Introduction to Marginal Structural Models

William D. Dupont

Recommended reading

Hernán MA, Robins JM (2016). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming. Preprint available at http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/ . Last accessed Aug. 20, 2015

Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Stat Med*. 2013; **32(9):**1584-618.

Copies of these slides are posted at
http://biostat.mc.vanderbilt.edu/WilliamDupontJan202016

Topics

- Counter factual events and the definition of causality
- What is a marginal structural model (MSM)?
- Simple simulation study with a confounding variable
  - Classical analysis vs MSM
- Simple longitudinal example
  - Classical analysis vs MSM
- More complex longitudinal example where classical methods do not work.
- Applying MSMs to real longitudinal studies

Counterfactual Events

Let Y_i be the outcome for the i^{th} patient,
a be some exposure, and
a_i be the actual exposure of the i^{th} patient.

Then Y_i^a is the potential outcome that would have been observed for the i^{th} patient if, possibly contrary to fact, she received exposure a.

For example:

Y_i denotes whether or not the i^{th} patient develops bronchiolitis

$$a = \begin{cases} 1 & \text{patient receives immunoprophylactic treatment} \\ 0 & \text{patient does not receive this treatment} \end{cases}$$

$$a_i = \begin{cases} 1 \text{ if } i^{th} \text{ patient receives treatment} \\ 0 \text{ if } i^{th} \text{ patient does not receive treatment} \end{cases}$$

$Y_i = Y_i^{a_i}$ is the bronchiolitis outcome that actually happened to the i^{th} patient given her true treatment.

Y_i^1 is the bronchiolitis outcome that would have happened to the i^{th} patient if she had been treated

Y_i^0 is the bronchiolitis outcome that would have happened to the i^{th} patient if she had not been treated.

Y_i^a is the counterfactual event that would have occurred given an exposure that may or may not have happened

Reality

Sally gets treated $a_i = 1$ → Sally gets bronchiolitis $Y_i = Y_i^{a_i} = Y_i^1 = 1$

Counterfactual land

Sally does not get treated $a = 0$ → Sally gets bronchiolitis $Y_i^0 = 1$

$\Pr\left[Y_i^0 = 1\right] = 1 - \Pr\left[Y_i^0 = 0\right]$

Sally does not get bronchiolitis $Y_i^0 = 0$

Causal Relationships

An treatment A has a causal effect on outcome Y if

$$\Pr\left[Y_i^a\right] \neq \Pr\left[Y_i^{a'}\right]$$

for exposures a and a'

For example, treatment has a **causal** effect on disease if

Pr[**Sally** gets disease given she **is** treated]
≠ Pr[**Sally** gets disease given she **is not** treated]

Note that this is very different from

Pr[**Sally** gets disease after **being** treated]
≠ Pr[**Johny** gets disease after **not being** treated]

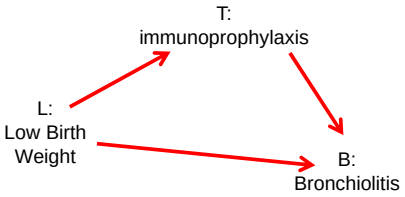
due to possible **confounding by indication**.

Causal inference gurus think of observational data as having **massive amounts of missing data**

If we are lucky and or diligent we could observe all of the data that actually occurred.

In addition there is at least twice as much missing data from **counterfactual outcomes that did not occur as a consequence of exposures not received.**

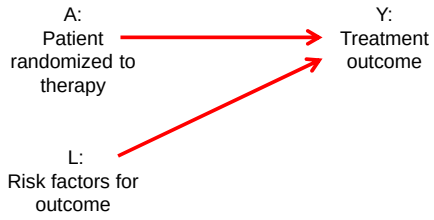
Directed Acyclic Graphs (DAGs)



Causally related events are indicted by arrows

Here L confounds the causal effect of T on B

Randomized Clinical Trials



Since treatment assignment is controlled by randomization there can be no causal path between L and A

Hence, any association between A and Y must be causal.

Exchangeability

A model with treatment A, covariate L and outcome Y is **exchangeable** if the counterfactual outcome $Y_i^a \mid L_i$ is independent of A.

That is $\Pr\left[Y_i^a \mid L_i, A_i\right] = \Pr\left[Y_i^a \mid L_i\right]$

In other words, knowing the actual treatment that a patient receives tells me nothing about her counterfactual outcome.

