Modern Statistical Learning Methods for Observational Data and Applications to Comparative Effectiveness Research

Chapter 5: Additional topics

David Benkeser
Emory Univ.

Marco Carone
Univ. of Washington

Larry Kessler
Univ. of Washington

Module 14

4th Annual Summer Institute for Statistics in Clinical Research

07/27/2017

Contents of this chapter

1. How can we analyze data with missing outcomes?
2. How can we analyze data with missing covariates?
3. What happens if we do not measure all confounders?
4. How can we analyze data with continuously-valued treatments?
5. How can we summarize subgroup effects and covariate/treatment interactions?

Missing outcomes

In observational studies and clinical trials, the outcome Y is typically not observed for every participant.

- People withdraw consent, move out of area, stop showing up for study visits,...
- The observed data unit is then $O := (W, A, \Delta, \Delta Y)$, with Δ denoting an indicator of having a measured outcome.

People who are lost may be meaningfully different from those who remain under study.

- What if every patient who had more serious disability three months after imaging did not show up for their one-year visit?
- If the analysis excludes these patients, we are making inference about treatment efficacy *in a possibly very different population*.

Can we modify our approaches to account for possibly informative missingness?

Missing outcomes

Ideally, for each patient in the population, we would like to:

1. Set treatment to $A = 1$ **and** observe patient at end of study (i.e., set $\Delta = 1$).
2. Set treatment to $A = 0$ **and** observe patient at end of study (i.e., set $\Delta = 1$).

The relevant counterfactual means are denoted by $E[Y(1,1)]$ and $E[Y(0,1)]$, where $Y(a,\delta)$ is the counterfactual outcome corresponding to setting $A = a$ and $\Delta = \delta$.

The difference $E[Y(1,1)] - E[Y(1,0)]$ is the ATE of interest in the full population.

What assumptions do we need to identify these parameters?

Missing outcomes

Previously, we required the randomization condition

$$Y(1) \perp A \mid W$$

which implies the treatment is randomized within strata of recorded covariates.

This will hold if:

- the study guarantees it by design (e.g., stratified randomized trial);
- all potential confounders have been recorded.

We now require the stronger randomization condition

$$Y(a,1) \perp A \mid W \quad \textbf{and} \quad Y(a,1) \perp \Delta \mid W$$

implying that treatment **and** missingness are randomized within strata.

It is generally difficult to ensure the latter through study design. We must hope that all potential confounders have been recorded.

Missing outcomes

Previously, we required the positivity assumption

$$P(A = a \mid W = w) > 0 \text{ for every plausible value } w$$

which implies all patients may potentially be assigned to treatment group $A = a$.

We now require the stronger positivity condition that, for all plausible value w,

$$P(A = a \mid W = w) > 0 \quad \textbf{and} \quad P(\Delta = 1 \mid A = a, W = w) > 0 \ .$$

Each patient must have had the chance to receive treatment **and** remain under study.

Missing outcomes

If the randomization and positivity conditions hold, then by the G-computation formula

$$\begin{aligned} E[Y(a,1)] &= E\left[E(\Delta Y \mid A=a, \Delta=1, W)\right] \\ &= \sum_w E(\Delta Y \mid A=a, \Delta=1, W=w)P(W=w) \ . \end{aligned}$$

We can define the subgroup-specific average treatment effect as

$$SATE(w) := E(\Delta Y \mid A=1, \Delta=1, W=w) - E(\Delta Y \mid A=0, \Delta=1, W=w) \ .$$

G-computation pools subgroup-specific treatment effects across target population:

$$\begin{aligned} ATE &= E[Y(1,1)] - E[Y(0,1)] \\ &= \sum_w SATE(w)P(W=w) \ . \end{aligned}$$

The G-computation excluding missing observations instead gives

$$\sum_w SATE(w)P(W=w \mid \Delta=1) \ .$$

Missing outcomes

The **inverse-probability-of-treatment-weighted (IPTW)** identification formula can also be modified to account for missing outcomes.

$$E[Y(a,1)] = E\left[\frac{I(A=a,\Delta=1)\Delta Y}{P(A=a,\Delta=1 \mid W)}\right] = E\left[\frac{I(A=a,\Delta=1)\Delta Y}{P(A=a \mid W)P(\Delta=1 \mid A=a,W)}\right]$$

This is a weighted average of the outcome of patients with $A=a$, weighted according to their propensity of being assigned to group $A=a$ **and** remaining under study.

If $P(A=1 \mid W=w)=.05$, a patient with $W=w$ had a 5% chance of being treated. Before, this patient stood in for approximately 19 similar patients not treated.

Now, this patient must also stand in patients who had a missing outcome.

If $P(\Delta=1 \mid A=1,W=w)=0.95$, a treated patient with $W=w$ had a 95% chance of completing the study. This patient has weight

$$\frac{1}{P(A=1 \mid W=w)P(\Delta=1 \mid A=1,W=w)} = \frac{1}{0.05\times 0.95} \approx 21\ .$$

Missing outcomes

The AIPTW estimator of ψ_1 can also be modified to account for missing outcomes:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(A_i=1,\Delta_i=1)}{g_n(W_i)}\right]Y_i}_{\text{IPTW estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[1-\frac{I(A_i=1,\Delta_i=1)}{g_n(W_i)}\right]\bar{Q}_n(1,W_i)}_{\text{augmentation term}} .$$

Here, g_n is an estimator of the extended propensity score

$$g(w) := P(\Delta = 1, A = 1 \mid W = w) ,$$

the conditional probability of being assigned $A = 1$ and remaining under observation.

