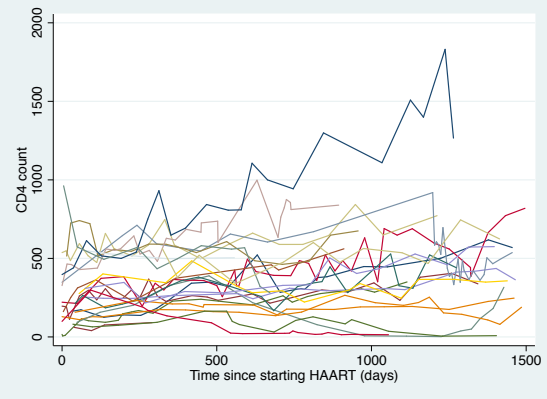
Longitudinal Data
Fall 2015



Chapter 10: Causal Inference and Marginal Structural Models

Instructors

Nick Jewell
(jewell@berkeley.edu)

GSI

Robin Mejia
(mejia@nasw.org)

References

Daniel, R.M., Cousens, S.N., Robins, J.M., De Stavola, B.L., Kenward, M.G., and Sterne, J.A.C. (2013) Methods for dealing with time-dependent confounding. Statistics in Medicine, 32, 1584-1618.

Hernán, M. and, Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5):550-560.

Hernán, M., Brumback, B., and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology, 11(5):561-570.

Mortimer, K. Neugebauer, R. van der Laan, M. and Tager, I. 2005. An Application of Model-Fitting Procedures for Marginal Structural Models. *Am. J. Epidemiology* **162**:382–388.

Prediction vs. Explanation

- Prediction – create a model that the clinician will use to help predict risk of a disease for an individual patient.

- Explanation – trying to investigate the causal association of a treatment or risk factor and a disease outcome.

- This chapter concerns studies where the goal is explanation (one can rarely derive accurate prediction for disease in humans).

The Caveat

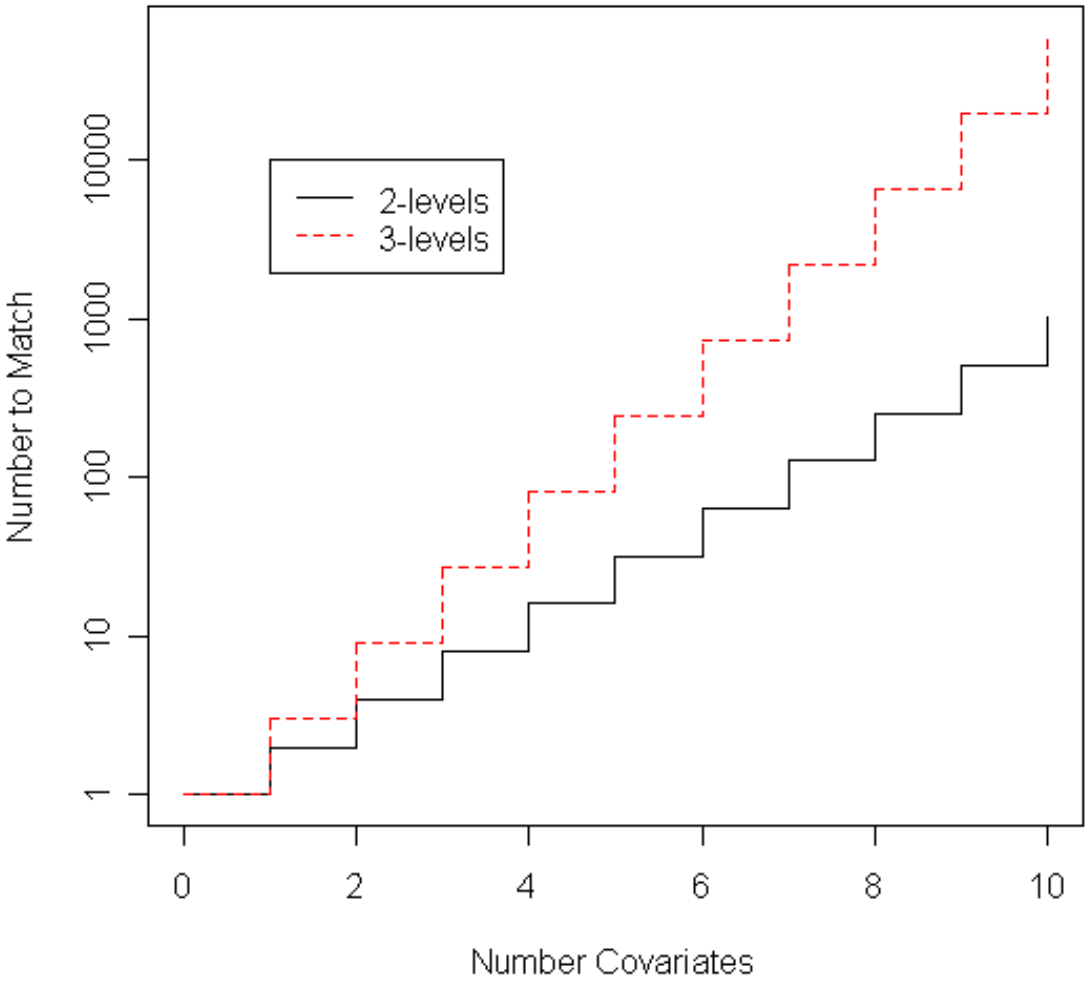
- Can't beat curse of dimensionality (unless one is lucky).

