Exercises

Contents

1 Exercises

1.1 Association vs causation

1. Table 1 shows the data collected in a study of 12 individuals. The goal was to estimate the effect of daily low-dose aspirin ($A = 1$) versus no aspirin ($A = 0$) on the risk of heart disease ($Y = 1$). The table also shows the values of the potential outcomes that would have been observed under treatment (Y_1) and under no treatment (Y_0) with aspirin.

ID	A	Y	Y_1	Y_0
1	0	0	0	0
2	0	0	0	0
3	0	1	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	1	1
7	1	0	0	1
8	1	0	0	1
9	1	0	0	0
10	1	1	1	0
11	1	1	1	1
12	1	1	1	1

Table 1: Potential outcome data

   (a) Compute the causal risk difference, the causal risk ratio, and the causal odds ratio.

   (b) Compute the causal effects listed in a) for the subpopulation that was factually exposed.

   (c) Compute the causal effects listed in a) for the subpopulation that was factually unexposed.

2. Consider the question 'is there a causal effect of A on Y?' applied to each of the pairs (A, Y) displayed in Table 2. For each pair,

A	$A=1$	$A=0$	Y
Aspirin	Daily use (150 mg)	No use	Coronary heart disease
Low salt diet	Yes	No	Stroke
sex	Female	Male	Lung cancer

Table 2: Causal research questions

(a) do you think the causal question above is appropriate for meaningful inference (i.e. is the causal effect of interest well defined?).

(b) If not, can you propose a more precise causal question (possibly involving a modified version of A) that reduces the vaguenesss of the original question?

3. In this course (as well as in the research field of causal inference in general) we assume that the potential outcome for each subject under each treatment is deterministic (i.e. non-random). This implies, for example, that with respect to a binary treatment A ($A=0$ for untreated and $A=1$ for treated) and a binary outcome Y ($Y=0$ for the favorable outcome and $Y=1$ for the unfavorable outcome), each subject *must* belong to one of the four groups displayed in Table 3. Subjects in

Y_a	'healthy'	'harmed'	'protected'	'doomed'
$A=0$	0	0	1	1
$A=1$	0	1	0	1

Table 3: Principal stratification

the first group do not develop the outcome, regardless of whether they are treated or not - we may call them 'healthy'. Subjects in the second group develop the outcome if and only if they are treated - we may call them 'harmed'. Subjects in the third group develop the outcome if and only if they are not treated - we may call them 'protected'. Subjects in the fourth group develop the outcome regardless of whether they are treated or not - we may call them 'doomed'. This classification of subjects based on joint potential outcomes is usually referred to as 'principal stratification'. Many people feel that the deterministic nature

of potential outcomes and principal stratification is unrealistic (this is especially true for statisticians, who are used to model everything in a random fashion). What is your opinion? Does your answer depend on what particular exposure and outcome we consider? Does your answer depend on what we mean with the word 'subject'?

4. A respectable researcher claims that she can compute individual causal effects in her research area. This claim implies that the fundamental problem of causal inference - the impossibility of observering a subject's outcome under two different values of the exposure - does not apply to her research. Do you think that such a claim can ever be correct? If yes, give an example of a study design that would allow for the computation of individual causal effects. If no, motivate.

5. Consider your research project. Is your project motivated by a causal question (i.e. is your aim to estimate a causal effect)? If so, can you formulate your research question in terms of potential outcomes?

1.2 Estimation of causal effects

1. Consider the data in Table 1. Are treated ($A = 1$) and untreated ($A = 0$) exchangeable?

2. Suppose that you are given the first three columns of Table 1, i.e. ID, A, and Y, but not the potential outcomes Y_0 and Y_1.

   (a) Given the first three columns, can you rule out the null hypothesis that the causal risk difference is equal to 0? If not, fill in example values of Y_0 and Y_1 for each subject so that the causal risk difference is equal to 0.

   (b) Given the first three columns, can you rule out that the causal risk difference is equal to 1? If not, fill in example values of Y_0 and Y_1 for each subject so that the causal risk difference is equal to 1.

   (c) Given the first three columns, which are the possible values for the causal risk difference? Given no data at all, which are the possible values for the causal risk difference?