A modified version of TMLE can be implemented, by fitting the augmentation regression using the covariate

$$Z^1 = \frac{I(A=1,\Delta=1)}{g_n(W)} .$$

The estimate g_n could simply be constructed by fitting a regression of the binary outcome $\tilde{A} := I(A = 1, \Delta = 1)$ on covariate vector W.

Missing outcomes

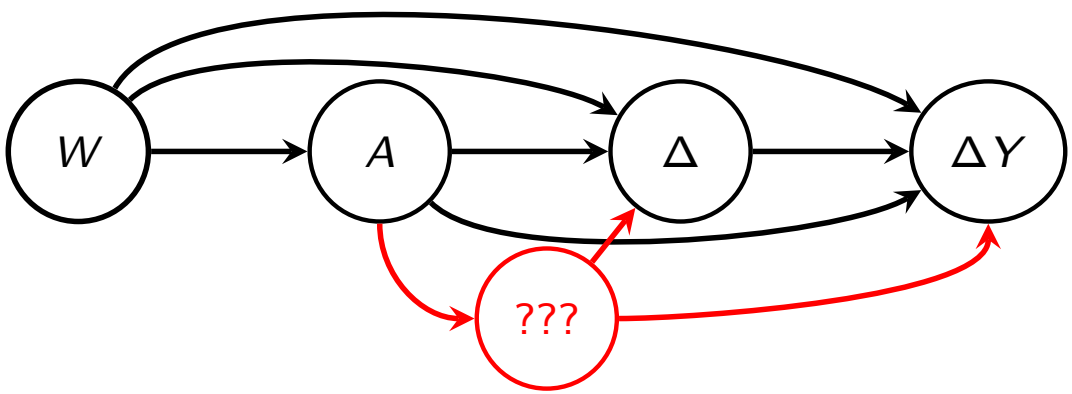
Recall that the randomization condition requires

$$Y(a,1) \perp A \mid W \quad \textbf{and} \quad Y(a,1) \perp \Delta \mid W$$

implying that treatment **and** missingness are randomized within strata.

This must preclude the possibility that no confounding happens after A is assigned.

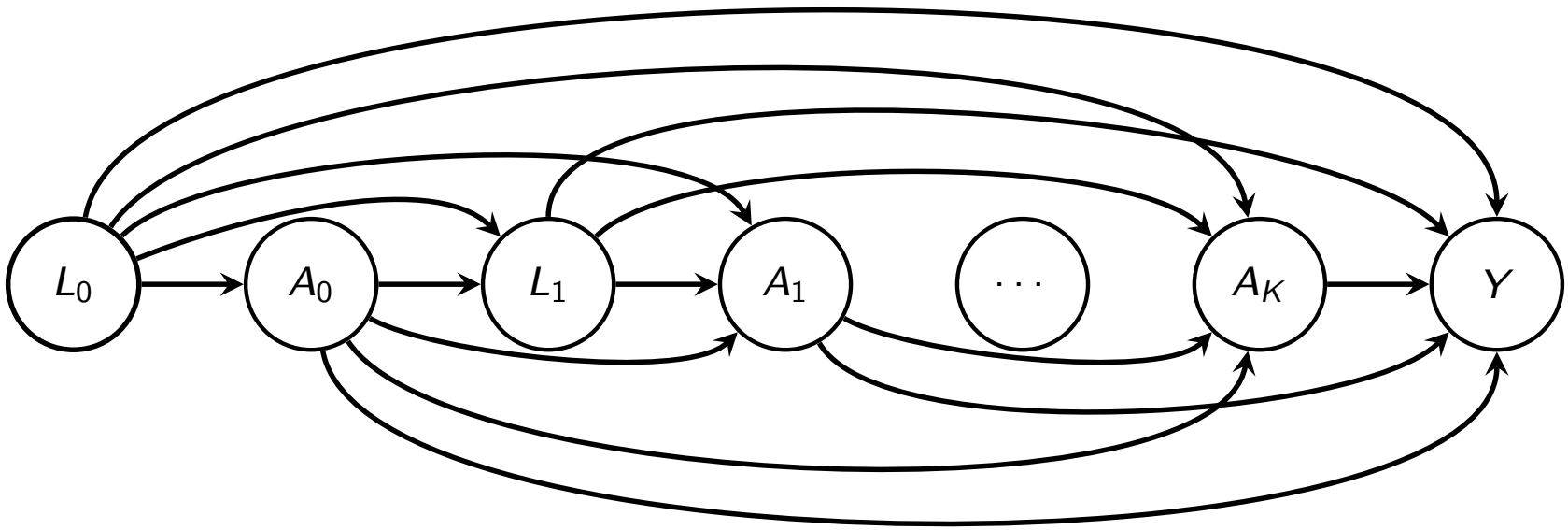
Do we believe that **nothing** happens between treatment assignment and end of study that influences a patient's outcome and probability of remaining under study?



Missing outcomes

We then have time-varying confounding, as is common in longitudinal data.

- $K + 1$ treatment decisions (or missingness indicators);
- Time-varying confounders measured between each treatment decision;
- Counterfactual outcome is $Y(\bar{a})$, where $\bar{a} := \{a_0, a_1, \ldots, a_K\}$;
- How do we identify and estimate $E[Y(\bar{a})]$?



Key points: missing outcomes

- Missingness is a problem in both observational data **and** clinical trials.
- Ignoring missing outcome data can bias estimates of causal effects.
- Typically, we can view censoring as another "treatment" decision and, if no time-varying confounding, we can easily modify G-computation, IPTW, and efficient approaches accordingly.
- If time-varying confounding, additional identification results and estimation strategies are necessary.

