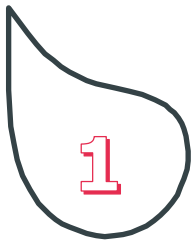
Causal Inference

June 23 2019

Contents



Overview. G-formula and inverse probability weighting - Miguel Hernan

Terminology

- Upper-case letters (e.g. A, Y): random variables
- Lower-case letters (e.g. a, y): possible values of random variables (i.e. fixed to individual)

- Potential outcomes = counterfactual outcomes: the situation would have been observed under A=a

- Consistency: If $A_i = a$, then $Y_i^a = Y_i^A = Y_i$ (same as Rubin's 'stable-unit-treatment-value assumption(SUTVA)')

- $\boldsymbol{Pr[Y^{a=1} = 1]}$: proportion of indivi. Y=1, had everybody been treated <- unconditional probability

$\boldsymbol{Pr[Y = 1 \,|A = 1]}$: proportion of indivi. Y=1 among who received treatment <- conditional probability

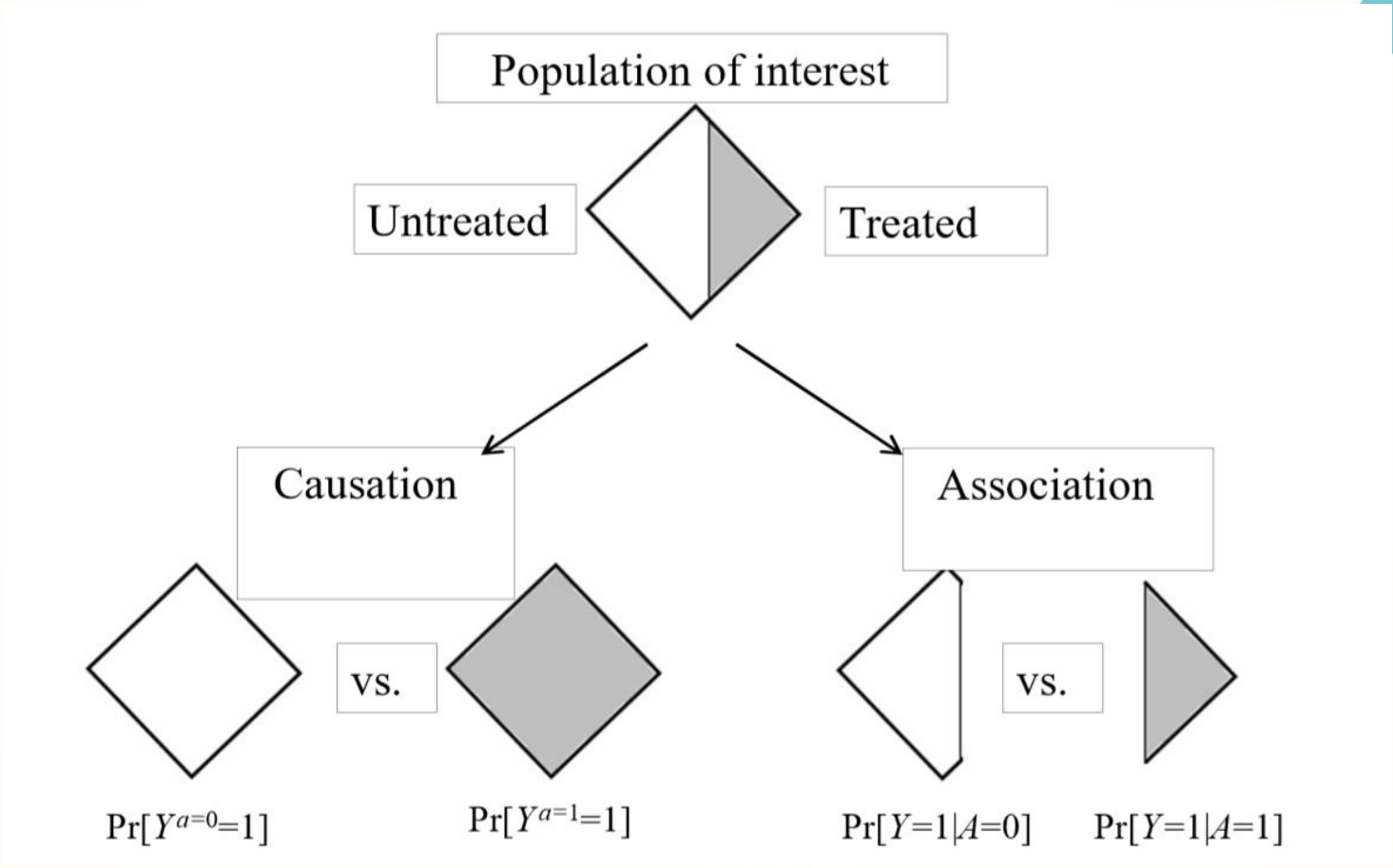
Association and Causation

- Fundamental problem of causal inference
  - Individual causal effects cannot be determined (= missing data problem)

