INVERSE PROBABILITY WEIGHTING
Marginal Structural Models

Barbra Dickerman, Joy Shi, 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

- Define marginal structural models
- Estimate the parameters of marginal structural models
- Estimate and use stabilized IP weights

- Key concepts
  - Marginal structural models
  - Stabilized IP weights

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
Dichotomous outcome D	Death by 1992 1: yes, 0: no
Baseline (pre-treatment) covariates	Age, sex, race, alcohol use, intensity of smoking, weight…

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?

- This is an informal statement of the questions

A more formal version of causal question #1 First define the counterfactual means

if everybody had quit smoking

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

if nobody had quit smoking

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

Then the formal question is:

- What is the average causal effect $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$?

Plan for today

A. Marginal structural models

B. Stabilized IP weights

If this were an ideal randomized experiment…

- … in which individuals had been randomly assigned to "smoking cessation"
- Then there would be no need for IP weights
  - The average outcome in the treated $\mathrm{E}[Y|A=1]$ is a consistent estimator of the mean outcome had everyone been treated $\mathrm{E}[Y^{a=1}]$
  - Same for the untreated
- The difference $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$ is consistently estimated by $\hat{\theta}_1$ from the regression model

$$\mathrm{E}[Y|A] = \theta_0 + \theta_1 A$$

This model is saturated

$E[Y|A] = \theta_0 + \theta_1 A$

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**

This model cannot possibly be misspecified

$E[Y|A] = \theta_0 + \theta_1 A$

0

(A) True

0%

(B) False

0%

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

If this were an observational study…

- … in which the treated and the untreated are exchangeable conditional on measured confounders?
  - e.g., age, sex, race
- Then IP weighting would create a pseudo-population with unconditional exchangeability
  - If no model misspecification
- The difference $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$ is consistently estimated by $\hat{\theta}_1$ from regression model

$$\mathrm{E}[Y|A] = \theta_0 + \theta_1 A$$

  fit to the pseudo-population

Models for both treatment and outcome

- So far we have used models to estimate the IP weights and weighted averages (no models) to estimate the mean outcome
- But we could also estimate the mean outcome via modeling
  - i.e., a linear regression model in which individuals are weighted by their W^A
- We can then compute a 95% confidence interval
  - By bootstrapping or computing analytic variance
  - By using robust variance (conservative 95% CI)

Weighted regression model
$\mathrm{E}[Y|A] = \theta_0 + \theta_1 A$

Estimates of IP weighted average of the outcome

- 5.2 kg in the treated = $\hat{\theta}_0 + \hat{\theta}_1$
- 1.8 kg in the untreated = $\hat{\theta}_0$

See ipw_msm.R, lines 5-37

Difference $\hat{\theta}_1$: 3.4 kg, conservative 95% CI: 2.4, 4.5

- This difference would have a causal interpretation as $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$ if
  - all confounders had been included in the IP weighting procedure
  - the model for treatment is correctly specified

A Marginal Structural Model (MSM)
$\mathrm{E}[Y^a] = \beta_0 + \beta_1 a$

- Interpretation of parameters
  - Mean weight gain if everybody untreated $\mathrm{E}[Y^{a=0}] = \beta_0$
  - Mean weight gain if everybody treated $\mathrm{E}[Y^{a=1}] = \beta_0 + \beta_1$
  - Difference $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] = \beta_1$
- The parameter β_1 is precisely what we have been trying to estimate all this time

This model cannot possibly be misspecified

$E[Y^a] = \beta_0 + \beta_1 a$

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**

Estimation of parameters of MSMs (Robins 1998)

$\mathrm{E}[Y^a] = \beta_0 + \beta_1 a$

- Fit a regression model $\mathrm{E}[Y|A] = \theta_0 + \theta_1 A$ to the pseudo-population
  - That is, fit a weighted regression model in which each individual contributes as many observations as their IP weight
- The parameter estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ from the IP weighted model are consistent for the parameters β_0 and β_1 of the MSM

Saturated vs. Nonsaturated MSMs

Our outcome model is saturated

- 2 parameters, 2 estimated quantities
- guaranteed to be correctly specified when the treatment is dichotomous

Because we didn't make any parametric assumptions

- The estimated effect of 3.4 kg was identical whether we used weighted sample averages or a weighted regression model

What if treatment is a continuous variable?