Missing covariates

In practice, covariates may also be missing for certain observations.

Consider the case of two covariates $W = (W_1, W_2)$, where W_2 is missing for some participants. So, in reality, we observe $W_* = (W_1, \Delta_2, \Delta_2 W_2)$.

Just as with missing outcomes, ignoring data with missing covariates could also cause bias in causal effect estimates.

We recommend this general strategy for handling missing covariates:

1. Does randomization holds given W_*?
   - If so, can proceed with standard approach with adjustment for W_1, Δ_2 and $\Delta_2 W_2$.
   - See Greenland & Finkle (1995) for limitations of this strategy.
2. Otherwise, does randomization hold for Δ_2 given W_1?
   - If so, treat missing indicator as intervention node, and use methods for multi time-point interventions. Intervention is on (Δ_2, A) in the sequence

$$W_1 \longrightarrow \Delta_2 \longrightarrow W_2 \longrightarrow A \longrightarrow Y \ .$$

3. Otherwise, assess which randomization assumption is more plausible, and refer to section on missing confounders.

Missing covariates

As an example, say W_2 represents a yes/no item on a health questionnaire to which some patients do not answer. In the observed data, W_2 can take values yes, no or NA.

Let $\Delta_2 = I(W_2 \neq \texttt{NA})$ and $\Delta_2 W_2 = \Delta_2 I(W_2 = \text{yes})$.

How can we determine whether $Y(a) \perp A \mid W_1, \Delta_2, \Delta_2 W_2$?

- Consider a stratum of patients who refuse to answer the questionnaire and have the same value of W_1.
- Reasonable to assume that treatment is randomized in this stratum?
- Or are there meaningful differences between people in this stratum who do and do not receive treatment?

If reasonable, then

$$E[Y(a)] = E[E(Y \mid A = a, W_1, \Delta_2, \Delta_2 W_2)]$$

and we can proceed using the standard methodology we have discussed.

Missing covariates

There are several options for estimating the outcome regression and propensity score. Considering the propensity score, how could we estimate $P(A = 1 \mid W_1, \Delta_2, \Delta_2 W_2)$?

The standard regression approach treats missingness as its own category, e.g.,

$$P(A = 1 \mid W_1, \Delta_2, \Delta_2 W_2) = \text{expit}(\alpha_0 + \alpha_1 W_1 + \alpha_2 \Delta_2 + \alpha_3 \Delta_2 W_2) \ .$$

Another option: we could "impute" missing values of W_2 (e.g., based on W_1) and fit a "full data" method on the imputed data. For example, we could

1. fit linear regression of W_2 on W_1 in subset of the data with $\Delta_2 = 1$;
2. fill in W_2 with predicted value from regression for observations with $\Delta_2 = 0$;
3. fit a random forest with outcome A and predictors W_1 and imputed W_2.

The choice of imputation technique can now be viewed as a tuning parameter for an algorithm. Let the super learner determine which is best!

Missing covariates

If there is residual confounding due to missingness in W_2, we can ask whether randomization holds for Δ_2 given W_1.

Can we explain the missingness in W_2 by the fully measured covariate W_1?

Example: We typically have some demographic information on all patients. Is this enough to explain why they did not answer questionnaire?

If so, then we might view Δ_2 as an intervention node and define our counterfactual of interest as $Y(1, a)$, the outcome we would observe if we forced $\Delta_2 = 1$ **and** $A = a$.

In such case, we have that:

- positivity requires positive probability of measuring W_2 for all strata of W_1;
- we require *sequential randomization*: Δ_2 randomized given W_1 **and** A randomized given $\Delta_2 = 1$, W_1 and W_2.
- Each of G-computation, IPTW, AIPTW and TMLE can be implemented in a longitudinal (i.e., multi time-point) framework – this will be covered in a two-day extension of this module next year!

Key points: missing covariates

- Ignoring missing covariate data can bias estimates of causal effects.
- Randomization may hold based on the observed missing data structure; however, this cannot be assessed using our observed data. Prior knowledge is key!
- If randomization fails to hold for missing data but we can predict missingness of covariates using available data, we can use approaches for multi time-point interventions.

Unmeasured confounding

In many observational studies, it is known that some confounders were not measured, and the ATE is therefore not identified by G-computation and IPTW formulas.

- Recall that this cannot generally be assessed just by inspecting the data.
- Be prepared to defend your interpretations against skeptics.

There are several ways to proceed in this situation:

1. modify interpretation of the results;
2. perform sensitivity analysis;
3. estimate bounds on causal effects;
4. explore alternative identification results.

Unmeasured confounding

Under appropriate conditions, we stated that the causal effect of interest was given by

$$E\left[E(Y|A=1,W)-E(Y|A=0,W)\right]$$

and that is what we sought to estimate.

However, we can also simply interpret this estimand as a covariate-adjusted marginal association, leaving aside claims of strict causality.

For example, in our BOLD analysis, we could have interpreted our estimand as

the difference in average disability score among patients with identical baseline covariate values who did versus did not receive early imaging, standardized to the baseline covariate distribution in the BOLD patient population.

A few comments:

- For this interpretation, no causal assumptions are required at all.
- This is arguably the closest we can get using the data to the true causal effect.
- Sensitivity analysis ties in scientific knowledge to help determine how close to the true effect we may be.

Unmeasured confounding

Sensitivity analysis is a commonly used term in associative analysis.

- How sensitive is a parametric model to including/removing a covariate, including/removing an interaction, etc?
- How well does this parametric model fit the observed data (e.g., model checking)?

