G-ESTIMATION OF STRUCTURAL NESTED MODELS

Joy Shi, Barbra Dickerman, Miguel Hernán

DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Learning objectives At the end of this lecture you will be able to

- Specify a structural nested model for a continuous outcome
- Conduct g-estimation to estimate the parameters of a structural nested model
- Describe the relation between g-estimation and instrumental variable estimation

- Key concepts
  - G-estimation
  - Structural nested mean models

Study population

- 1629 cigarette smokers
- Aged 25-74 years when interviewed in 1971-75 (baseline)
- Interviewed again in 1982
- Known sex, age, race, weight, height, education, alcohol use, and smoking intensity at both baseline and follow-up visits, and who answered the general medical history questionnaire at baseline

Key variables

Treatment A	Quit smoking between baseline and 1982 1: yes, 0: no
Continuous outcome Y	Weight gain, kg Weight in 1982 minus baseline weight Available for 1566 individuals
Baseline (pre-treatment) covariates	Age, sex, race, alcohol use, intensity of smoking, weight…
Censoring C	Missing weight in 1982 1: yes, 0: no

Formal version of causal question: First define the counterfactual means

if everybody had quit smoking

- $E[Y^{a=1}]$
- $Y^{a=1}$ is an individual's outcome under $a=1$

if nobody had quit smoking

- $E[Y^{a=0}]$
- $Y^{a=0}$ is an individual's outcome under $a=0$

Then the formal question is:

- What is the average causal effect $E[Y^{a=1}] - E[Y^{a=0}]$?
  - ignore censoring for now

Average causal effect can be defined marginally or conditionally

- Marginal: Effect in the entire population
  - $E[Y^{a=1}] - E[Y^{a=0}]$
  - Methods: Standardization, IP weighting of MSMs, Instrumental variable estimation (under homogeneity)
- Conditional: Effects in subsets defined by effect modifiers
  - $E[Y^{a=1}|V=v] - E[Y^{a=0}| V=v]$ for all values l
  - All of the above combined with stratification
- Conditional: Effects in subsets defined by *all* confounders
  - $E[Y^{a=1}|L=l] - E[Y^{a=0}|L=l]$ for all values l
  - Methods: stratification, outcome regression, and g-estimation of structural nested models

Footnote: Average causal effects can also be conditional on PS

- Just replace the individual covariates L by the propensity score PS
  - $E[Y^{a=1}|PS=p] - E[Y^{a=0}|PS=p]$ for all values p
  - Methods: stratification and outcome regression using PS rather than L

Some new terms

- Structural nested models:
  - Models with parameter(s) that can be interpreted as the magnitude of the causal effect conditional on covariates
- G-estimation:
  - Method to estimate the parameters of structural nested models