In a RCT treatment is assigned at random, and it does not matter which group receives which treatment. Hence, the treatment groups are exchangeable and

$$\Pr\left[Y_i^a = 1\right] = \Pr\left[Y_i^a = 1 \mid A_i\right]$$

Example of Non-Exchangeability

Here, immunoprophylaxis is much more likely to be given to babies at high risk of bronchiolitis.

Hence, if I know Johny did not receive T then I know that he is less likely to develop B than if I did not know his actual treatment.

In other words, babies who did, and did not, receive T are **not exchangeable**, since treated babies are inherently at greater risk than untreated babies.

Marginal Structural Models

A model is

structural if it regresses a counterfactual outcome Y_i^a against treatment and other covariates.

marginal if it only involves treatment covariates.

In a RCT if I use simple logistic regression to regress outcome Y_i against treatment A_i then I have a **marginal structural model (MSM)**

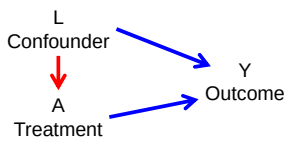
$$\underbrace{\text{logit}\left[\pi\left(Y_i^a\right)\right] = \text{logit}\left[\pi\left(Y_i\right)\right]}_{\text{True due to exchangeability}} = \underbrace{\alpha + \beta \times A_i}_{\text{No confounders in model}}$$

Simple Confounder Example

$\Pr[L=1]=0.6$

$\Pr[A=1 \mid L=1]=0.9$

$\Pr[A=1 \mid L=0]=0.3$



		A		
		0	1	Total
L	0	278,998	120,050	399,048
				39.9%
	1	59,728	541,224	600,952
				60.1%
Total		338,726	661,274	1,000,000
				100%

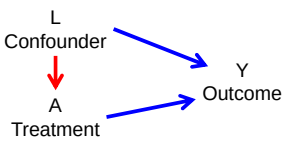
Simple Confounder Example

Sample Size = 1,000,000

$\Pr[L=1]=0.6$

$\Pr[A=1 \mid L=0]=0.3$

$\Pr[A=1 \mid L=1]=0.9$

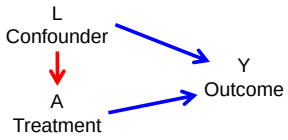


		A		
		0	1	Total
L	0	278,998	120,050	399,048
		69.9%	30.1%	100%
	1	59,728	541,224	600,952
		9.9%	90.1%	100%
Total		338,727	661,274	1,000,000

Odds ratio for A given L = 21.1

Simple Confounder Example

We can't do anything about the relationship between L and Y or A and Y.



The key idea of MSMs is that we take a weighted sample of subjects in which the relationship between L and A is destroyed.

Inverse Probability Treatment Weights

We do this using inverse probability treatment weights (IPTWs)

		A		
		0	1	Total
L	0	278,998	120,050	399,048
	1	59,728	541,224	600,000
Total		338,726	661,274	1,000,000

Inverse Probability Treatment Weights

L	A	True Sample Size	Pr[A|L] True	Pr[A|L] Estimated	Estimated IPTW		IPTW × N
0	0	278,998	0.7	0.6992	1/0.6992 =	1.430	399,048
0	1	120,050	0.3	0.3008	1/0.3008 =	3.324	399,048
0		399,048					
1	0	59,728	0.1	0.0994	1/0.0994 =	10.061	600,952
1	1	541,224	0.9	0.9006	1/0.9006 =	1.110	600,952
1		600,952					
Total		1,000,000					2,000,000

Inverse Probability Treatment Weights

L	A	True Sample Size	Pr[A|L] True	Pr[A|L] Estimated	Estimated IPTW		Weighted Sample Size
0	0	278,998	0.7	0.6992	1/0.6992 =	1.430	199,524
0	1	120,050	0.3	0.3008	1/0.3008 =	3.324	199,524
0		399,048					
1	0	59,728	0.1	0.0994	1/0.0994 =	10.061	300,476
1	1	541,224	0.9	0.9006	1/0.9006 =	1.110	300,476
1		600,952					
Total		1,000,000					1,000,000

True Sample

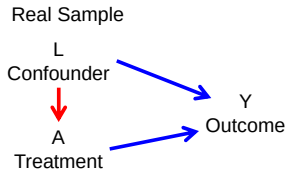
		A		
		0	1	Total
L	0	278,998	120,050	399,048
	1	59,728	541,224	600,952
Total		338,726	661,274	1,000,000