For causal analysis, we take an alternative approach to sensitivity analysis.

- How much unmeasured confounding would there need to be in order to make an observed association disappear?

A rich literature exists on this topic (see additional reading). We provide an example from Díaz & van der Laan (2013).

- Chagas disease affects 8 million people in Latin America.
- Long incubation periods of the disease (up to 30 years) make clinical trials to study treatment efficacy prohibitively expensive.
- This also makes observational studies difficult as many participants are lost to follow-up over these long periods of time.
- Analysis combines data from 19 observational studies that did not measure participant-level confounders.

Unmeasured confounding

The observed data unit is $O := (A, \Delta, \Delta Y)$, where Y is a binary cure status and Δ indicates whether a patient's cure status was observed at the end of the study.

The causal parameter of interest is the average treatment effect amongst the treated,

$$ATT = E[Y(1) - Y(0)|A = 1] \ .$$

However, we know that sicker patients tend to receive treatment and also may drop out of studies more often due to their health. As such, ATT cannot be identified based on the observed data.

Nevertheless, consider estimating

$$ATT_{\text{obs}} = E(\Delta Y|A = 1) - E(\Delta Y|\Delta = 1, A = 0) \ .$$

Can we bound the difference between ATT and ATT_{obs} by a quantity that we understand scientifically?

Unmeasured confounding

The worst-case scenario for the treatment would be if all participants with missing outcomes who received treatment were not cured.

In this case, $E(\Delta Y|A=1)$ is a conservative gauge of $E[Y(1)|A=1]$, in the sense that

$$E(\Delta Y|A=1) \;\leqslant\; E[Y(1)|A=1]\;.$$

This gives us the inequality

$$\begin{aligned} ATT_{\text{obs}} - ATT &= \{E(\Delta Y|A=1) - E(\Delta Y|\Delta=1, A=0)\} \\ &\qquad - \{E[Y(1)|A=1] - E[Y(0)|A=1]\} \\ &\leqslant E[Y(0)|A=1] - E(\Delta Y|\Delta=1, A=0) \\ &\leqslant E[Y(0)|A=1]\;. \end{aligned}$$

The difference between ATT_{obs} and ATT is at most the probability of "spontaneous cure" amongst the treated, an interpretable quantity.

Unmeasured confounding

We can now study tests of $H_0 : ATT \leqslant 0$ by studying $H_0 : ATT_{\text{obs}} \leqslant \delta$ for different choices of δ, a hypothesized spontaneous cure probability.

We would reject H_0 at level $\alpha = 0.05$ for any value of $\delta \leqslant 0.44$.

- Because sicker patients receive treatment, it is likely that $E[Y(0)|A = 1] \leqslant E[Y(0)|A = 0]$.
- $\hat{E}(\Delta Y|\Delta = 1, A = 0) = 0.01$, so it is likely that $E(Y(0)|A = 1) \leqslant 0.44$.

An interpretable bound allows us to conclude that there is likely at least some causal effect of treatment, even with confounding present.

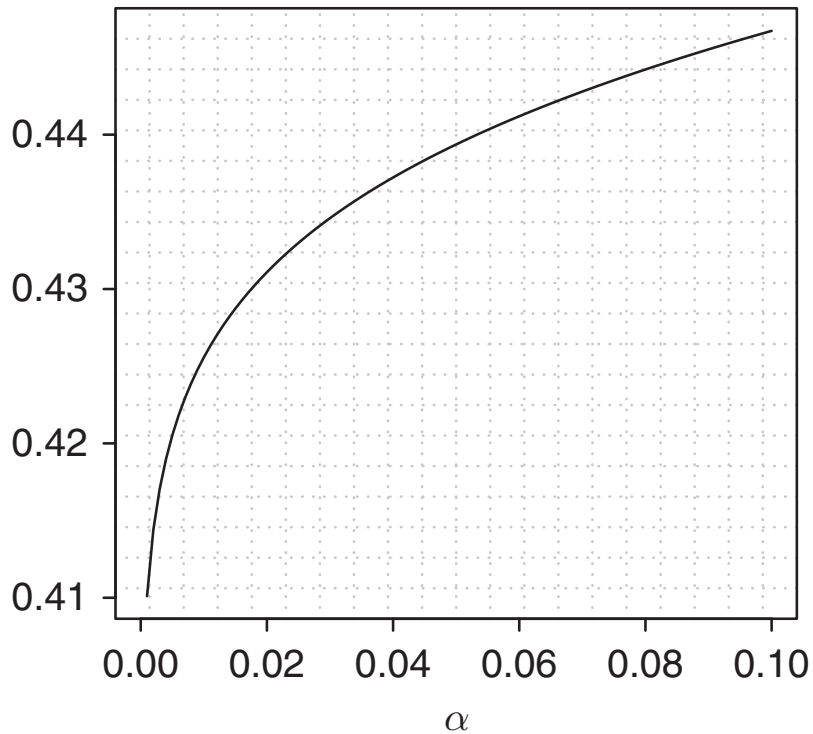


Figure: Choices of δ for which $H_0 : ATT \leqslant 0$ is rejected at different type I error probabilities (α)

Unmeasured confounding

This form of sensitivity analysis is closely related to approaches in the "partial identifiability" literature (Manski, 1990), wherein bounds on the difference between the causal effect and the observed association are derived.

Usually, very few (if any) assumptions are required to derive bounds, though this sometimes results in highly conservative bounds. For example, we may have generated bounds on the ATE that always include zero.

Interested? Take the Causal Inference module at the UW Summer Institute in Statistics and Modeling in Infectious Diseases.

