An Introduction to Causal Inference, with Extensions to Longitudinal Data

Tyler VanderWeele
Harvard Catalyst Biostatistics Seminar Series
November 18, 2009

Plan of Presentation

- Association and Causation
- Counterfactual Framework
- Confounding and Regression
- Causal Inference for Longitudinal Data
- Marginal Structural Models and Inverse Probability of Treatment Weighting
- Example: The Persistence of the Effect of Loneliness on Depression

Association and Causation

Causal Inference attempts to articulate the assumptions needed to move from conclusions about association to conclusion about causation

Association: Two variables are associated if information about one tells you something about the likelihood of the other (statistical correlation)

Causation: Two variables are causally related if an intervention on one has the potential to change the other

Example: The United Nations studied governmental failure and found that the best indicator that a government was about to fail was the infant mortality rate... is this causal?

Association and Causation

Association does not imply causation

Many research studies will appropriately qualify their findings, noting that their results concern association amongst variables and do not necessarily imply causal relationships

However: Whenever these finding are interpreted, the interpreter will almost inevitably interpret the findings causally

We need the discipline of causal inference to be able to articulate what is being assumed when we go about interpreting our findings causally (moving from association to causation) and to be able to discuss whether these assumptions are reasonable

Association and Causation

Charig et al. (1986) used observational data to study the treatment of kidney stones; the treatments were not randomized

	Number Administered	Success Rate	Proportion
Treatment A	350	273	78%
Treatment B	350	289	83%

Was treatment B better?

Do the proportions reflect causal relationships?

If we gave everyone treatment B would this be better than if we gave everyone treatment A?

Association and Causation

SMALL STONES

	Number Administered	Success Rate	Proportion
Treatment A	87	81	93%
Treatment B	270	234	87%

LARGE STONES

	Number Administered	Success Rate	Proportion
Treatment A	263	192	73%
Treatment B	80	55	69%

More individuals with treatment A had large kidney stones

Now treatment A looks better

Do these stratified proportions reflect causal relationships?

How do we know?

Counterfactuals

The most important idea in causal inference is that of a counterfactual

Counterfactual: The basic idea of a counterfactual is what would have happened if, contrary to fact, we had done something other than what we did?

E.g. what would have happened if we had we given treatment A to a particular individual instead of treatment B?

Lewis (1973): “If c and e are two actual events such that e would not have occurred without c, then c is a cause of e.”

Idea of tying causation to counterfactuals goes at least as far back as Hume (1748)

Counterfactuals

In the kidney stone example, for each individual we have two counterfactual outcomes (or “potential outcomes”, Rubin 1974 cf. Neyman 1923)

Y_1 = Would the individual have been cured if given treatment A

Y_0 = Would the individual have been cured if given treatment B

For each individual we only get to observe one of Y_1 and Y_0

We observe Y_1 if the individual received treatment A

We observe Y_0 if the individual received treatment B

We have no way to observe the other counterfactual outcome

Counterfactuals

Ind	Y_1	Y_0
1	1	0
2	0	1
3	1	1
4	0	0
5	1	1
6	0	0
7	1	1
8	1	0

Each individual has two counterfactual outcomes: Y_1 and Y_0

There may be some individuals (like #1) who are cured only if they are given treatment A

There may be some individuals (like #2) who are cured only if they are given treatment B

There may be some individuals (like #3) who are cured regardless of the treatment given

There may be some individuals (like #4) who are not cured regardless of the treatment

Counterfactuals

Ind	Y_1	Y_0
1	1	0
2	0	1
3	1	1
4	0	0
5	1	1
6	0	0
7	1	1
8	1	0
Total	5/8	4/8

$E[Y_1] - E[Y_0] = 5/8 - 4/8$

$= 1/8$

If we knew all counterfactual outcomes we could just compare the totals:

How many are cured if everyone is given treatment A?

How many are cured if everyone is given treatment B?