3. Table 4 shows the data from a study to compute the causal effect of antiretroviral therapy A on death Y in subjects infected with HIV. Individuals were classified as treated ($A = 1$) if they received antiretroviral therapy, and as untreated ($A = 0$) otherwise. Death is coded as $Y = 1$. The variable L represents CD4 count (1:low; 0:high). Assume that the treated and untreated are conditionally exchangeable, given L.

	$L = 0$		$L = 1$	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
$A = 1$	20	30	108	252
$A = 0$	40	10	24	16

Table 4: HIV data

   (a) Compute the conditional causal risk difference, given $L = 0$ and $L = 1$.

   (b) Compute the marginal causal risk difference.

   (c) Can you compute the causal risk difference for those who actually received the treatment? Is this a relevant parameter?

4. Consider your research project. If your aim is to estimate a causal effect, do you think that exposed and unexposed are exchangeable with respect to the outcome of interest in your study population? If not, which covariates do you think you need to adjust for in order to achieve exchangeability?

1.3 Directed Acyclic Graphs

1. Consider your research project. Draw a DAG that describes your study. What do you need to adjust for? Anything you should not adjust for?

2. (*Non-compliance in randomized experiments*). A common feature of randomized experiments is that subjects do not always adhere to their assigned treatment. In this exercise we investigate why this feature

may be problematic. We consider a study in which each subject is randomly assigned either to a new treatement or a standard treatment. The study is unblinded, i.e. the participants are aware of which treatment they are asigned to. Some subjects who are assigned to the new treatment decide to take the standard treatment and vice versa. Thus, we distinguish between *treatment assignment*, which we denote with R (0 for 'assigned to standard treatement', 1 for 'assigned to new treatement'), and *treatment actually taken*, which we denote with A (0 for 'taking standard treatement', 1 for 'taking new treatement'). For each subject, a binary outcome Y is measured (0 for 'unfavorable outcome', 1 for 'favorable outcome').

(a) Draw a DAG that represents this study.

(b) One possible way to analyze this data is to compare the outcome for those who actually took the new treatment ($A = 1$) versus those who actually took the standard treatment ($A = 0$). In the literature, this analysis is usually referred to as the 'as-treated' (AT) analysis. Use the DAG to explain why the AT analysis may be problematic from a causal inference point of view.

(c) An alternative approach is to compare the outcome for those who were randomized to the new treatment ($R = 1$) versus those who were randomized to the standard treatment ($R = 0$). This analysis is usually referred to as the 'intention-to-treat' (ITT) analysis. Use the DAG to explain the rationale behind the ITT analysis.

3. (*Intrumental variables*). Many observational studies suffer from confounding. In this exercise we investigate a method of 'confounding adjustment' which, under certain assumptions, has the remarkable property of producing causal inference even in the presence of unmeasured confounding. Let A be the exposure of interest, let Y be the outcome of interest, and let U be all unmeasured variables (confounders) that affect both A and Y. Let Z be a measured variable which have the following properties: a) U does not affect Z, b) Z does not affect U, c) Z and U don't have common causes, d) Z affects A, e) Z has no effect on Y, apart from an indirect effect mediated through A. A variable Z which have properties a)-e) is called an *instrumental variable*.

(a) Draw a DAG that connects A, Y, U, and Z.

(b) Show that an observed association between Z and Y implies that A has a causal effect on Y (that is, we can test whether A has a causal effect on Y by testing whether Z an Y are associated).

(c) Try to come up with a real epidemiological scenario which may be represented by your DAG, to a reasonable degree of approximation.

4. (*Measurement errors*). A randomized study was carried out to investigate whether high protein diet is beneficial with respect to various health indicators, compared to an ordinary diet. Each subject in a large cohort was randomized to high protein diet ($A = 1$) or ordinary diet ($A = 0$). After 6 months each subject was asked to fill in a detailed questionaire. One question concerned weight loss during the last 6 months. Unfortunately there was reason to believe that not all subjects were totally honest when answering this specific question. Thus, we distinguish between true weight loss (Y) and reported weight loss (Y^*).