- Average causal effects of A on Y in the population exist if $\boldsymbol{Pr\left[Y^{a=1}=1\right] \neq Pr\left[Y^{a=0}=1\right]}$
- Association between A and Y in the population exist if $\boldsymbol{Pr[Y=1 \mid A=1] \neq Pr[Y=1 \mid A=0]}$
  - If there is no association, they are independent $A \coprod Y$
- If outcome is nondichotomous, $\mathbf{E}\left[Y^{a=1}\right]$(population mean or expectation) can be substituted

- There is confounding when $\boldsymbol{Pr[Y^{a}=1] \neq Pr[Y=1 \mid A=a]}$

Definition of causal effects



Conditions for causal inference

- Ideal randomized experiment
  - No loss to follow-up
  - Full compliance with assigned treatment
  - One version of treatment (well-defined treatment)
  - Double blind assigment (neither subjects nor investigators know)

1. Exchangeability <- marginal randomization
   - $\boldsymbol{Pr[Y^a = 1 \,|A = 1] = Pr[Y^a = 1 \,|A = 0] \Leftrightarrow A[\!]Y^a \; for \; all \; a \; (\neq A[\!]Y)}$
   - Lack of confounding
   - If not holds, we need counfouding adjustment
2. Positivity (explained later)
   - No treatment that no one take
3. Consistency (explained later)
   - SUTVA

Case study 1

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…

Exchangeability…

- Holds in mariginally randomized experiments => no need of confounding adjustment : **association is causation**
- Nonparametric estimation with saturated linear model $E[Y|A] = \theta_0 + \theta_1 A$
  - If not treated, i.e. A=0, $E[Y|A = 0] = \theta_0 + \theta_1 \times 0 = \theta_0$
  - If treated, i.e. A=1, $E[Y|A = 1] = \theta_0 + \theta_1 \times 1 = \theta_0 + \theta_1$
  - Then, average effects estimate is $(\theta_0 + \theta_1) - \theta_0 = \theta_1$ (achievable nonparametrically)

- Not hold in conditionally randomized experiments, observational study => need of confounding adjustment
- What if randomized conditioning on L?
  - Need to estimate $E[Y^{a=1}]$ and $E[Y^{a=0}]$, i.e. The standardized mean in the treated and in the untreated

Standardization

$$E[Y^{a=1}] = \mathrm{E}[Y^{a=1}|L=1] \times \Pr[L=1] + \mathrm{E}[Y^{a=1}|L=0] \times \Pr[L=0]$$
$$= \mathrm{E}[Y|L=1, A=1] \times \Pr[L=1] + \mathrm{E}[Y|L=0, A=1] \times \Pr[L=0]$$
$$= \sum_{l=1,0} E[Y|L=l, A=1] \times \Pr[L=l]$$