Here we would see that treatment A is better on average

Counterfactuals

Ind	Y_1	Y_0	Trt
1	1	???	A
2	0	???	A
3	???	1	B
4	0	???	A
5	???	1	B
6	???	0	B
7	???	1	B
8	1	???	A
Total	???	???	
Obs	2/4	3/4	

Apparent effect seems to be -1/4

In practice, we only observe one counterfactual outcome for each individual

We can observe the numbers who are cured who got treatment A and who got treatment B

These might not reflect what would happen to the population

For example, those who got treatment B may be healthier

Confounding

We would like it to be the case that those who had treatment A and those who had treatment B are comparable (in their counterfactual outcomes)

If that were the case then the outcomes of those who had treatment A would be similar to the outcomes if the whole population had been given treatment A

And the outcomes of those who had treatment B would be similar to the outcomes if the whole population had been given treatment B

As we have seen already in the kidney stone example, however, this will often not be the case (those who received treatment A had larger stones)

Confounding

Even if the groups who received treatment A and those who received treatment B are not comparable…

It is possible that within strata of other variables (e.g. kidney stone size) those who received treatment A and those who received treatment B are comparable

If so, then the proportions within strata of kidney stone size will reflect average counterfactual outcomes for the strata

We will use $X \perp\!\!\!\perp Y \mid Z$ to denote that X is independent of Y conditional on Z

Confounding: Formally, we say that the effect of treatment A on outcome Y is unconfounded given covariates C if for all values a:

$$Y_a \perp\!\!\!\perp A \mid C$$

Confounding

If $Y_a \perp\!\!\!\perp A \mid C$ then within strata of the confounding variables, the treatment groups are comparable (i.e. they have similar counterfactual outcomes) we can draw causal conclusions because:

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

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

so…

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

We can compute causal effects from the data

Confounding

SMALL STONES (C=0)

	Number Administered	Success Rate	Proportion
Treatment A	87	81	93%
Treatment B	270	234	87%

LARGE STONES (C=1)

	Number Administered	Success Rate	Proportion
Treatment A	263	192	73%
Treatment B	80	55	69%

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

$E[Y_1|C=0] - E[Y_0|C=0] = 93\% - 87\% = 6\%$

$E[Y_1|C=1] - E[Y_0|C=1] = 73\% - 69\% = 4\%$

Confounding

In practice to make the assumption of no-unmeasured confounding reasonable we try to collect data on as many variables as possible that affect both the treatment/exposure under consideration and the outcome

Sometimes we don't know whether a particular variables affects both the treatment and the outcome; we may be confident it affects one but unsure about the other; often we control for these as well

Thus, in practice, often all "pre-treatment variables" are controlled for

However, if we are interested in the total effect of treatment, we don't want to control for variables which occur after the treatment

Controlling for a variable occurring after the treatment which is a consequence of treatment can bias our estimates

Effectively, by controlling for such "post-treatment" variables, one blocks part of the effect of treatment

Randomized Trials and Observational Studies

With randomized trials, who gets which treatment (treatment A vs. treatment B) is determined randomly

We thus know that, at least in expectation, the treatment groups are comparable; we will have that:

$Y_a \perp\!\!\!\perp A$

Treatment is determined randomly and so it is independent of any background characteristics and it is independent of the counterfactual outcomes

If treatment is assigned randomly with probabilities that depend on C then we will have that

$Y_a \perp\!\!\!\perp A \mid C$

Randomized Trials and Observational Studies

With observational data we must control for covariates to control for confounding

However, with observational data we are never sure our assumptions hold and so we are never certain of our conclusions about causation

Randomized trials are advantageous because we know the assumptions needed to draw causal conclusions in fact hold

We do not need to control for covariates to address confounding

However, in practice, we often do control for covariates to improve efficiency (or to attempt to address random imbalances in the treatment groups) or because we are interested in subgroup analyses

Regression and Causation

Regression and Causation: For regression coefficients to have a **causal** interpretation we need both that the linear regression assumptions (linearity, normality, independence, homoskedasticity) hold 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 β_1 units.

