Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Motivating example

3

Example: from Robins and Hernan (2009)

- Study of the effect of highly antiretroviral therapy (HAART) on a global health score at the end of followed-up in 32000 HIV infected subjects (trx naïve) followed for one year.
  - A_0: HAART at start of follow-up (1:yes, 0:no)
  - A_1: HAART at month 6 (1:yes, 0:no)
  - L_1: Indicator that Cd4 count > 200 cells/ml at month 6 just before making decision about HAART assignment (1:yes, 0:no).
  - Y global health score (continuous)
- Assume for now, that HAART was randomized at baseline but not at time month 6.
- We will later drop the assumption of randomization at baseline, and assume that the study was observational.

4

Data analyst table

Example taken from Robins and Hernan (2009) See reference list for exact reference

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2
row	**n**	**A_0**	**L_1**	**A_1**	**$E(Y \mid A_0, L_1, A_1)$**
1	2000	0	1	0	200
2	6000	0	1	1	220
3	6000	0	0	0	50
4	2000	0	0	1	70
5	3000	1	1	0	130
6	9000	1	1	1	110
7	3000	1	0	0	230
8	1000	1	0	1	250

Time 0 = start of follow-up
Time 1 = month 6
Time 2 = month 12

Example: there were 1000 patients who received HAART during first and second semester and had CD4 count below 200 cells/ml . The mean of the global health score at month 12 of these 1000 subjects was 250.

5

Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Static vs dynamic treatment regimes

- Static regime: everybody receives $A_0=a_0$ and $A_1=a_1$ regardless of the patient characteristics,
  - e.g. everybody receives HAART the second semester but not the first
- Dynamic regime: subject receives HAART depending on the values of recorded covariates
  - E.g. nobody receives HAART the first semester and only those whose CD4 count are below 200 receive HAART the second semester.
- Today we will study the analysis of the effects of static regimes.

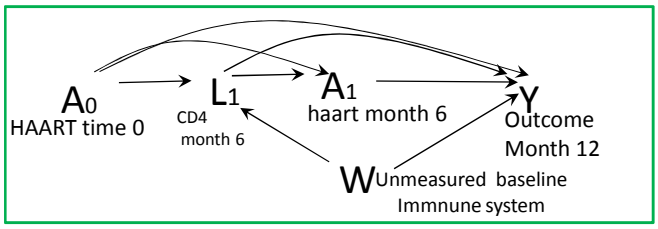
7

Counterfactual means for static regimes

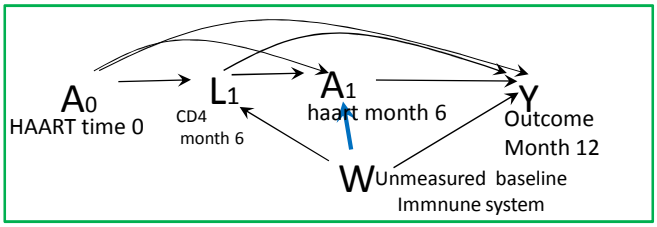
- $E(Y_{a_0=0,a_1=0})$ mean global score in hypothetical world in which nobody takes HAART for the entire year.
- $E(Y_{a_0=1,a_1=1})$ mean global score in hypothetical world in which everybody takes HAART for the entire year.
- $E(Y_{a_0=1,a_1=0})$ mean global score in hypothetical world in which everybody takes HAART the first semester and nobody takes HAART the second semester.
- $E(Y_{a_0=0,a_1=1})$ mean global score in hypothetical world in which nobody takes HAART the first semester and everybody takes HAART the second semester.

8

A_0 randomized $E(Y_{a0,a1})$ identified



A_0 randomized $E(Y_{a0,a1})$ not identified



9

Identification

- In the preceding graphs $E(Y_{a0,a1})$ is identified in the graph without the blue arrow but it is not identified in the graph with the blue arrow.
- Identification of a counterfactual expectation $E(Y_{a0,a1})$ (a quantity that depends on the oracle's data distribution) means that you can hope to estimate it with the data in the data analyst table!
  - However, it does not mean that you can compute the counterfactual quantity naively with crude means among those that received $A_0=a_0$ and $A_1=a_1$.
  - Depending on the graph structure, you may need more sophisticated calculations, like standardization or the two-stage IPW procedures that we saw in earlier lectures.

10

Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

When are the crude means unbiased estimators of the counterfactual means?

12

Data analyst table

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2	Sum of global Score at time 2
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	$n*E[Y\|A_0,L_1,A_1]$
1	2000	0	1	0	200	400,000
2	6000	0	1	1	220	1,320,000
3	6000	0	0	0	50	300,000
4	2000	0	0	1	70	140,000
5	3000	1	1	0	130	390,000
6	9000	1	1	1	110	990,000
7	3000	1	0	0	230	690,000
8	1000	1	0	1	250	250,000

13

Data analyst table

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2	Sum of global Score at time 2
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	$n*E[Y\|A_0,L_1,A_1]$
1	2000	0	1	0	200	400,000
2	6000	0	1	1	220	1,320,000
3	6000	0	0	0	50	300,000
4	2000	0	0	1	70	140,000
5	3000	1	1	0	130	390,000
6	9000	1	1	1	110	990,000
7	3000	1	0	0	230	690,000
8	1000	1	0	1	250	250,000

Crude means

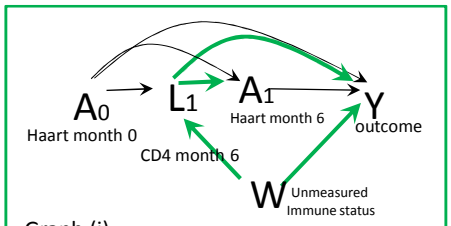
$E(Y|A_0=0,A_1=0)$ =
(400,000 + 300,000)/(2000+6000)
= 87.5

$E(Y|A_0=0,A_1=1)$ =
(1,320,000 + 140,000)/(6000+2000)
= 182.5

$E(Y|A_0=1,A_1=0)$ =
(390,000 + 690,000)/(3000+3000)
= 180

$E(Y|A_0=1,A_1=1)$ =
(990,000 + 250,000)/(9000+1000)
= 124

14



Graph (i)

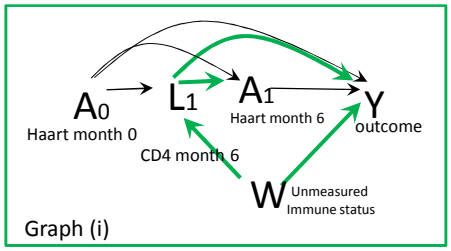
In graph (i)

$E(Y_{a0,a1}) \neq E(Y|A_0=a_0,A_1=a_1)$

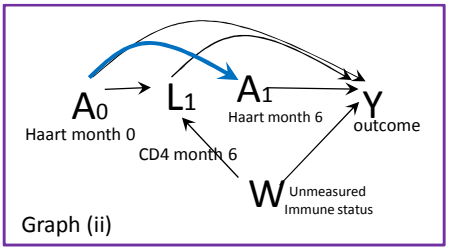
Because of open back-door paths A_1- L_1 – W – Y and A_1- L_1 – Y . Then

$E(Y_{a_0=0,a_1=0}) \neq 87.5$, etc

15



Graph (i)



Graph (ii)

In graph (i)

$E(Y_{a0,a1}) \neq E(Y|A_0=a_0,A_1=a_1)$

Because of open back-door paths A_1- L_1 – W – Y and A_1- L_1 – Y . Then

$E(Y_{a_0=0,a_1=0}) \neq 87.5$, etc

In graph (ii)

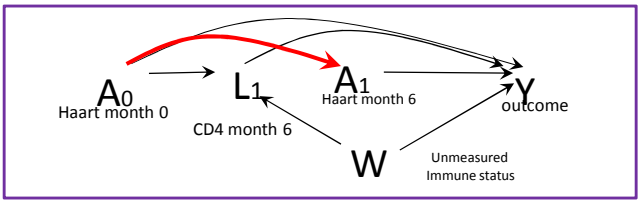
$E(Y_{a0,a1}) = E(Y|A_0=a_0,A_1=a_1)$

Crude mean = counterfactual mean because the open back door path from A_1 to Y goes only through A_0