Odds ratio for A given L = 21.1

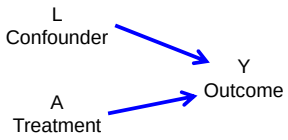
Weighted Sample

		A		
		0	1	Total
L	0	199,524	199,524	399,048
	1	300,476	300,476	600,952
Total		500,000	500,000	1,000,000

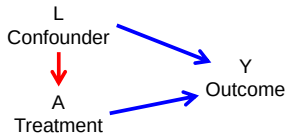
Odds ratio for A given L = 1



In the weighted sample the DAG is



Regressing Y against A allows me to estimate the marginal causal effect of A on Y



Suppose the true model is

$$\text{logit}\left[\pi(Y)\right] = \alpha + \beta A + \lambda L$$

where $\alpha = \text{logit}[.01]$
$\beta = \log[2]$
$\lambda = \log[3]$

Then the OR for Y associated with A adjusted for L = 2
the OR for Y associated with L adjusted for A = 3

Simulate number of patients with the indicated values of Y given L and A

L	A	Y	Sample Size
0	0	0	276,183
0	0	1	2,815
0	0		278,998
0	1	0	117,656
0	1	1	2,394
0	1		120,050
1	0	0	57,978
1	0	1	1,750
1	0		59,728
1	1	0	510,294
1	1	1	30,930
1	1		541,224
Total			1,000,000

Classical Model

$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta A_i + \lambda L_i$$

applied to real data

	Odds Ratio	Standard Error
L	2.973	0.052
A	2.003	0.037

MSM Model

$$\text{logit}\left[\pi(Y_i^a)\right] =$$
$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta A_i$$

applied to IPTW data with robust standard error estimate

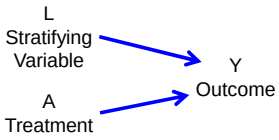
	Odds Ratio	Robust Std. Err.
A	1.998	0.042

Robust standard errors are calculated with the Huber-White sandwich estimator

$$\text{logit}\left[Y_i = 1\right] = \alpha + \beta A_i$$

applied to real data gives OR for Y associated with A = 3.88

Collapsible Statistics



If the relative risks of Y given A are identical in strata defined by L, then these relative risks also equal the overall (marginal) risk of Y associated with A

Statistics with this property are called **collapsible**

Relative risks are collapsible, odds ratios are not unless they approximate relative risks.

L=0 RR = (4/10)/(2/10) = 2

		A = 0	A = 1	Total
Y	0	8	6	14
	1	2	4	6
Total		10	10	20

L=1 RR = (6/10)/(3/10) = 2

		A = 0	A = 1	Total
Y	0	7	4	11
	1	3	6	9
Total		10	10	20

All Subjects RR = (10/20)/(5/20) = 2

		A = 0	A = 1	Total
Y	0	15	10	25
	1	5	10	15
Total		20	20	40

Marginal relative risk = within-strata relative risks.

L=0 OR = (8/2)/(4/2) = 2

		A = 0	A = 1	Total
Y	0	2	2	4
	1	4	8	12
Total		6	10	16

L=1 OR = (2/4)/(2/8) = 2

		A = 0	A = 1	Total
Y	0	8	4	12
	1	2	2	4
Total		10	6	16

All Subjects OR = (10/6)/(6/10) = 2.78

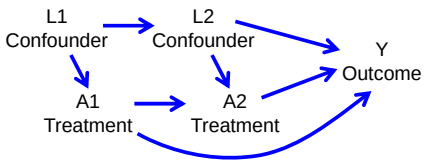
		A = 0	A = 1	Total
Y	0	10	6	16
	1	6	10	16
Total		16	16	32

Marginal odds ratio ≠ within-strata odds ratios.

In Simulated example I chose the probabilities of Y to be < 6% so that the odds ratios can be interpreted as relative risks

L	A	Y	Sample Size
0	0	0	276,183
0	0	1	2,815
0	0		278,998
0	1	0	117,656
0	1	1	2,394
0	1		120,050
1	0	0	57,978
1	0	1	1,750
1	0		59,728
1	1	0	510,294
1	1	1	30,930
1	1		541,224
Total			1,000,000

MSMs in Longitudinal Data Analysis



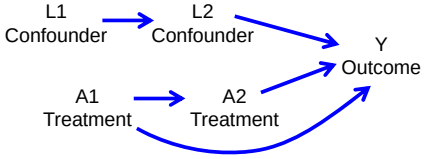
Here A1 affects Y directly and through A2

L1 and L2 confound the effects of A1 and A2 on Y

This model can be analyzed with either classical methods or MSMs.