Regression and Causation

Regression and Association: If we do not have all confounding variables in the model, regression coefficients do not have a causal interpretation but still have an associational interpretation provided the linear regression assumptions hold.

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

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 β_1 units higher

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

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.

In the remainder of the presentation we will 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 sides effects and on CD4 counts

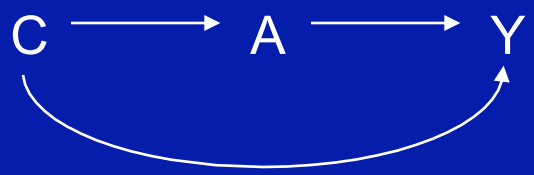
Example 2: We might be interested in the cumulative effects of loneliness, which varies over time, on depression

We will first summarize and review the principles of confounding control that we have discussed thus far

Causal Inference Principle I

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

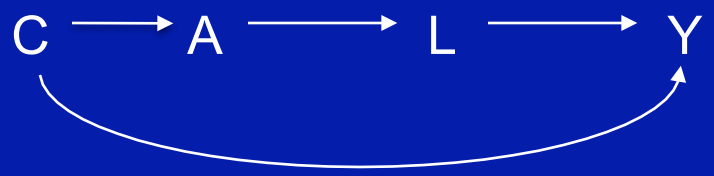
Causal Inference Principle I: If C is a common cause of A and Y then we should control for C



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

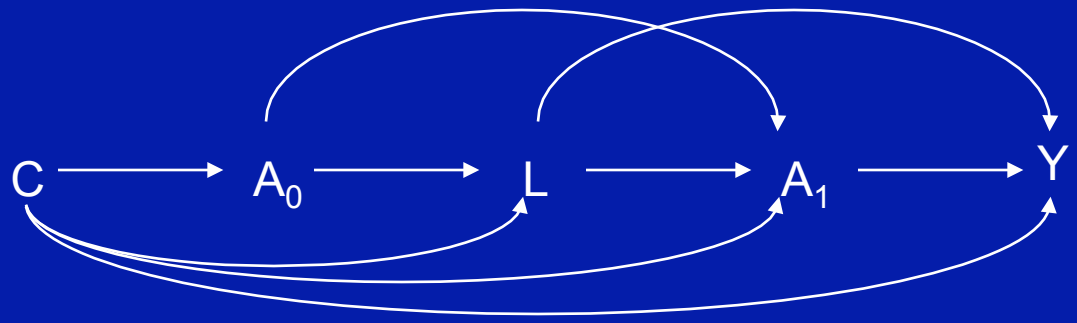
Causal Inference Principle II: If there is an intermediate variable between A and Y, we should not control for it.



If we 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

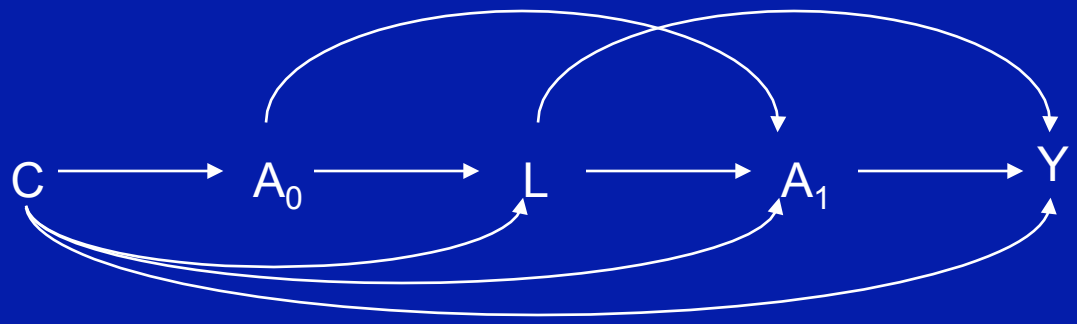
Suppose we want to know what the effects of interventions on loneliness at times 0 and 1 (denoted by A_0 and A_1) are on depression at time 2 (denoted by Y) with baseline covariates denoted by C and L the level of depressive symptoms between the two intervention times



Clearly we need to control for C as this is a common cause of treatment A_0 and outcome Y

Causal Inference with Longitudinal Data

Should we control for L?