16

Exogenous treatments

17



This is a graph of a sequentially randomized trial conducted as follows:

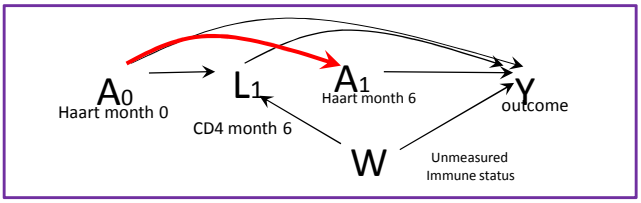
At Baseline:

- subjects are randomized to A0 =1 with prob P(A0=1), say 2/3, and to A0=0 with prob. 1-P(A0=1), say 1/3

At month 6:

- those that were assigned to A0=1 are randomized to A1=1 with probability P(A1=1|A0=1), say ½, and to A1=0 with prob. 1- P(A1=1|A0=1), say ½
- those that were assigned to A0=0 are randomized to A1=1 with probability P(A1=1|A0=0), say 1/4, and to A1=0 with prob. 1- P(A1=1|A0=1), say ¾

18

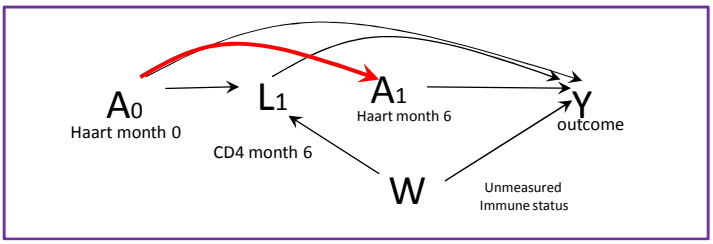


In the trial depicted in this graph (and described in the preceding slide), at each randomization stage (either at baseline or at month 6) subjects randomized to one of the two arms are exchangeable with subjects randomized to the other arm.

Then, the crude means are equal to the counterfactual means:

$$E(Y_{a_0,a_1}) = E(Y \mid A_0=a_0, A_1=a_1)$$

19



A set of treatments (A_0 , A_1,...,A_k) is exogenous when each A_j receives arrows from all, some or none of the preceding A_0,…,A_{j-1}, but does not receive arrows from any other variable in the DAG.

Result: When a set of treatments (A_0 , A_1,...,A_k) is exogenous the crude means are equal to the counterfactual means, i.e.

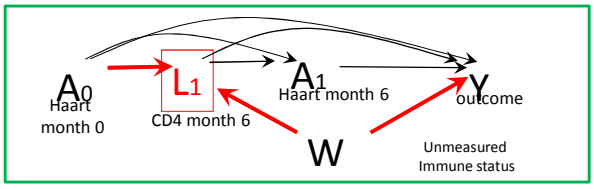
$$E(Y_{a_0,\ldots,a_k}) = E(Y \mid A_0=a_0, A_1=a_1, \ldots, A_k=a_k)$$

20

Problems with analysis that stratify or standardize on L_1

21

Stratified and standardized analysis on L_1 yields biased results



Stratification on L_1 induces selection bias: it opens the red path between A_0 and Y.

Thus, the stratified the crude mean differences stratified on L_1 :

$$E(Y|A_0=a_0,A_1=a_1, L_1=l_1) - E(Y|A_0=a'_0,A_1=a'_1, L_1=l_1)$$

are not equal to the stratified causal mean differences

$$E(Y_{a_0,a_1}|L_1=l_1) - E(Y_{a'_0,a'_1}|L_1=l_1)$$

As a consequence, the standardized mean differences

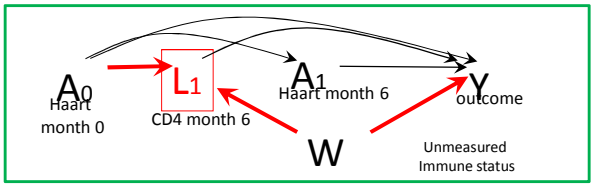
$$\sum_{l_1} [E(Y|A_0=a_0,A1=a_1, L_1=l1) - E(Y|A_0=a'_0,A1=a'_1, L_1=l_1)]P(L_1=l_1)$$

is not equal to the crude mean difference

$$E(Y_{a_0,a_1}|L_1=l_1) - E(Y_{a'_0,a'_1}|L_1=l_1)$$

22

Stratified and standardized analysis on L_1 yields biased results



Stratification on L_1 induces selection bias **even if the treatments are exogenous!!!**: stratification opens the red path between A_0 and Y even if A_1 only receives no arrow from L_1 and W.

Thus, the stratified the crude mean differences stratified on L_1 :

$$E(Y|A_0=a_0,A_1=a_1, L_1=l_1) - E(Y|A_0=a'_0,A_1=a'_1, L_1=l_1)$$

are not equal to the stratified causal mean differences

$$E(Y_{a_0,a_1} |L_1=l_1) - E(Y_{a'_0,a'_1} |L_1=l_1)$$

As a consequence, the standardized mean differences

$$\sum_{l_1} [E(Y|A_0=a_0,A1=a_1, L_1=l1) - E(Y|A_0=a'_0,A1=a'_1, L_1=l_1)]P(L_1=l_1)$$

is not equal to the crude mean difference

$$E(Y_{a_0,a_1} |L_1=l_1) - E(Y_{a'_0,a'_1} |L_1=l_1)$$

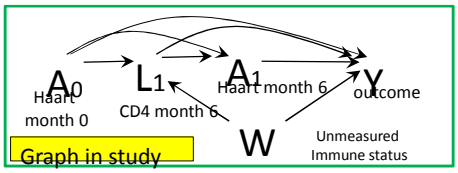
23

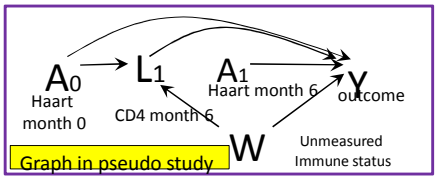
Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Inverse Probability Weighted (IPW) estimation

25





IPW estimation.

(a) Stage 1: Create (by IPW) a pseudo study in which $E(Ya_0,a_1)$ remains the same as in the original study, but the prob. of receiving A_1 does not depend on anything because everyone receives trx $A_1=0$ and everyone also receives trx $A_1=1$.

To create this pseudo study:

- weight those that in the real study received $A_1=0$ by $w = P(A_1=0|A_0,L_1)^{(-1)}$ so that they represent those that did not receive $A_1=0$
- weight those that in the real study received $A_1=1$ by $w = P(A_1=1|A_0,L_1)^{(-1)}$ so that they represent those that did not receive $A_1=1$

(a) Stage 2: Analyze your data as if the treatments (A_0,A_1) ***were exogenous***, i.e. as if the graph was like the one on the right. 26

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score at Time 2	weight	n in peudo study
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	w	n-pseudo
1	2000	0	1	0	200	4	8000
2	6000	0	1	1	220	4/3	8000
3	6000	0	0	0	50	4/3	8000
4	2000	0	0	1	70	4	8000
5	3000	1	1	0	130	4	12000
6	9000	1	1	1	110	4/3	12000
7	3000	1	0	0	230	4/3	4000
8	1000	1	0	1	250	4	4000
sum	**32000**						**64000**