- Consider simplistic following scenario
  - W are the covariates (confounders) and all are categorical with the same number of level, e.g., 2 or 3 or 4...

  - In order to get *nonparametric causal inference*, one must have a perfectly matched unexposed person for every exposed person.

  - Given the number of confounders, how many subjects does one have to sample for every exposed person?

Number of unexposed per exposed subject one needs to sample to get perfect matching.



Theoretical Experiment

- Start with some hypothesis (changing risk factor for everyone in population changes the marginal risk of disease).

- Define a theoretical experiment that would address the hypothesis of interest.
  1. Makes explicit the specific hypothesis
  2. Defines the Full Data
  3. Defines the specific parameter of interest
  4. Leads ultimately to estimators from the observable data and the necessary identifiability assumptions

The problem with observational studies: lack of randomization

- If one has a treatment, or risk factor, with two levels (A and B), no guarantee that study populations (those getting A and B) will be roughly equivalent (in risk of the disease of interest).

- In a perfect world can given everyone in study level A, record outcome, reset clock and then give level B.

- Randomization means one can interpret estimates as if this is precisely what was done.

Counterfactuals

- Even defining statistically what a "causal" effect is, is not trivial.

- One way that leads to practical methods to estimate causal effects is to define ***COUNTERFACTUALS***

- Assume that the "full" data would be, for every subject, one could observe the outcome of interest for each possible level of the treatment (or risk factor) of interest.

Counterfactuals, CONT.

- So, if *Y* is the outcome, *A* is the tx of interest, then the best statistical situation is one where one observes, for each subject, Y_a, for each treatment level A=*a*.

- For example if there is simply two levels of exposure (eg, cigarettes *A=1* mean yes and *=0* is no), then each subject has in theory two counterfactuals, Y_0 and Y_1.

- These are called counterfactuals because, they are the outcomes on might observe if, counter to fact, one could set the clock back and re-start the whole experiment with a new *a*.

- To estimate specific causal effects, we then define parameters that relate, for instance, how the means of these counterfactuals differ as one changes *a.*

Causal Parameters in Point Tx studies

- Point treatment studies are those where the effect of interest refers to a treatment, risk factor, ... at one point in time.

- Time-dependent treatment studies are those where the effect of interest is a time-course of treatment or exposure (more in a bit).

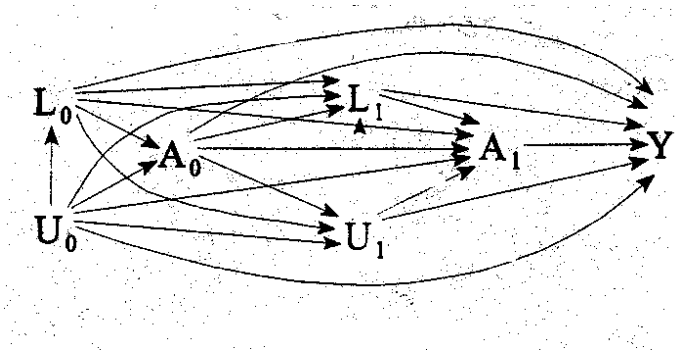
- Types of effects (or parameters) one might want to estimate are: total effects, direct and indirect effects, dynamic treatment regimes....

Why Longitudinal Studies create Unique Challenges for Causal Inference

- We've talked about how longitudinal studies provide opportunities to disentangle cause and effect that cross-sectional studies lack.

- However, if we are interested in how the history of treatment (or exposure) is causally associated with a future outcome, a new problem arises, not inherent in cross-sectional studies (that usually are interested in how exposure or treatment at one time affect an outcome).

- Let's say we have the following data structure:



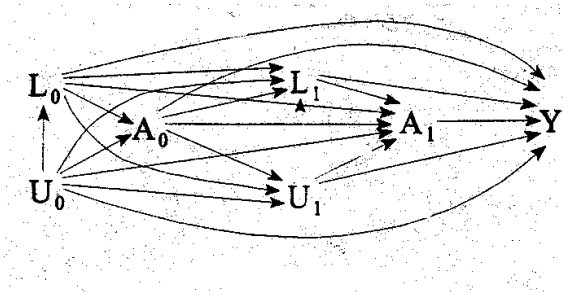
with L_j the confounders (current CD4 count), A_j the exposures of (anti-retroviral therapy) and Y the outcome of interest (say viral load). For now ignore the U_j's.

- Also, assume the association of interest is on the total amount of anti-retroviral therapy ($A_0+A_1=0,1$ *or* 2) and the viral load at the end of the study, Y.