- For example, 'change in smoking intensity' rather than 'smoking cessation'
  - where smoking intensity is measured as number of cigarettes per day
- If treatment can take 100 values, then there are 100 counterfactual means
  - One per level of treatment
- Consider the MSM $\mathrm{E}[Y^a] = \beta_0 + \beta_1 a + \beta_2 a^2$
  - to estimate those 100 means

This model is saturated

$E[Y^a] = \beta_0 + \beta_1 a + \beta_2 a^2$

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**

This model cannot possibly be misspecified

$E[Y^a] = \beta_0 + \beta_1 a + \beta_2 a^2$

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**

Nonsaturated MSMs

- The MSM may be misspecified when treatment has more than a few levels
  - e.g., a continuous variable, a dichotomous time-varying variable
- Estimating IP weights for treatments with more than two levels requires models other than the logistic model
  - For continuous treatments we need to estimate the density, which is hard
  - See Chapter 12

Why MSMs are "structural"?

- The outcome variable of a model can be
  - observed
  - counterfactual
- Models for counterfactual outcomes are referred to as "structural" or "causal"
- Parameters for treatment in structural models
  - have direct causal interpretation
  - can be estimated by IP weighting or G-estimation

Why MSMs are "marginal"?

- MSMs are marginal because they are models for functionals of the marginal distribution of the counterfactual outcome
- MSMs do not impose any restrictions on the joint distribution of counterfactual outcomes under different treatment values
  - MSMs model the mean of $Y^{a=1}$ and the mean of $Y^{a=0}$, but MSMs are agnostic as to what the relation is (if any) between these two counterfactual outcomes

Nonstabilized vs. Stabilized IP weights

$$W^A = \frac{1}{f(A|L)} \qquad SW^A = \frac{f(A)}{f(A|L)}$$

- Nonstabilized weights are often unusable because they result in very unstable estimates; stabilized weights are preferable
- However, stabilized and nonstabilized weights yield the same estimates when using saturated outcome models
  - Check it out in our smoking cessation example

See ipw_msm.R, lines 40-88

When using W^A, treatment is independent of measured confounders in the pseudo-population

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**

When using SW^A, treatment is independent of measured confounders in the pseudo-population

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**

What is the mean of the nonstabilized weights W^A

1 — 0%

0 — 0%

Number of treatment levels — 0%

Other — 0%

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

What is the mean of the stabilized weights SW^A

1 — 0%

0 — 0%

Number of treatment levels — 0%

Other — 0%

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

MSMs offer the choice to estimate marginal or conditional effects

- MSMs can include covariates to evaluate effect modification by baseline variables
- For example, for V= age (in years), we can consider the nonsaturated MSM

$$\mathrm{E}[Y^a \mid V] = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 a V$$

where the parameter β_3 measures modification of the effect of smoking cessation by age

MSMs 29

This model is saturated

$E[Y^a|V] = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$

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**

This model cannot possibly be misspecified
$E[Y^a|V] = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$

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**

V-stabilized weights

$$SW^A(V) = \frac{f(A|V)}{f(A|L)}$$

Baseline variables V included in the MSM also need to be included in

- the denominator of the weights (as part of L)
- the numerator of the weights

- In practice, we will estimate the probabilities in the numerator by fitting a logistic model for treatment with V as the only covariates
- V-stabilization results in IP weights that are more stabilized than the ones without V