$P(A_1=1|A_0=0,L_1=1) = 6000/8000 = 3/4$

$P(A_1=0|A_0=0,L_1=1) = 1 - ¾ = 1/4$

$P(A_1=1|A_0=0,L_1=0) = 2000/8000 = 1/4$

$P(A_1=0|A_0=0,L_1=0) = 1 - ¼ = 3/4$

$P(A_1=1|A_0=1,L_1=1) = 9000/12000 = 3/4$

$P(A_1=0|A_0=1,L_1=1) = 1 - 3/4 = 1/4$

$P(A_1=1|A_0=0,L_1=1) = 1000/4000 = 1/4$

$P(A_1=0|A_0=0,L_1=1) = 1 - 3/4 = 3/4$

27

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score at Time 2	weight	n in peudo study
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	w	n-pseudo
1	2000	0	1	0	200	4	8000
2	6000	0	1	1	220	4/3	8000
3	6000	0	0	0	50	4/3	8000
4	2000	0	0	1	70	4	8000
5	3000	1	1	0	130	4	12000
6	9000	1	1	1	110	4/3	12000
7	3000	1	0	0	230	4/3	4000
8	1000	1	0	1	250	4	4000

Prob. of A_1 in pseudo study

$P^*(A_1=1|A_0=0,L_1=1) = 8000/(8000+8000) = 1/2$

$P^*(A_1=1|A_0=0,L_1=0) = 8000/(8000+8000) = 1/2$

$P^*(A_1=1|A_0=1,L_1=1) = 12000/(12000+12000) = 1/2$

$P^*(A_1=1|A_0=1,L_1=0) = 4000/(4000+4000) = 1/2$

28

		HAART Time 1	CD4 Time 1	HAART Time 1	Mean global score at Time 2	weight	n in peudo study	Sum of global score in pseudo study
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	w	n-pseudo	n-pseudo*$E[Y\|A_0,L_1,A_1]$
1	2000	0	1	0	200	4	8000	1,600,000
2	6000	0	1	1	220	4/3	8000	1,760,000
3	6000	0	0	0	50	4/3	8000	400,000
4	2000	0	0	1	70	4	8000	560,000
5	3000	1	1	0	130	4	12000	1,560,000
6	9000	1	1	1	110	4/3	12000	1,320,000
7	3000	1	0	0	230	4/3	4000	920,000
8	1000	1	0	1	250	4	4000	1,000,000

Crude means in pseudo-study $E^*(Y|A_0,A_1) = E(Ya_0,a_1)$

$E^*(Y|A_0=0,A_1=0)$ = (1,600,000 + 400,000) / (8000+ 8000) = 125

$E^*(Y|A_0=0,A_1=1)$ = (1,760,000 + 560,000) / (8000+ 8000) = 145

$E^*(Y|A_0=1,A_1=0)$ = (1,560,000 + 920,000) / (12000+ 4000) = 155

$E^*(Y|A_0=1,A_1=1)$ = (1,320,000 + 1,000,000) / (12000+ 4000) = 145

		HAART Time 1	CD4 Time 1	HAART Time 1	Mean global score at Time 2	Sum of global Score at time 2	weight	n in peudo study	Sum of global score in pseudo study
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$	n*$E[Y\|A_0,L_1,A_1]$	w	n-pseudo	n*$E[Y\|A_0,L_1,A_1]$*w
1	2000	0	1	0	200	400,000	4	8000	1,600,000
2	6000	0	1	1	220	1,320,000	4/3	8000	1,760,000
3	6000	0	0	0	50	300,000	4/3	8000	400,000
4	2000	0	0	1	70	140,000	4	8000	560,000
5	3000	1	1	0	130	390,000	4	12000	1,560,000
6	9000	1	1	1	110	990,000	4/3	12000	1,320,000
7	3000	1	0	0	230	690,000	4/3	4000	920,000
8	1000	1	0	1	250	250,000	4	4000	1,000,000

Crude means in actual study $E(Y|A_0,A_1)$

$E(Y|A_0=0,A_1=0)$ = (400,000 + 300,000) / (2000+ 6000) = 87.5

$E(Y|A_0=0,A_1=1)$ = (1,320,000 + 140,000) / (6000+ 2000) = 182.5

$E(Y|A_0=1,A_1=0)$ = (390,000 + 690,000) / (3000+ 3000) = 180

$E(Y|A_0=1,A_1=1)$ = (990,000 + 250,000) / (9000+ 1000) = 124

30

Summary of results

HAART Time 0	HAART Time 1	Crude Mean in Actual study	Crude Mean in Pseudo study
A_0	A_1	$E[Y\|A_0,A_1]$	$E^*[Y\|A_0,A_1]$
0	0	87.5	125
0	1	182.5	145
1	0	180	155
1	1	124	145

Additive causal effect of always vs never take HAART during the 12 months

$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0})$

IPW estimate: 145 – 125 =20 **Crude estimate (biased) : 124 – 87.5 = 36.5**

Direct effect of HAART the first semester in hypothetical world in which we force HAART the second semester

$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=1})$

IPW estimate: 145 – 145 = 0 **Crude estimate (biased) : 124 – 182.5 = -58.5**

Effect of early vs late one semester of HAART

$E(Y_{a_0=1,a_1=0}) - E(Y_{a_0=0,a_1=1})$

IPW estimate: 155 – 145 = 10 **Crude estimate (biased) : 180 – 182.5 = -2.5**

31

		HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2
row	n	A_0	L_1	A_1	$E[Y\|A_0,L_1,A_1]$
1	2000	0	1	0	200
2	6000	0	1	1	220
3	6000	0	0	0	50
4	2000	0	0	1	70
5	3000	1	1	0	130
6	9000	1	1	1	110
7	3000	1	0	0	230
8	1000	1	0	1	250

Recall, the IPW estimator of $E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=1})$ was equal to 0.

Stratified crude mean differences

$E(Y|A_0=1,A_1=1, L_1=1) - E(Y|A_0=0,A_1=1, L_1=1)$
110 -220 = -110

$E(Y|A_0=1,A_1=1, L_1=0) - E(Y|A_0=0,A_1=1, L_1=0)$
250 – 70 = 180

Standardization adjusting for L1

$P(L_1=1) = (2000+6000+3000+9000) / 320000 = 5/8$

Standardized mean difference:
-110*(5/8) + 180 *(3/8) = -1.25

Standardization adjusting for L_1 yields a negative value (-1.25) but the true causal effect estimated by IPW is equal to 0.

Rationale of the IPW procedure: summary

- We can pretend that the pseudo study is formed by two copies ("clones") of each person, one clone receives $A_1=0$ and the other receives $A_1=1$. So,
  - we can pretend that to assign A_1 we have flipped one same coin for everyone.
- Since we have also flipped one same coin for everyone to assign A_0
  - a coin that might be different from the imaginary coin for assigning A_1, because, for example, we may have assigned A_0 with probability different from ½.
- Then, in the pseudo study, the subjects assigned to each of the four treatment arms

  $(A_0=0,A_1=0)$, $(A_0=0,A_1=1)$, $(A_0=1,A_1=0)$, $(A_0=1,A_1=1)$

  are exchangeable, so we can estimate the counterfactual means $E(Y_{a0,a1})$ with the crude means $E(Y|A_0=a_0,A_1=a_1)$

33

Stabilized IPW estimation

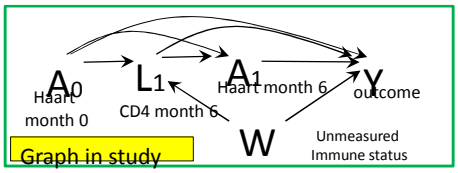
34

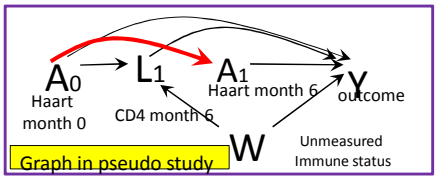
Non –stabilized vs stabilized IPW

- The IPW procedure we have just seen creates a pseudo-study in which
  1. has size equal to the double of the actual study size
  2. $P^*(A_1=1|L_1=l_1, A_0=a_0)= ½$
  3. The crude mean in the pseudo-study $E^*(Y|A_0=a_0,A_1=a_1)$ is the counterfactual mean $E(Y_{a0,a1})$

- We will now see a way a "trick" to modify the IPW computation (referred to as "stabilized" IPW) that creates a new pseudo-study in which
  1. has size equal to the actual study size
  2. $P^*(A_1=1|L_1=l_1, A_0=a_0)$ is equal to $P(A_1=1| A_0=a_0)$ in the actual study
  3. Because in the new pseudo study, A_1 only depends on A_0 , the treatments (A_0,A_1) remain exogenous as with the previous IPW procedure. Thus,
     - In the new pseudo-study, the crude means are equal to the counterfactual means
  4. The crude mean $E^*(Y|A_0=a_0,A_1=a_1)$ in the new pseudo-study is EXACTLY THE SAME NUMBER as the crude mean in the IPW pseudo-study

35





Stabilized IPW estimation.