Plan for this lecture

A. Review
- Conditional exchangeability

B. Structural nested models

C. G-estimation

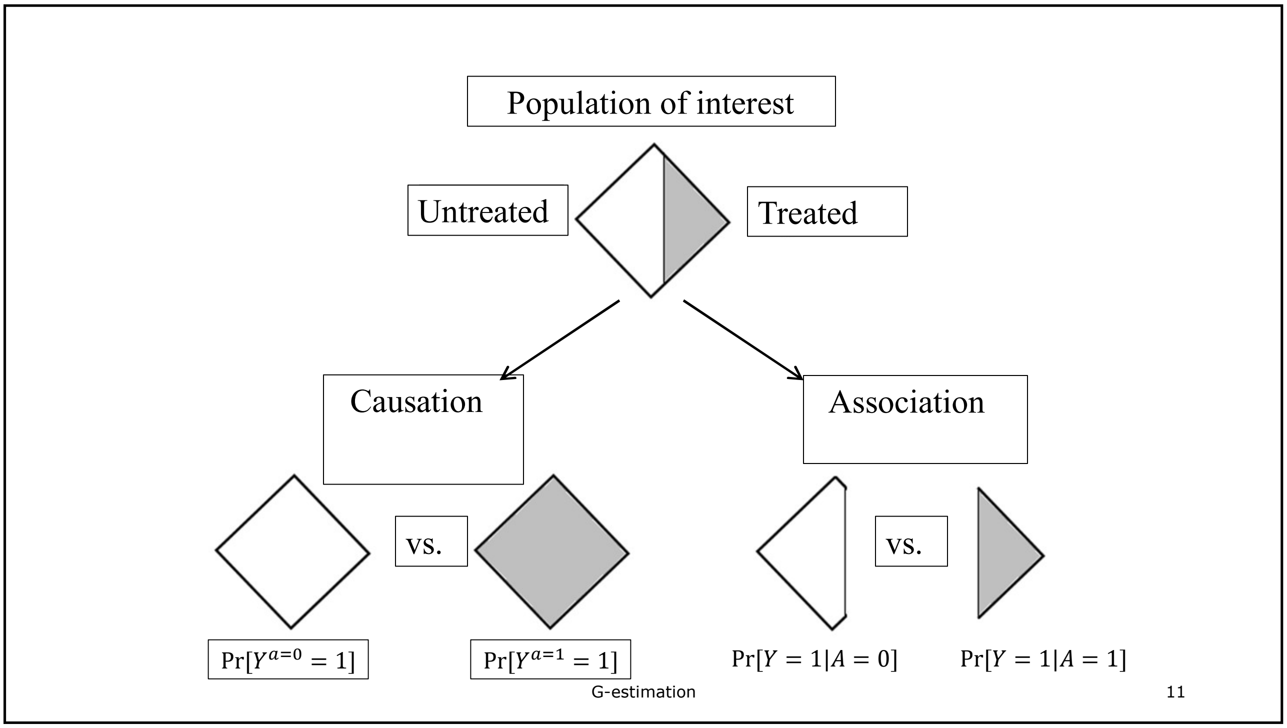
D. Adjustment for selection bias

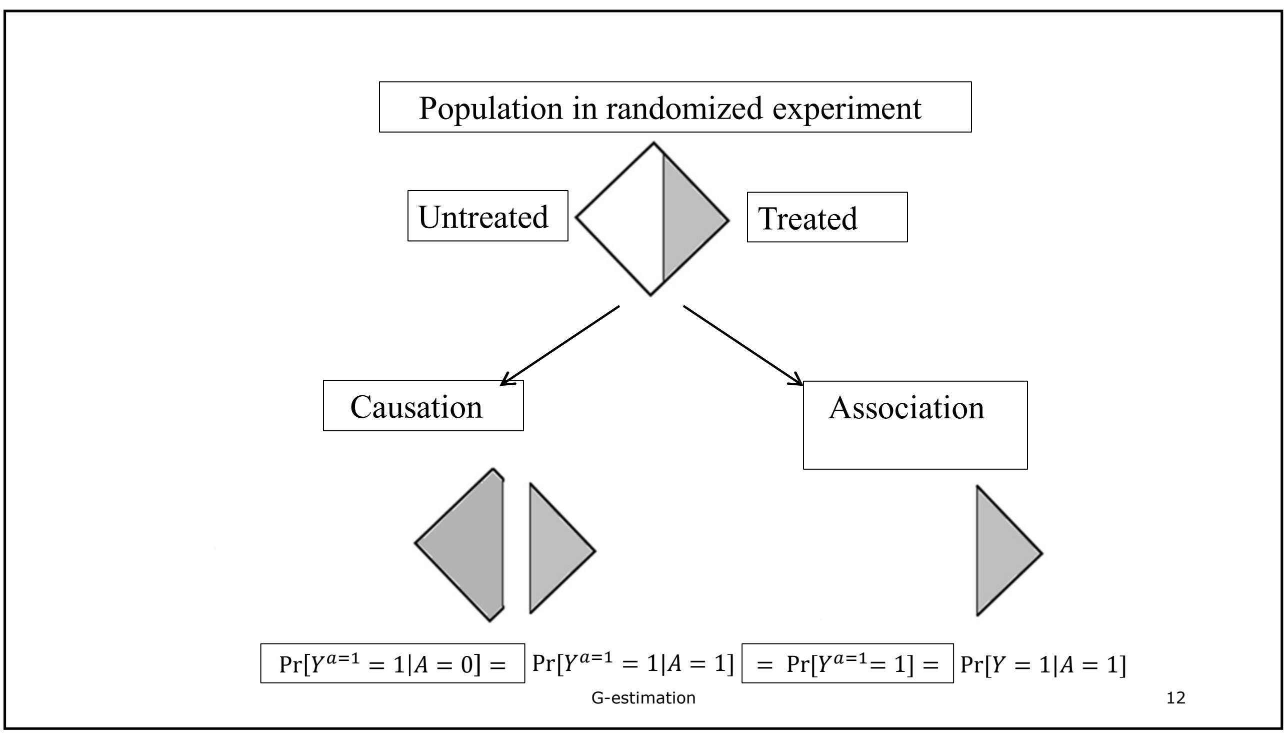
E. G-estimation and IV estimation

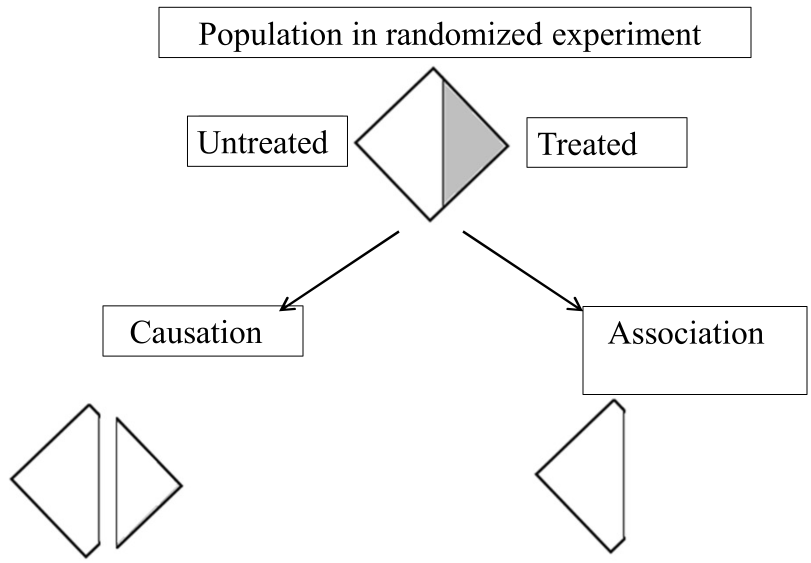
Conditional exchangeability
Formal definition

$$Y^a \coprod A | L = l \text{ for all a}$$

- Conditional exchangeability is equivalent to randomization within levels of the variables in L
  - e.g., smoking cessation occurred at random within levels of age, sex, race, smoking intensity, etc.
- Also known as the assumption of no unmeasured confounding


Population of interest
Untreated
Treated
Causation
Association
vs.
vs.
$\Pr[Y^{a=0} = 1]$
$\Pr[Y^{a=1} = 1]$
$\Pr[Y = 1|A = 0]$
$\Pr[Y = 1|A = 1]$
G-estimation
11


Population in randomized experiment
Untreated
Treated
Causation
Association
$\Pr[Y^{a=1} = 1|A = 0] =$
$\Pr[Y^{a=1} = 1|A = 1]$
$= \Pr[Y^{a=1} = 1] =$
$\Pr[Y = 1|A = 1]$
G-estimation
12



$\Pr[Y^{a=0} = 1 | A = 0] \; \boxed{= \Pr[Y^{a=0} = 1 | A = 1]} \; \boxed{= \Pr[Y^{a=0} = 1] =} \; \Pr[Y = 1 | A = 0]$

