# Time-Varying Exposures and Marginal Structural Models

Tyler J. VanderWeele
Departments of Epidemiology and Biostatistics
Harvard T.H. Chan School of Public Health

# Plan of Presentation

(1) Regression and Causation
(2) Study Design and Causal Inference
(3) Challenges with Time-Varying Exposures
(4) Marginal Structural Models
(5) Empirical Examples
(6) Brief Introduction to Structural Mean Models
(7) Brief Introduction to the Parametric G-formula

# Regression and Causation

Regression and Association: If the linear regression is correctly specified, but we do not have all confounding variables in the model, regression coefficients do not have a causal interpretation but do have an associational interpretation:

$E[Y|A,C] = \beta_0 + \beta_1 A + \beta_2' C$

i.e. If we randomly select two individuals from a population and both have the same value of C but the second individual has a value of A one unit higher than the first then, on average, the second individual will have a value of Y which is $\beta_1$ units higher

Again, this is true even if there are unmeasured confounders which are not in the model

# Regression and Causation

Regression and Causation: For regression coefficients to have a **causal** interpretation we need both that the linear regression be correctly specified and that all confounders of, e.g., the relationship between treatment A and Y be in the model.

$E[Y|A,C] = \beta_0 + \beta_1 A + \beta_2' C$

If $Y_a \perp\!\!\!\perp A \mid C$ then:

$E[Y_1|C=c] - E[Y_0|C=c] = \beta_1$

i.e. intervening to increase A by one unit will, on average, increase Y by $\beta_1$ units, conditional on covariates

# Marginal Structural Models

If we do have no unmeasured confounding so that $Y_a \perp\!\!\!\perp A \mid C$ then instead of controlling confounding by regression we can control confounding by use of "inverse probability of treatment weighting" (Robins et al., 2000)

This technique estimates the parameters of what is often called a "marginal structural model"

A marginal structural model is a model for the counterfactual outcomes:

$E[Y_a] = \beta_0 + \beta_1 a$

The model is "marginal" (rather than conditional - we are not conditioning on C)

The model is "structural" (it is for the counterfactual outcomes)

Because it is a model for counterfactuals we cannot fit it directly

# Marginal Structural Models

But under the assumption of no unmeasured confounding that
$Y_a \perp\!\!\!\perp A \mid C$ we can fit the model using a weighting technique

To estimate the parameters of the MSM: $E[Y_a] = \beta_0 + \beta_1 a$
we use a weighting approach

We first estimate the inverse probability of treatment weights
$w^i = 1/P(A=a^i \mid C=c^i)$
where $a^i$ and $c^i$ are the values for individual i of A and C respectively
For treated subjects this is $1/P(A=1|C=c^i)$; for controls $1/P(A=0|C=c^i)$
We then regress observed Y on A (without C) i.e. $E[Y|a] = \beta_0 + \beta_1 a$
but with each subjected weighted by $w^i$

If there is no unmeasaured confounding the weighted regression consistently estimates the parameters of the MSM
This is essentially using a 'propensity score' but by weighting

# Feedback

Confounding control becomes more difficult when earlier levels of the outcome can affect subsequent levels of the exposure

Example:

Religious participation has been associated with lower depression rates in numerous studies (Koenig et al., 2012)

But the vast majority (86%) are cross-sectional

This is problematic…

Maselko et al. (2012) examine, using longitudinal data, evidence for an effect in the reverse direction

Among women, depression at age 18 predicts lower service attendance subsequently, controlling for baseline service attendance

# Feedback

There is evidence for reverse causation

This result essentially renders cross-sectional data useless for assessing causality

Even if there were no effect of service attendance on depression, one would find a “protective association” simply because those who became depressed stopped attending

To help begin to rule out this possibility we would need to control for baseline outcome (depression) in the analyses

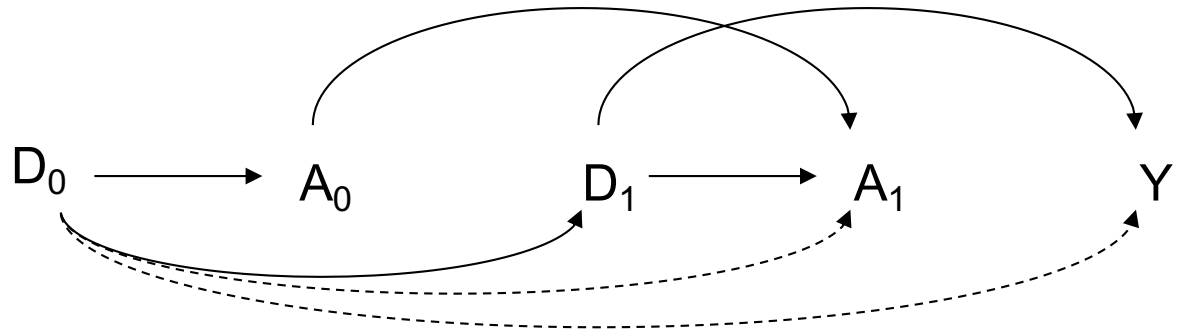
# Feedback

Control for baseline outcome helps rule out reverse causation

It does so provided

- Depression measurement two measurement periods back does not affect current depression independent of depression one period back

- - If so we need to control for prior exposure and/or earlier outcome as well




- Control for prior exposure also helps rule out confounding by other factors as the confounder would have to affect current attendance independent of past attendance

# Study Design

Different study designs allow for different levels of robustness to confounding and reverse causation; we can establish a certain hierarchy (Lash et al., 2021):

Cross-sectional studies
Cohort / follow-up with adjustment for baseline confounders
Cohort / follow-up with adjustment for baseline outcome also
Studies looking at change in exposure i.e. which also allow for adjustment for prior exposure
Longitudinal studies allowing for time-varying exposures and outcomes and for feedback
Randomized controlled trial

Different fields and subfields have very different distributions across these study design types
For evidence for causation control for baseline outcome should generally be considered a minimum threshold



VanderWeele, T.J. (2021). Can sophisticated study designs with regression analyses of observational data provide causal inferences? JAMA Psychiatry, 78:244-246.

# Causal Inference with Longitudinal Data

Thus far we have considered the effect of treatment at a single point in time on some outcome at a single point in time.

We will now consider a setting in which the treatment/exposure may vary over time:

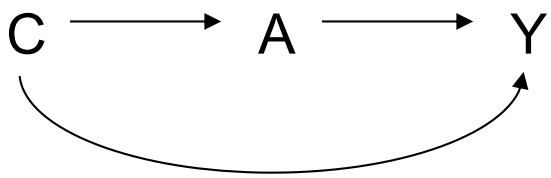
- Example 1: HIV/AIDS patients may or may not receive HAART at each visit depending on side effects and on CD4 counts
- Example 2: The cumulative effects of religious service attendance over time on health or depression
- Example 3: We might be interested in the cumulative effects of loneliness, which varies over time, on depression

We will first summarize and review a couple of principles of confounding control

# Causal Inference Principle I

Suppose we wish to estimate the *total* effect of A on Y.