(a) Stage 1: Create (by IPW) a pseudo study in which $E(Y_{a0,a1})$ remains the same as in the original study, but the prob. of receiving A_1 depends ONLY on the treatment A_0 actually received and is equal to the $P(A_1|A_0)$ in the original study. To create this pseudo study:

- weight those that in the real study received **$A_1=0$** by
  $ws = P(A_1=0|A_0) / P(A_1=0|A_0,L_1) = P(A_1=0|A_0)$ w
- weight those that in the real study received **$A_1=1$** by
  $ws = P(A_1=1|A_0) / P(A_1=1|A_0,L_1) = P(A_1=1|A_0)$ w

b) Stage 2: Analyze your data as if the treatments (A_0,A_1) were exogenous, i.e. as if the graph was like the one on the right.

36

Stabilized vs non-stabilized IPW means

- **Unstabilized weights**

$$1 / P(A_1|A_0,L_1)$$

- **Stabilized weights**

$$P(A_1|A_0) / P(A_1|A_0,L_1)$$

- It is a mathematical fact if you compute the crude means in the two pseudo studies (the one obtained by stabilized IPW and the one obtained by non-stabilized IPW) you obtain exactly the same numbers. So why even bother computing the stabilized weights?
- I can't give you the answer yet, because stabilized weights don't make a difference for estimating crude means but do make a difference (yield less variables estimators) for estimating the parameters of *Marginal Structural Models*.
- So I will postpone giving you the answer until we discuss Marginal Structural Models

37

Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Estimation of time dependent treatment effects from longitudinal observational studies

A_0 and A_1 observed, not randomized

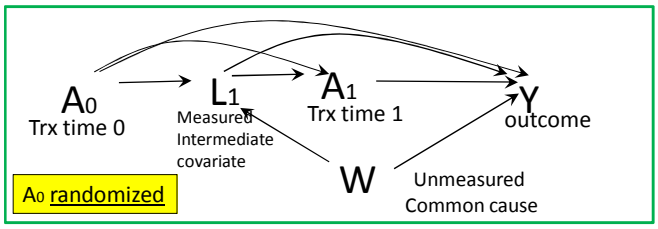
39

Example: adapted from Robins and Hernan (2009)

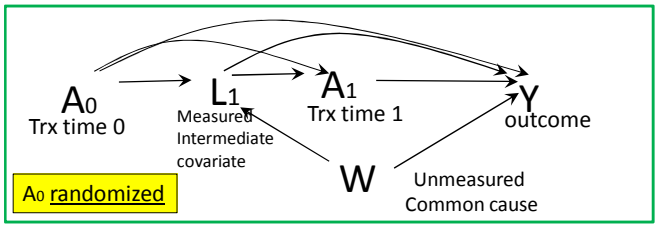
- Suppose that we have **<u>an observational longitudinal</u>** study of the effect of highly antiretroviral therapy (HAART) on a global health score at the end of followed-up we now have data on 96000 subjects that have been recently found to be infected with HIV and who are followed for one year.
-
- We measure the following variables
  - A_0: HAART at start of follow-up (1:yes, 0:no) (visit at which HIV is diagnosed)
  - A_1: HAART at month 6 (1:yes, 0:no)
  - L_0: Indicator that Cd4 count > 200 cells/ml at baseline just before making decision about HAART assignment (1:yes, 0:no).
  - L_1: Indicator that Cd4 count > 200 cells/ml at month 6 just before making decision about HAART assignment (1:yes, 0:no).
  - Y global health score (continuous)

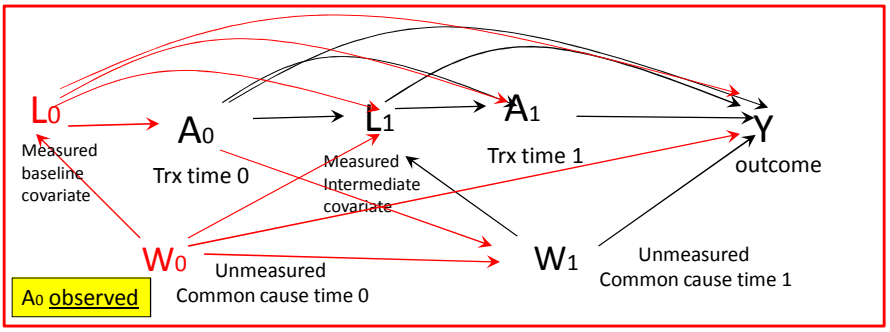
40

A_0 randomized vs A_0 observed: $E(Y_{a0,a1})$ identified



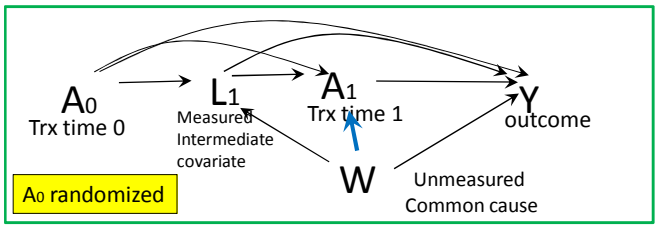
A_0 randomized vs A_0 observed: $E(Y_{a0,a1})$ identified



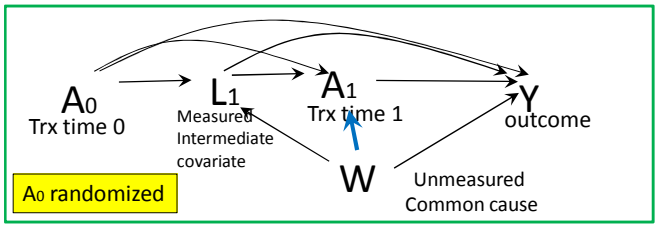


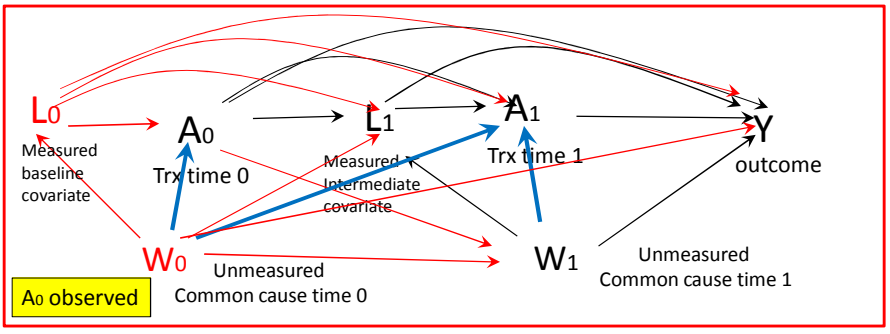
42

A_0 randomized vs A_0 observed: $E(Y_{a0,a1})$ NOT identified



A_0 randomized vs A_0 observed: $E(Y_{a0,a1})$ NOT identified





44

Identification

- In the preceding graphs $E(Y_{a0,a1})$ is NOT identified if any of the blue arrows is present.

- If all the blue arrows are absent, then $E(Y_{a0,a1})$ is identified.