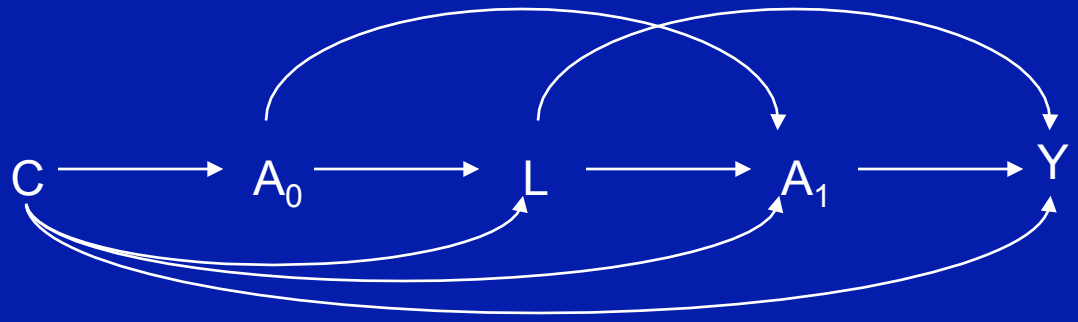


If we don't control for L, then we have an uncontrolled confounder because L is a common cause of treatment A_1 and outcome Y

This would violate causal inference principle I

Causal Inference with Longitudinal Data

What about L?



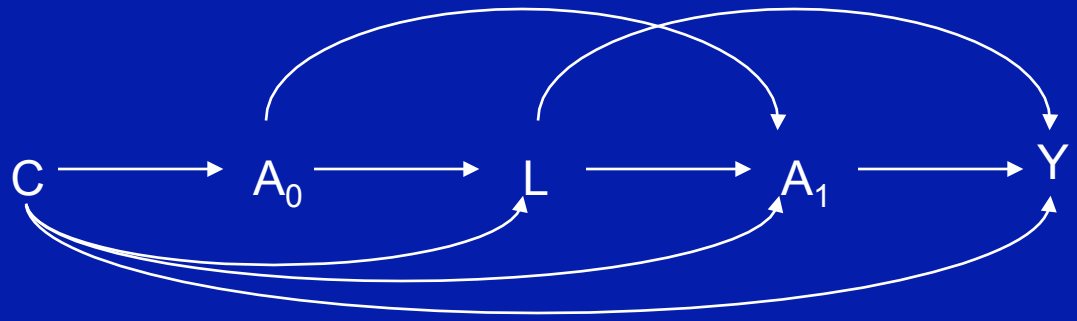
But if we do control for L then we have controlled for an intermediate variable between A_0 and Y

<u>This would violate causal inference principle II</u>

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_0 and A_1 on Y in this case



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

Causal Inference with Longitudinal Data

Instead of regression (i.e. a model for the outcome conditional on the covariates) we will use what is called a "marginal structural model" (a model for the counterfactual outcomes):

Let $Y_{a^0a^1}$ be the counterfactual value of Y for an individual under an intervention to set A_0 to a_0 and A_1 to a_1

Regression: $E[Y|A_0=a_0, A_1=a_1, C=c] = \mu + \beta_0 a_0 + \beta_1 a_1 + \beta_2' c$

MSM: $E[Y_{a^0a^1}] = \kappa + \gamma_0 a_0 + \gamma_1 a_1$

The MSM is for the counterfactual outcomes, not the observed outcomes, and the expectation is marginalized over the entire population (not conditional on the covariates)