Clarification:
Average causal effect for continuous outcomes

- Only mean exchangeability is required
  - within levels of the covariates L, treated individuals would have had the same mean outcome as the untreated individuals had they being untreated, and vice versa
  - mean counterfactual outcome is the same in the treated and the untreated with the same value of L
- For dichotomous outcomes, exchangeability and mean exchangeability are the same

An equivalent expression of conditional exchangeability

$$Y^a \coprod A | L = l \text{ for all a}$$

$$\Pr[A = 1 | L = l, Y^{a=1}] = \Pr[A = 1 | L = l,\ Y^{a=0}]$$

$$= \Pr[A = 1 | L = l]$$

- That is, conditional on the measured covariates, the probability of treatment does not depend on the value of the counterfactual outcome

Dream with me for a minute

- Suppose conditional exchangeability given L holds
- Suppose we know the values of the counterfactual outcomes $Y^{a=1}$ and $Y^{a=0}$ for all subjects
- Suppose that we fit a logistic model for treatment that includes one of the counterfactual outcomes as a covariate
  - $\text{logit} \Pr[A=1|L, Y^a] = \alpha_0 + \alpha_1 Y^a + \alpha_2 L$
  - For either $a=0$ or $a=1$

What would be the value of the coefficient α_1 for $Y^{a=0}$?

$\text{logit Pr}[A = 1|L, Y^{a=0}] = \alpha_0 + \alpha_1 Y^{a=0} + \alpha_2 L$

0

Zero — 0%

Greater than zero — 0%

Less than zero — 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

What would be the value of the coefficient α_1 for $Y^{a=1}$?

$\text{logit Pr}[A = 1|L, Y^{a=1}] = \alpha_0 + \alpha_1 Y^{a=1} + \alpha_2 L$

0

Zero — 0%

Greater than zero — 0%

Less than zero — 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

We can test whether conditional exchangeability holds by fitting a propensity score model with counterfactual outcomes as covariates and then checking the value of α_1.

0

True

0%

False

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Reminder:
Different notation, different concepts

- Conditional exchangeability of treated and untreated is represented by

$$Y^a \coprod A | L = l$$

- Conditional independence of treatment and outcome is represented by

$$Y \coprod A | L = l$$

Under the sharp null, these two statements are expected to be true:

$Y^a \perp\!\!\!\perp A | L = l$

$Y \perp\!\!\!\perp A | L = l$

where L is the vector of measured variables in our study

0

True

0%

False

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Plan

A. Review
- Conditional exchangeability assumption

B. Structural nested models

C. G-estimation

D. Adjustment for selection bias

E. G-estimation and IV estimation

Average causal effect within levels of the confounders

- These conditional effects are represented by
  - $\mathrm{E}[Y^{a=1}|L=l] - \mathrm{E}[Y^{a=0}|L=l]$ for all values l
- Or, equivalently
  - $\mathrm{E}[Y^{a=1} - Y^{a=0}|L=l]$
- Outcome regression
  - models for the conditional mean outcome $\mathrm{E}[Y^{a}|L=l] = \mathrm{E}[Y|A=a, L=l]$
- Structural nested models
  - models for the conditional average effect $\mathrm{E}[Y^{a=1} - Y^{a=0}|L=l]$

Outcome regression One dichotomous confounder L

- If effect modification

$$\mathrm{E}[Y|A=a, L=l] = \beta_0 + \beta_1 A + \beta_2 AL + \beta_3 L$$

  - Effect in stratum $L=0$: β_1
  - Effect in stratum $L=1$: $\beta_1 + \beta_2$

- If no effect modification

$$\mathrm{E}[Y|A=a, L=l] = \beta_0 + \beta_1 A + \beta_3 L$$

  - Effect in all strata : β_1

Structural nested mean model
One dichotomous confounder L

- ☐ If effect modification

$$E[Y^a - Y^{a=0} | A=a, L=l] = \beta_1\, a + \beta_2\, aL$$

  - Effect in stratum L=0 : β_1
  - Effect in stratum L=1 : $\beta_1 + \beta_2$

- ☐ If no effect modification

$$E[Y^a - Y^{a=0} | A=a, L=l] = \beta_1\, a$$

  - Effect in all strata : β_1

Structural nested models have fewer parameters than outcome regression models

- ☐ Because SNMs model the causal effect directly
  - That is, the difference in mean outcomes rather than the mean outcomes
- ☐ No intercept and no parameters for the confounders
- ☐ Fewer opportunities of having bias due to model misspecification

