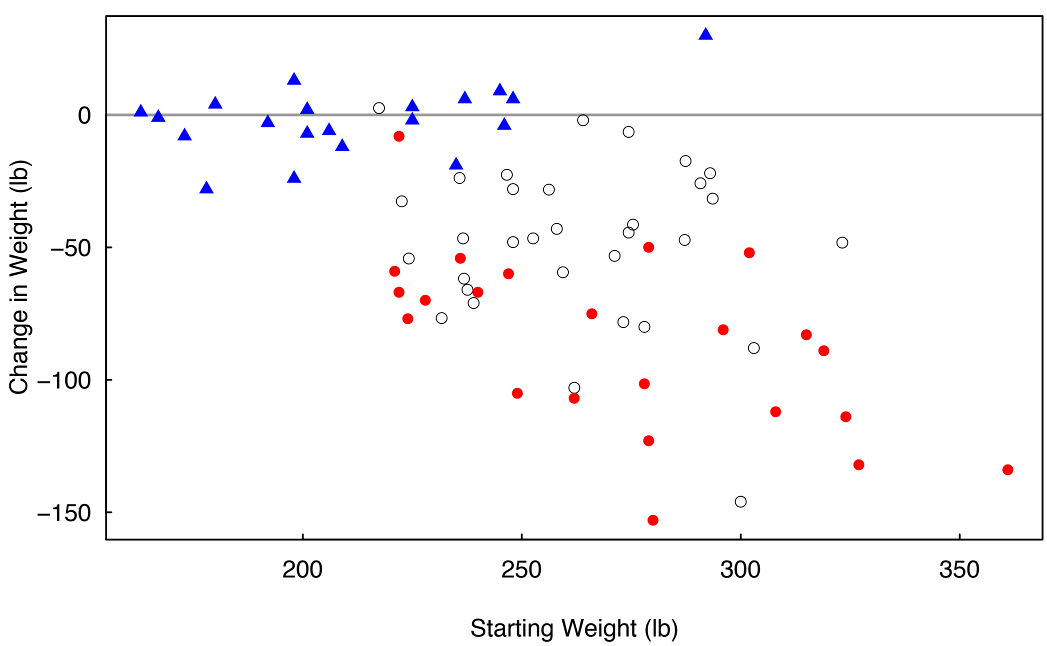
Challenge for observational studies:

show subjects in treatment groups alike—as though randomized

Observational data (1992): retrospectively collected weight losses in very overweight diabetic patients who received one of three treatments:

- • gastric-bypass surgery
- ○ very-low-calorie liquid diet
- ▲ standard medical care

Treatments: standard medical care (▲), very-low-calorie liquid diet (○), or gastric-bypass surgery (•).



Must adjust comparison for baseline weight.

1. Use regression (weight change on baseline weight) to adjust for baseline weight.

2. Stratify on baseline weight, use strata that contain large enough samples from at least 2 treatment groups.

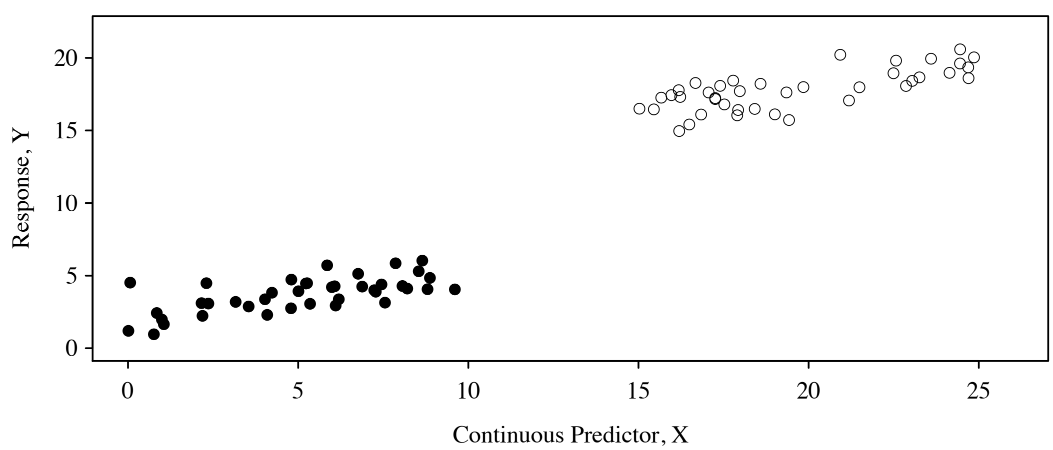
This can fix differences in baseline weight.