   (a) Draw a DAG that represents this study.

   (b) Given your DAG, can you use the association between A and Y^* to *test* for a causal effect of A on Y?

   (c) If your answer is 'no' to the previous questions, what additional assumptions would you have to make (i.e. how would you have to modify your DAG) in order for you to give an affirmative answer? If your answer is yes, can you then see any problem at all with the fact that the study suffers from measurement errors?

5. (*Post treatment selection bias; due to Hernandez-Diaz et al 2006*). Birth weight is a strong predictor of neonatal and infant mortality. Probably for that reason, and because birthweight data are readily available, investigators have frequently stratified on birth weight when evaluating the effect of other risk factors (e.g., maternal smoking, multiple pregnancies, placenta previa) on infant mortality. This stratification often produces a rather counter-intuitive result: infants who are exposed to the particular risk factor (e.g. infants whose mother smoked during pregnancy) have a lower mortality rate than infants who are not exposed to the risk factor. This phenomenon is known as

the 'birth weight paradox', and it has been a source of controversy for decades.

(a) Draw a DAG that represents the causal relations betwen smoking during pregnancy, birth weight, and infant mortality.

(b) Use the DAG to discuss possibly explanations of the birth weight paradox.

6. Read the article 'Cigarette smoking and the incidence of Parkinson's disease in two prospective studies' (*Annals of Neurology* 2001; 50:780-786).

(a) Summarize the designs and methods of the study.

(b) Summarize the results of the study, i.e. what associations did the authors find?

(c) In the first column of page 784, a paragraph begins with 'The key question is whether this strong inverse association reflects a truly protective effect of smoking on the risk of developing PD.' Draw at least one DAG that is consistent with the explanations given in this paragraph.

(d) In the first column of page 785, a paragraph begins with 'There are also several versions of the argument claiming the existence of a causal effect of PD on smoking behavior...' Draw at least one DAG that is consistent with the explanations given in this paragraph.

(e) In the first column of page 785, a paragraph begins with 'Confounding...' Draw at least one DAG that is consistent with the explanations given in this paragraph.

(f) In the first column of page 784, a paragraph begins with 'The information bias...' Draw at least one DAG that is consistent with the explanations given in this paragraph.

(g) In the second column of page 784, a paragraph begins with 'There are a number of variations of the hypothesis that selection bias...' Draw at least one DAG that is consistent with the explanations given in this paragraph.

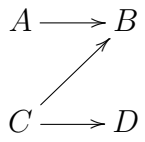

Figure 1:

7. Find all pairwise marginal and conditional independencies in the DAG in Figure 1.

8. For each of exercies a), b) and c) below, draw a DAG that contains four variables A, B, C, and D. Each DAG should imply the (conditional) independencies listed in the corresponding exercise, and *only* these independencies.

   (a)

   $$A \amalg C|B$$

   (b)

   $$A \amalg C|B, D$$

   (c)

   $$A \amalg C|B, D$$

   and

   $$B \amalg D|A$$

1.4 Multiple exposures

1. Consider the DAG in Figure 2. Suppose that the observed number of subjects having each combination of (A^0, L, A^1, Y) are as in Table 5.

   (a) Given the population proportions of (A^0, L, A^1, Y), which of the arrows in Figure 2 can you test the presence of, using 'standard' association tests? Carry out all feasible tests.

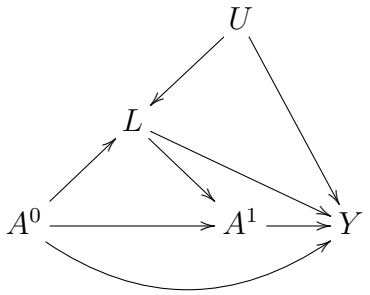

Figure 2:

A^0	L	A^1	$Y=0$	$Y=1$
0	0	0	192	48
0	0	1	48	12
0	1	0	96	64
0	1	1	24	16
1	0	0	10	10
1	0	1	40	40
1	1	0	16	64
1	1	1	64	256

Table 5: Data for the DAG in Figure 2.