45

Identification

- The big distinction that allows identification when the blue arrows are absent but that prevents it when they are present is that
  - When all the blue arrows are absent, the measured variables suffice to block
    - All the back door paths from A_1 to Y
    - All the back door paths from A_0 to Y
  - When any of the blue arrows is present, there is at least one back door path from
    - A_1 to Y, and/or
    - A_0 to Y
  - that cannot be blocked

46

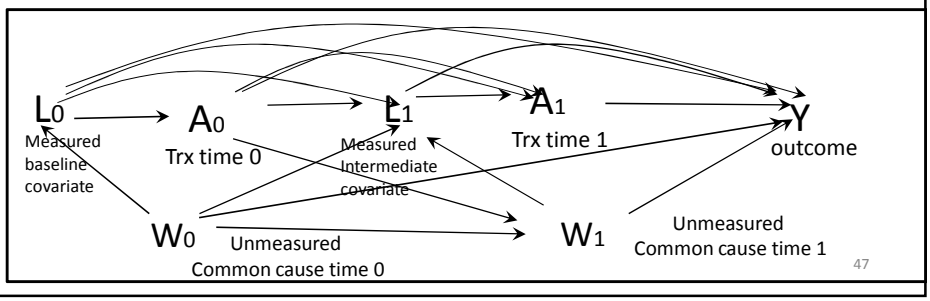
The graph below implies that

$Y_{a0,a1} \perp\!\!\!\perp A_1 \mid L_1, A_0, L_0$ ($Y_{a0,a1}$ and A_1 are conditionally independent given L_1, A_0, L_0)

$Y_{a0,a1} \perp\!\!\!\perp A_0 \mid L_0$ ($Y_{a0,a1}$ and A_0 are conditionally independent given L_0)

These two conditions together are often referred to as

NO UNMEASURED CONFOUNDING



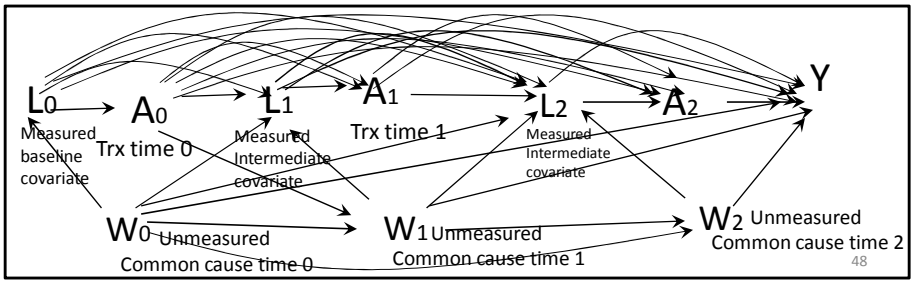
47

The graph below implies that

$Y_{a0,a1,a2} \perp\!\!\!\perp A_2 \mid L_2, A_1, L_1, A_0, L_0$
$Y_{a0,a1,a2} \perp\!\!\!\perp A_1 \mid L_1, A_0, L_0$
$Y_{a0,a1,a2} \perp\!\!\!\perp A_0 \mid L_0$

These three conditions together are often referred to as

NO UNMEASURED CONFOUNDING

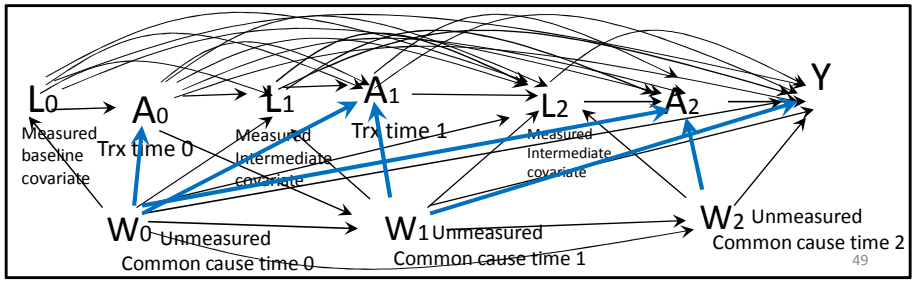


48

No unmeasured confounding does not hold if any of the blue arrows is present in the true causal diagram.

If any of the blue arrows are present then the counterfactual means $E(Y_{a0,a1,a2})$ are not identified

If all the blue arrows are absent then the counterfactual means $E(Y_{a0,a1,a2})$ are identified



Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

IPW estimation of the effects of two time dependent treatments from longitudinal observational studies

51

IPW for longitudinal observational studies

- Let's return to the case of two trxs A_0 and A_1, neither of which has been randomized.

- We will now see that if A_0 was not randomized, then under the no-unmeasured confounding assumption, we can estimate the counterfactual means by IPW weighting using
  - non-stabilized weights

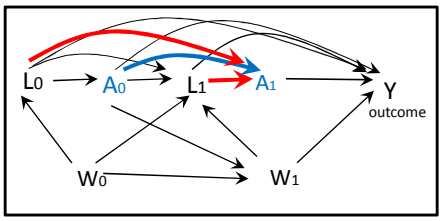
    $$w = 1 / [\, P(A_1|L_1,A_0,L_0) * P(A_0|L_0) \,]$$
  - or stabilized weights

    $$w = [P(A_1|A_0) * P(A_0)\,] / [\, P(A_1|L_1,A_0,L_0) * P(A_0|L_0) \,]$$

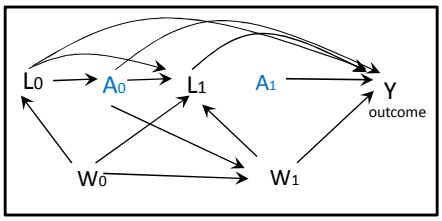
52

Non-stabilized IPW estimation

53



Graph of actual study



Graph of pseudo study after first IPW weighting

If we weight each person by

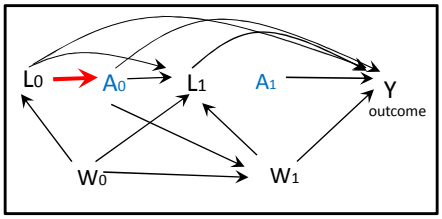
$$1/P(A_1|L_1,A_0,L_0)$$

that is:

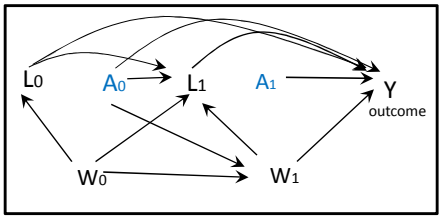
- those that received $A_1=1$ are weighted by $1/P(A_1=1|L_1,A_0,L_0)$
- those that received $A_1=0$ are weighted by $1/P(A_1=0|L_1,A_0,L_0)$

we create a pseudo-study in which $E(Y_{a0,a1})$ remains the same as in the original study and, such that the probability of receiving $A_1=1$ is the same for everybody and equal to 1/2.

54



Graph of pseudo study after first IPW weighting



Graph of pseudo study after second IPW weighting

If, in the pseudo-study we weight each person by

$$1 / P(A_0 | L_0)$$

that is:

- those that received $A_0=1$ are weighted by $1 / P(A_0=1 | L_0)$
- those that received $A_0=0$ are weighted by $1 / P(A_0=0 | L_0)$

we create a new pseudo-study in which $E(Y_{a0,a1})$ remains the same as in the first pseudo study and, such that A_0 is randomized with the same (fair) coin flipped for everyone.

55

Analysis in the final pseudo study

- We can pretend that the final pseudo study is formed by four copies ("clones") of each person, and each clone receives one of the four trx combinations

  $(A_0=1,A_1=1)$, $(A_0=1,A_1=0)$, $(A_0=0,A_1=1)$, $(A_0=0,A_1=0)$

- So, we can pretend that we have flipped the same coin for everyone, and that there is unconditional exchangeability for comparing the four trx combinations.