Why Longitudinal Studies create Unique Challenges for Causal Inference

- What would normally do to adjust for confounding? Simply put the confounders in the model:

$$E\left[Y \mid A_1 + A_0 = a, L_0 = l_0, L_1 = l_1\right] =$$
$$\beta_0 + \beta_1 a + \beta_2 l_0 + \beta_3 l_1$$



where β_1 is the coefficient of interest now.

- The problem is, this doesn't work, because, L_1 is both a confounder of A_1, but is on the causal pathway of A_0 - so, one rule says adjust for it, the other says don't

- The bottom line is that adjusting for the time-dependent confounders in this case does not generally work, i.e., does not result in an interpretable causal parameter.

- This lead to researchers to think of 1) parameters (other than β_1 that have a causal interpretation and 2) ways of estimating these "new" parameters.

Start with Point Tx Case

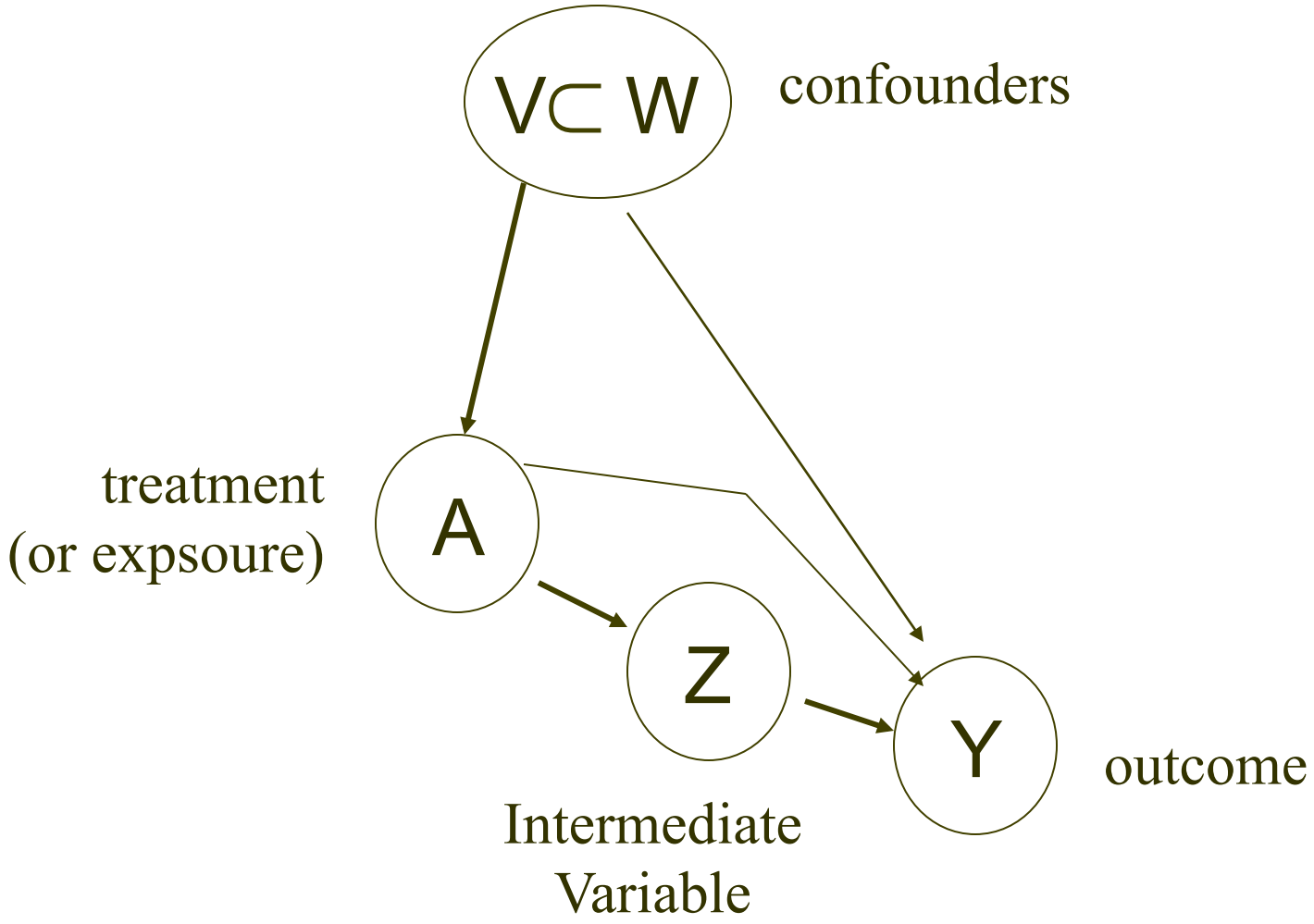
- Although the problem just discussed is unique to longitudinal data (and specifically longitudinal exposures of interest), easiest to understand the techniques developed to get around this problem in the so-called *point treatment* case.