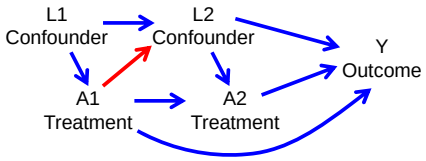
MSMs in Longitudinal Data Analysis

In the IPT weighted data the model becomes



The MSM analysis consists of regressing Y against A1 and A2 in the weighted sample and using robust standard errors.

Classical methods fail if A1 has a causal effect on L2



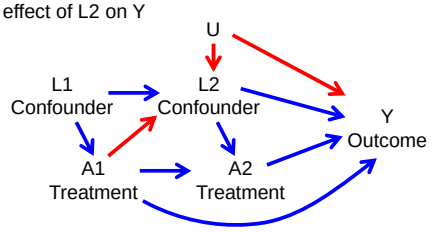
Here L2 confounds the effect of A2 on Y

I need to adjust for L2 to estimate the causal effect of A2 on Y.

But L2 is on the causal pathway from A1 to Y

If I adjust for L2 I block the effect of A1 on Y through L2

Moreover, suppose that unknown variables U confound the effect of L2 on Y

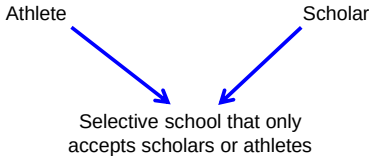


Now the path from A1 to Y through U collides at L2.

This means that as long as we do not adjust for L2 that the effect of A1 on Y through U is blocked

If I adjust for L2, I open a path from A1 through U to Y, which means that U confounds the effect of A1 on Y.

Stratifying on (adjusting for) a **collider** induces a relationship (**opens a path**) between its parents



If I attend the school and I am not an athlete then I must be a scholar.

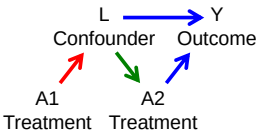
More Complicated Longitudinal Example

Sample Size = 1,000,000

$\Pr[A1=1]=0.5$

$\Pr[L=1 \mid A1=0]=0.4$

$\Pr[L=1 \mid A1=1]=0.8$



$\alpha=\text{logit}[.01]$

$\beta=\log[2]$

$\lambda=\log[3]$

$\text{logit}[\pi(Y_i)]=\alpha+\beta A2_i+\lambda L_i$

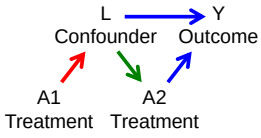
L is on the causal pathway from A1 to Y

$\Pr[A2=1 \mid L=0]=0.3$

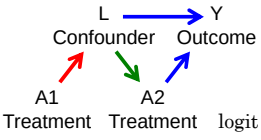
$\Pr[A2=1 \mid L=1]=0.9$

L Strongly confounds the effect of A2 on Y

Simulated number of patients with indicated values of A1, L, A2 and Y



A1	L	A2	Y	Sample Size
0	0	0	0	207,779
0	0	0	1	2,119
0	0	1	0	88,567
0	0	1	1	1,758
0	1	0	0	19,420
0	1	0	1	573
0	1	1	0	169,410
0	1	1	1	10,390
1	0	0	0	69,208
1	0	0	1	694
1	0	1	0	29,216
1	0	1	1	612
1	1	0	0	38,778
1	1	0	1	1147
1	1	1	0	339600
1	1	1	1	20729



$\text{logit}[\pi(Y_i)]=\alpha+\beta_1 A1_i+\beta_2 A_{2i}+\lambda L_i$

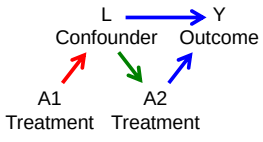
	Odds Ratio	95% Confidence Interval
A1	1.00	0.98 – 1.02
A2	2.03	1.96 – 2.11
L	3.00	2.89 – 3.11

$\beta_2=\log[2]$

$\lambda=\log[3]$

The effect of A1 on Y through L is blocked by stratifying on L

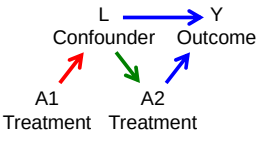
But A1 has an appreciable effect on Y through L