- Then in the final pseudo study we can estimate the counterfactual means $E(Y_{a0,a1})$ with the crude means $E(Y | A_0=a_0,A_1=a_1)$

56

IPW procedure for two time dependent trxs in a longitudinal study

- Stage 1. Compute for each value of a_0, a_1, l_0 and l_1 the propensity scores

  $P(A_1=a_1 | L_1=l_1, A_0=a_0, L_0=l_0)$ and $P(A_0=a_0 | L_0=l_0)$

- Stage 2. Create a pseudo study by weighting each subject who got $A_0=a_0$, $A_1=a_1$, $L_0=l_0$ and $L_1=l_1$ with

  $w = 1 / [\, P(A_1=a_1 | L_1=l_1, A_0=a_0, L_0=l_0) * P(A_0=a_0 | L_0=l_0) \,]$

- Stage 3. In the pseudo study, compute for each a_0 and a_1, the crude means $E(Y | A_0=a_0, A_1=a_1)$. These are your IPW estimators of $E(Y_{a0,a1})$

57

			HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2
Row	n	L0	A0	L1	A1	E[Y|A0,L1,A1,L0]
1	2000	0	0	1	0	200
2	6000	0	0	1	1	220
3	6000	0	0	0	0	50
4	2000	0	0	0	1	70
5	3000	0	1	1	0	130
6	9000	0	1	1	1	110
7	3000	0	1	0	0	230
8	1000	0	1	0	1	250
9	9000	1	0	1	0	250
10	3000	1	0	1	1	300
11	1000	1	0	0	0	80
12	3000	1	0	0	1	100
13	31500	1	1	1	0	160
14	10500	1	1	1	1	120
15	1500	1	1	0	0	280
16	4500	1	1	0	1	350

Computations of $P(A_1=1 | A_0,L_0,L_1)$ and $P(A_1=0 | A_0,L_0,L_1)$

l0	a0	l1	Propensity $A_1=1$	Propensity $A_1=0$
0	0	0	2000/8000=1/4	6000/8000=3/4
0	0	1	6000/8000=3/4	2000/8000=1/4
0	1	0	1000/4000=1/4	3000/4000=3/4
0	1	1	9000/12000=3/4	3000/12000=1/4
1	0	0	3000/4000=3/4	1000/4000=1/4
1	0	1	3000/12000=1/4	9000/12000=3/4
1	1	0	4500/6000=3/4	1500/6000=1/4
1	1	1	10500/42000=1/4	31500/42000=3/4

58

			HAART Time 0	CD4 Time 1	HAART Time 1	Mean global score Time 2
Row	n	L0	A0	L1	A1	E[Y|A0,L1,A1,L0]
1	2000	0	0	1	0	200
2	6000	0	0	1	1	220
3	6000	0	0	0	0	50
4	2000	0	0	0	1	70
5	3000	0	1	1	0	130
6	9000	0	1	1	1	110
7	3000	0	1	0	0	230
8	1000	0	1	0	1	250
9	9000	1	0	1	0	250
10	3000	1	0	1	1	300
11	1000	1	0	0	0	80
12	3000	1	0	0	1	100
13	31500	1	1	1	0	160
14	10500	1	1	1	1	120
15	1500	1	1	0	0	280
16	4500	1	1	0	1	350

Computations of P(A0=1|L0) and P(A0=0|L0)

l0	Propensity A0=1	Propensity A0=0
0	16000 /32000=1/2	16000/32000=1/2
1	48000 /64000=3/4	16000/64000=1/4

59

		CD4 time 0	HAART Time 0	CD4 Time 1	HAART Time 1	Mean score Time 2	weight	n Pseudo study	sum in pseudo study
Row	n	L0	A0	L1	A1	E[Y|A0,L1,A1,L0]	1/Product of ppensities	n*w	n*w*E[Y|A0,L1,A1,L0]
1	2000	0	0	1	0	200	2* 4	8*2000= 16,000	16,000*200= 3,200,000
2	6000	0	0	1	1	220	2* 4/3	16,000	3,520,000
3	6000	0	0	0	0	50	2* 4/3	16,000	800,000
4	2000	0	0	0	1	70	2* 4	16,000	1,120,000
5	3000	0	1	1	0	130	2* 4	24,000	3,120,000
6	9000	0	1	1	1	110	2* 4/3	24,000	2,640,000
7	3000	0	1	0	0	230	2* 4/3	8,000	1,840,000
8	1000	0	1	0	1	250	2* 4	8,000	2,000,000
9	9000	1	0	1	0	250	4* 4/3	48,000	12,000,000
10	3000	1	0	1	1	300	4* 4	48,000	14,400,000
11	1000	1	0	0	0	80	4* 4	16,000	1,280,000
12	3000	1	0	0	1	100	4* 4/3	16,000	1,600,000
13	31500	1	1	1	0	160	4/3* 4/3	56,000	8,960,000
14	10500	1	1	1	1	120	4/3 * 4	56,000	6,720,000
15	1500	1	1	0	0	280	4/3 * 4	8,000	2,240,000
16	4500	1	1	0	1	350	4/3 * 4/3	8,000	2,800,000

60

		HAART Time 0	HAART Time 1	n Pseudo study	sum in pseudo study
Row	n	A0	A1	n*w	n*w*E[Y|A0,L1,A1,L0]
1	2000	0	0	8*2000= 16,000	16,000*200= 3,200,000
2	6000	0	1	16,000	3,520,000
3	6000	0	0	16,000	800,000
4	2000	0	1	16,000	1,120,000
5	3000	1	0	24,000	3,120,000
6	9000	1	1	24,000	2,640,000
7	3000	1	0	8,000	1,840,000
8	1000	1	1	8,000	2,000,000
9	9000	0	0	48,000	12,000,000
10	3000	0	1	48,000	14,400,000
11	1000	0	0	16,000	1,280,000
12	3000	0	1	16,000	1,600,000
13	31500	1	0	56,000	8,960,000
14	10500	1	1	56,000	6,720,000
15	1500	1	0	8,000	2,240,000
16	4500	1	1	8,000	2,800,000

Crude means in pseudo study

$\bar{E}^*(Y|A_0=0,A_1=0)$ =
(3,200,000 +800,000+12,000,000+1,280,000)/
(16,000+16,000+48,000+16,000)
=180

$E^*(Y|A0=0,A1=1)$ =
(3,520,000 +1,120,000+14,400,000+1,600,000)/
(16,000+16,000+48,000+16,000)
=215

$E^*(Y|A0=1,A1=0)$ =
(3,120,000 +1,840,000+8,960,000+2,240,000)/
(24,000+8,000+56,000+8,000)
=168.33

$E^*(Y|A0=1,A1=1)$ =
(2,640,000 +2,000,000+6,720,000+2,800,000) /
(24,000+8,000+56,000+8,000)
=147.5

61

Stabilized IPW estimation of the effects of two time dependent treatments from longitudinal observational studies

62

Stabilized IPW procedure for two time dependent trxs in a longitudinal study

- Stage 1. Compute for each value of a_0, a_1, l_0 and l_1 the propensity scores

$$\frac{P(A_1=a_1|L_1=l_1, A_0=a_0, L_0=l_0)}{P(A_1=a_1|A_0=a_0)} \quad \text{and} \quad \frac{P(A0=a_0|L0=l_0)}{P(A0=a_0)}$$

- Stage 2. Create a pseudo study by weighting each subject who got $A_0=a_0$, $A_1=a_1$, $L_0=l_0$ and $L_1=l_1$ with

$$w = [P(A_1=a_1|A_0=a_0) * P(A0=a_0)] / [P(A1=a_1|L1=l_1, A0=a_0, L0=l_0) * P(A0=a_0|L0=l_0)]$$

- Stage 3. In the pseudo study, compute for each a_0 and a_1, the crude means $E(Y|A_0=a_0, A_1=a_1)$. These are your IPW estimators of $E(Y_{a0,a1})$

63

Friday afternoon

- Time dependent treatments.
- Dynamic vs static regimes
- Analysis of randomized followed up studies
  - Why are conventional associational methods incorrect?
  - When is the associational analysis valid? Exogenous treatments
  - IPW methodology
- Analysis of longitudinal observational studies
  - Why are conventional associational methods incorrect?
  - IPW methodology
  - Marginal structural mean models for time dependent treatments

Marginal Structural Mean Models

65

Marginal structural mean model

- A Marginal Structural Mean Model is a formula for the dependence of the counterfactual means $E(Y_{a_0,a_1,\ldots,a_k})$ on $a_0,a_1,\ldots, a_k$

$$E(Y_{a_0,a_1,\ldots,a_k}) = m(\theta_0,\ldots,\theta_p, a_0,a_1,\ldots, a_k)$$

- the function m( ) is a known function and the parameters $\theta_0,\ldots,\theta_p$ are unknown