- That is, when one has an exposure or treatment (or whatever) of interest measured at one time point, has confounders measured as well and wants to find the causal association of the factor of interest on an outcome (assumed to come after).

- We will discuss the different kinds of causal parameters in the pt. treatment case first (counterfactual distributions).

- The methods are nearly identically in the time-dependent treatment (exposure) setting.

General Causal Graph For a Point Treatment Study



Total Effects in Point Tx Studies

- Parameters of the distribution of the counterfactuals: Y_a
- Examples for binary A (0 or 1):
  - $E[Y_1]-E[Y_0]$
  - $E[Y_1]/E[Y_0]$
- Examples for binary A and Y (Causal OR)
  - $E[Y_1](1-E[Y_0])/\{E[Y_0](1-E[Y_1]\}=$

$$P[Y_1=1](1-P[Y_0=1])/\{P[Y_0=1](1-P[Y_1=1]\}$$

Total Effects in Point Tx Studies, cont.

- Regression models (marginal structural models – MSM's) relating mean of Y_a vs *a:* $E[Y_a]=m(a|\beta)$ *(*continuous or ordered categorical *a),* e.g.,

$$m(a \mid \beta) = \beta_0 + \beta_1 a$$

or

$$m(a \mid \beta) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 a))}$$

- Stratified MSM's

$$m(a, V \mid \beta) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$$

Dynamic Treatment Regimes

- $\mathbf{D} = \{w \rightarrow d(w) \in \mathbf{A} : d\}$ is the abstract set of possible dynamic treatment regimes.

- $\mathbf{D} = \{w \rightarrow d_\theta(w) : \theta\}$ is slightly less abstract (that is, your rule will controlled by some constant (or perhaps a vector if W is a vector), θ.

- Finally, the specific rule could be:

$$d_\theta(w) = I(w > \theta)$$

A Dynamic Treatment Regime Parameter of Interest

- Parameter of interest for this might be:

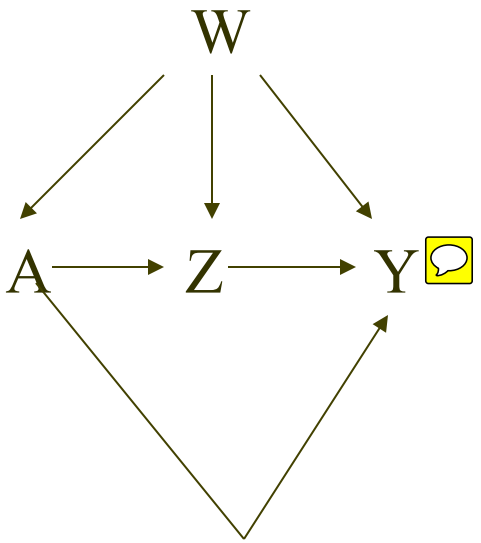
$$E[Y_{d_\theta}]=P(Y_{d_\theta}=1)$$

  which represents the expected outcome in a population where all subjects are subject to treatment rule, $d_{\theta.}$

- Example is to put patient on cholesterol lowering drugs (the A = yes or no) if cholesterol (the W) is > 200 ($\theta_{.}=200$) and the outcome is heart disease (the Y).

Direct Effects

- Need to define counterfactuals for both endogenous variables, *Z* and *Y*.
- Y_a = the counterfactual outcome if receives *A*=*a*.
- Y_{az} = the counterfactual if receives *A*=*a*, *Z*=*z*.
- Z_a = counterfactual *Z* if receives *A*=*a*.
- Note, $Y_a = Y_{aZa}$
- Finally, Y_{aZa^*}, $a^* \neq a$ is the counterfactual if receives *A*=*a* but if the *Z*, counter to fact, the subject had received a different *a* (*a**).



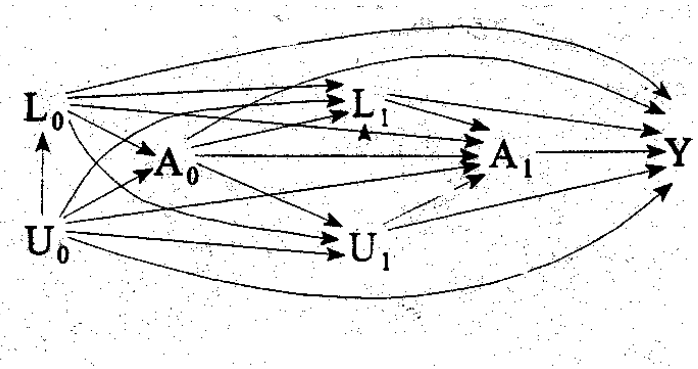
Direct Effects

- First, let A =0 be a reference group
- Total Effect: $E[Y_1]-E[Y_0]$
- Pure Direct Effect: is the difference between the exposed (A=*a*) and unexposed (*A=0*) if the intermediate variables is "set" to its value when *A=0*:

$$E\left[Y_{aZ_a}\right] - E[Y_0] =$$
$$E\left[Y_{aZ_a}\right] - E\left[Y_{0Z_0}\right]$$

Time-Dependent Treatments

- Measurements made at regular times, 0,*1,2,...,K*.
- *A(j)* is the treatment (or exposure) dose on the *j*th day.
- Y is outcome measured at the end (only once) at day *K+1*.
- $\overline{A}(j) = (A(0), A(1), ..., A(j))$ is the history of treatment as measured at time *j.*
- $\overline{L}(j) = (L(0), L(1), ..., L(j))$ is the history of the potential confounders measured at time *j*.