Structural nested models vs. Marginal structural models

- For time-varying treatments, structural nested models allow product terms between treatment and time-varying covariates
  - marginal structural models do not
- For point treatments (like in our data), both classes of models can be used
  - MSMs estimate marginal effect or conditional effects, SNMs estimate conditional effects only

Plan

A. Review
   - Conditional exchangeability assumption

B. Structural nested models

C. G-estimation

D. Adjustment for selection bias

E. G-estimation and IV estimation

G-estimation

- The procedure to estimate the parameters of structural nested models
- Next we describe, in 5 steps, g-estimation to estimate β_1 in a structural nested mean model
  - $\mathrm{E}[Y^a - Y^{a=0} | A=a, L=l] = \beta_1 a$
  - under no effect modification by any confounders
    - Aside: in the presence of EM, there are multiple parameters to be identified: $\mathrm{E}[Y^a - Y^{a=0} | A = a, L = l] = \beta_1 a + \beta_2 aL$

Step 1: Mean model transformed into an individual-level, rank-preserving model

$$Y^a - Y^{a=0} = \psi_1 a$$

- This step is not really necessary for g-estimation but simplifies teaching the method
- This rank-preserving model has stronger assumptions than the mean model
- The estimate of ψ_1 obtained via g-estimation for the rank-preserving model is also valid for the mean model

This model assumes constant treatment effect
$Y^a - Y^{a=0} = \psi_1 a$
That is, the individual causal effect ψ_1 is the same for all individuals and equal to the average causal effect β_1

0

True
0%

False
0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Step 2: Some rearranging

$$Y^{a=0} = Y^a - \psi_1\, a$$

- We'll see why later
- Still too many unknowns in the model

Step 3: Model linked to the observed data

$$Y^{a=0} = Y - \psi_1 A$$

- The observed outcome Y equals the counterfactual outcome Y^a when a is equal to the observed treatment A
  - By consistency of counterfactuals

- How can we validly estimate ψ_1?

Step 4: Suppose that a friend of yours knows the value of ψ_1

- But they only tell you that ψ_1 is one of the following
  - $\psi^\dagger = -1.11$, $\psi^\dagger = 0$, or $\psi^\dagger = 3.46$
- How can you identify the true value ψ_1 among the 3 possible values $\psi^\dagger$?
- For each subject, compute $H(\psi^\dagger) = Y - \psi^\dagger A$ for each of the three possible values $\psi^\dagger$
- The newly created variables $H(-1.11)$, $H(0)$, and $H(3.46)$ are candidate counterfactuals
  - One of them is the counterfactual outcome $Y^{a=0}$
  - $H(\psi^\dagger) = Y^{a=0}$ if $\psi^\dagger = \psi_1$

Computing $H(\psi^{\dagger})$ from our observed data

$Y^{a=0} = Y - \psi_1 A$

$H(\psi^{\dagger}) = Y - \psi^{\dagger} A$

Recall that ψ_1 is one of the following: $\psi^{\dagger} = -1.11$, $\psi^{\dagger} = 0$, or $\psi^{\dagger} = 3.46$

ID	A	L	Y
1	0	0	3.5
2	1	1	1.2
3	1	1	2.7
…	…	…	…

Computing $H(\psi^{\dagger})$ from our observed data

$Y^{a=0} = Y - \psi_1 A$

$H(\psi^{\dagger}) = Y - \psi^{\dagger} A$

Recall that ψ_1 is one of the following: $\psi^{\dagger} = -1.11$, $\psi^{\dagger} = 0$, or $\psi^{\dagger} = 3.46$