What if there were 20 other baseline variables that were also different?

Example with no overlap:

Observational study to compare treatments A and B.

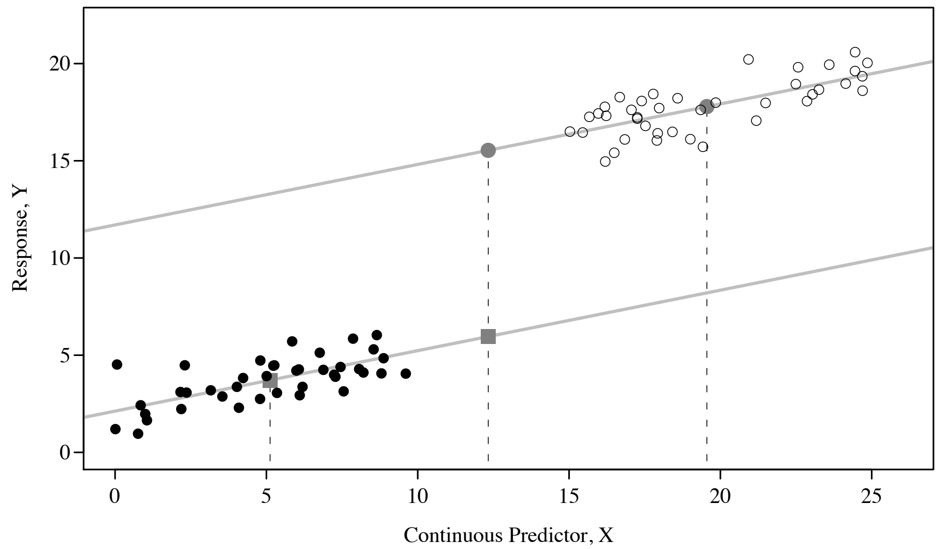
We have baseline characteristic X on each participant.



How about strata here?

Regression adjustment depends on assuming model is correct in a region *where we have no data.*

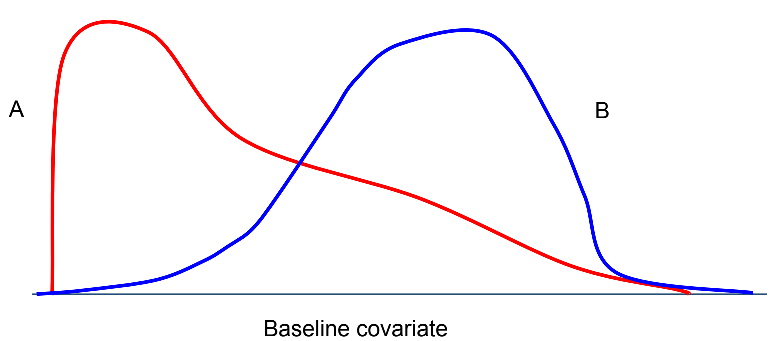
Unadjusted difference $\bar{x}_A - \bar{x}_B = 14$; controlling X, difference of LSmeans is 9.6.



Differences in baseline characteristics

Baseline characteristics may differ between observational treatment groups

- **partial overlap** of range, or no overlap (eg examples 1 and 2)
- **imbalance**: similar range, but different distributions



Constructed observational study

National Supported Work (NSW) was an experiment that randomly assigned adults to job training program or no program, in 1976–1977.

Demographic characteristics, earnings in 1974 and 1975.
Outcome: earnings in 1978.