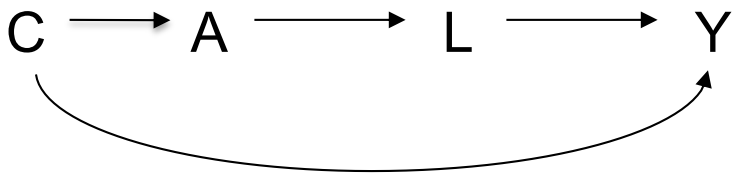
Causal Inference Principle I: If C is a common cause of A and Y then we should control for C




If we do not control for C, then the association we observe between A and Y may not be due to the causal effect of A on Y but rather due to the association between A and Y induced by C

# Causal Inference Principle II

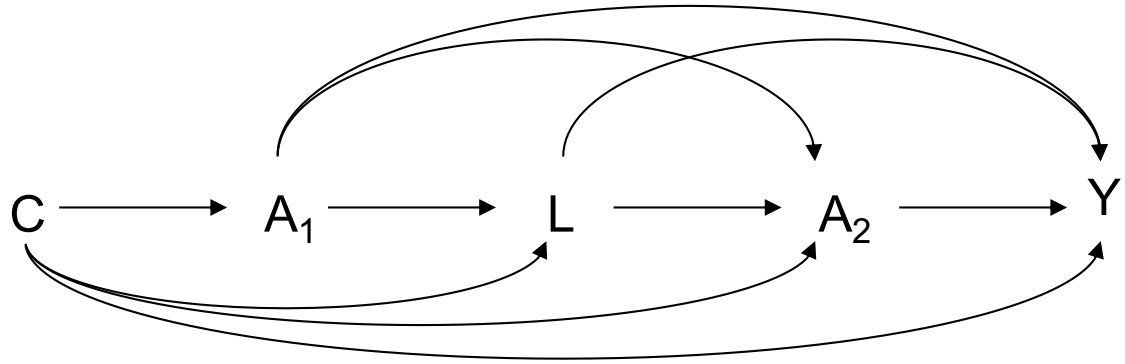
Causal Inference Principle II: If there is an intermediate variable between A and Y, we should not control for it.




If we want to assess the *total* effect of A on Y and we do control for L then some of the association between A and Y due to the causal effect of A and Y may be blocked by controlling for L.

# Causal Inference with Longitudinal Data

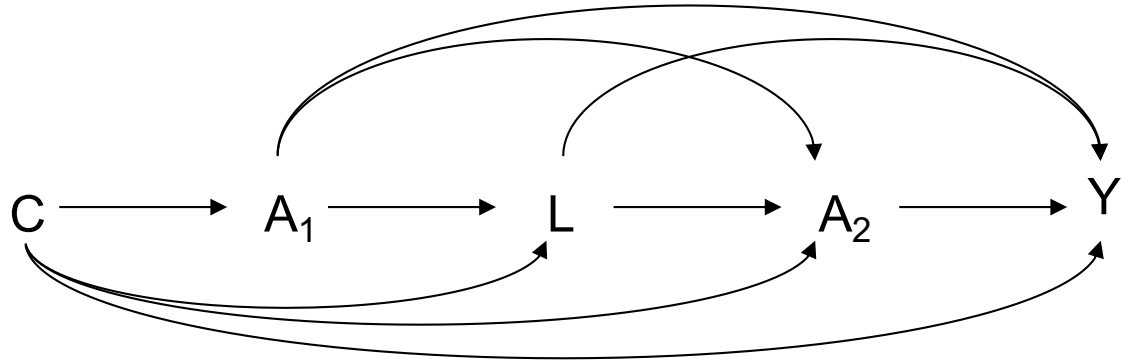
Suppose we want to know what the joint effects of interventions on loneliness at times 1 and 2 (denoted by $A_1$ and $A_2$) are on depression at time 3 (denoted by Y) with baseline covariates denoted by C and L the level of depressive symptoms between the two intervention times




Clearly we need to control for C as this is a common cause of treatment $A_1$ and outcome Y

# Causal Inference with Longitudinal Data

Should we control for L?

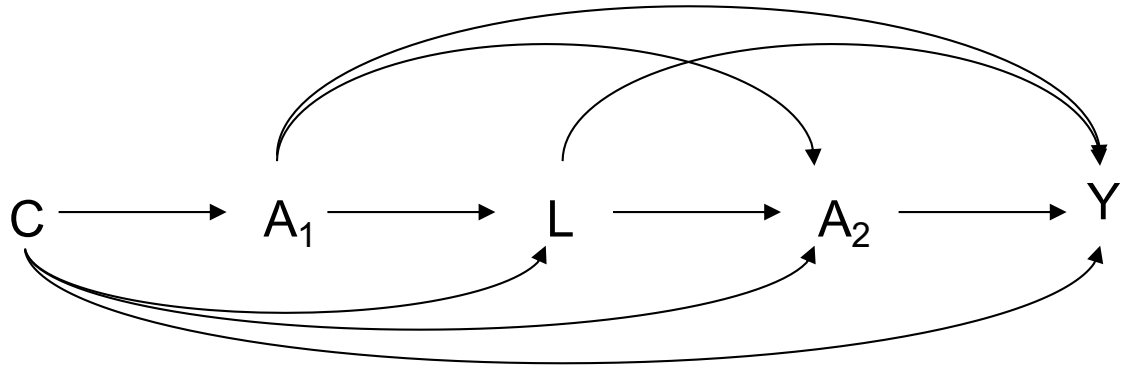



If we don't control for L, then we have an uncontrolled confounder because L is a common cause of treatment $A_2$ and outcome Y

**This would violate causal inference principle I**

# Causal Inference with Longitudinal Data

Should we control for L?



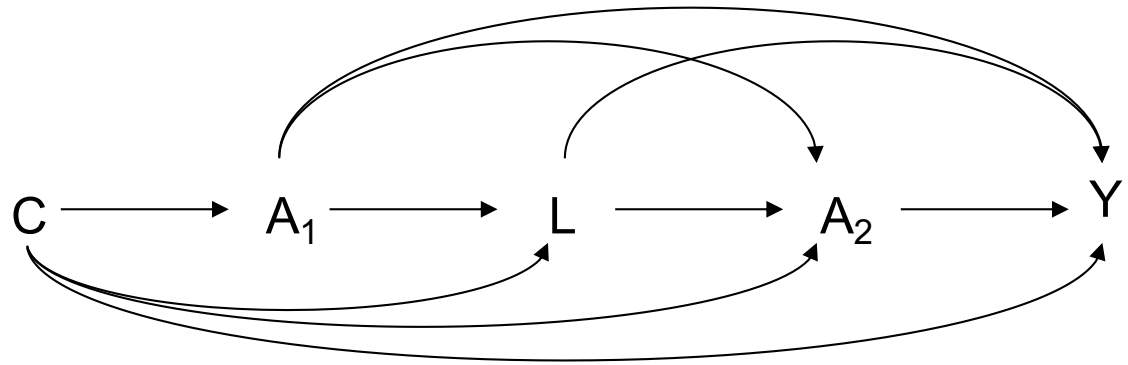

But if we do control for L then we have controlled for an intermediate variable between $A_1$ and Y

**This would violate causal inference principle II**

# Causal Inference with Longitudinal Data

Our two causal inference principles conflict!