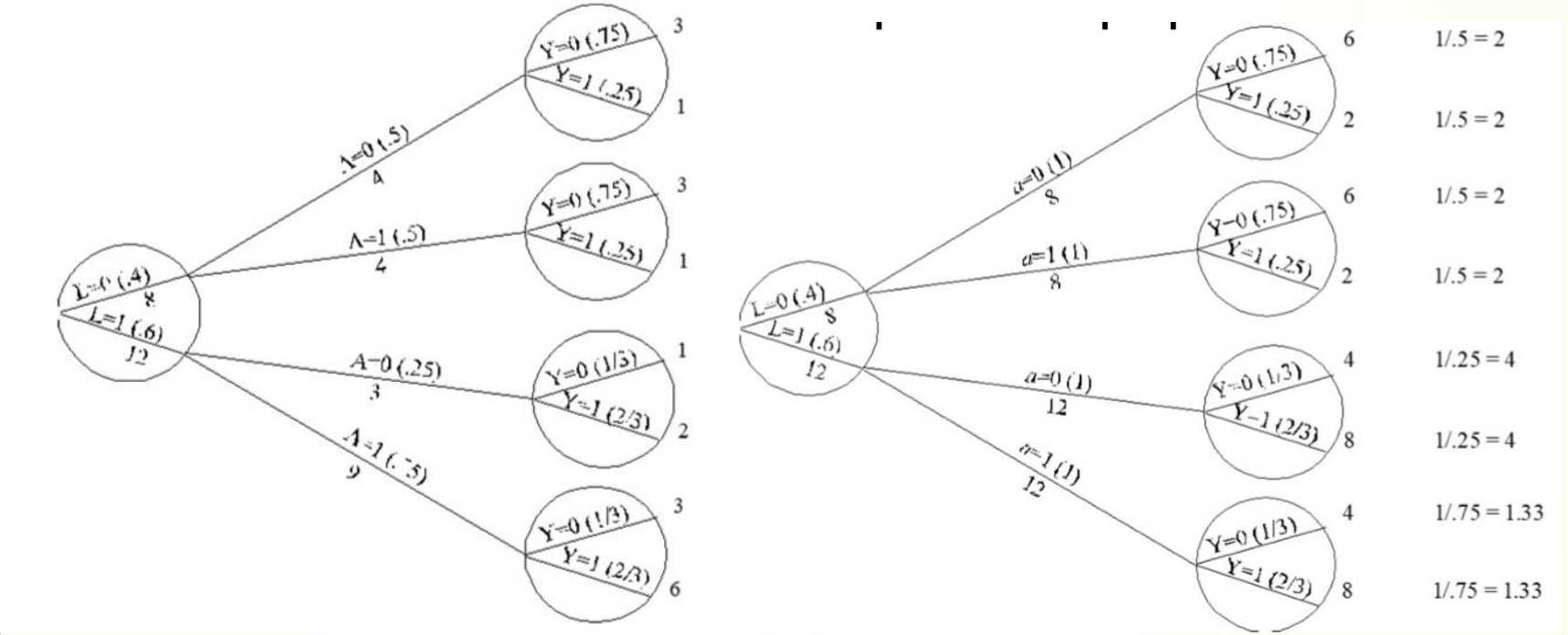
Exchangeability

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

- Then, we can estimate the average causal effect
- Without model (refer above)
- With model (bias-variance trade-off)
  - Nonparametric estimation = saturated linear model $E[Y|A,L] = \theta_0 + \theta_1 A + \theta_2 L + \theta_1 A\mathrm{L}$
    : allow difference in treatment effect between L
    : the curse of dimensionality
  - Parametric estimation = nonsaturated linear model $E[Y|A,L] = \theta_0 + \theta_1 A + \theta_2 L$
    : restriction that treatment effect is the same
    : variance become larger

Inverse probability(IP) weighting

- IP weighting is anther methods to gain the average causal effects under conditional exchangeability, i.e. adjustment



The goal is to make the pseudo-population with weighting

The way of IP weighting and interpretation

- IP weights: $W^A = \frac{1}{f(A|L)}$
  - $f(a)$ is the probability density function (pdf) of the random variable A evaluated at the value a
- Stabilized IP weights: $SW^A = \frac{f(A)}{f(A|L)}$
  - get the same size of pseudo-population as the original one
  - Lead to smaller variance (by avoiding weighting too much on small number of people)
- Generalized IP weights: $GW^A = \frac{g(A)}{f(A|L)}$

Violations of positivity

- Structural
  - No causal inferences for subsets w/ structural non-positivity
  - Causal inference by restricting the study population
- Random
  - Causal inference with parametric models to smooth over the subsets w/ non-positivity

- …So,
- Standardization vs IP weighting ?
  - In nonparametric estimate, same outcome is acquired
  - In parametric estimate, the outcome is different
  - Recommend doubly-robust methods

Suppose we want to estimate the causal effect of A on Y…

- If everybody had been treated: $E[Y^{a=1}]$
- If everybody had been untreated: $E[Y^{a=0}]$
- Then, average causal effect: $E[Y^{a=1}]$ - $E[Y^{a=0}]$

- Weighted regression model $E[Y|A] = \theta_0 + \theta_1 A$
  - Associational model
  - The difference, i.e. θ_1 would have a causal interpretation as $E[Y^{a=1}]$ - $E[Y^{a=0}]$ when **all confounders are included** in calculation (e.g. Estimate ini the pseudo-population)
- Marginal Structural Model (MSM) $E[Y^a] = \beta_0 + \beta_1 a$
  - Causal model
  - $\theta_1 = \beta_1$ if θ_1 would have a causal interpretation

Marginal structural model (MSM)

- Structural?
  - Structural = causal
  - The outcome variable is counterfactual
  - i.e. parameters for treatment in MSM have direct causal interpretation
- Marginal?
  - Marginal = unconditional
  - No need to include the confounders as covariates in the model
  - i.e. effect may be estimated in the entire population
  - If include covariates, allow to estimate conditional effect