Causal Inference with Longitudinal Data

MSM: $E[Y_{a^0a^1}] = \kappa + \gamma_0 a_0 + \gamma_1 a_1$

Because we do not observe $Y_{a^0a^1}$ for all possible values of a_0 and a_1 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:

(1) $Y_{a^0a^1} \perp\!\!\!\perp A_0 \mid C$

(i.e. the effect of A_0 on the final outcome Y is unconfounded given C)

(2) $Y_{a^0a^1} \perp\!\!\!\perp A_1 \mid \{C, A_0, L\}$

(i.e. the effect of A_1 on Y is unconfounded given baseline C, A_0 and the potential intermediate(s) denoted by L)

Causal Inference with Longitudinal Data

MSM: $E[Y_{a_0 a_1}] = \kappa + \gamma_0 a_0 + \gamma_1 a_1$

Robins showed that under these no-unmeasured-confounding assumptions we can obtain consistent estimators of κ , γ_0 and γ_1 (the parameters of the MSM) by fitting the regression model:

$E[Y|A_0=a_0, A_1=a_1] = \kappa + \gamma_0 a_0 + \gamma_1 a_1$

where each subject i is weighted by

$$\frac{1}{P(A_0=a_0^i|C=c^i)} \times \frac{1}{P(A_1=a_1^i|A_0=a_0^i, C=c^i, L=l^i)}$$

where a^i_0, a^i_1, c^i, l^i are the values for individual i of A_0, A_1, C and L respectively

Control for confounding is addressed by weighting rather than regression

(the weighted regression should use “sandwich” estimators of the standard errors to be valid; see SAS code later)

Causal Inference with Longitudinal Data

The weights

$$\frac{1}{P(A_0=a_0^i|C=c^i)} \times \frac{1}{P(A_1=a_1^i|A_0=a_0^i,C=c^i,L=l^i)}$$

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_0 and A_1 are binary then the probabilities could be obtained using a logistic regression

First a regression of A_0 on C

Second a regression of A_1 on $\{A_0,C,L\}$

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

$E[Y_{a_0a_1}] = \kappa + \gamma_0 a_0 + \gamma_1 a_1$

Causal Inference with Longitudinal Data

This approach to fitting the MSM still works if so called “stabilized weights” are used:

$$\frac{P(A_0=a_0^i)}{P(A_0=a_0^i|C=c^i)} \times \frac{P(A_1=a_1^i|A_0=a_0^i)}{P(A_1=a_1^i|A_0=a_0^i,C=c^i,L=l^i)}$$

These stabilized weights often result in reduced variance

If the exposure/treatment A_0 and A_1 are continuous then the probabilities are replaced by probability density functions (which we will use in the application below)

The approach described above extends to more than two times of treatment; an additional set of weights is calculated for each treatment time

Loneliness and Depression

The relationship between loneliness and depression as psychological constructs is complex, both constructs indicating negative affect, loneliness about one's social relationships and depression more generally

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

We use data 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

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 assess 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