Time-dependent counterfactuals of Interest

$$Y_{\bar{a}}, \quad \bar{a} = (a(0), a(1), \ldots, a(K))$$

- Must know define counterfactuals with regard to a whole vector of possible treatments.

$$\bar{a} = (0,0,0\ldots,0), \quad \bar{a} = (1,1,1\ldots,1)$$

- If A is binary (e.g., yes/no) there are 2K possible counterfactuals, e.g.,

Example of MSM for time-dependent Tx

- Choose a reasonable model that relates counterfactual mean to treatment history.

- Example:

$$E[Y_{\bar{a}}] = \beta_0 + \beta_1 sum(\bar{a})$$

- where

$$sum(\bar{a}) = \sum_{j=1}^{K} a(j)$$

Estimators
of MSM's in Point Treatment Studies

- A denotes a "treatment" or exposure of interest – assume categorical.
- W is a vector (set) of confounders
- Y is an outcome
- Define observed data is X= (Y,W)
- Y_a are the counterfactual outcomes of interest
- The "full data" is: $X^{FULL} = (X_a, a \in A)$

Key Assumptions

1. Consistency Assumption: observed data, O is $O=(A,X_A)$ – i.e., the data for a subject is simply one of the counterfactual outcomes from the full data.

2. Randomization Assumption: $A \perp Y_a \mid W, \forall a$ so no unmeasured confounders for treatment. In other words: within strata of W, A is randomized

Key Assumptions, cont.

1. Experimental Treatment Assignment: all treatments are possible for all members of the target population, or:

$$P(A = a \mid W) > 0,$$

for all W.

Likelihood of Data in simple Point Treatment

- Given the assumptions, the likelihood of the data simplifies to:

$$L(O) = P(Y \mid A,W)P(A \mid W)$$

- Factorizes into the distribution of interest and the treatment assignment distribution.

- The G-computation formula works specifically with parameters of $P(Y|A,W)$ estimated with maximum likelihood and ignores the treatment assignment distribution.

G-computation Approach

- Given assumptions, note that $P(Y|A=a,W) = P(Y_a|W)$ or $E(Y|A=a,W) = E(Y_a|W)$.

- Then, $$E[Y_a] = \int_w E[Y_a \mid W = w] dP(w)$$

- Which leads to our G-comp. estimate of the counterfactual mean in this simple context.

$$\hat{E}[Y_a] = \sum_{i=1}^{n} \frac{1}{n} \hat{E}[Y \mid A = a, W = W_i]$$

- Regress $\hat{E}[Y_a]$ vs. a to get an estimate of MSM.

Inverse Probability of Treatment Weighted (IPTW) estimator

- The G-comp. approach models is three-steps
  1. Model E[Y|A,W]
  2. Estimate $E[Y_a]$
  3. Regress $\hat{E}[Y_a]$ vs. *a* to get an estimate of MSM (e.g., $m(a \mid \beta) = \beta_0 + \beta_1 a$)

- The IPTW uses a different approach that instead models treatment assignment to adjust for confounding and uses these as weights in regression.

- Define *g(a|W)* to be the *P(A=a|W)*.

IPTW Estimating Function

- General Estimating Function is (for stratified by V MSM's):

$$\frac{h(A,V)}{g(A|W)}(Y - m(A,V|\beta))$$

- For unstratified MSM's

$$\frac{h(A)}{g(A|W)}(Y - m(A|\beta))$$

- Example (linear model) $\frac{\begin{pmatrix}1\\A\end{pmatrix}}{g(A|W)}(Y - m(A|\beta))$

More details on linear model

- Ignoring confounding by W and assuming:

$$m(a \mid \beta) = \beta_0 + \beta_1 a$$

then finding the estimates of β can be equivalent to minimizing the residual sum of squares:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - [\beta_0 + \beta_1 A_i])^2$$

- Minimize by taking the derivative with respect to β, and setting to 0, solving for β

$$\frac{d}{d\beta} \sum_{i=1}^{n} (Y_i - [\beta_0 + \beta_1 A_i])^2 = 0 \propto \begin{pmatrix} \sum_{i=1}^{n} (Y_i - [\beta_0 + \beta_1 A_i]) = 0 \\ \sum_{i=1}^{n} A_i (Y_i - [\beta_0 + \beta_1 A_i]) = 0 \end{pmatrix}$$

More details on linear model

- For each observation: get

$$\begin{pmatrix} (Y_i - [\beta_0 + \beta_1 A_i]) \\ A_i (Y_i - [\beta_0 + \beta_1 A_i]) \end{pmatrix} = 0, \text{ or, } \begin{pmatrix} 1 \\ A_i \end{pmatrix} (Y_i - [\beta_0 + \beta_1 A_i])$$

or,

$$h(A_i)(Y_i - [\beta_0 + \beta_1 A_i]) = 0, \; h(A_i) = \begin{pmatrix} 1 \\ A_i \end{pmatrix}$$

- Then, one option to estimate this MSM with confounding (by W) is:

$$\frac{h(A_i)}{g(A_i \mid W_i)} (Y_i - m(A_i \mid \beta))$$

with weight $1/g(A_i|W_i)$.