- Again, recommend doubly-robust estimators

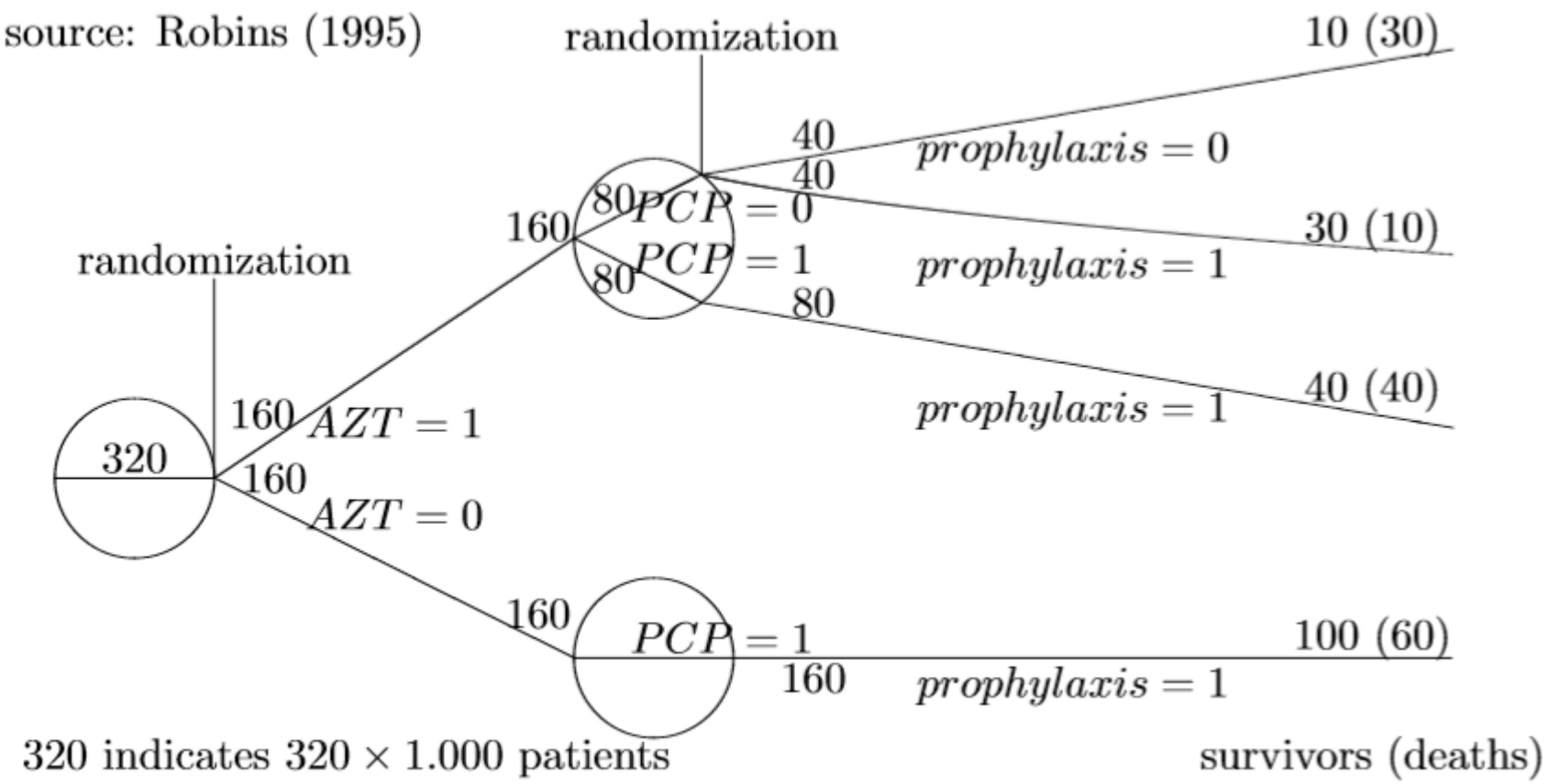
Advantage of MSM with IP weighting or standardization

- **IP weighting and standardization** are able to control time-varying treatment and confounders
  - Can handle treatment-confounder feedback

- **Outcome regression and propensity score methods** introduce bias if treatments and confounders are time-variant
  - Can not handle treatment-confounder feedback

Marginal Structural Models and G-estimation of Structural Nested Models
- Judith Lok

Case study 2



What is the causal effect of AZT on mortality?

- AZT (Zidovudine, Retrovir) : anti-HIV drug
- PCP : Pneumocystis pneumonia
- Prophylaxis : anti-HIV drug

Taking account of the treatment regime

- Treatment Regime is
  - AZT was followed by “no prophylaxis” if no PCP
  - AZT was followed by “prophylaxis” if PCP

- Intention-to-treat assessment?
- Conditioning on PCP (might lead to selection bias)?
- Also, positivity violation?
- Conditioning on prophylaxis?

-> important to talk with subject-matter experts

MSM conditioning on past treatment

- Under the assumption that
  - Independent, identically distributed full data (i.i.d.)
  - No unmeasured confounding
  - Missing At Random(MAR) : censoring depend on past observed characteristics but not on forther prognosis
  - Positivity

- MSM investigate the effect of "static" treatment regimes
  - Meaning treatment would not be patient-specific or be affected by previous outcomes
  - IP was the probability of receiving actual treatment for each patient, i.e. $P(A_k = a_k|\overline{L_k}, \overline{A_{k-1}})$

How to model multistate?