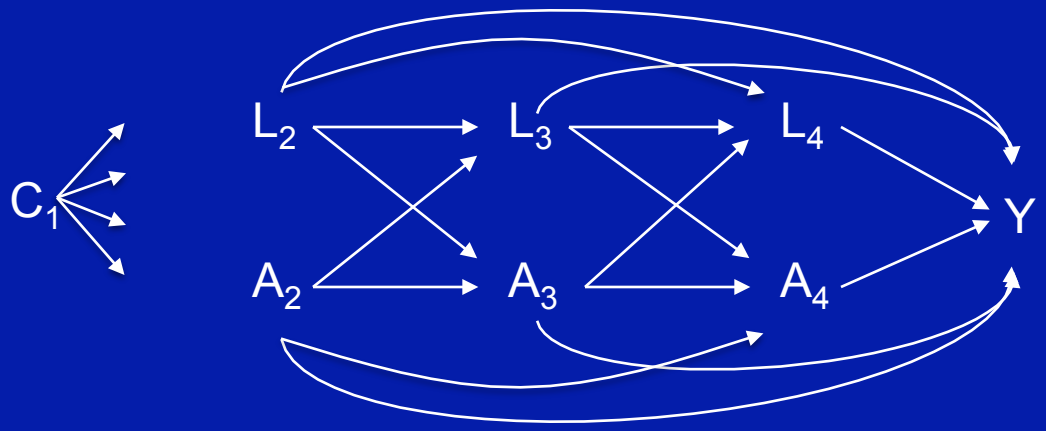
Loneliness and Depression

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

We first fit models for the ITP weights (loneliness is considered as a continuous exposure so we use linear regression for the weights):

```
proc reg data=depres;
   model uclaY2=uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1 age gender race1 race2 bincome years bmar;
   output out=depres student=rd2;
run;

proc reg data=depres;
   model uclaY3=uclaY2 cesdY2 swlssumY2 pmedY2 pcondY2 uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1
              age gender race1 race2 bincome years bmar;
   output out=depres student=rd3;
run;

proc reg data=depres;
   model uclaY4=uclaY3 cesdY3 swlssumY3 pmedY3 pcondY3 uclaY2 cesdY2 swlssumY2 swlssumY2
              pmedY2 pcondY2 uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1 age gender race1
              race2 bincome years bmar;
   output out=depres student=rd4;
run;
```

Loneliness and Depression

The PROC REG procedures in SAS gives standardized residuals and now we evaluate, for each individual, the normal probability density function at the value of the residual to obtain e.g. $1/P(A_2=a^i_2|C=c^i)$, $1/P(A_3=a^i_3|A_2=a^i_2,C=c^i,L_2=l^i_2)$, etc.

```
data depres;
  set depres;
  wd2=(2.718**(-.5*rd2*rd2))/2.506;
  wd3=(2.718**(-.5*rd3*rd3))/2.506;
  wd4=(2.718**(-.5*rd4*rd4))/2.506;
  ww=(1/wd2)*(1/wd3)*(1/wd4);
run;
```

See code at the end of the handout for the estimation of weights if treatment is binary rather than continuous

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

```
proc genmod data=depres;
   class caseid;
   model cesdY5 = uclaY2 uclaY3 uclaY4 / error=normal link=id;
   weight ww;
   repeated subject = caseid/ type = unstr;
run;
```

For the MSM $E[Y_{a^2a^3a^4}] = \kappa + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4$

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

Limitations

The analysis is subject to the following limitations/caveats:

MSMs work best with discrete treatment times

Both loneliness and depressive symptomatology vary continuously over time whereas data is only available on an annual basis

MSMs are subject to “no unmeasured confounding assumptions” described earlier; these will at best hold only approximately with observational data; the importance of potential violations can be assessed to a certain extent in sensitivity analysis

The IPTW technique can behave somewhat erratically when exposures are continuous; the technique is best suited for dichotomous or categorical treatments

Extensions

The IPTW technique for fitting MSMs can also be used to address censoring and drop out (see Robins et al., 2000 for an overview)

Marginal structural models can be used with other data types:
(1) Dichotomous outcomes (Robins et al., 2000)
(2) Time-to-event data (Hernán et al., 2000)
(3) Repeated measures data (Hernán et al., 2002)
(4) Mediation analysis (VanderWeele, 2009)

A good introductory article on MSMs is:

Robins JM, Hernán MA, Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11:550-560.

Note: The loneliness-depression data was re-analyzed adjusting for censoring, using a repeated measures marginal structural model (which tends to give more stable results with continuous exposures), and using stabilized weights and very similar results were obtained.

References

Cacioppo J, Hawkley L, Crawford L, Ernst J, Burleson M, Berntson G, Nouriani B, Spiegel D (2006). Loneliness within a nomological net: An evolutionary perspective. Journal of Research in Personality 40: 1054-1085.

Cacioppo J, Hughes M, Waite L, Hawkley L, Thisted R (2006). Loneliness as a specific risk-factor for depressive symptoms: Cross-sectional and longitudinal analyses. Psychology and Aging 21: 140-151.