What it means?

- For generalized linear models (linear, logistic, Poisson) can use the estimating equation the programs already have built in by just adding a weight which can involve only an estimate of: $1/g(A_i|W_i)$.

- Of course, that also means one needs an estimate of $g(a|W) \equiv P(A=a|W)$ to plug in.

- Easiest to derive when A is discrete (most stat papers and applications of MSM's have categorical A's).

- Intuitive interpretation of why the procedure works (get's rid of confounding by *W*) is that is changes the distribution of *W* in strata defined by *A* so that all strata of *A* having an equal distribution of *W*.

or,

- Then, one option to estimate this MSM with confounding (by W) is:

with weight $1/g(A_i|W_i)$.

Simple Example (A = Tx, Y = Outcome W = Confounder) – all binary

-> W = 0

A	Y = 0	Y = 1	Total
0	10	40	50
	20.00	80.00	100.00
1	30	20	50
	60.00	40.00	100.00
Total	40	60	100
	40.00	60.00	100.00

-> W = 1

A	Y = 0	Y = 1	Total
0	16	24	40
	40.00	60.00	100.00
1	252	108	360
	70.00	30.00	100.00
Total	268	132	400
	67.00	33.00	100.00

Crude association

```
. cs Y A [freq=pop]

                 | A                       |
                 |   Exposed   Unexposed   |     Total
-----------------+-------------------------+----------
           Cases |       128          64   |       192
        Noncases |       282          26   |       308
-----------------+-------------------------+----------
           Total |       410          90   |       500
                 |                         |
            Risk |  .3121951    .7111111   |      .384
                 |                         |
                 |      Point estimate     |  [95% Conf. Interval]
                 |-------------------------+----------------------
 Risk difference |         -.398916        | -.5027443   -.2950877
      Risk ratio |         .4390244        |  .3612844    .5334922
 Prev. frac. ex. |         .5609756        |  .4665078    .6387156
 Prev. frac. pop |              .46        |
                 +-------------------------------------------------
                           chi2(1) =    49.65  Pr>chi2 = 0.0000
```

Getting estimate of P(A|W) and weights

```
. logit A W [freq=pop]


Logit estimates                                   Number of obs   =        500
                                                  LR chi2(1)      =      72.70
                                                  Prob > chi2     =     0.0000
Log likelihood = -199.34791                       Pseudo R2       =     0.1542

------------------------------------------------------------------------------
           A |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |   2.197225   .2603417     8.44   0.000     1.686964    2.707485
       _cons |   3.79e-16          .2     0.00   1.000    -.3919928    .3919928
------------------------------------------------------------------------------

. predict p
(option p assumed; Pr(A))

. replace p = 1-p if A==0
(2 real changes made)

. gen wt = 1/p

. gen newpop = int(wt*pop)
```

Look at “new” population

	A	W	Y	pop	p	wt	newpop
1.	0	0	0	10	.5	2	20
2.	0	0	1	40	.5	2	80
3.	1	0	0	30	.5	2	60
4.	1	0	1	20	.5	2	40
5.	0	1	0	16	.1	9.999998	159
6.	0	1	1	24	.1	9.999998	239
7.	1	1	0	252	.9	1.111111	280
8.	1	1	1	108	.9	1.111111	120

Association of A and W in old and new population

```
. cs A W [freq=pop]

                | W                     |
                |   Exposed   Unexposed |     Total
----------------+-----------------------+----------
          Cases |       360          50 |       410
       Noncases |        40          50 |        90
----------------+-----------------------+----------
          Total |       400         100 |       500
           Risk |        .9          .5 |       .82
                |                       |
                |      Point estimate   |  [95% Conf. Interval]
                |-----------------------+----------------------
     Risk ratio |               1.8     |  1.475633    2.195669
                +-------------------------------------------------
                           chi2(1) =    86.72  Pr>chi2 = 0.0000

. cs A W [freq=newpop]

                | W                     |
                |   Exposed   Unexposed |     Total
----------------+-----------------------+----------
          Cases |       400         100 |       500
       Noncases |       398         100 |       498
----------------+-----------------------+----------
          Total |       798         200 |       998
           Risk |  .5012531          .5 |   .501002
                |                       |
                |      Point estimate   |  [95% Conf. Interval]
                |-----------------------+----------------------
     Risk ratio |         1.002506      |  .8586389    1.170479
                +-------------------------------------------------
                           chi2(1) =     0.00  Pr>chi2 = 0.9747
```