Regression methods will not allow us to estimate the joint causal effects of $A_1$ and $A_2$ on Y in this case




This problem will generally arise with time-varying treatment if there is a variable, such as L, that is both a confounder and an intermediate variable

We then speak of "time-*dependent* confounding" because whether a variable L is a confounder depends on time (e.g. time 1 or time 2)

# Causal Inference with Longitudinal Data

Instead of regression (i.e. a model for the outcome conditional on the covariates) we will use a "marginal structural model" (again, a model for the counterfactual outcomes, Robins et al., 2000):

Let $Y_{a1a2}$ be the counterfactual value of Y for an individual under an intervention to set $A_1$ to $a_1$ and $A_2$ to $a_2$

Regression: $E[Y|A_1=a_1, A_2=a_2,C=c] = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3' c$

MSM: $E[Y_{a1a2}] = \kappa + \gamma_1 a_1 + \gamma_2 a_2$

Because we do not observe $Y_{a^1a^2}$ for all possible values of $a_1$ and $a_2$ for all individuals we cannot fit the MSM directly

However we can fit the MSM using a weighting technique under certain assumptions. Specifically we need that:

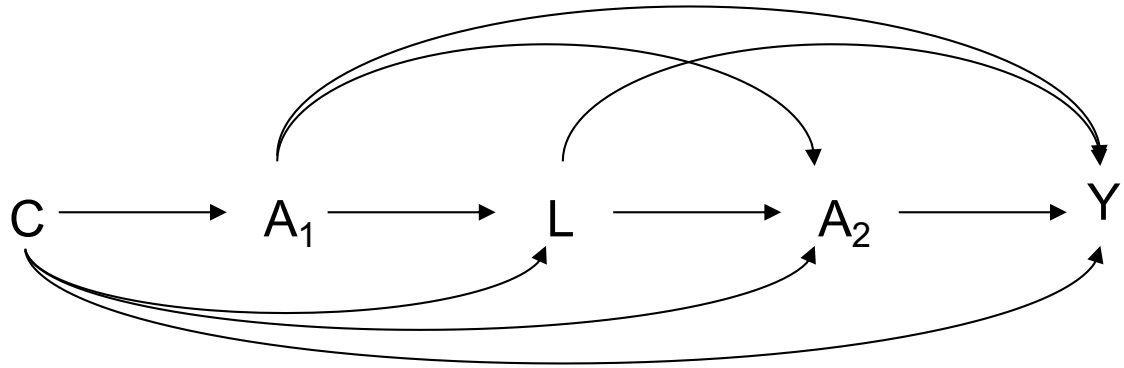
# Causal Inference with Longitudinal Data

(1) $Y_{a_1 a_2} \perp\!\!\!\perp A_1 \mid C$

(i.e. the effect of $A_1$ on the final outcome Y is unconfounded given C)

(2) $Y_{a_1 a_2} \perp\!\!\!\perp A_2 \mid \{C, A_1, L\}$

(i.e. the effect of $A_2$ on Y is unconfounded given baseline C, $A_1$ and the potential intermediate(s) denoted by L)




For the reasons given earlier, we generally wanted to include in C a measurement of the outcome (and possibly of the exposure $A_0$ as well) prior to the first measurement we are consider interventions on

# Causal Inference with Longitudinal Data

MSM: $E[Y_{a1a2}] = \kappa + \gamma_1 a_1 + \gamma_2 a_2$

Robins (1999) showed that under these no-unmeasured-confounding assumptions we can obtain consistent estimators of $\kappa$ , $\gamma_1$ and $\gamma_2$ (the parameters of the MSM) by fitting the regression model:

$E[Y|A_1=a_1, A_2=a_2] = \kappa + \gamma_1 a_1 + \gamma_2 a_2$

where each subject i is weighted by

$$\frac{1}{P(A_1 = a_1^i | C = c^i)} \times \frac{1}{P(A_2 = a_2^i | A_1 = a_1^i, C = c^i, L = l^i)}$$

where $a^i_1$, $a^i_2$, $c^i$, $l^i$ are the values for individual i of $A_1$, $A_2$, C and L respectively (C and L can be multivariate)

Control for confounding is addressed by weighting rather than regression

(the weighted regression should use "sandwich"/"robust" estimators of the standard errors to be valid; see code in lab)

# Causal Inference with Longitudinal Data

The weights

$$\frac{1}{P(A_1 = a_1^i | C = c^i)} \times \frac{1}{P(A_2 = a_2^i | A_1 = a_1^i, C = c^i, L = l^i)}$$

are referred to as "inverse probability of treatment weights" (IPTW) because they correspond, for each subject, to the inverse of the probability of their receiving the treatment they in fact received, conditional on their covariate history

If the treatments $A_1$ and $A_2$ are binary then the probabilities could be obtained using a logistic regression

First a regression of $A_1$ on C

Second a regression of $A_2$ on $\{A_1, C, L\}$

We break the confounding control task into two parts

Again, the weighted regression estimates the parameters of the MSM;

$E[Y_{a_1 a_2}] = \kappa + \gamma_1 a_1 + \gamma_2 a_2$

# Extension: Stabilized Weights
# Extension: Ordinal and Continuous Exposures

Extension 1: The same approach to fitting the MSM still works if so called "stabilized weights" are used instead:

$$\frac{P(A_1 = a_1^i)}{P(A_1 = a_1^i | C = c^i)} \times \frac{P(A_2 = a_2^i | A_1 = a_1^i)}{P(A_2 = a_2^i | A_1 = a_1^i, C = c^i, L = l^i)}$$

These stabilized weights often result in reduced standard errors

Extension 2: Ordinal and Continuous Exposures

If the exposure is *categorical* one again estimates the probability of each individual receiving the treatment that was in fact received (e.g. using *multinomial logistic regression*)

If the exposure/treatment $A_1$ and $A_2$ are *continuous* then the probabilities are replaced by *probability density functions* (which we will use in an application below)

# Extension 3: Additional Time Points

The approach extends to more than two times of treatment
With three exposure periods to intervene on we can use the MSM:

$E[Y_{a1a2a3}] = \kappa + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3$

Confounding Assumption: At each period k, the baseline covariates C, and the history of the time-varying covariates, $L_1,\ldots,L_{k-1}$ and exposures up through time k-1, $A_1,\ldots A_{k-1}$ suffice to control for confounding of the effect of the exposure, at time k, $A_k$, on the outcome at each subsequent time i.e. $Y_{a1a2\ldots aT} \perp\!\!\!\perp A_k \mid (C, L_1,\ldots,L_{k-1}, A_1,\ldots A_{k-1})$