Charig CR, Webb DR, Payne SR, Wickham OE. (1986). Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. BMJ 292:879-82.

Hernán M.A., Brumback B., Robins J.M. (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11:561-570.

Hernán M.A., Brumback B., Robins J.M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine*, 21:1689- 1709.

Hume, D. (1748). An Enquiry Concerning Human Understanding. Reprinted, 1958, LaSalle, IL: Open Court Press.

References

Lewis, D. (1973). Causation. Journal of Philosophy, 70:556-567.

Lewis, D. (1973). Counterfactuals. Harvard University Press, Cambridge.

Neyman, J. (1923). Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, Trans.) in Statist. Sci. 5, 463--72.

Robins JM, Hernán MA, Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11:550-560.

Robins J.M. (1999). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Halloran, M.E. and Berry, D., eds. NY: Springer-Verlag, pp. 95-134.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies, Journal of Educational Psychology, 66:688-701.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20:18-26.

SAS Code for Binary Exposures

Suppose the exposures of interest uclaY2, uclaY3, uclaY4 were binary then the following code could be used for the probabilities for the weights $1/P(A_2=a^i_2|C=c^i)$, $1/P(A_3=a^i_3|A_2=a^i_2,C=c^i,L_2=l^i_2)$, etc.

```
proc logistic data=depres descending;
   model uclaY2=uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1 age gender race1 race2 bincome years bmar;
   output out=depres predicted=pd2;
run;

proc logistic data=depres descending;
   model uclaY3=uclaY2 cesdY2 swlssumY2 pmedY2 pcondY2 uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1
              age gender race1 race2 bincome years bmar;
   output out=depres predicted=pd3;
run;

proc logistic data=depres descending;
   model uclaY4=uclaY3 cesdY3 swlssumY3 pmedY3 pcondY3 uclaY2 cesdY2 swlssumY2 swlssumY2
              pmedY2 pcondY2 uclaY1 cesdY1 swlssumY1 pmedY1 pcondY1 age gender race1
              race2 bincome years bmar;
   output out=depres predicted=pd4;
run;
```

For the weights one can then use the following code:

```
data depres;
  set depres;
   if uclaY2=1 then wd2=pd2; else wd2=(1-pd2);
   if uclaY3=1 then wd3=pd3; else wd3=(1-pd3);
   if uclaY4=1 then wd4=pd4; else wd4=(1-pd4);
   ww=(1/wd2)*(1/wd3)*(1/wd4);
run;
```

To fit the MSM one can use the following code

```
proc genmod data=depres;
    class caseid;
    model cesdY5 = uclaY2 uclaY3 uclaY4 / error=normal link=id;
    weight ww;
    repeated subject = caseid/ type = unstr;
run;
```

If stabilized weight were going to be used one would also fit models for the numerator of the weights $1/P(A_2=a^i_2)$, $1/P(A_3=a^i_3|A_2=a^i_2)$, etc. and could use the following code

```
proc logistic data=depres descending;
   model uclaY2=;
   output out=depres predicted=pn2;
run;
proc logistic data=depres descending;
   model uclaY3=uclaY2;
   output out=depres predicted=pn3;
run;
proc logistic data=depres descending;
   model uclaY4=uclaY3 uclaY2;
   output out=depres predicted=pn4;
run;

data depres;
  set depres;
   if uclaY2=1 then sw2=pn2/pd2; else sw2=(1-pn2)/(1-pd2);
   if uclaY3=1 then sw3=pn3/pd3; else sw3=(1-pn3)/(1-pd3);
   if uclaY4=1 then sw4=pn3/pd4; else sw4=(1-pn4)/(1-pd4);
   sw=w2*w3*w4
run;


proc genmod data=depres;
    class caseid;
    model cesdY5 = uclaY2 uclaY3 uclaY4 / error=normal link=id;
    weight sw;
    repeated subject = caseid/ type = unstr;
run;
```