Association of A and Y in new population

```
. cs Y A [freq=newpop], or

                |  A                      |
                |   Exposed   Unexposed   |     Total
----------------+-------------------------+----------
          Cases |       160         319   |       479
       Noncases |       340         179   |       519
----------------+-------------------------+----------
          Total |       500         498   |       998
                |                         |
           Risk |       .32    .6405622   |  .4799599
                |                         |
                |      Point estimate     |  [95% Conf. Interval]
                |-------------------------+----------------------
 Risk difference |        -.3205622        | -.3792806   -.2618439
     Risk ratio |         .4995611        |  .4326862     .576772
 Prev. frac. ex. |         .5004389        |   .423228    .5673138
 Prev. frac. pop |         .2507209        |
     Odds ratio |         .2640605        |  .2031715    .3431976  (Cornfield)
                +-------------------------------------------------
                          chi2(1) =   102.72  Pr>chi2 = 0.0000
```

Diarrhea among HIV+ patients

- Of interest was the contribution of medication to the prevalence of diarrhea among AIDS patients.
- The assumption was that the remainder of the diarrhea was caused by pathogens.
- So, the parameters of interest were $E[Y_a]=P(Y_a=1)$, *a* represents the levels of medication, and $Y_a = 1$ if patient has diarrhea, 0 otherwise.

Step 1: Choosing Marginal Model

- First, one must choose a model for the marginal treatment (risk) probabilities.
- In this case, the outcome, Y, is binary (diarrhea) and the risk variable of interest, A, is categorical (medical risk, 4 levels: very low, low, medium and high).
- What we want is a model that will describe the marginal risk of diarrhea for each of these 4 levels.

Step 1, cont.

- One choice is:

$$\text{logit}\{E[Y_a]\}=\text{logit}\{P(Ya=1)\}=\beta_0+\beta_1 a_1+\beta_2 a_2+\beta_3 a_3$$

- or,

$$P(Y_a=1)=m(a\mid\beta)=\frac{1}{1+\exp(-(\beta_0+\beta_1 a_1+\beta_2 a_2+\beta_3 a_3))}$$

- a_1 - a_3 are dummy variables representing the levels of medical risk.

- a_1=1 if a=low (0 otherwise), a_2=1 if a=medium, and a_3 =1 if a = high.

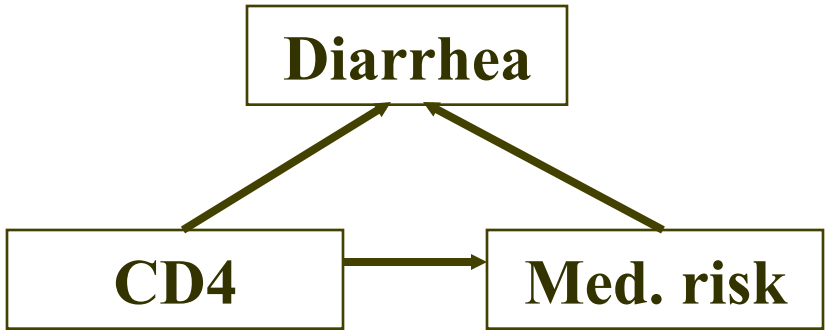
Step 1, cont.

- So, $P(Y_0=1) = 1/(1+exp(-(\beta_0)))$ = probability of diarrhea if everyone had level 0 of medication (lowest risk).

- $P(Y_3=1) = 1/(1+\exp(-(\beta_0+\beta_3)))$ = probability of diarrhea if everyone had level 3 of medication (highest risk).

Step 2: Choosing confounders.

- This study has the possibility of confounding because patients with more advanced illness are both given medications with higher risk of diarrhea and suffer from greater chance of diarrhea from infection.



Step 3: Adjusting for confounders by modeling probability of treatment.

- Let *W* be the list of variables that, based on causal graph, one believes confounds the causal effect of *A* on *Y*.
- In the diarrhea example, W is CD4 count.
- MSM works (adjusts the estimates) by including estimated sampling weights which are the 1/ P(A|W=w).

Modeling probability of treatment, X.

- In our case, A is a categorical variable with four levels and W is continuous.
- One way to model P(A|W) is with multinomial logistic regression:

$$P(A = a \mid W) = \frac{\exp(W\beta^{(a)})}{\exp(W\beta^{(0)}) + \exp(W\beta^{(1)}) + \exp(W\beta^{(2)}) + \exp(W\beta^{(3)})}$$