We add an additional inverse probability of treatment weight for each exposure/treatment period:

$$W = \prod_{k=1}^{t} \frac{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i)}{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i, L_1 = l_1^i, \ldots, L_{k-1} = l_{k-1}^i)}$$

# Extension 4: Conditional MSMs

Sometimes the marginal structural model is estimated conditional on the baseline covariates C (prior to the first treatment time)

$$\mathbb{E}[Y_{a_1 a_2 a_3} | C = c] = \kappa + \gamma c + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3$$

This can help increase efficiency and stability of estimates; it can also allow interactions between baseline covariates and exposure

When conditioning on baseline covariates C, these covariates are then also included in the numerator of the weights (since control is then made by regression rather than weighting)

$$\frac{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i)}{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i, L_1 = l_1^i, \ldots, L_{k-1} = l_{k-1}^i)}$$

The weight for the first exposure period is thus just 1 (since numerator and denominator cancel)

One can alternatively condition on a subset of baseline covariates, which then go in both numerator and denominator; others denominator only

# Additional Assumption

For the inverse probability of treatment weighting procedure to give consistent estimates of the parameters, in addition to the confounding assumptions we need that the probability of treatment transitions

$$p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i, L_1 = l_1^i, \ldots, L_{k-1} = l_{k-1}^i)$$

are strictly between 0 and 1 for all values of the past (i.e. never exactly 0 or exactly 1)

This is sometimes referred to as a “positivity” assumption
If this does not hold, bias can result
See Robins et al. (2000) for an example

# Trimming of Weights

In some cases if the certain treatment probabilities are small given particular values of past covariates, the weight

$$\frac{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i)}{p(A_k = a_k^i \mid A_1 = a_1^i, \ldots, A_{k-1} = a_{k-1}^i, C = c^i, L_1 = l_1^i, \ldots, L_{k-1} = l_{k-1}^i)}$$

may be very large; this can make estimates very unstable

Sometimes, in practice to deal with this the weights are "trimmed" (discarded) at the 1st and 99th percentiles, or "truncated" (set to the values at those percentiles) of the overall weight distribution (i.e. product of all of the weights at each time) for more stable estimates
If more extreme trimming is done e.g. 10th/90th percentile the interpretation (relevant population) may be substantially affected

There is a bias-variance tradeoff here (Cole and Hernan, 2008)
"Trimming" changes support; or "truncating" weights introduces bias

# Example: Loneliness and Depression

The relationship between loneliness and depression as psychological constructs is complex, both conceptually and causally

However, conceptual and empirical work suggests that loneliness and depression are distinct constructs (Cacioppo et al., 2006ab)

Both constructs indicating negative affect, loneliness about one's social relationships and depression more generally

Data come from a longitudinal study with measurements on loneliness and depression over 5 years to assess both the magnitude and persistence of the effect of loneliness on depression

# Loneliness and Depression

Analyses might attempt to address whether or not it is just most recent loneliness or an entire history of loneliness (e.g. multiple interventions on loneliness) that affect depressive symptoms

The analyses were used to help determine whether a single intervention or temporally separated two-part intervention would be implemented in a randomized trial in the US Army

<u>Limitations:</u>

The data do not fully capture loneliness and depressive symptoms, which are continuously time-varying; the measurements here are annual

We will define counterfactual depression outcomes with respect to hypothetical interventions on loneliness

But there is more than one way to intervene on loneliness; our counterfactuals will potentially be ill-defined

But for the purposes of making decisions about the design of the trials this is perhaps the best we can do

# Loneliness and Depression

Data were obtained from the Chicago Health, Aging, and Social Relations Study (CHASRS), a population-based study of non-Hispanic Causasians, African Americans and Latino Americans born between 1935 and 1952 living in Cook County, Illinois (n=228)

Data in CHASRS is available on age, gender, ethnicity, marital status, education, income at baseline and also on depression, loneliness, subjective well-being, psychiatric conditions and psychiatric medications measured at baseline and at each of the four subsequent years.

Loneliness was assessed using the UCLA-R (a 20-item questionnaire with scores that range from 20 to 80)

Depressive symptomatology was assessed using the CES-D (a 20-item questionnaire with scores that range from 0 to 60)

One CES-D item asks about loneliness and this was excluded and the resulting measure (CES-D-ML) ranges from 0 to 57

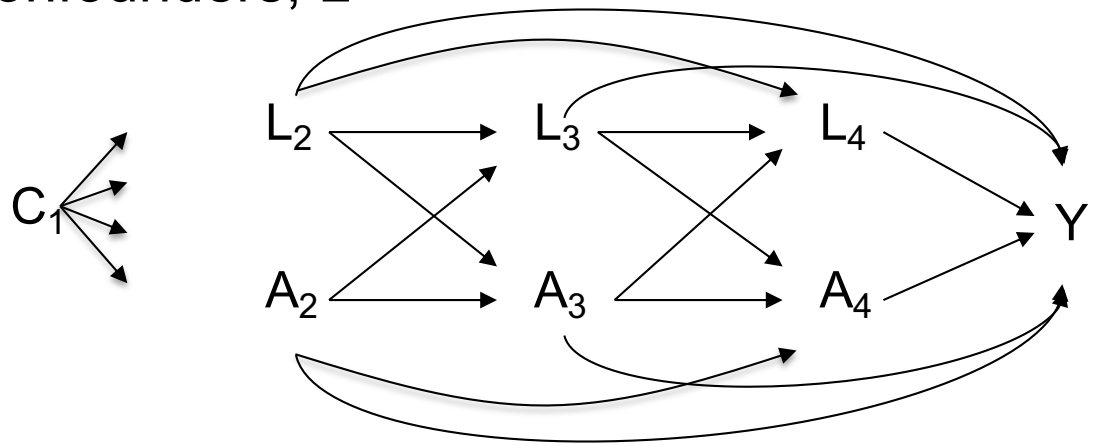
# Loneliness and Depression

All measures in year 1 were considered as baseline covariates, C

We consider the effects of hypothetical interventions on loneliness, A, during visits 2, 3 and 4 on final depressive symptomatology, Y, at visit 5

The baseline covariates included age, gender, ethnicity, marital status, education, and income and initial values of loneliness, depression, subjective well-being, and psychiatric conditions and medications

Subsequent values of depression, well-being, and psychiatric conditions/medications were considered as potential time-dependent confounders, L



# Loneliness and Depression

If the exposures of interest uclaY2, uclaY3, uclaY4 were binary we could estimate weights by fitting logistic regressions for each of the probabilities of the exposure:

$1/P(A_2=a^i_2|C=c^i)$,

$1/P(A_3=a^i_3|A_2=a^i_2,C=c^i,L_2=l^i_2)$,

$1/P(A_4=a^i_4|A_2=a^i_2,A_3=a^i_3,C=c^i,L_2=l^i_2,L_3=l^i_3)$,

Here loneliness is considered as continuous in this analysis

In this case for the IPT weights we use linear regression for the weights using the normal probability density function

This will give standardized residuals and we then evaluate, for each individual, the normal probability density function at the value of the residual to obtain the weight