ID	A	L	Y	$H(\psi^{\dagger}=-1.11)$
1	0	0	3.5	3.5 – (-1.11)(0) = 3.5
2	1	1	1.2	1.2 – (-1.11)(1) = 2.31
3	1	1	2.7	2.7 – (-1.11)(1) = 3.81
…	…	…	…	…

Computing $H(\psi^\dagger)$ from our observed data

$Y^{a=0} = Y - \psi_1 A$

$H(\psi^\dagger) = Y - \psi^\dagger A$

Recall that ψ_1 is one of the following: $\psi^\dagger = -1.11$, $\psi^\dagger = 0$, or $\psi^\dagger = 3.46$

ID	A	L	Y	$H(\psi^\dagger$=-1.11)	$H(\psi^\dagger$=0)
1	0	0	3.5	3.5 – (-1.11)(0) = 3.5	3.5 – 0(0) = 3.5
2	1	1	1.2	1.2 – (-1.11)(1) = 2.31	1.2 – 0(1) = 1.2
3	1	1	2.7	2.7 – (-1.11)(1) = 3.81	2.7 – 0(1) = 2.7
…	…	…	…	…	…

Computing $H(\psi^\dagger)$ from our observed data

$Y^{a=0} = Y - \psi_1 A$

$H(\psi^\dagger) = Y - \psi^\dagger A$

Recall that ψ_1 is one of the following: $\psi^\dagger = -1.11$, $\psi^\dagger = 0$, or $\psi^\dagger = 3.46$

ID	A	L	Y	$H(\psi^\dagger$=-1.11)	$H(\psi^\dagger$=0)	$H(\psi^\dagger$=3.46)
1	0	0	3.5	3.5 – (-1.11)(0) = 3.5	3.5 – 0(0) = 3.5	3.5 – 3.46(0) = 3.5
2	1	1	1.2	1.2 – (-1.11)(1) = 2.31	1.2 – 0(1) = 1.2	1.2 – 3.46(1) = -2.26
3	1	1	2.7	2.7 – (-1.11)(1) = 3.81	2.7 – 0(1) = 2.7	2.7 – 3.46(1) = -0.76
…	…	…	…	…	…	…

Step 5: Test all the candidates, choose the true value

- To identify which candidate $H(\psi^{\dagger})$ is the counterfactual $Y^{a=0}$, fit 3 models

$$\text{logit } \Pr[A=1|L, H(\psi^{\dagger})] = \alpha_0 + \alpha_1 H(-1.11) + \alpha_2 L$$
$$\text{logit } \Pr[A=1|L, H(\psi^{\dagger})] = \alpha_0 + \alpha_1 H(0) + \alpha_2 L$$
$$\text{logit } \Pr[A=1|L, H(\psi^{\dagger})] = \alpha_0 + \alpha_1 H(3.46) + \alpha_2 L$$

- Remember the assumption of conditional exchangeability?
  - and its relation to the parameter α_1 in the model

$$\text{logit } \Pr[A=1|L, Y^{a=0}] = \alpha_0 + \alpha_1 Y^{a=0} + \alpha_2 L$$

Which candidate $H(\psi^{\dagger})$ is equal to the counterfactual outcome?
Recall: $\text{logit } \Pr[A=1|L, H(\psi^{\dagger})] = \alpha_0 + \alpha_1 H(\psi^{\dagger}) + \alpha_2 L$

0

Come on. This is hopeless. 0%

The candidate $H(\psi^{+})$ with α_1=0 0%

The candidate $H(\psi^{+})$ with α_1>0 0%

The candidate $H(\psi^{+})$ with α_1<0 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Which $\psi^\dagger$ is equal to the causal parameter ψ_1?
Recall: $\text{logit Pr}[A = 1|L, H(\psi^\dagger)] = \alpha_0 + \alpha_1 H(\psi^\dagger) + \alpha_2 L$