66

Marginal Structural Mean Model

- Saturated model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 a_0 + \theta_2 a_1 + \theta_3 a_0 a_1$$

a_0	a_1	$E(Y_{a_0,a_1})$
0	0	θ_0
0	1	$\theta_0 + \theta_2$
1	0	$\theta_0 + \theta_1$
1	1	$\theta_0 + \theta_1 + \theta_2 + \theta_3$

67

Marginal Structural Mean Model

- Saturated model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 a_0 + \theta_2 a_1 + \theta_3 a_0 a_1$$

  - Additive causal effect of always vs never take HAART during the 12 months

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0}) = \theta_1 + \theta_2 + \theta_3$$

  - Direct effect of HAART the first semester in hypothetical world in which we force HAART the second semester

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=1}) = \theta_1 + \theta_3$$

  - Effect of one semester HAART early vs late

$$E(Y_{a_0=1,a_1=0}) - E(Y_{a_0=0,a_1=1}) = \theta_1 - \theta_2$$

68

Marginal Structural Mean Model

- Example of a non-saturated model. Effect depends on cumulative exposure

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1)$$

a_0	a_1	$E(Y_{a_0,a_1})$
0	0	θ_0
0	1	$\theta_0 + \theta_1$
1	0	$\theta_0 + \theta_1$
1	1	$\theta_0 + 2\,\theta_1$

69

Marginal Structural Mean Model

- Non-Saturated model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1)$$

  - Additive causal effect of always vs never take HAART during the 12 months

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0}) = 2\,\theta_1$$

  - Direct effect of HAART the first semester in hypothetical world in which we force HAART the second semester

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=1}) = \theta_1$$

  - Effect of one semester HAART early vs late

$$E(Y_{a_0=1,a_1=0}) - E(Y_{a_0=0,a_1=1}) = 0$$

70

Saturated vs non-saturated models

- A non-saturated model can be incorrectly specified
  - E.g. consider the model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1).$$

  - The model encodes assumptions made by the data analyst
    - For example, the model assumes that receiving trx for six months has the same effect on the outcome whether trx was received the first semester or the second
  - But nature may have a different plan! May be, for example, that it does matter when the patients receive trx!! In such case the model is wrong. It is incorrectly specified.
- A saturated model is always correctly specified
  - E.g. consider the model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 a_0 + \theta_2 a_1 + \theta_3 a_0 a_1$$

  - A saturated model never contradicts any plan of nature!. You can always find values of the parameters (the θ's) that agree with any possible plan that nature may have!!!

71

Saturated vs non-saturated models

- When you analyze your data assuming a non-saturated model
  - E.g. you assume model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1).$$

  - Then
    - If the model is correctly specified, your conclusions will be correct.
    - If the model is incorrectly specified, then your conclusions will be wrong. You will get biased estimators of the treatment effects of interest.
- When you analyze your data assuming a saturated model
  - E.g. you assume model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 a_0 + \theta_2 a_1 + \theta_3 a_0 a_1$$

  - Then,
    - If you use the correct algorithms for computing the estimators, you will arrive at the right conclusions.
- So, why assume a non-saturated model if you risk arriving at wrong conclusions???
  - Because saturated models have very many parameters and you may not have enough data to estimate all the parameters.

72

Saturated vs non-saturated models

- If you assume a saturated model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 a_0 + \theta_2 a_1 + \theta_3 a_0 a_1$$

- You need to have in your study subjects that followed all four possible trx combinations:
  1. Subjects that took HAART in both semesters
  2. Subjects that took HAART the first semester but not the second
  3. Subjects that took HAART the second semester but not the first
  4. Subjects that did not take HAART in any of the two semesters

- If you assume the non-saturated model

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1).$$

- It suffices to have in your study subjects that took two of the four possible trx combinations (with the exception of the combinations 2 & 3 alone).

73

Estimation of the parameters of marginal structural models

- To teach you how to estimate the parameters of non-saturated marginal structural mean models we need to review briefly a **statistical** model: the regression model.

- It will turn out that you can estimate the parameters of non-saturated MSM with just one small modification to the standard estimators of the parameters of regression models.

74

Regression models for conditional means

- A regression model for a conditional mean of an outcome Y given covariates $A_0,\ldots,A_k$ is a formula for the dependence of the **conditional (crude) means** $E(Y|A_0=a_0,A_1=a_1,\ldots,A_k=a_k)$ on $a_0,a_1,\ldots,a_k$

$$E(Y|A_0=a_0,A_1=a_1,\ldots,A_k=a_k) = g(\gamma_0,\ldots,\gamma_p, a_0,a_1,\ldots,a_k)$$

- the function g( ) is a known function and the parameters $\gamma_0,\ldots,\gamma_p$ are unknown

75

Regression (associational) model

- Saturated model

$$E(Y|A_0=a_0,A_1=a_1) = \gamma_0 + \gamma_1 a_0 + \gamma_2 a_1 + \gamma_3 a_0 a_1$$

a_0	a_1	$E(Y\|A_0=a_0,A_1=a_1)$
0	0	γ_0
0	1	$\gamma_0 + \gamma_2$
1	0	$\gamma_0 + \gamma_1$
1	1	$\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3$

76

Regression (associational) model

- Example of a non-saturated model. Conditional mean depends on cumulative exposure.

$$E(Y|A_0=a_0,A_1=a_1) = \gamma_0 + \gamma_1 (a_0 + a_1)$$

a_0	a_1	$E(Y\|A_0=a_0,A_1=a_1)$
0	0	γ_0
0	1	$\gamma_0 + \gamma_1$
1	0	$\gamma_0 + \gamma_1$
1	1	$\gamma_0 + 2\gamma_1$

77

Estimation of the parameters of regression models

- The **method of ordinary least squares** which is implemented in all statistical packages can be used to obtain unbiased estimators of the parameters γ of the regression model.

- Most statistical packages allow the possibility of estimating the parameters γ of the regression model by the **method of weighted least squares** which is just like the method of ordinary least squares except that each subject is allowed to carry a different weight

78

Caution about the causal interpretation of the parameters of the regression model

- Suppose the following regression model is true

$$E(Y|A_0=a_0,A_1=a_1) = \gamma_0 + \gamma_1 (a_0 + a_1)$$

  - Then,

$$E(Y|A_0=1,A_1=1) - E(Y|A_0=0,A_1=0) = 2\gamma_1$$

- Suppose that we can estimate γ_1 . Do we then know anything about the causal mean difference $E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0})$?
  - When (A_0,A_1) are exogenous treatments , YES. The crude means are equal to the counterfactual means. Consequently, in such case we have

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0}) = 2\gamma_1$$

  - When (A_0,A_1) are NOT exogenous treatments, NO. The crude means are, in general, different from the counterfactual means so we cannot conclude that

$$E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0}) \neq 2\gamma_1$$

79

Lesson

- You may propose a regression model which may be correctly specified.
  - You may find that your all statistical model checking techniques confirm that your model fits the data well
- Yet, this does not imply that you can ascribe to the regression parameter estimators a causal interpretation
  - For example, if the non-saturated model

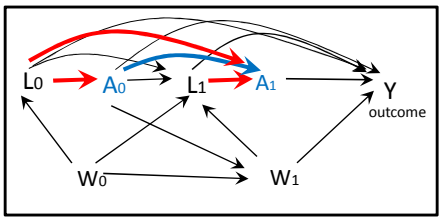
$$E(Y|A_0=a_0,A_1=a_1) = \gamma_0 + \gamma_1 (a_0 + a_1)$$

  - Is correctly specified, then the least squares estimator of $2\gamma_1$ is an unbiased estimator of the crude mean difference $E(Y|A_0=1,A_1=1) - E(Y|A_0=0,A_1=0)$. However, it is NOT an unbiased estimator of the causal mean difference $E(Y_{a_0=1,a_1=1}) - E(Y_{a_0=0,a_1=0})$