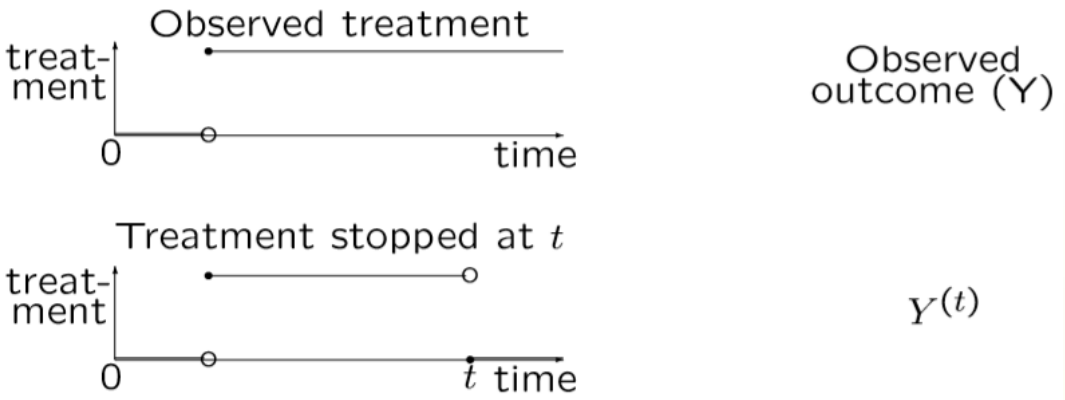
- Problem is
  - Computationally involved
  - Necessity to model each transition given the past
  - No specific parameter to indicate whether treatment affects the outcome of interest -> no standard test for treatment effect

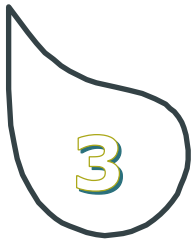
⇒ Structural Nested Mean Model and Structural Failure Time Models

  - To estimate the effect of treatment on the final outcome

Structural Nested Mean Model (SNMM)

- Assumptions
  - No unmeasured confounding
  - Consistency
- Treatment effects, or difference bw observed outcomes and counterfactual is defined as,
  - $\gamma_k(\overline{l_k}, \overline{a_k}) = E[Y^{(\overline{a_k},\overline{0})} - Y^{(\overline{a_{k-1}},\overline{0})} | \overline{L_k} = \overline{l_k}, \overline{A_k} = \overline{a_k}]$
  - The outcome had treatment stopped at k+1 versus at k

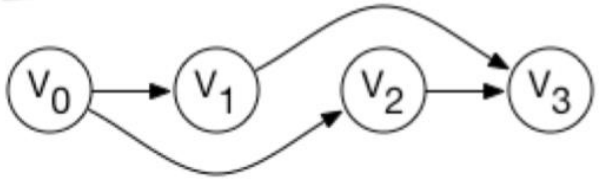


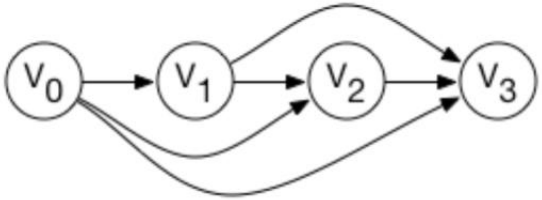


Single World Intervention Graphs and other recent developments in causal inference
- James Robins

Directed Acyclic Graphs (DAGs)

- Whose nodes (vertices) are random variables with directed edges (arrows) and no directed cycles
- Parents of variables (e.g. $PA_1 = V_0$)
- A path is closed if it contains a collider, otherwise path is open
- Complete DAG
  - There is an arrow between every pair of nodes
  - $f(v) = f(v_3|v_1, v_2)f(v_2|v_0)f(v_1|v_0)f(v_0)$
  - Nonparametric (saturated) model
- Incomplete DAG
  - $f(v) = f(v_3|v_0, v_1, v_2)f(v_3|v_0, v_1)\, f(v_1|v_0)f(v_0)$





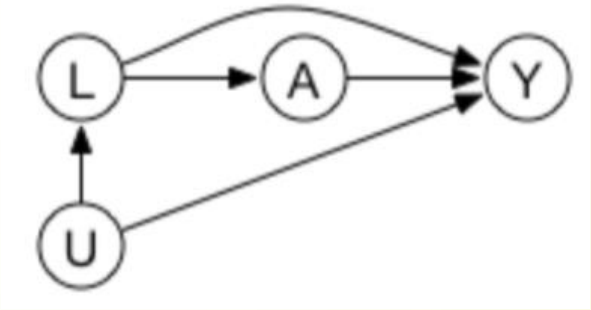
d-separation and d-connected

- D-separation
  - When no open path between two variables along which probability can flow, we call two variables are d-separated
  - Otherwise, we call they are d-connected
  - We also think d-separation and d-connected with condition

- If two sets of nodes are d-separated, they will be independent in every distribution in DAG (soundness)
- If two sets of nodes are not d-separated, there will be some distribution that they are not independent in DAG (completeness)

Causal DAGs

1. Lack of arrows = the absence of direct causal effects
2. Any variables are causes of all its descendants (vise versa)
3. All common causes must be on the graph even if they are not measured
4. Causal Markov Assumption (CMA) is hold: the causal DAG = a statistical DAG = distribution of factors
5. CMA = conditional on its direct causes, a variable is independent of any variable it does not cause
   - d-separation implies statistical independence
   - d-connection does not imply statistical dependence (but generally we assume dependence)