0

(A) The ψ^+ with α_1=0 when used to compute the candidate $H(\psi^+)$
0%

(B) The ψ^+ with α_1>0 when used to compute the candidate $H(\psi^+)$
0%

(C) The ψ^+ with α_1<0 when used to compute the candidate $H(\psi^+)$
0%

Done: That was g-estimation

- Search over all possible values of $\psi^\dagger$ until you find one that results in $H(\psi^\dagger)$ with α_1=0
  - The candidate $H(\psi^\dagger)$ with α_1=0 is the counterfactual and the corresponding $\psi^\dagger$ is the true value ψ_1
- In practice, use more sophisticated search methods
  - Need some additional coding
- Sometimes the estimator has closed form
  - No need to search or optimize
  - e.g., linear models like the ones considered here

To get the 95% confidence interval for ψ_1

- Invert the test for α_1=0
  - Find the set of values of $\psi^\dagger$ that result in a *P*-value>0.05 when testing for α_1=0
  - The 95% confidence limits are the limits of that set of values
  - Can use Wald test, Score test…
- Bootstrapping

Back to our data: Point estimate for ψ_1

- Point estimate: 3.5 *See gestimation.R, lines 19-42*
  - Value of $\psi^\dagger$ that produces the value of α_1 closest to 0
  - Or, equivalently, p-value closest to 1
- Adjusting for baseline covariates:
  - Sex, age, race, education, smoking intensity, duration of smoking, exercise, active lifestyle, baseline weight

95% confidence interval for estimate of ψ_1

- 95% CI: 2.6, 4.4 approx
- By inversion of the test: values of $\psi^\dagger$ for which the test for α_1=0 has *P*-value>0.05 in model
  - We might not be able to find the value of $\psi^\dagger$ that corresponds to exactly p-value=0.05
  - Select next closest $\psi^\dagger$ with p-value<0.05 to give us a (conservative) 95% CI

See gestimation.R, lines 19-42

ψ_1	**p-value**
2.5	0.028
2.6	0.049
2.7	0.081
…	…
3.4	0.888
3.5	0.929
3.6	0.750
…	…
4.2	0.091
4.3	0.055
4.4	0.032

More efficient g-estimators exist (narrower 95% CIs)

- Doubly-robust estimators are more efficient
  - Require a model for the mean of $H(\psi^\dagger)$ given the confounders L
- G-estimation based on functions of $H(\psi^\dagger)$ may be more efficient
  - e.g., we could use the model

$$\text{logit Pr}[A=1|L, H(\psi)] = \alpha_0 + \alpha_1 [H(\psi^\dagger)]^3 + \alpha_2 L$$

Causal questions of interest

1. What is the effect of smoking cessation on weight gain?

2. What is the effect of smoking cessation on risk of death? ⇨

- G-estimation cannot be generally used with logistic models, particularly in the time-varying setting
- Can use log models under rare disease assumption

Plan

A. Review
  - Conditional exchangeability assumption

B. Structural nested models

C. G-estimation

D. Adjustment for selection bias

E. G-estimation and IV estimation

Key variables

Treatment A	Quit smoking between baseline and 1982 1: yes, 0: no
Continuous outcome Y	Weight gain, kg Weight in 1982 minus baseline weight Available for 1566 individuals
Baseline (pre-treatment) covariates	Age, sex, race, alcohol use, intensity of smoking, weight…
Censoring C	Missing weight in 1982 1: yes, 0: no

Causal question

- Estimate the counterfactual mean weight gain if nobody's outcome had been censored
  - $E[Y^{a=1,c=0}]$
  - $E[Y^{a=0,c=0}]$
- Estimate the average causal effect
  - $E[Y^{a=1,c=0}] - E[Y^{a=0,c=0}]$
- Conditional on confounders when using g-estimation

G-estimation + IP weighting

- G-estimation to adjust for confounding
- IP weighting to adjust for selection bias

- First, use IP weighting to estimate IP weights W^C
  - The weights create a pseudo-population without selection bias by measured covariates
  - i.e., no arrows from *L* and *A* into *C*