# Loneliness and Depression

Finally we run a regression of the final outcome (depressive symptomatology at visit 5) on loneliness at visits 2, 3 and 4, where each subject is weighted by the inverse probability of treatment weights

For the MSM $E[Y_{a2a3a4}] = \kappa + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4$

| Parameter | Estimate | Standard Error | 95% Confidence Limits | | Z | Pr > \|Z\| |
|---|---|---|---|---|---|---|
| uclaY2 | -0.1212 | 0.0981 | -0.3135 | 0.0711 | -1.23 | 0.2169 |
| uclaY3 | 0.3413 | 0.1532 | 0.0411 | 0.6414 | 2.23 | 0.0259 |
| uclaY4 | 0.2618 | 0.1222 | 0.0223 | 0.5013 | 2.14 | 0.0322 |

# Loneliness and Depression

| Parameter | Estimate | Standard Error | 95% Confidence Limits | | Z | Pr > \|Z\| |
|---|---|---|---|---|---|---|
| uclaY2 | -0.1212 | 0.0981 | -0.3135 | 0.0711 | -1.23 | 0.2169 |
| uclaY3 | 0.3413 | 0.1532 | 0.0411 | 0.6414 | 2.23 | 0.0259 |
| uclaY4 | 0.2618 | 0.1222 | 0.0223 | 0.5013 | 2.14 | 0.0322 |

The analysis suggests that a hypothetical intervention to change loneliness by 1 point at visit 3 and by 1 point at visit 4 would decrease depressive symptomatology by about 0.34+0.26 = 0.6 points at visit 5

e.g. if an intervention changed loneliness at visits 3 and 4 from 45 at each visit to 35 at each visit then the CES-D-ML score at visit five would be expected to be 10*0.34+10*0.26 = 6 points lower

The magnitude of the effect is fairly large but it is also persistent

Loneliness 2 years prior appears to have an effect on present depressive symptomatology even if also intervening on loneliness 1 year prior

# Other Modeling Forms

Thus far we have considered MSMs that have a different coefficient for each time point e.g.:

$$E[Y_{a2a3a4}] = \kappa + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4$$

But we can also consider other forms e.g. cumulative effects

$$E[Y_{a2a3a4}] = \kappa + \gamma \, cum(a)$$

where $cum(a) = a_2 + a_3 + a_4$

or we could consider weighted averages where more recent exposures are given more weight, or interaction terms e.g. $a_3 a_4$

For binary outcomes, we can also fit logistic MSM's:

$$logit[P(Y_{a2a3a4}=1)] = \kappa + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4$$

# Religion and Depression

We will consider different functional forms of an MSMs examining religious service attendance and depression from the Nurses Health Study data (Li et al., 2016)

Analyses used 48,984 U.S. nurses with mean age 58 years Followed up from 1996 to 2008

Religious services attendance was self-reported in 1992, 1996, 2000, 2004, etc.

Depression was self-reported as being physician-diagnosed clinical depression or anti-depressant use in 1992, 1996, 2000, 2004, etc.

Depressive symptoms were measured in 1992, 1996, 2000, 2004, etc. Prior depression may affect subsequent attendance

We will fit MSMs examining effects of service attendance in 1996 and 2000 on depression (binary and continuous) in 2004, controlling for covariates, baseline outcome, and baseline exposure in 1992

# Religion and Depression

The following covariates in 1992 are adjusted for in the analysis:

baseline service attendance,
baseline depression,
age,
race/ethnic groups,
geographic region,
employment status,
nurses education level,
husband education level,
census trace level income,
marital status,
living arrangement,
number of close friends,
having someone close to talk to,
physical limitations or disability

body mass index at age 18,
weight change since age 18,
height
physical activity,
diet quality,
smoking status,
alcohol consumption,
family history of MI,
family history of cancer,
family history of diabetes,
hypertension,
high cholesterolemia,
diabetes,
post-menopausal hormone use

# Service Attendance and Depression

| | Outcome: Depression in 2004 | | | | | |
|---|---|---|---|---|---|---|
| | Binary | | | GDS continuous | | |
| | OR | Lower 95% CI | Upper 95% CL | Estimate | Std Err | P value |
| Religious service attendance in 1996 (Never) | 1.00 | 1.00 | 1.00 | - | 1.00 | 1.00 |
| Religious service attendance in 1996 (< 1/wk.) | 1.00 | 0.89 | 1.12 | 0.03 | 0.09 | 0.70 |
| Religious service attendance in 1996 (1/wk.) | 1.09 | 0.95 | 1.24 | 0.11 | 0.10 | 0.27 |
| Religious service attendance in 1996 (> 1/wk.) | 1.15 | 0.98 | 1.36 | 0.12 | 0.12 | 0.29 |
| | | | | | | |
| Religious service attendance in 2000 (Never) | 1.00 | 1.00 | 1.00 | - | 1.00 | 1.00 |
| Religious service attendance in 2000 (< 1/wk.) | 0.87 | 0.79 | 0.97 | -0.20 | 0.08 | 0.01 |
| Religious service attendance in 2000 (1/wk.) | 0.75 | 0.67 | 0.84 | -0.48 | 0.09 | <.0001 |
| Religious service attendance in 2000 (> 1/wk.) | 0.71 | 0.62 | 0.82 | -0.53 | 0.10 | <.0001 |

There is a dose-response relationship the effect of service attendance in 2000 on depression in 2004

Results are consistent across outcomes measures (binary or continuous)

Past service attendance (in 1996) does not seem to have much effect beyond current service attendance, but this may obscure transitions…

# Service Attendance and Depression

| | Marginal structural model | | | | |
|---|---|---|---|---|---|
| | Depression[a] | | Continuous level of depressive symptoms[b] | | |
| Religious service attendance | Odds ratios | 95 % CI | Estimate | Std Err | *P* value |
| <1/week· in 1996 and never in 2000 | Reference | | Reference | Reference | Reference |
| <1/week· in 1996 and <1/week· in 2000 | 0.87 | 0.78, 0.96 | −0.21 | 0.08 | 0.01 |
| <1/week in 1996 and 1/week in 2000 | 0.90 | 0.77, 1.05 | −0.31 | 0.12 | 0.01 |
| <1/week in 1996 and >1/week in 2000 | 0.99 | 0.70, 1.39 | 0.14 | 0.32 | 0.65 |
| ≥ 1/week in 1996 and never in 2000 | 1.33 | 1.09, 1.62 | 0.30 | 0.18 | 0.09 |
| ≥ 1/week in 1996 and <1/week in 2000 | 1.05 | 0.91, 1.22 | 0.02 | 0.11 | 0.88 |
| ≥ 1/week in 1996 and 1/week in 2000 | 0.82 | 0.72, 0.92 | −0.38 | 0.09 | <0001 |
| ≥ 1/week in 1996 and >1/week in 2000 | 0.80 | 0.70, 0.92 | −0.44 | 0.10 | <0001 |