IP weighted model (with *V*-stabilization)

$$E[Y|A,V] = \theta_0 + \theta_1 A + \theta_2 V + \theta_3 AV$$

- This model allows us to estimate the IP weighted means in the treated and the untreated for each value of age
- The estimate of effect modification is $\hat{\theta}_3 = -0.025$ (conservative 95% CI: -0.11, 0.61)

See ipw_msm.R, lines 92-122

- Little evidence of effect modification by age

Some approaches to causal inference that rely on models

Outcome regression

- Conditional on all confounders or the propensity score
- Confounding adjustment via stratification/conditioning
- $E[Y|A,L] = \theta_0 + \theta_1 A + \theta_2 L$

Marginal structural models

- Unconditional or conditional on some effect modifiers
- Confounding adjustment via IP weighting
- $E[Y^a] = \beta_0 + \beta_1 a$
- $E[Y^a|V] = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$

Parametric g-formula (standardization)

- Uses outcome regression averaged over the confounders to estimate average causal effects

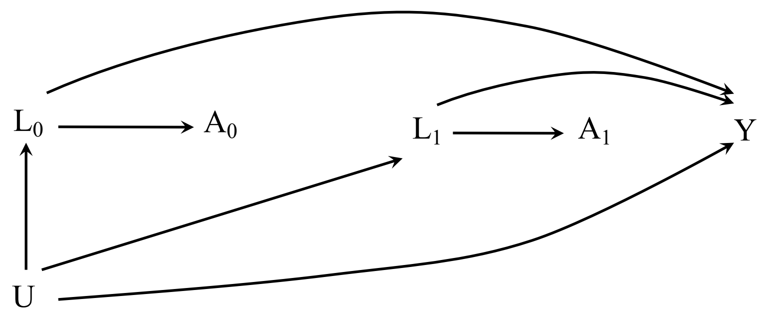
IP weighting of Marginal Structural Models or Parametric g-formula?

- Advantages of IP weighting
  - Less computationally intensive
  - Often easier to model treatment than outcome (and, in time-varying setting, confounders)
  - Parameter for null hypothesis (no g-null paradox)
  - Easier to identify positivity violations
- Disadvantages of IP weighting
  - More sensitive to violations (or quasi-violations) of positivity
- When possible, we can use doubly-robust estimators

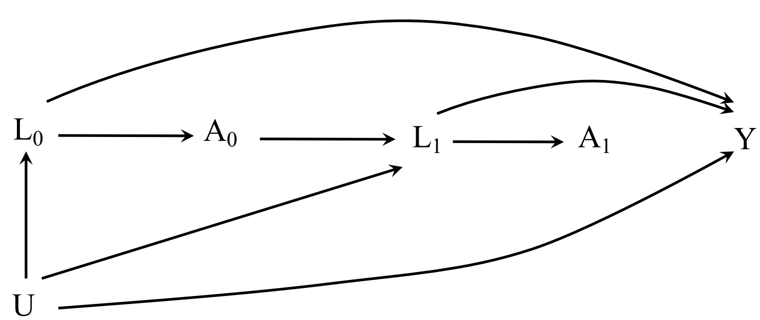
Key advantage of IP weighting and standardization over stratification-based methods

- In many studies, treatment is time-varying
  - Medical therapies, lifestyle, diet…
- and therefore the confounders are time-varying too
  - there may be treatment-confounder feedback
- If time-varying treatments and confounders, and confounders are affected by prior treatment
  - IP weighting and standardization/g-formula control confounding because they can handle treatment-confounder feedback
  - Outcome regression and propensity score methods introduce bias because they cannot handle treatment-confounder feedback

Preview: time-varying confounding



Preview: time-varying confounding with treatment-confounder feedback



Readings

- *Causal Inference: What If*. Chapter 12

Progress report

1. Introduction to modeling
2. Stratified analysis:
   - outcome regression
   - propensity scores
3. Standardization
4. IP weighting
   - Marginal structural models

   (to be continued…)