G-estimation + IP weighting

- Second, use g-estimation to adjust for confounding in the pseudo-population
  - Find the value of ψ that produces the value of α_1 closest to 0 in the IP weighted model

    $\text{logit } \Pr[A=1|L, H(\psi^\dagger)] = \alpha_0 + \alpha_1 H(\psi^\dagger) + \alpha_2 L$
  - using estimates of the weights W^C

- Need to use robust variance estimator
  - Conservative 95% confidence interval

Back to our data: Point estimate for ψ_1

- Point estimate: 3.4 *See gestimation.R, lines 44-104*
- 95% CI: 2.5, 4.5 approx
- Adjusting for baseline covariates:
  - Sex, age, race, education, smoking intensity, duration of smoking, exercise, active lifestyle, baseline weight

Plan

A. Review
  - Conditional exchangeability assumption

B. Structural nested models

C. G-estimation

D. Adjustment for selection bias

E. G-estimation and IV estimation

G-estimation to consistently estimate the parameter ψ_1

- From the model $Y^{a=0} = Y - \psi_1 A$
- Compute $H(\psi^\dagger) = Y - \psi^\dagger A$
  - for a range of possible values $\psi^\dagger$
  - The variables $H(\psi^\dagger)$ are candidate counterfactuals
  - One of them is the counterfactual outcome $Y^{a=0}$
    - $H(\psi^\dagger) = Y^{a=0}$ if $\psi^\dagger = \psi_1$
- We just need to identify the value $\psi^\dagger$ that makes $H(\psi^\dagger) = Y^{a=0}$

How to identify the value $\psi^\dagger$ that makes $H(\psi^\dagger) = Y^{a=0}$?

- We said that **IF** conditional exchangeability holds

$$Y^a \coprod A | L = l$$

- Then ψ_1 is the value $\psi^\dagger$ that makes α_1=0 in the model
  $\text{logit } \Pr[A=1|L, H(\psi^\dagger)] = \alpha_0 + \alpha_1 H(\psi^\dagger) + \alpha_2 L$
- But there is another way
  - If we have an instrumental variable Z

How to identify the value of ψ that makes $H(\psi^\dagger) = Y^{a=0}$?

- **By definition** of IV, (unconditional) exchangeability is expected to hold

$$Y^a \coprod Z$$

- Then the true value of ψ is the one that makes $\alpha_1=0$ in the model $\text{logit}\ \Pr[Z=1|H(\psi)] = \alpha_0 + \alpha_1 H(\psi^\dagger)$

IV estimation as a particular case of g-estimation

- G-estimation based on the IV is mathematically equivalent to the standard IV estimand
  - Defined in last week's notes
- The standard IV estimator consistently estimates the parameter ψ^* from our structural mean model
  - Under instrumental assumptions (i)-(iii) and assumption of no additive effect modification (iv)
  - See Chapter 16 of *Causal Inference: What If.*

Within the g-estimation framework, IV becomes a more general method that allows

- Use of baseline covariates
  - To increase efficiency
    - Especially if covariates are strongly associated with Y and weakly associated with A
  - To address violations of condition (iii)
- Use of multiplicative models and assumption of no multiplicative effect modification (iv)
- Extensions to time-varying treatments and time-varying instruments, survival analyses

Conclusion

- We have studied (essentially) all available methods to adjust for confounding
- Of those, IP weighting, standardization/g-formula, and g-estimation can be generally used with time-varying treatments

IP weighting, standardization, or g-estimation?

- Why do we have to choose?
  - If possible, use all methods separately and doubly-robust versions when available
  - Similar effect estimates increase our confidence in the results
  - Different effect estimates will make us examine our modeling assumptions
- See EPI207 for more about these methods

Readings

- *Causal Inference: What If*. Chapter 14

Progress report

1. Introduction to modeling
2. Stratified analysis:
   - Outcome regression
   - Propensity scores
3. Standardization
4. Inverse probability weighting
   - Marginal structural models
5. Instrumental variable estimation
6. G-estimation
7. Causal survival analysis