- **Intervention**: data from participants assigned to job training program
- **Control**: data from Current Population Survey, Panel Study in Income Dynamics

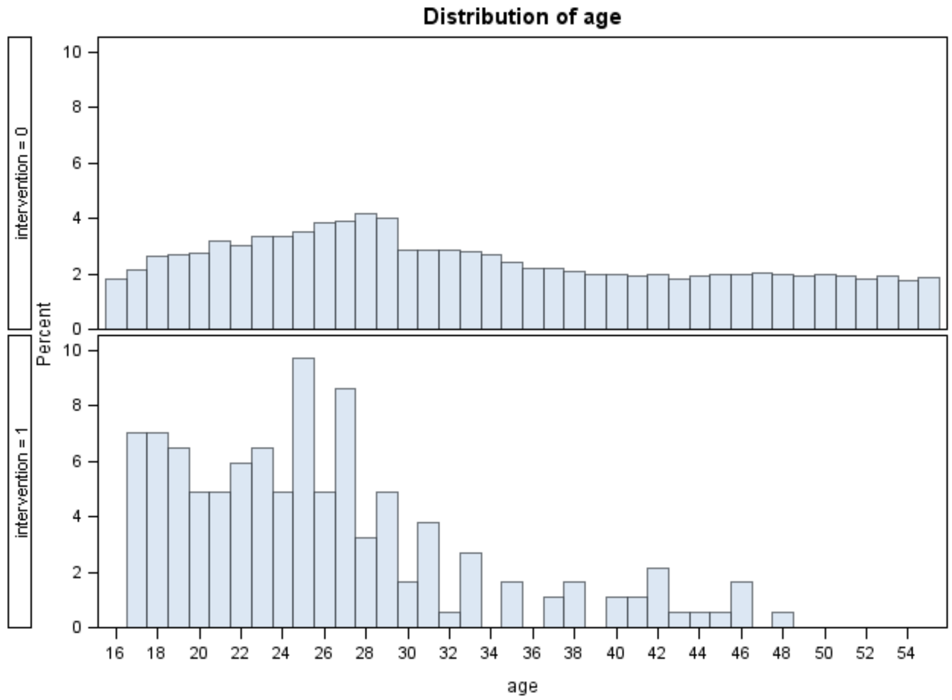
LaLonde RJ, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review,* 1986; 76(4) 604–620.

Comparing baseline characteristics (mean ± SD):

	Intervention $n = 185$	**Control** $n = 18482$	p-value	Standardized difference
Age (years)	25.8 ± 7	33.4 ± 11	<.0001	−0.7
Education (years)	10.3 ± 2	12.0 ± 3	<.0001	−0.6
Baseline earnings ($)	1,814 ± 270	14,563 ± 73	<.0001	−1.3
Black	84%	10%	<.0001	
Hispanic	6%	7%	.694	
Married	19%	73%	<.0001	
No HS degree	70%	30%	<.0001	

Standardized difference in means (standardized bias) = $\left(\bar{x}_A - \bar{x}_B\right)/\text{pooled SD}$

Age at baseline: imbalance, partial overlap



Proc Univariate: Use CLASS statement to get panel of histograms:

```
ODS graphics on;
Proc Univariate data= NSW ;
  var age;
  class intervention;
  histogram age / nrows=2 ;
  run;
ODS graphics off;
```

Propensity Score

Adapt stratification idea: restrict to intevention/control pairs matched on

baseline earnings, age, education, Black/Hispanic status, marriage, HS degree.

Instead of very difficult matching of pairs on all these variables,

compute each subject's **propensity score** =

chance subject assigned to control, according to baseline characteristics.
(logistic regression)

Form matched pairs based on propensity score.

Calculate propensity scores in logistic regression:

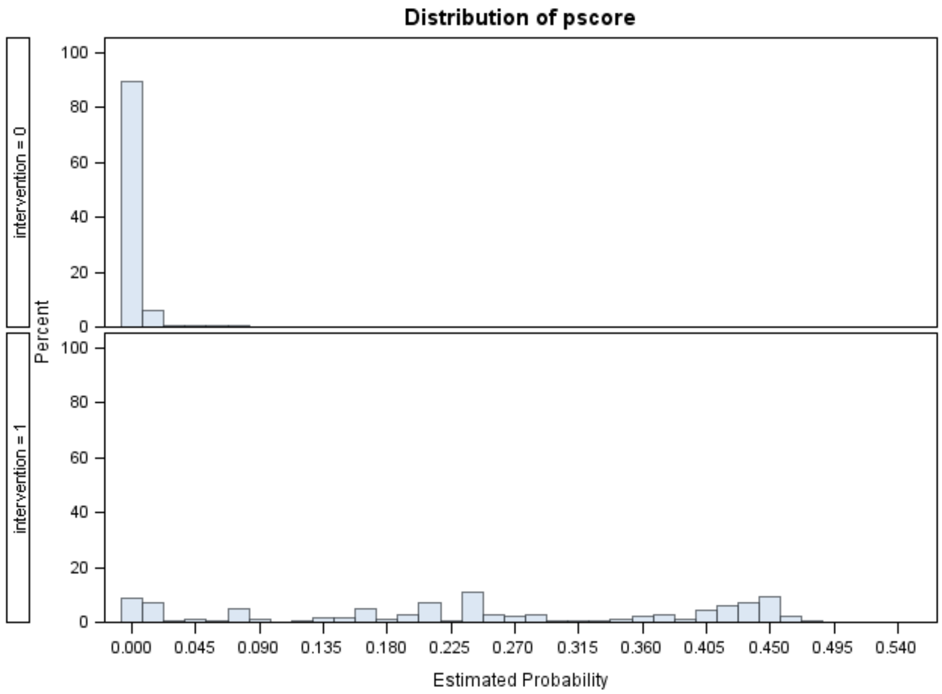
```
Proc Logistic  descending  data=NSW;
   model intervention = educ black hisp  married
        nodegree|age|earnings1974|earnings1975 @2 ;
```

@2 gives all main effects + 2-factor interactions

```
   output out=NSW_P   pred = pscore ;
```

pscore = estimated chance of being in intervention group = *propensity score*

Severe imbalance:



Forming matched pairs

Use macro **%PSmatching** adapted from M. Coca-Perraillon (1987).

Treatment and Control observations must be in separate datasets such that:

- Control data includes: idC = subject_id, pscoreC = propensity score

  Treatment data includes: idT, pscoreT

- *method of matching:* NN (nearest neighbor), caliper, or radius

```
data T C;   make 2 datasets
  set NSW_P ;
  if intervention=0 then do;
    idC = subject; pscoreC = pscore;  output C  ; end;
  if intervention=1 then do;
    idT = subject; pscoreT = pscore;  output T  ;  end;

 %include "PSmatching.sas";   path to macro in separate file

 %PSMatching (datatreatment= T, datacontrol= C, method= NN,
 numberofcontrols= 1, caliper=, replacement= no, out=PS_match_NN);

 Proc Print data=PS_match_NN(obs=10);
```