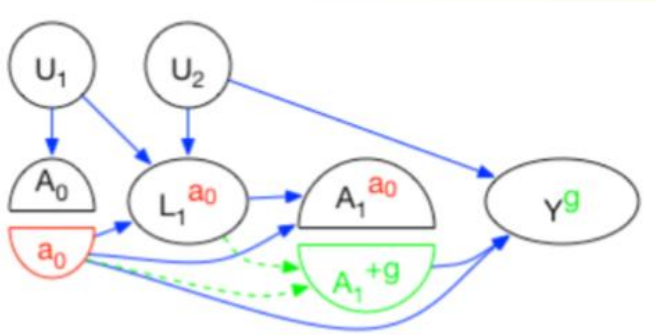
Single-World Intervention Graphs (SWIGs)

- The way to represent counterfactuals on the graphs
- SWIG G(0) represents $\Pr(A, Y^{a=0})$
- SWIG G(1) represents $\Pr(A, Y^{a=1})$

- Since we cannot show $Y^{a=0}$ and $Y^{a=1}$ on the same SWIG, the name Single-World Intervention Graphs is appropriate

SWIGs for dynamic regimes

- The treatment at time t is determined by g_t
- Under the regime $g = (a_0, L_1)$
  - $A_0^{+g} = a_0$
  - $A_1^{+g} = g_1(L_1^{a0}) = L_1^{a0}$
- For any regime g, static or dynamic, the g-formula will identify the counterfactual outcome if :
  - $Y^g \coprod A_0$
  - $Y^g \coprod A_1 | L_1, A_0 = a_0$
- And, the g-formula will have a causal interpretation
  - $E[Y^g] \sum_{l_1} E[Y|A_1 = l_1, L_1 = l_1, A_0 = a_0] \Pr[L_1 = l_1 | A_0 = a_0]$



We do not consider the treatment A is d-co

Causal Mediation Analysis - Tyler VanderWeele

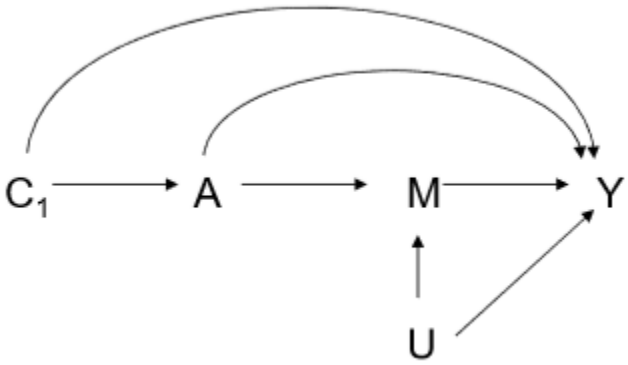
Standard approach to investigate mediation

1. Difference method
   - Using the difference between the two coefficient
   - $E[Y|A=a, C=c] = \phi_0 + \phi_1 a + \phi_2 c$
   - $E[Y|A=a, M=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4 c$
   - Indirect effect = $\phi_1 - \theta_1$
   - Direct effect = θ_1
2. Product method
   - $E[Y|A=a, C=c] = \beta_0 + \beta_1 a + \beta_2 c$
   - $E[Y|A=a, M=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4 c$
   - Indirect effect = $\beta_1 \theta_1$
   - Direct effect = θ_1

- Product method and difference method
  - coincide for continuous outcomes
  - Will not coincide for binary outcomes

Limitation1: Mediator-outcome confounding

- Just as unmeasured exposure-outcome confounders can generate confounding bias of estimates of overall effects, mediator-outcome confounders can generate bias of estimates of direct and indirect effects
- Meaning, we might get paradoxical result!
- Approach 1) pay attention to mediator-outcome confounding variables even during study design stage
- Approach 2) conduct sensitivity analysis



Limitation2:exposure-mediator interactions

- Even if we include an interaction term, often analysis goes:
  - $E[Y|A=a, C=c] = \phi_0 + \phi_1 a + \phi_2 c$
  - $E[Y|A=a, M=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c$
  - Indirect effect = $\phi_1 - \theta_1$
  - Direct effect = θ_1

- Approach 1) consider the causal definitions f direct and indirect effects for mediation analysis and required unmeasured confounding assumptions
- Approach 2) describe the regression methods which can be used in accord with the above definition
- Approach 3) provide sensitivity analysis techniques to assess the possible violations to unmeasured confounding assumptions

Approach 1) Definitions

- Controlled direct effect:

$$CDE|m = Y(A = 1|M = m) - Y(A = 0|M = m)$$
$$E[CDE|m] = E[Y|A = 1, m] - E[Y|A = 0, m]$$

- Natural direct effect:

$$NDE = Y(A = 1|M = M_0) - Y(A = 0|M = M_0)$$
$$E[NDE] = \sum_m \{E[Y|A = 1, m] - E[Y|A = 0, m]\}\Pr(M = m|A = 0)$$