Unmeasured confounding

In special situations, there may be alternative ways to identify causal effects.

Instrumental variables are a common tool to estimate causal effects with unmeasured confounding (Balke & Pearl, 1997).

- An instrument is a variable Z that is related to A but not to Y.
- Example: Encouragement to seek early imaging.

Under some conditions (Angrist, Imbens & Rubin, 1996), the average treatment effect is identified by

$$ATE = E\left\{ \frac{\overbrace{E(Y|Z=1,W) - E(Y|Z=0,W)}^{\text{conditional effect of } Z \text{ on } Y}}{\underbrace{P(A=1|Z=1,W) - P(A=1|Z=0,W)}_{\text{conditional effect of } Z \text{ on } A}} \right\}.$$

Key points: unmeasured confounding

- If you work with observational data (or even clinical trials data with missing outcomes!), you should worry about unmeasured confounding.
- Ideally, this is corrected at the design stage by collecting a rich set of covariates on trial participants.
- There are several possible approaches to tackle (to the extent possible) unmeasured confounding at the analysis stage.

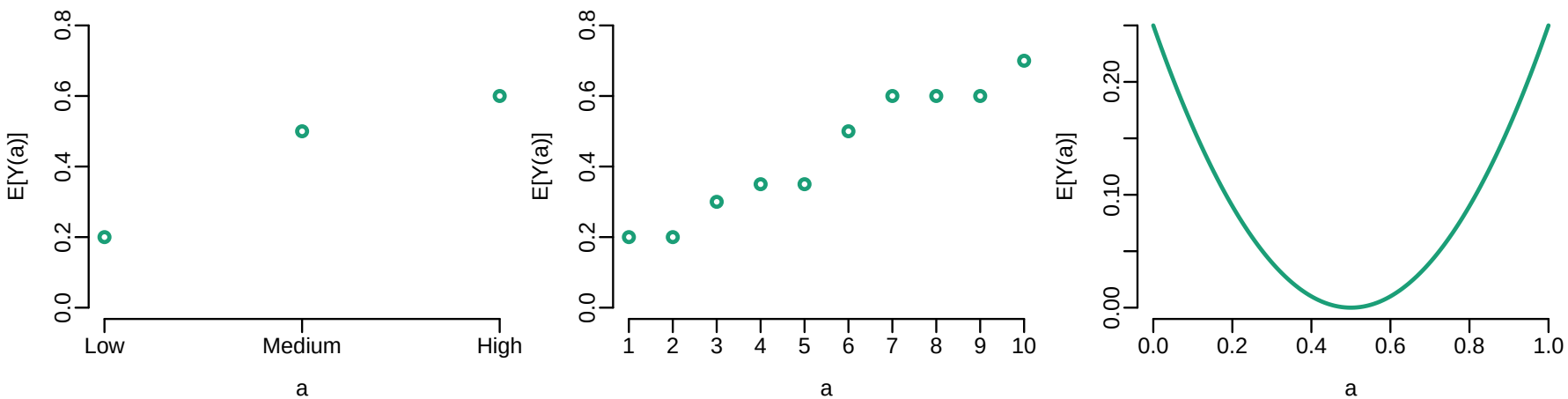
Continuous-valued treatments

In some cases, the treatment of interest A is not binary, but continuous rather.

- Examples: time of early imaging, dose of drug, level of exposure.

We are often interested in a counterfactual outcome under a range of interventions.

- What is the mean outcome if we intervene to set $A = a$ for various $a \in \mathcal{A}$?
- We use $\mathcal{A}$ to denote the set of interventions, which could take only a few values, many values, or an infinite number of values.



Continuous-valued treatments

If $\mathcal{A}$ includes only a few values, all the tools we have developed apply directly.

- We have tools to estimate $E[Y(a)]$ for any particular a.
- Test, e.g., the null hypothesis $H_0 : E[Y(\text{low})] = E[Y(\text{medium})] = E[Y(\text{high})]$.

For statistical reasons beyond the scope of this course, it is not possible to estimate a continuous dose-response curve without making strong assumptions.

- Heuristically, we simply do not observe enough patients with every intervention level we might care about.
- What can we do in this situation?

If $\mathcal{A}$ contains many values, we could still use a multiple degree-of-freedom test to test the null hypothesis that the mean is the same across all $\mathcal{A}$.

- Could we summarize the change in mean across a using a single parameter?

Continuous-valued treatments

We can describe trends using parameters of working models, also known as marginal structural models.

- "Working model" implies that we do not necessarily believe this model to be true.
- Be careful: often, marginal structural models are used in a parametric context in which they are assumed to be true (i.e., not simply working models).