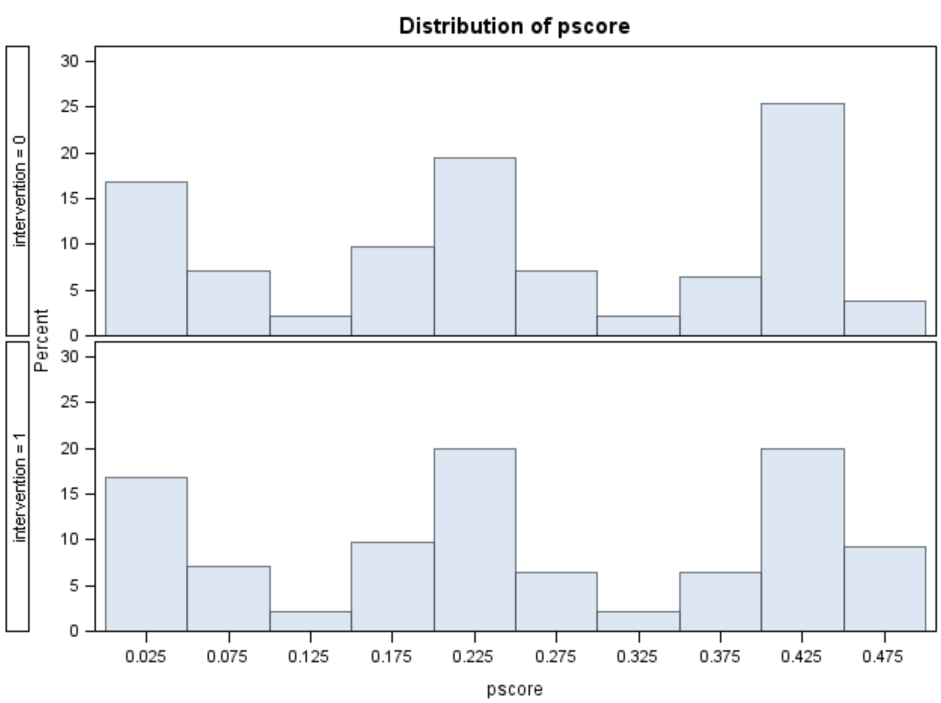

Obs	Id Selected Control	PScore Control	Matched To TreatID	PScore Treat
1	18659	0.45266	16143	0.45266
2	15025	0.27739	16158	0.27764
3	15921	0.46539	16112	0.47071
4	15475	0.24408	16024	0.24408
5	15078	0.01349	16043	0.01350
6	15769	0.28393	16152	0.28338
7	11515	0.01247	16087	0.01247
8	18641	0.23639	16099	0.23639
9	18663	0.44155	16140	0.43946
10	15734	0.27471	16170	0.27466

All other variables gone; need to separate observations for merging.

```
data pairs_NN;
  set  PS_match_NN;
  subject = IdSelectedControl; pscore = PScoreControl;
  pair = _N_; intervention=0;
  output;

  subject = MatchedToTreatID; pscore = PScoreTreat;
  pair = _N_; intervention=1 ;
  output;
  keep subject pscore pair intervention;
```


Make histogram of propensity scores by treatment group to check balance:



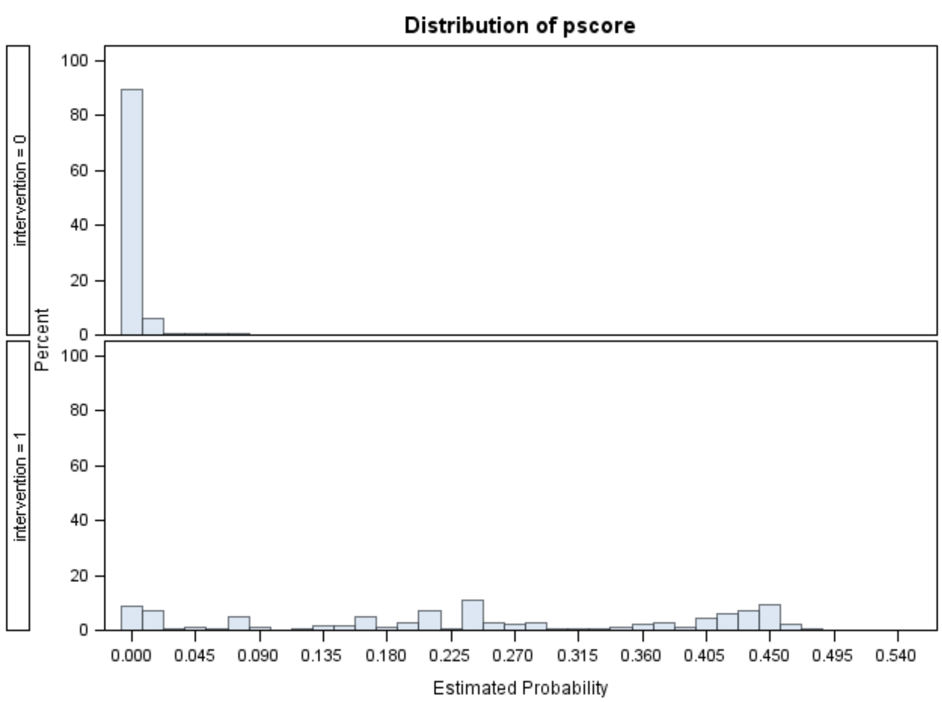
Nearest-neighbor matching found pairs for all 185 intervention subjects.

Caliper matching restricts the matches to have propensity scores within the caliper size.