- Natural indirect effect:

$$NIE|m = Y(M = M_1|A = 1) - Y(M = M_0|A = 1)$$
$$E[NIE] = \sum_m E[Y|A = 1, m]\{\Pr(M = m|A = 1) - \Pr(M = m|A = 0)\}$$

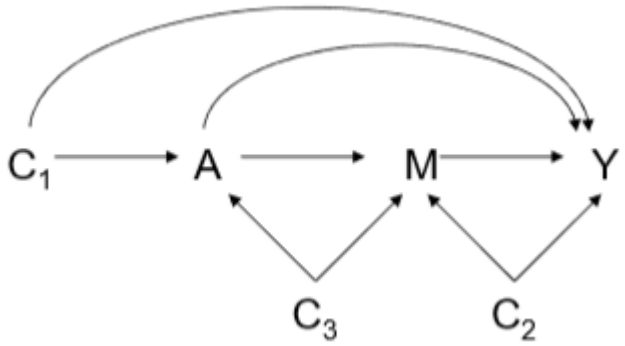
- Total effect:

$$Y_1 - Y_0 = NIE + NDE$$

- No presuppose that no interactions between the exposure and mediator

Approach 1) no unmeasured confounder assumption

1. No unmeasured exposure-outcome confounders given C
2. No unmeasured mediator-outcome confounders given (C,A)
3. No unmeasured exposure-mediator confounders given C
4. No unmeasured mediator-outcome confounder affected by exposure



Approach 2) Regression model for causal mediation analysis

- Similar concepts apply to treatment levels A=a to A=a*, then get the expression of regression
  - $E[Y|A=a, M=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c$
  - $E[M|A=a, C=c] = \beta_0 + \beta_1 a + \beta_2 c$

  - $CDE = (\theta_1 + \theta_3 m)(a - a*)$
  - $NDE = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a + \beta_2 E[C])\}(a - a*)$
  - $NIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a*)$

- SE can be obtained using the delta
- Proportion mediated is the indirect effect divided by the total effect (SAS, STATA and SPSS can do automatically for continuous, binary, count, and time-to-event outcomes

Approach 2) cautions for binary outcomes

- Difference method for a dichotomous outcomes and logistic regression will give valid estimates, provided
  - Model without the interaction is correctly specified
  - No unmeasured confounding assumptions are satisfied
  - Outcome is rare (can be relaxed by using log-linear)

- With common outcome, the difference method fails with logistic regression due to non-collapsibility

- Monte Carlo approach, a simulation-based approach give more flexibility

Approach 3) sensitivity analysis

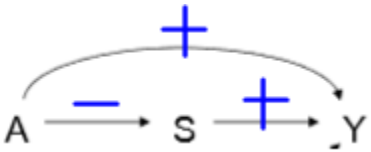
- In order to examine the extent to which the unmeasured confounder would have to affect both the mediator and the outcome to invalidate conclusions about NDE and NIE

- With an observed NDE or NIE of RR, we have unmeasured confounder if RR_{UY} and $RR_{AU|M}$ are greater than:
  - $E-value = RR + sqrt[RR \times (RR-1)]$
  - $RR_{UT}|A=1,m = \max \frac{\max \Pr(Y=1|A=1,m)}{\min \Pr(Y=1|A=1,m)}$: the max effect among exposed of U on Y, not through M
  - $RR_{AU}|m = max \frac{\Pr(u|A=1,m)}{\Pr(u|A=0,m)}$: smaller in magnitude on the RR scale than magnitude of max effect U on M across strata of A

  - We can apply this in a routine manner to both the estimate and the confidence interval limit closest to the null

Surrogate paradox

- "surrogate paradox" is manifest if
  - The surrogate and outcome are strongly positively correlated
  - The treatment has a positive effect on the surrogate
  - The treatment has a negative effect on the outcome



- Might happen if
  - E[Y|a,s,u] is NOT non-decreasing in a, i.e. a negative direct effect of A on Y, OR
  - E[Y|a,su] is NOT non-decreasing in s, i.e. the positive correlation of S and Y is not because of the actual effect but because of confounding, OR
  - P(S>s|a,u) is NOT non-decreasing in a, i.e. transitivity fails; A affects S for different people than S affects Y

Unification of Mediation and Interaction

- Assess mediation in the presence of interaction to get direct and indirect effects
- Under the composition assumption that $Y_a = Y_{aMa}$, total effects can be decomposed into four components
  - $Y_1 - Y_0 = (Y_{10} - Y_{00}) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_0) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1 - M_0) + (Y_{10} - Y_{00})(M_1 - M_0)$
  - CDE : effect of A in the absence of M
  - INTref : interaction that operates only if the mediator is present in the absence of exposure
  - INTmed : interaction that operates only if the exposure changes the mediator
  - PIE : effect of the mediator in the absence of the exposure times the effect of the exposure on the mediator