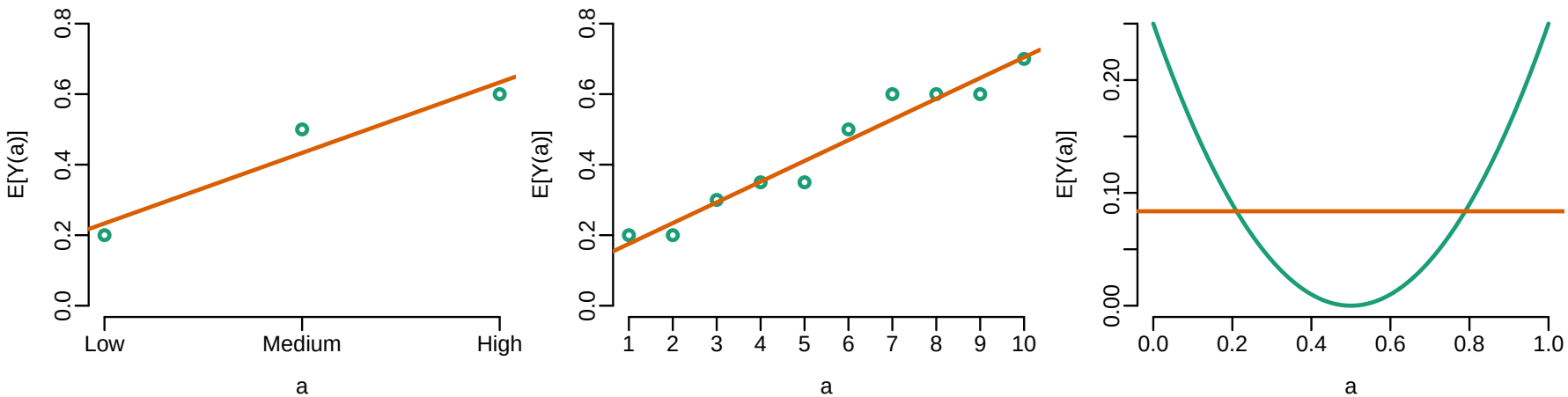
We choose working model $m(a, \beta)$ and define the target parameter as a projection onto this model (Neugebauer, van der Laan, 2007).

- working model: $m(a, \beta) = \beta_0 + \beta_1 a$;
- projection: $\beta_0 = \text{argmin}_\beta \sum_{a \in \mathcal{A}} \{E[Y(a)] - m(a, \beta)\}^2$;
- interpretation: the "best-fitting straight line" to the true causal dose-response curve. If $\beta_1 > 0$, then increasing a tends to cause increases in the average outcome.

Continuous-valued treatments

How interesting a given target parameter is depends on the underlying causal curve.

- Linear working models for curves that are monotone may be highly relevant...
- ...but if the curve is parabolic?



Continuous-valued treatments

In order to efficiently estimate parameters of marginal structural models, we need propensity scores $P(A = a \mid W)$ *for all* $a \in \mathcal{A}$.

- If A has multiple levels, we could use multinomial logistic regression.
- It is also possible to use sequential super learning.

The positivity assumption can be troublesome when treatments have many levels.

- We need positive probability of receiving *every* treatment in *every* strata.
- Choosing a weighted projection onto MSM can help alleviate these problems – this is akin to using 'stabilized weights':

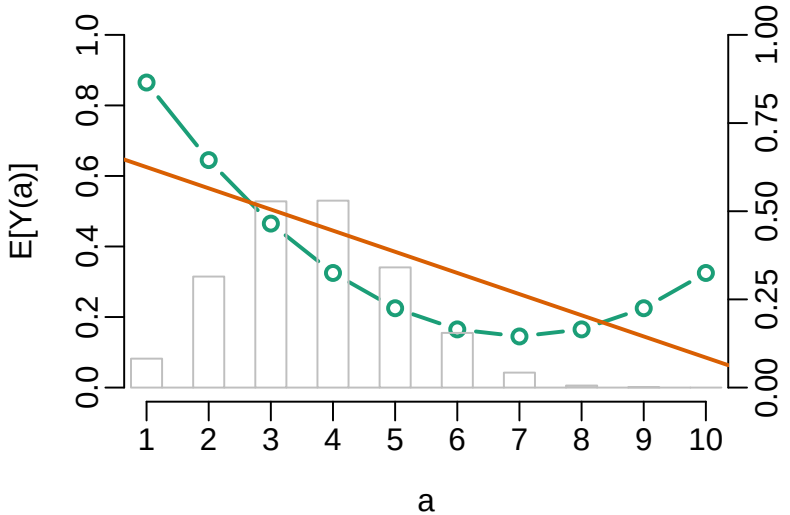
$$\beta_0 = \text{argmin}_\beta \sum_{a \in \mathcal{A}} w(a) \{E[Y(a)] - m(a, \beta)\}^2$$

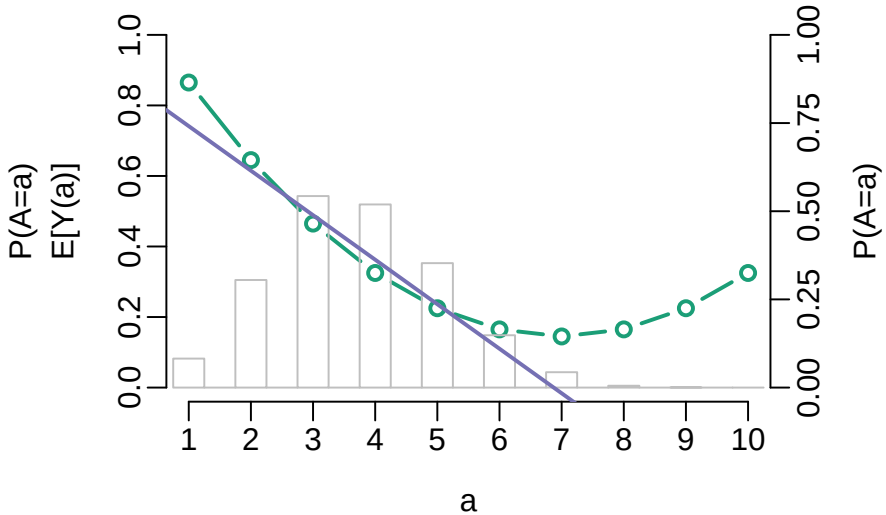
- If $w(a) = P(A = a)$ and a is rare, this level is down-weighted when choosing the best-fitting line.