# Depression and Service Attendance

| | Outcome: Service Attendance once per week or more in 2004 |
|---|---|
| Depression in 1996 | OR=0.91, (95% CI: 0.83, 1.00 ) |
| Depression in 2000 | OR=0.74, (95% CI: 0.68, 0.80) |

There also appears to be fairly strong evidence for an effect of depression in 2000 decreasing subsequent service attendance in 2004 as well

Here there does also seem to be a further direct effect of depression in 1996 independent of depression in 2000

# Service Attendance and Depression

The results confirm evidence for an effect of service attendance on depression but…

- Controlling for reverse causation/feedback and time-dependent confounding
- And confirming also an effect in the reverse direction, of depression on lowering attendance

The results on depression decreasing service attendance may be just as interesting in terms of implications for religious communities:

- Those who become depressed tend to cease attendance, which may exacerbate depression yet further
- Further support could be extended to such persons before they leave

Allowing for feedback and accounting for both directions may give further important insights into the dynamics governing different relationships; it can be helpful to fit these models in both directions

# Study Design and Research Practice

While hundreds of studies on religion and depression have been published, only a small minority are longitudinal with control for baseline outcome

For service attendance and life satisfaction >99% use cross-sectional

When effects may be in both directions, cross-sectional data is almost useless

Similar issues certainly arise in other areas

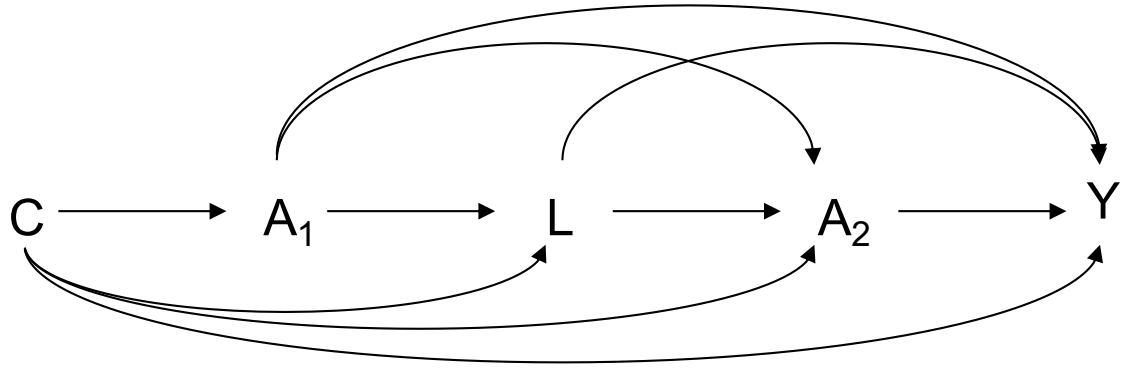
Perhaps particularly problematic with “softer” exposures and outcomes

Perhaps the biggest threat to applied causal inference today is study design!

- Researchers assessing causality should try to avoid studies with cross-sectional data (only hypothesis-generating if e.g. never been studied before)
- Journals should otherwise begin to reject such cross-sectional studies
- Cross-sectional studies should be discarded from systematic reviews
- Systematic reviews and meta-analysis restricted to longitudinal studies with control for baseline outcome would be valuable
- The evidence base often comes from too few studies

# Ignoring Time-Dependent Confounding

When can we look at the effects of trajectories of exposure and ignore time-varying confounders?




The problem arises when a variable L confounds the effect of subsequent exposure and is itself affected by prior exposure

The problems with traditional regression methods go away if:

- There is no effect of L on subsequent exposure, or
- There is no effect of L on the outcome, or
- There is no effect of prior exposure on L

We can examine evidence for these empirically; when one of the three effects is absent we do not need to use causal methods for time-dependent confounding

# Cox MSM

A similar approach can be taken for a time to event outcome using a Cox proportional hazard model

Consider a time-to-event outcome T, with baseline covariates V, time-varying confounders L, time-varying treatment A

Let $\bar{a} = (A_1,\ldots A_T)$ denote the whole treatment history and a(t) treatment at time t

For the hazard at time t, with baseline covariates V the Cox MSM is:

$$\lambda_{T_{\bar{a}}}(t|V) = \lambda_0(t)\exp(\beta_1 a(t) + \beta_2 V)$$

The confounding assumptions are as before:

Confounding Assumption: At each period k, the baseline covariates V, and the history of the time-varying covariates, $L_1,\ldots,L_{k-1}$ and exposures up through time k-1, $A_1,\ldots A_{k-1}$ suffice to control for confounding of the effect of the exposure, at time k, $A_k$, on the outcome at each subsequent time

# Cox MSM

The weights vary over time and are given by

$$\prod_{k=0}^{\mathrm{int}(t)} \frac{pr(A(k) = a_i(k) | \bar{A}(k-1) = \bar{a}_i(k-1), V = v_i)}{pr(A(k) = a_i(k) | \bar{A}(k-1) = \bar{a}_i(k-1), \bar{L}(k) = \bar{l}_i(k))}$$

And these can, for each time, for a binary treatment, be estimated by logistic regression (in the notation above L(0)=V so is included in the denominator also)

The Cox MSM can be fit by fitting a regular Cox proportional hazard model

$$\lambda_T(t|\bar{A}(t), V) = \lambda_0(t)\exp(\gamma_1 A(t) + \gamma_2 V).$$

weighted by the inverse probability of treatment weights (this is somewhat more complex to implement with software)

# Censoring

If there is censoring this can also be taken into account through inverse probability of censoring weighting (IPCW)

At each time point one estimates the probability of remaining in the study conditional on the past exposure and covariates and being in the study throughout the past; the numerator of the weight does not condition on the time-varying confounders

The overall IPCW is the product of the weights at each time

$$\prod_{k=0}^{t} \frac{pr[C(k) = 0 | \bar{C}(k-1) = 0, \bar{A}(k-1) = \bar{a}_i(k-1), V = v_i]}{pr[C(k) = 0 | \bar{C}(k-1) = 0, \bar{A}(k-1) = \bar{a}_i(k-1), \bar{L}(k-1) = \bar{l}_i(k-1)]}$$

# Censoring

To fit the MSM under censoring we also need a sequential ignorable censoring assumption

Ignorable Censoring: At each period k, the counterfactual hazard is independent of censoring conditional on the the baseline covariates V, and the history of the time-varying covariates, $L_1,\ldots,L_{k-1}$ and exposures up through time k-1, $A_1,\ldots A_{k-1}$

One again fits a weighted proportional hazard model

$$\lambda_T(t|\bar{A}(t), V) = \lambda_0(t)\exp(\gamma_1 A(t) + \gamma_2 V).$$

where the weights are now the product of the weights for treatment and the weights for censoring

# Zidovudine and Survival

Hernan et al. (2000) consider the effect of zidovudine on survival for HIV positive men using data from MACS (Multicetner AIDs Cohort Study)

Zidovudine is time-varying

Time-varying confounding is also likely present: those with low CD4 counts are more likely to initiate treatment

This may make zidovudine look harmful

We need to use MSMs to control for time-varying treatment

Analysis of MACS data:
2178 HIV+ homosexual or bisexual men (w/o AIDS at baseline)

Follow-up every 6 months from 1986-1994

# Zidovudine and Survival

Baseline Covariates:

model with baseline covariates includes also: age, calendar year (1985, 1986, 1987–89, or 1990–1993), CD4 (<200, 200–499, or ≥500/$\mu$l), CD8 (<500; 500–999; or ≥1,000 per $\mu$l), WBC (<3,000; 3,000–4,999; or ≥5,000 per $\mu$l), RBC (<35, 35–44, or ≥45 × $10^5$ per $\mu$l), platelets (<150, 150–249, or ≥250 × $10^3$ per $\mu$l), presence of symptoms (yes if fever, oral candidiasis, diarrhea, weight loss, oral hairy leukoplakia, or herpes zoster, or no if otherwise).