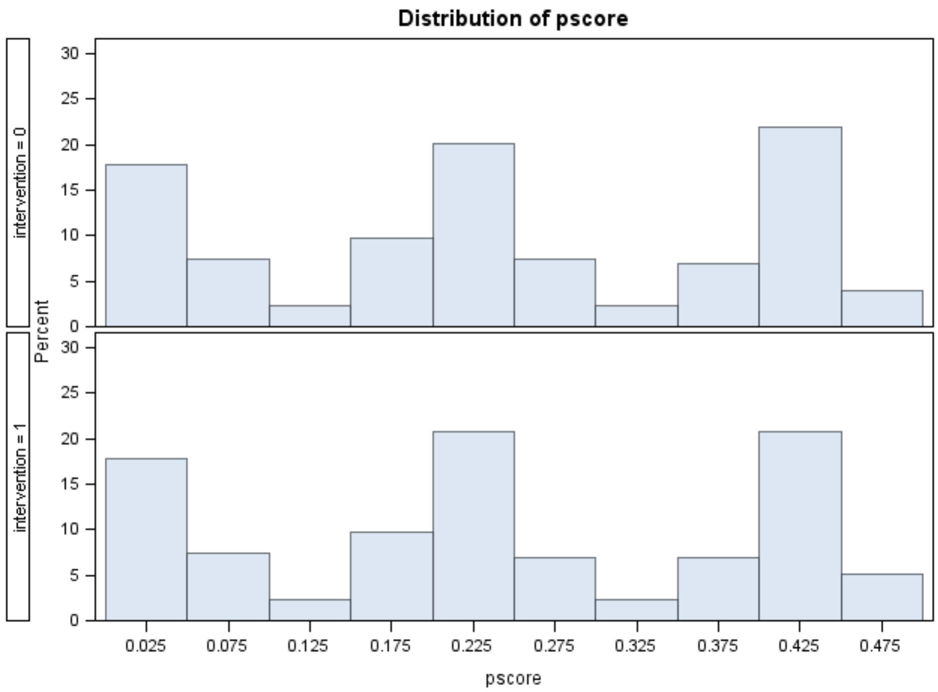
Setting caliper = .005 (based on range of propensity scores) gave pairs for 174 intervention subjects.

```
%PSMatching(datatreatment= T, datacontrol= C, method= caliper,
 numberofcontrols= 1, caliper=.005 , replacement=no,
 out=PS_match_cal);
```


How large should caliper be?



Comparison after caliper matching:



Estimated treatment effect, based on propensity score matching

First, merge matched pairs with baseline predictors.

Then model response

```
d_earnings = earnings1978 - mean(earnings1974, earnings1975);
```

on treatment, adjusting for baseline predictors, restricted to matched pairs.

Do not include pair in the model.

```
proc sort data=pairs_cal; by subject;
proc sort data=NSW; by subject;

data matched;
  merge pairs_cal NSW;
  by subject;
  if (pair NE .);

Proc GLM data=matched;
  class intervention;
  model d_earnings = intervention educ black hisp  married
    nodegree|age|earnings1974|earnings1975 @2 ;
  lsmeans intervention / stderr pdiff;
  estimate "trt - control" intervention -1 1 ;
```


intervention	d_earnings LSMEAN	Standard Error	H0:LSMEAN=0 Pr > |t|	H0:LSMean1= LSMean2 Pr > |t|
0	3481.02999	536.50504	<.0001	0.1961
1	4469.56756	536.50504	<.0001	

Dependent Variable: d_earnings

Parameter	Estimate	Standard Error	t Value	Pr > |t|
trt - control	988.537570	763.197677	1.30	0.1961

Estimates of treatment effect

$851 From experiment groups (LaLonde, 1986)

$989 From sample matched on propensity score, regression adjusted

$640 From all data, regression adjusted

References

Gelman and Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models,* §10.1–10.3

RB D'Agostino: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist. Med.* 1998; 17, 2265–2281.

PR Rosenbaum and DB Rubin: The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika,* 1983; 70: 41–55.

M. Coca-Perraillon (1987) "Local and Global Optimal Propensity Score Matching."

Large literature on causal inference from observational studies:
see §10.8 in Gelman and Hill.