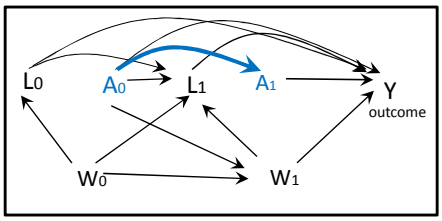
80

Estimation of the parameters of Marginal Structural Mean models

81



Graph of actual study



Graph of final pseudo study after IPW weighting

- Suppose that we postulate a non-saturated MSM, say, for example

$$E(Y_{a_0,a_1}) = \theta_0 + \theta_1 (a_0 + a_1).$$

If we weight each subject in the study that has $A_1=a_1$, $A_0=a_0$, $L_1=l_1$, $L_0=l_0$ with either

$$ws = [P(A_1=a_1|A_0=a_0) * P(A0=a_0)] / [P(A1=a_1|L1=l_1,A0=a_0, L0=l_0)* P(A0=a_0|L0=l_0)]$$

or with

$$w = 1 / [P(A_1=a_1|L_1=l_1,A_0=a_0, L_0=l_0)* P(A_0=a_0|L_0=l_0)]$$

We create a pseudo-study such that

- (A_0,A_1) are exogenous treatments, so $E^*(Y_{a_0,a_1}) = E^*(Y|A_0=a_0,A_1=a_1)$
- $E(Y_{a_0,a_1})$ in the original study is the same as $E^*(Y_{a_0,a_1})$ in the pseudo study

82

Analysis in the final pseudo study

- In the final pseudo study, after IPW weighting by w or ws, we have

$$E^*(Ya_0,a_1) = E^*(Y|A_0=a_0,A_1=a_1)$$
$$E^*(Ya_0,a_1) = \theta_0 + \theta_1(a_0 + a_1)$$

- So, in the final pseudo-study we have that

$$E^*(Y|A_0=a_0,A_1=a_1) = \theta_0 + \theta_1(a_0 + a_1)$$

- Consequently, we can estimate θ_0 and θ_1 by least squares in the pseudo-study !!!

83

Non-stabilized IPW procedure for estimating the parameters of a marginal structural model

- Postulate the marginal structural model of your choice

$$E(Y_{a0,a1}) = m(a_0,a_1,\theta)$$

- Stage 1: for each subject in the study that has $A_1=a_1$, $A_0=a_0$, $L_1=l_1$, $L_0=l_0$ with

$$w = 1 / [\, P(A_1=a_1|L_1=l_1,A_0=a_0,L_0=l_0)^* P(A_0=a_0|L0=l_0)\,]$$

- Stage 2. estimate θ with any statistical package that runs weighted least squares using as weights the values w computed in stage 1, outcome Y and covariates A_0, A_1.

84

Stabilized IPW procedure for estimating the parameters of a marginal structural model

- Postulate the marginal structural model of your choice

$$E(Y_{a0,a1}) = m(a_0,a_1,\theta)$$

- Stage 1: for each subject in the study that has $A_1=a_1$, $A_0=a_0$, $L_1=l_1$, $L_0=l_0$ with

$$ws = [P(A_1=a_1|A_0=a_0) * P(A_0=a_0)] / [P(A_1=a_1|L_1=l_1,A_0=a_0, L_0=l_0)* P(A_0=a_0|L_0=l_0)]$$

- Stage 2. estimate θ with any statistical package that runs weighted least squares using as weights the values ws computed in stage 1, outcome Y and covariates A_0 and A_1.

85

Why is stabilized IPW a better idea than non-stabilized IPW for estimating MSM's?

86

Stabilized vs Non-stabilized IPW

- In realistically sized studies, if a subject receives a very large weight relative to the others then the resulting estimators of the θ's are very imprecise.
- Intuition:
  - Imagine a study with, say 201 subjects.
  - Suppose one subject, say John, receives a weight of 100,000 and all others receive a weight of roughly 2.
  - In such study the estimation of θ is essentially driven by John's data because John counts for 100,000 people whereas all others together just count for 400=200*2
- This is bad ! You don't want your estimate of to be driven by one person's data!!!
- But, with non-stabilized IPW, if by chance a person with a very tiny chance of being assigned to a given trx, say with $P(A_1=1|A_0,L_1)$ very small, did indeed receive trx $A_1=1$, then that person would carry a very large weight and the IPW estimator of θ will be driven by this person's data!

- **IPW with stabilized weights are a way to fix this problem.**
- **The idea is to downweight the subjects that are most likely to receive large weights**

Intuition of why stabilized IPW is a good idea

- Suppose that in your study $P(A_1=1|A_0=0)$ is small, then you would expect that the subjects that received $A_0=0,A_1=1$ will have small propensity $P(A_1=1|A_0=0,L_1=l_1)$ for nearly all the values of l_1 and hence be the more likely to cause trouble.

- So, when you estimate θ by least squares in your pseudo-study, you would like to downweight the contributions from subjects that have "covariates" $A_0=0$ and $A_1=1$.

- You can effectively implement the downweighting by re-weighting the subjects that have "covariates" $A_0=0$ and $A_1=1$ with $P(A_1=1|A_0=0)$

88

Estimation in realistically sized studies

Estimation in realistically sized studies

- We have argued that the reason for adopting a non-saturated MSM is that in realistically sized studies, we may run out of subjects receiving each of the possible treatment regimes of interest to be able to estimate them separately.
- But in realistically sized studies we will also run out of subjects to estimate separately the propensity scores that are needed to carry out IPW!!!

Estimation in studies with realistic sample sizes

- The stabilized IPW procedure requires that you compute for each a_1, a_0, l_1, l_0 estimators of the propensity scores

$$P(A_1=a_1|L_1=l_1,A_0=a_0, L_0=l_0),\ P(A_0=a_0|L_0=l_0)$$

- To estimate, e.g., $P(A_1=a_1|L_1=l_1,A_0=a_0, L_0=l_0)$, with the observed proportion of subjects in your study taking $A_1=a_1$ in stratum $(A_0=a_0,L_1=l_1,L_0=l_0)$ you will run into the problem of thin or empty cells, because you may have very few, if any, subjects in that stratum
- The problem will be aggravated, when some of the L's are continuous variables, or if some of the L's are comprised by several variables.
- Then, you will have no choice but to estimate the propensity scores under parametric models…
- We will discuss how to do this briefly next

91

Logistic regression models for the propensity scores

- To estimate the propensity scores in realistically sized studies, you will need to assume models for

$$P(A_1=a_1|L_1=l_1,A_0=a_0, L_0=l_0),\ P(A_0=a_0|L_0=l_0)$$

- If each of A_0, A_1 are 0/1 variables, often, it is ok to assume models that share parameters across time, and that depend on just one or two lagged variables, e.g.

$$\text{logit}P(A_1=1|A_0, L_1, L_0)=\alpha_0+\alpha_1 L_1$$
$$\text{logit}P(A_0=1|L_0)=\alpha_0+\alpha_1 L_0$$

- If you assume a model with shared parameters across times, you can estimate the parameters of the models by a pooled logistic regression in which each subject contributes k times to the estimation

92

Standard errors

- When you estimate the parameters θ of the marginal structural model with **weighted least squares** using weights equal to either
  - Non-stabilized w, estimated assuming some models for the propensity scores
  - Stabilized ws, estimated assuming some models for the propensity scores in the numerator and denominator
- any statistical package will output standard error estimators of your estimators of θ
- These standard error estimators are actually conservative, i.e. larger than they should be.
- The reason for this is because the statistical software does not know that you have "estimated" the weights. It acts as if the weights were just fixed numbers, not estimators of anything.
- And it turns out that pretending that the weights are fixed known numbers rather than estimated values, leads to over estimating the variance of the estimator of θ. The reason for this is a bit technical and beyond the scope of this course.
- The lesson, though, is that it is ok to use the standard errors from the output of the weighted least squares software because they are conservative.

93

The end.
Thank you!

Andrea Rotnitzky
arotnitzky@utdt.edu

94