$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta A1_i$$

	Odds Ratio	95% Confidence Interval
A1	1.59	1.56 – 1.62

And L confounds the effect of A2 on Y

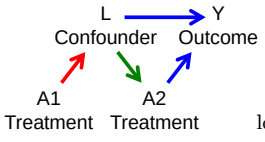


$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta A2_i$$

	Odds Ratio	95% Confidence Interval
A2	3.95	3.83 – 4.08

$\beta_2 = \log[2]$

And L confounds the effect of A2 on Y



$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta A_{2i} + \lambda L_i$$

	Odds Ratio	95% Confidence Interval
A2	2.03	1.96 – 2.11
L	3.00	2.89 – 3.10

$\beta_2 = \log[2]$
$\lambda = \log[3]$

Inverse Probability Treatment Weights

L	A2	True Sample Size	Estimated Pr[A2|L]	Estimated IPTW		Weighted Sample Size
0	0	279,800	0.6996	1/0.6996 =	1.429	199,976
0	1	120,153	0.3004	1/0.3004 =	3.329	199,976
0		399,953				
1	0	59,918	0.0999	1/0.0999 =	10.014	300,024
1	1	540,129	0.9001	1/0.9001 =	1.111	300,024
1		600,047				
Total		1,000,000				1,000,000

True Sample

		A2 = 0	A2 = 1	Total
L	0	279,800	120,153	399,953
	1	59,918	540,129	600,047
Total		339,718	660,282	1,000,000

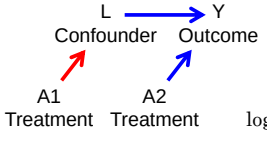
Odds ratio for A2 given L = 21.0

Weighted Sample

		A2 = 0	A2 = 1	Total
L	0	199,976	199,976	399,952
	1	300,024	300,024	600,048
Total		500,000	500,000	1,000,000

Odds ratio for A2 given L = 1

MSM Analysis in IPT Weighted Sample



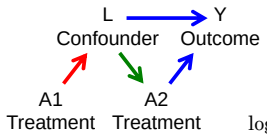
$$\text{logit}\left[\pi(Y_i)\right] = \alpha + \beta_1 A_{1i} + \beta_2 A_{2i}$$

	Odds Ratio	95% Confidence Interval
A1	1.46	1.41 – 1.50
A2	2.04	1.96 – 2.13

Unweighted regression of Y against A1 gives OR = 1.59

Unweighted regression of Y against A2 and L gives OR = 2.03 for A2 adjusted for L

Classical Regression of Y against A1 and A2 without Adjustment for L



$$\text{logit}\left[\pi\left(Y_i\right)\right] = \alpha + \beta_1 A_{1i} + \beta_2 A_{2i}$$

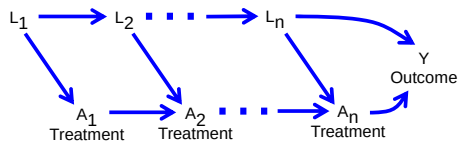
	Odds Ratio	95% Confidence Interval
A1	1.25	1.23 – 1.28
A2	3.72	3.60 – 3.84

This OR is too high

Applying MSMs to Real Data

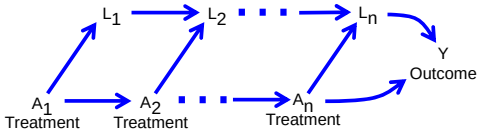
- For real data we are unlikely to have a model with a perfect fit in which the IPT weighting perfectly blocks the causal link between confounders and treatment.
- Often use propensity scores to adjust for indication bias.
- IPT weights are stabilized by multiplying by the unadjusted probability of treatment
- The results from MSMs are less powerful and are harder to interpret than classical methods. This is particularly true when common events are modeled by odds ratios.
- My advice is to avoid using them whenever classical methods are viable. They should only be used with some longitudinal studies.

Longitudinal Data Does not Always Require MSMs



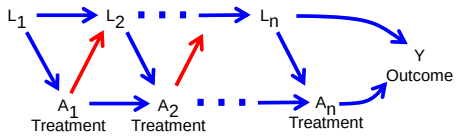
Use classical methods regressing Y against A_1, A_2, … , A_n adjusting for L_1, L_2, … , L_n

Longitudinal Data with Covariates on the Causal Pathway to Y



Use classical methods regressing Y against A_1, A_2, … , A_n

Longitudinal Data with Covariates that are both Confounders and on the Causal Pathway to Y



Use marginal structural models to regress Y against A_1, A_2, … , A_n in an IPT weighted sample

Thanks

Advice on MSMs
Bryan Shepherd
Wayne Ray

Help with the PRIMA RSV & Asthma database
Pingsheng Wu

Programming and data management
Tan Ding
Dale Plummer