Continuous-valued treatments

Compare the true value of the parameter for the unweighted projection to the weighted projection:

- The choice between the two must consider the tradeoff between scientific interest and statistical stability.
- The unweighted projection may be more interesting, but we may obtain more precise estimators of the weighted projection.





Key points: continuous-valued treatments

- If there are only several discrete categories of treatment, then standard techniques may be applied to each level separately.
- If many levels of ordered treatment, marginal structural models can be useful for summarizing causal dose response curves.
- Positivity issues can be handled by choosing appropriately weighted projections.
- Parameters based on other types of intervention can also be used to study the effect of a continuous treatment on an outcome.

Subgroup effects and interactions

The treatment of interest may have a different effect in different subpopulations.

- Example: Early imaging beneficial in younger patients? In older patients?

Describing whether and how the treatment's efficacy differs across subpopulations is often of interest. For example, what are the policy implications of a treatment with null average effect because:

- it results in worse outcomes in half the population, and better outcomes in half the population, or;
- it has no effect on anyone in the population?

The average treatment effect is not the "wrong" parameter to study, as many people believe. It is still describing the effect of the treatment in the population *on average*.

- G-computation: subgroup-specific treatment effect averaged over subgroups.
- No need for those subgroup-specific effects to be the same!

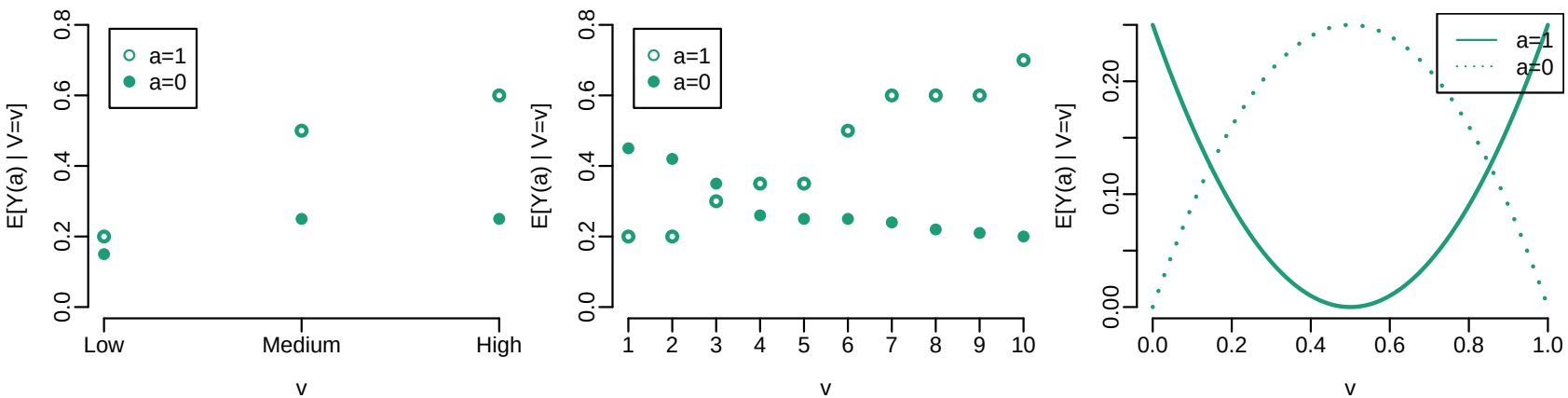
Subgroup effects and interactions

We use $\mathcal{V}$ to denote all the subgroups (defined by baseline covariates) in which we are interested clinically.

- Example: $\mathcal{V} = \{\text{older than 75}, \text{younger than 75}\}$
- Example: $\mathcal{V} = \{65\text{yo}, 66\text{yo}, \dots\}$
- Example: $\mathcal{V} = \{65\text{yo men}, 65\text{yo women}, 66\text{yo men}, 66\text{yo women}, \dots\}$

The counterfactual parameters of interest are $E[Y(a)|V = v]$ for $a \in \mathcal{A}$, $v \in \mathcal{V}$.

- For every subgroup v, what is the average outcome under each treatment a.
- We might test whether the average treatment effect is the same for patients across all subgroups.



Subgroup effects and interaction

Identification and estimation of the counterfactual mean in patients with $V = v$ when we assign treatment $A = a$ are straightforward modifications of previous techniques.

G-computation: $E[Y(a)|V = v] = E[E(Y|A = a, V, W)|V = v]$

Averaging is performed relative to distribution of W *only in* stratum $V = v$.

IPTW: $E[Y(a)|V = v] = E\left[\frac{I(A = a, V = v)}{g(v, W)} Y \middle| V = v\right]$

Reconstructing a population of patients with $A = a$ *only in* stratum $V = v$.

Using the observed data, we can estimate

$$\begin{aligned} \text{the outcome regression :} \quad & \bar{Q}(a, v, w) := E(Y \mid A = a, V = v, W = w) \\ \text{the propensity score :} \quad & g(v, w) := P(A = a \mid V = v, W = w) \quad . \end{aligned}$$

Subgroup effects and interactions

In this case, the G-computation estimator of $E[Y(1) \mid V = v]$ is given by

$$\psi_{n,G,1,v} := \frac{1}{\sum_{i=1}^{n} I(V_i = v)} \sum_{i=1}^{n} I(V_i = v) \bar{Q}_n(1, v, W_i) \ .$$

Even though, we can use all patients to fit regressions, we are computing the outer empirical mean only using patients with $V = v$.

- If there are few of these (e.g., v is a given age), this estimator is unstable.
- If $\mathcal{V}$ assumes an infinite number of values (e.g., v is a given BMI), then we cannot estimate this parameter without making strong assumptions.