Time-Varying Covariates:

† Weights calculated as described in the text using data on baseline covariates plus most recent CD4, CD8, WBC, RBC (<30, 30–39, or ≥40 × $10^5$ per $\mu$l), platelets, presence of symptoms, presence of AIDS-defining illness, and previous zidovudine use.

Fit Cox model with just baseline zidovudine use, and baseline covariates

Also fit Cox MSM to adjust for time-dependent confounding (adjusted also for censoring)

# Zidovudine and Survival

**TABLE 1. Inverse-Probability-of-Treatment Weighted Estimates of the Causal Effect of Zidovudine Therapy on Mortality in the Multicenter AIDS Cohort Study**

| Unweighted estimates* | RR | 95% CI |
|---|---|---|
| Unadjusted | 3.55 | 2.95–4.27 |
| Only baseline covariates | 2.32 | 1.92–2.81 |

| Weighted estimates† | RR | Valid 95% Conservative CI | Invalid Model-Based‡ 95% CI |
|---|---|---|---|
| Stabilized weights | 0.74 | 0.57–0.96 | 0.62–0.87 |
| Nonstabilized weights | 0.76 | 0.54–1.05 | 0.71–0.80 |

# Time-Varying Exposures and Regression

MSMs can be used to examine the effects of entire trajectories of time-varying exposures

However, certain causal questions can be addressed with a single regression even with a time-varying exposure

Even if our exposure is time-varying if we include in the baseline covariates C past outcome and past exposure we can get the effect of a single change in the exposure

If $Y_a \perp\!\!\!\perp A \mid C$ and we fit the regression: $E[Y|A,C] = \beta_0 + \beta_1 A + \beta_2' C$

Then: $E[Y_1|C=c] - E[Y_0|C=c] = \beta_1$

# Time-Varying Exposures and Regression

Regression and Time-Varying Exposures:

From the regression we get the effect of a single change in the exposure

This also includes the effects that the change in the present exposure may have on future exposures

Cox Model and Time-Varying Exposures:

We can do the same in a Cox model by updating the exposure, and also updating prior exposure (as a time-varying covariate) to control for confounding

This is not often done, but could be

This then captures the instantaneous effect of current exposure on the hazard

MSM and Time-Varying Exposures:

What we cannot do with just a basic regression or Cox model is assess the effect of a trajectory (e.g. treat 3 times in a row) or assess direct effects (e.g. is there an effect of exposure two periods back not through 1 period back?)



For these we need MSMs, or… other methods for time-varying exposures

# Structural Mean Models

MSMs:

Marginal structural models are relatively easy to implement

But the weighting approach reduces efficiency

It is also often less stable with continuous exposures

Structural Mean Models:

Another class of models, structural nested models, has more flexibility

A special sub-class of these is structural mean models

See Hernán and Robins (2009), and Vansteelandt and Sjolander (2016) for recent overviews

Unfortunately, model fitting is more difficult

One special case that is relatively straightforward can be employed when the joint effects of the exposure over just two time-periods, $A_1$ and $A_2$, on the outcome is of interest.

# Structural Mean Models

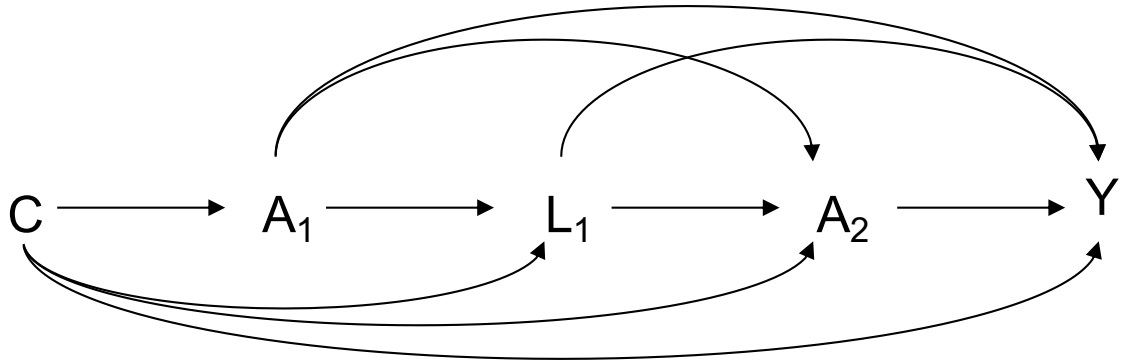
Confounding assumptions are the same as before:

At each period k, the baseline covariates and the history of the time-varying exposure and covariates up through time k-1 suffice to control for confounding of the effect of the exposure, at time k, on the final outcome.

With two time periods, $A_1$ and $A_2$, we assume:

The effect of $A_1$ on Y is unconfounded conditional on C,

The effect of $A_2$ on Y is unconfounded conditional on C, $A_1$ and $L_1$



# Structural Mean Models

The structural mean model here takes the form:

$$E[Y_{a_1 a_2} \mid C = c] = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_1 a_2 + \beta_4 c$$

Instead of weighting, this can be fit by using a two-stage regression approach (Vansteelandt, 2009).

In the first stage we fit a regular linear regression model of Y on $A_1$, $A_2$, C, and $L_1$, allowing for potential interaction between $A_1$ and $A_2$:

$$E[Y \mid a_1, a_2, c, l] = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_1 a_2 + \gamma_4 c + \gamma_5 l$$

# Structural Mean Models

We then use coefficient estimates of this regression model in two ways

First, we take the estimate of regression coefficient for the exposure $A_2$, $\gamma_2$, and the coefficient for the interaction, $\gamma_3$, as the estimates for $\beta_2$ and $\beta_3$ in the structural mean model.