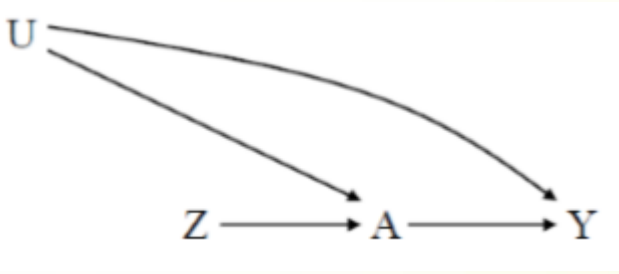
CAUSAL METHODS TO TAME UNMEASURED CONFOUDING - ERIC TCHETGEN TCHETGEN

Instrumental variable (IV)

- IV approach refers to a particular set of methods that allow one to recover a causal effect of an exposure in the presence of unmeasured confounding
- Key assumption : one has observed a pre-exposure unconfounded IV, which affects the outcome only through its effects on the exposure

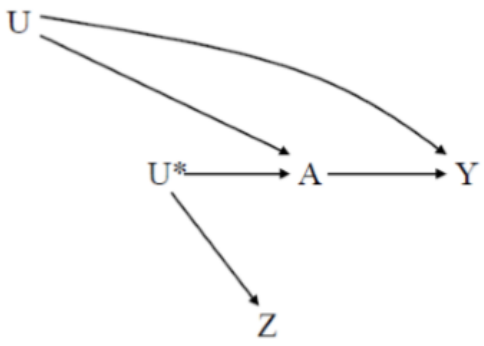
$$Y \coprod Z | U, A$$

- Mendelian randomization is also kind of instrumental variable approach



Formal definition of IV

- Assumption 1) Z and A are associated, Z has a causal effect on A, or Z and A share common causes
- Assumption 2) Z affects the outcome Y only through A, i.e. no direct effect of Z on Y (exclusion restriction)
- Assumption 3) Z does not share common causes with the outcome Y, i.e. no confounding for the effect of Z on Y
- Assumption 4) Monotonicity : there are no defiers, that is, there is only never takers ($A_0 = 0, A_1 = 0$), always takers ($A_0 = 1, A_1 = 1$), compliers ($A_0 = 0, A_1 = 1$), but defiers ($A_0 = 1, A_1 = 0$),



Complier average causal effect(CACE)

- The causal effect for individuals who would adhere to their assignment
- OR, the effect for individuals for whom treatment is manipulable
- Instrumental variable estimand $\beta_A = \frac{\beta_Z}{\alpha_Z} = \frac{causal\ effect\ of\ Z\ on\ Y}{causal\ effect\ of\ Z\ on\ A}$
- Wald estimand
  - $\frac{effect\ of\ randomization\ on\ Y = ITT\ effect}{effect\ of\ randomization\ on\ compliance} = \frac{E(Y|Z=1) - E(Y|Z=0)}{\Pr(A=1|Z=1) - \Pr(A=1|Z=0)}$
- CACE
  - $CACE = E[(Y_1 - Y_0)|A_1 > A_0] = \frac{E(Y|Z=1) - E(Y|Z=0)}{\Pr(A=1|Z=1) - \Pr(A=1|Z=0)}$
  - However, both counterfactuals are never observed for a person, thus compliers are not identified

Effect of treatment on the treated (ETT)

- Alternative assumption 4) no current treatment value interaction
- The advantage is that it does not require the monotonicity assumption, BUT requires ruling out the possibility of effect heterogeneity of the effect of A in the treated
- $ETT = \frac{E(Y|Z=1) - E(Y|Z=0)}{\Pr(A=1|Z=1) - \Pr(A=1|Z=0)} = E[(Y_1 - Y_0)|A=1]$

Average Causal Effect (ACE)

- Required extra assumption 5) homogeneity assumption
- $ACE = E(Y_{a=1} - Y_{a=0}) = \frac{E(Y|Z=1) - E(Y|Z=0)}{\Pr(A=1|Z=1) - \Pr(A=1|Z=0)}$

Covariates

- In the IV approach, we must account for covariates to
  - Account for confounding of the effects of Z on Y, and preventing a violation of the exclusion restriction by C
  - Partially account for confounding of the effects of A on Y
  - Explain variation in the outcome Y to improve efficiency

Two stage least square (2SLS)

- Another methods to estimate the causal effect instead of IV estimand or wald estimand

- Stage 1: fit a linear regression of A on Z and C, and compute the predicted value

$$\hat{A} = \hat{E}(A|Z, C) = \widehat{\alpha_0} + \widehat{\alpha_1} Z + \widehat{\alpha_2} C$$

- Stage 2:fit a linear regression of Y on predicted A and C

$$E\left(Y\middle|\hat{A}, C\right) = \mu_0 + \mu_1 \hat{A} + \mu_2 C$$

Reference

Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press. https://www.amazon.com/Explanation-Causal-Inference-Mediation-Interaction/dp/0199325871

Thank you for your attention!