We can again make use of nonparametric marginal structural models to summarize these interactions.

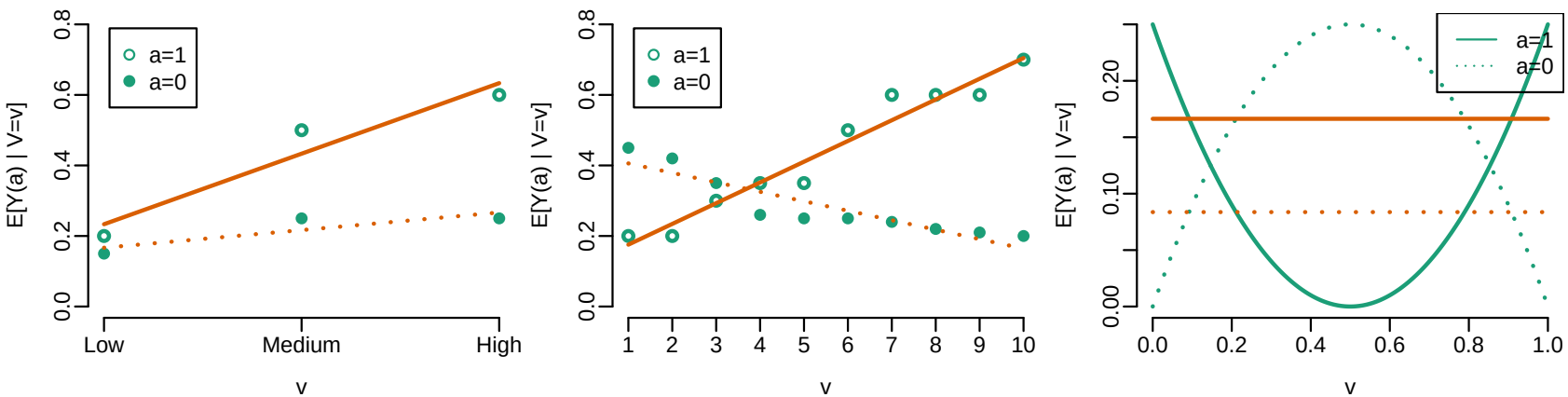
Subgroup effects and interactions

We choose working model $m(a, v, \beta)$ and define the target parameter as a projection onto this model.

- Working model: $m(a, v, \beta) = \beta_0 + \beta_1 a + \beta_2 av$;
- Weighted projection:

$$\beta_0 = \text{argmin}_\beta \sum_{a \in \mathcal{A}} \sum_{v \in \mathcal{A}} w(a, v)\{E[Y(a)|V = v] - m(a, v, \beta)\}^2;$$

- Interpretation: The "best-fitting working model" to the true causal dose-response curve across different levels of v. If $\beta_2 > 0$, then there is a trend suggestive of increasing average treatment effect for increasing v.



Key points: subgroup effects and interactions

- Average treatment effect is often relevant even when interactions are present.
- Straightforward extensions of previous results allow us to estimate the average treatment effect within particular strata.
- Comparison of the average treatment effect across these strata constitutes testing for an interaction.
- Marginal structural models may be helpful (often necessary) to estimate interactions between treatment and many subgroups (e.g., defined by a continuous variable).

References and additional reading

References:

Angrist J, Imbens G, Rubin D (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444-455. doi: 10.2307/2291629

Balke A, Pearl J (1997). Bounds on Treatment Effects from Studies with Imperfect Compliance. *JASA*; 92(439): 1171–76. doi: 10.1080/01621459.1997.10474074.

Díaz I, van der Laan M (2013). Sensitivity Analysis for Causal Inference under Unmeasured Confounding and Measurement Error Problems. *Int J Biost*; 9(2): 149–160. doi: 10.1515/ijb-2013-0004.

Greenland S, Finkle W. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses (1995). *Am J Epidemiol*; 142 (12): 1255-1264. doi: 10.1093/oxfordjournals.aje.a117592.

Manski C (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review: Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association*; 80(2): 319–323. JSTOR: http://www.jstor.org/stable/20065922006592.

Neugebauer R, van der Laan M (2007). Nonparametric causal effects based on marginal structural models. *J Stat Plan and Inf*; 137(2):419-434. doi: 10.1016/j.jspi.2005.12.008

References and additional reading

Additional reading:

Ertafaie A, Small D, Flory J, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharma and Drug Safety*; 26(4): 357–67. 10.1002/pds.4158.

Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M (2014). Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *J Causal Inference* ; 2:147185. doi: 10.1515/jci-2013-0007.

Rosenblum, M (2011). *Targeted Learning: Causal Inference for Observation and Experimental Data*; Chapter 9: 145-160. Springer New York. doi: 10.1007/978-1-4419-9782-1.

Robins J, Hernán M, Brumback B (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*; 11(5): 550–60. doi: 10.1097/00001648-200009000-00011.

Robins J, Rotnitzky A, Scharfstein D. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. *Statistical Models in Epidemiology, the Environment and Clinical Trials*; 1–94. Springer New York. doi: 10.1007/978-1-4612-1284-3.

Rotnitzky A, Scharfstein D, Su S, Robins J (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. Biometrics; 57:10313. doi: 10.1111/j.0006-341X.2001.00103.x.

Zheng W, Petersen M, van der Laan M. Doubly Robust and Efficient Estimation of Marginal Structural Models for the Hazard Function. *Int J Biostat*: 12(1); 233–52. doi: 10.1515/ijb-2015-0036.