Second, we take the coefficient estimates of $\gamma_2$ and $\gamma_3$, and for each individual i we calculate the following outcome residuals:

$$\hat{Y} = Y - \hat{\gamma}_2 a_2 - \hat{\gamma}_3 a_1 a_2$$

We then regress these residuals on the exposure A and covariates C:

$$E[\hat{Y} \mid a, c] = \mu_0 + \mu_1 a_1 + \mu_2 c$$

We then take the estimate of the coefficient $\boldsymbol{\mu}_1$ from this regression as our estimate of $\beta_1$ in the structural mean model

# Structural Mean Models

Provided the confounding assumptions described above hold, this approach will give consistent estimates for the parameters of the structural mean model (Vansteelandt, 2009)

Standard errors can be obtained by bootstrapping

Although the procedure is somewhat involved, it can be carried out using standard software and it tends to be more stable if the exposures under study are continuous rather than binary or categorical

This is not the standard way to fit structural mean / structural nested models but the typical approaches are computationally more involved (EPI207)

For other more complex settings and structural nested models in survival contexts see EPI207

# G-Formula

The approaches described above using marginal structural models and structural mean models have come to be routinely used methods for assessing the effects of time-varying exposures

The marginal structural model approach, due its relative straightforward implementation, has become especially popular

However, these tools were developed somewhat after the original theory for causal inference with time-varying exposures appeared

The initial theory of causal inference for time-varying exposures (Robins, 1986) instead gave formulas for the causal effects based on a generalization of standardization

This generalization is often referred to as the g-formula

The whole suite of methods as “g-methods” (“generalized” causal effects)

# G-Formula

From before we have that if $Y_a \perp\!\!\!\perp A \mid C$ then

$$E[Y_1|C=c] - E[Y_0|C=c] = E[Y|A=1,C=c] - E[Y|A=0,C=c]$$

This gives us conditional causal effects

If we want the marginal effect for the population we standardize by the distribution of C:

$$E[Y_1] - E[Y_0] = \Sigma_c \{E[Y|A=1,C=c] - E[Y|A=0,C=c]\}P(C=c)$$

The G-Formula is a generalization of this to exposures that vary over time

# G-Formula

Confounding Assumption: At each period k, the baseline covariates C, and the history of the time-varying covariates, $L_1,\ldots,L_{k-1}$ and exposures up through time k-1, $A_1,\ldots A_{k-1}$ suffice to control for confounding of the effect of the exposure, at time k, $A_k$, on the outcome at each subsequent time

Specifically, the effect comparing two exposure trajectories $(a_1,...,a_T)$ and $(a'_1,...,a'_T)$ (for example, for a binary exposure we might take $(a_1=1,...,a_T=1)$ and $(a'_1=0,...,a'_T=0)$) is then given by:

$$\sum_{l,c} E[Y \mid a_1,...,a_T,c,l_1,...,l_T]\prod_{k=1}^{T} p(L_k=l_k \mid a_1,...,a_{k-1},c,l_1,...,l_{k-1})p(C=c)$$

$$-\sum_{l,c} E[Y \mid a'_1,...,a'_T,c,l_1,...,l_T]\prod_{k=1}^{T} p(L_k=l_k \mid a'_1,...,a'_{k-1},c,l_1,...,l_{k-1})p(C=c)$$

# G-Formula

$$\sum_{l,c} E[Y \mid a_1,\ldots,a_T,c,l_1,\ldots,l_T] \prod_{k=1}^{T} p(L_k = l_k \mid a_1,\ldots,a_{k-1},c,l_1,\ldots,l_{k-1}) p(C = c)$$

$$-\sum_{l,c} E[Y \mid a'_1,\ldots,a'_T,c,l_1,\ldots,l_T] \prod_{k=1}^{T} p(L_k = l_k \mid a'_1,\ldots,a'_{k-1},c,l_1,\ldots,l_{k-1}) p(C = c)$$

These formulas are standardizations of the conditional expectation of the final outcome Y conditional on exposure history, baseline covariates, and time-varying covariate history standardized by the probability of the time-varying and baseline covariates conditional on the past values of these covariates and the past values of the exposure history

# Parametric G-Formula

$$\sum_{l,c} E[Y \mid a_1,\ldots,a_T,c,l_1,\ldots,l_T]\prod_{k=1}^{T} p(L_k = l_k \mid a_1,\ldots,a_{k-1},c,l_1,\ldots,l_{k-1})p(C=c)$$

$$-\sum_{l,c} E[Y \mid a'_1,\ldots,a'_T,c,l_1,\ldots,l_T]\prod_{k=1}^{T} p(L_k = l_k \mid a'_1,\ldots,a'_{k-1},c,l_1,\ldots,l_{k-1})p(C=c)$$

In principle, with large amounts of data, one could estimate each of the conditional expectations and each of the probabilities by sample averages

However, with relatively a lengthy covariate and/or exposure history this would require a very large amount of data to do accurately

An alternative strategy would be to use parametric models for the conditional expectation and for the covariate probabilities.

# Parametric G-Formula

The approach of using parametric models for the expectations and probabilities is sometimes referred to as the parametric g-formula

After fitting the parametric models with the data and using the g-formula to obtain estimates of causal effects, one can use bootstrapping to obtain standard errors and confidence intervals

G-Null Paradox:

While intuitively attractive, the parametric g-formula approach was shown to suffer from certain theoretical problems

In particular, it can be shown that under certain specifications of the parametric models, when some of the variables are binary and some are continuous, it can be possible that the models are specified in such a way that it is impossible for the parametric g-formula to obtain an effect of 0 when in fact the null of no effect is true.

This problem is sometimes referred to as the g-null paradox (Robins and Wasserman, 1997; Robins and Hernán, 2009)

# Parametric G-Formula

For this reason, for some time, the use of parametric g-formula approaches had been avoided in the literature and it was in part this problem that motivated some of the developments described above e.g. MSMs / SMMs

More recently, the use of parametric g-formula approach has been re-evaluated

<u>G-Null Paradox – Practice vs. Theory:</u>

While the problem of the g-null paradox is relevant in theory, it has been suggested that the bias that results in practice may often be very small

Even if one cannot obtain an exact numerical estimate of zero under the null, one might get very close

Moreover, the problems of the g-null paradox may be attenuated further by using flexible parametric models for the conditional expectations and probabilities such as splines

# Parametric G-Formula

The magnitude of bias implied by the g-null paradox may not be as substantial a concern as was thought and, depending on the setting, may not be large

See Young and Tchetgen Tchtegen (2014) for further discussion and examples

As a result of these considerations, the parametric g-formula approach has been used somewhat more frequently in recent years than in the past and macros have been developed for its implementation (Daniel et al., 2011; HSPH Causal Inference Program, 2016)

# G-Methods

Relative Advantages and Disadvantages:

MSMs are easy to implement, but lose efficiency

SNMs are flexible but difficult to implement

Parametric g-formula is flexible and can be implemented with macros but suffers from certain theoretical problems (but may not be severe in practice)

# Conclusions

(1) Time-varying exposures are common in practice

(2) When confounders also vary over time and we are interested in the effects of exposure trajectories traditional regression methods fail

(3) Marginal structural models can be used to estimate such effects

(4) Structural mean models provide an alternative and can be somewhat more efficient but generally less easy to implement

(5) The use of parametric models (or flexible splines) and the g-formula can be another way to estimate effects

(6) Regression methods with control for past outcome and past exposure can also be used to estimate effects of single interventions