Modeling P(A=a | W) in Stata

```
. mlogit medcat cd4, basecategory(0)

Multinomial regression                          Number of obs =     211
                                                chi2(3)       =    1.97
                                                Prob > chi2   = 0.5780
Log Likelihood = -238.54152                     Pseudo R2     = 0.0041

------------------------------------------------------------------------------
  medcat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
0.07-0.21|
     cd4 |  -.0008203   .0012283    -0.668   0.504    -.0032277    .0015871
   _cons |   .8739337   .5995797     1.458   0.145    -.3012209    2.049088
---------+--------------------------------------------------------------------
0.21-0.43|
     cd4 |  -.0008316   .0010207    -0.815   0.415     -.002832    .0011688
   _cons |    2.28173   .5122226     4.455   0.000     1.277792    3.285668
---------+--------------------------------------------------------------------
> 0.43   |
     cd4 |  -.0015786   .0011637    -1.356   0.175    -.0038595    .0007023
   _cons |    1.63735    .556692     2.941   0.003     .5462543    2.728447
------------------------------------------------------------------------------
(Outcome medcat== < 0.07 is the comparison group)
```

Getting Estimates of 1/P(A=a|W)

```
. predict p1 p2 p3 p4, p

. gen pwght=1/p1

. replace pwght = 1/p2 if medcat==2

. replace pwght = 1/p3 if medcat==3

. replace pwght = 1/p4 if medcat==4
```

Get the probabilities for above model

.	cd4	diarrhea	p1	p2	p3	p4
1.	794	0	.1139346	.1422799	.5684277	.1753577
2.	200	0	.0663043	.1348798	.5329911	.2658248
3.	430	0	.0824113	.1387831	.550747	.2280587
4.	415	1	.0812754	.1385672	.5497382	.2304192
5.	716	0	.1065159	.1418175	.5657657	.1859009
6.	213	0	.0671422	.1351335	.5341216	.2636027
7.	174	1	.0646534	.1343609	.5306862	.2702994
8.	310	0	.0736634	.1369031	.5420853	.2473482
9.	264	1	.0705111	.1360912	.5384135	.2549842
10.	9	0	.0549289	.1307218	.5147436	.2996057
11.	436	1	.0828691	.1388679	.5511445	.2271186
12.	51	0	.0572841	.1317035	.5190113	.2920011
13.	705	0	.1054997	.1417395	.5653399	.1874209
14.	579	0	.0943769	.1406234	.5595861	.2054137
15.	437	1	.0829456	.1388819	.5512104	.2269621
16.	848	0	.1192923	.1425067	.5699013	.1682997
17.	196	0	.0660481	.134801	.5326403	.2665106
18.	218	0	.0674667	.13523	.5345525	.2627507
19.	928	0	.1275697	.1427017	.5715237	.158205
20.	486	.	.0867616	.1395399	.5543227	.2193758

Step 4: Do probability weighted regression.

- For each individual, create weights:

$$WGHT_i = \frac{1}{\hat{P}(A_i \mid W_i)}.$$

- This has the effect (sort of) of adding many observations when the estimated probability is low and few if the probability is high.

Step 4, cont.
Confounding as Sampling Bias

- One can think of confounding as a form of sampling bias.

- Consider the study of a treatment for AIDS patients where the physicians considers the health of the patients when deciding the Tx to prescribe.

- By re-weighting the sample, based on the probability of being sampled on factors related to the health, one gets a sample evenly distributed with respect to variables that affect the outcome.

Doing this for the Diarrhea Data

```
logit diarrhea a1 a2 a3 [pweight=pwght]
```

```
Logit Estimates                                   Number of obs =    210
                                                  chi2(3)       =   4.76
                                                  Prob > chi2   = 0.1900
Log Likelihood = -140.46701                       Pseudo R2     = 0.0269

------------------------------------------------------------------------------
             |               Robust
diarrhea | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          a1 |   1.279149   .8527155     0.369   0.712      .3463273    4.724494
          a2 |   1.612001   .9248658     0.832   0.405      .5235987    4.962857
          a3 |   2.867675   1.765869     1.711   0.087       .857775     9.58708
------------------------------------------------------------------------------
```

Note pwght is $1/\hat{P}(A_i \mid W_i)$

Comparing no adjustment to MSM

Estimates of $P(Y_a=1)$ and OR's assuming no confounding and adjusted for CD4 count using MSM.

	Assuming No Confounding		MSM	
Medical Risk	**OR(95% CI)**	**P(Ya=1)**	**OR(95% CI)**	**P(Ya=1)**
<0.07	1(ref.)	0.29	1(ref.)	0.34
0.07-0.21	1.55(0.43,5.64)	0.39	1.27(0.35,4.72)	0.40
0.21-0.43	1.98(0.65,5.99)	0.45	1.61(0.52,4.96)	0.45
>0.43	3.60(1.10,11.80)	0.60	2.87(0.86,9.59)	0.60

Inference

- To get standard errors for the coefficient estimates, use BOOTSTRAPPING

- Repeat 1000 times
  1. randomly re-sample the data with replacement
  2. estimate weights
  3. perform weighted regression
  4. store coefficient estimates
  5. go back up to 1

- Getting the sample standard deviation of these 1000 sets of coefficients provides a standard error.

Other Data Types

- Can apply this general methodology to any regression scenario (linear, Cox regression, etc.).
- Almost always one can do with existing software, as long as program allows user-supplied regression weights.
- Works very similarly in time-dependent case (but, wts a bit more complicated)