(b) Use the G-formula to compute the direct effect of A^0 on Y, at $A^1 = 0$ and $A^1 = 1$ separately. Given your answer to (a), can you test for a direct effect of A^0 on Y (i.e. not mediated through A^1) using a simpler method than the G-formula? If so, do it and compare to the result of the G-formula.

1.5 Regression models

1. Load the data set `aids` in the file `single.txt`. It contains the data set used in the lecture to illustrate regression models for point exposures.

   (a) Replicate the results showed in the lecture. Make sure you understand what the code does. If you are not an `R`-user, think about how you would do these analyses in your favorite software.

   (b) For both the outcome model and the exposure model, add a square term for CD4 count. In `R`, this is done by adding the term `I(L^2)` to the right-hand side of the models in the `glm` function. Use these more elaborate models to estimate the marginal causal effect of A on Y. Compare with the simpler model; does the inclusion of the second order term have any substantial impact on the results?

2. Load the data set `aids` in the file `multiple.txt`. It contains the data set used in the lecture to illustrate regression models for multiple exposures.

   (a) Replicate the results showed in the lecture. If you are no an `R`-user, think about how you would do these analyses in your favorite software.

   (b) Add an interaction term between a^0 and a^1 in the MSM. Use this more elaborate model to estimate the joint effect of A^0 and A^1. Compare with the simpler model; does the inclusion of the second order term have any substantial impact on the results?

2 Solutions

2.1 Association vs causation

1. (a) $\Pr(Y_1 = 1) = 4/12 = 1/3$. $\Pr(Y_0 = 1) = 8/12 = 2/3$. Causal risk difference:

$$\Pr(Y_1 = 1) - \Pr(Y_0 = 1) = -1/3$$

Causal risk ratio:

$$\frac{\Pr(Y_1 = 1)}{\Pr(Y_0 = 1)} = 1/2$$

Causal odds ratio:

$$\frac{\Pr(Y_1 = 1)}{\Pr(Y_1 = 0)} \Big/ \frac{\Pr(Y_0 = 1)}{\Pr(Y_0 = 0)} = 1/4$$

(b) $\Pr(Y_1 = 1|A = 1) = 3/6 = 1/2$. $\Pr(Y_0 = 1|A = 1) = 4/6 = 2/3$. Conditional causal risk difference, given $A = 1$:

$$\Pr(Y_1 = 1|A = 1) - \Pr(Y_0 = 1|A = 1) = -1/6$$

Conditional causal risk ratio, given $A = 1$:

$$\frac{\Pr(Y_1 = 1|A = 1)}{\Pr(Y_0 = 1|A = 1)} = 3/4$$

Conditional causal odds ratio, given $A = 1$:

$$\frac{\Pr(Y_1 = 1|A = 1)}{\Pr(Y_1 = 0|A = 1)} \Big/ \frac{\Pr(Y_0 = 1|A = 1)}{\Pr(Y_0 = 0|A = 1)} = 1/2$$

(c) $\Pr(Y_1 = 1|A = 0) = 1/6$. $\Pr(Y_0 = 1|A = 0) = 4/6 = 2/3$. Conditional causal risk difference, given $A = 0$:

$$\Pr(Y_1 = 1|A = 0) - \Pr(Y_0 = 1|A = 0) = -1/2$$

Conditional causal risk ratio, given $A = 0$:

$$\frac{\Pr(Y_1 = 1|A = 0)}{\Pr(Y_0 = 1|A = 0)} = 1/4$$

Conditional causal odds ratio, given $A = 0$:

$$\frac{\Pr(Y_1 = 1|A = 0)}{\Pr(Y_1 = 0|A = 0)} \Big/ \frac{\Pr(Y_0 = 1|A = 0)}{\Pr(Y_0 = 0|A = 0)} = 1/10$$

2. Discuss in class.

3. Discuss in class.

4. Discuss in class.

5. Discuss in class.

2.2 Estimation of causal effects

1. $\Pr(Y_1 = 1|A = 1) \neq \Pr(Y_1 = 1|A = 0)$, so exposed an unexposed are not exchangeable.

2. (a) The causal null hypothesis cannot be ruled out by the first three columns. An example is given by Table 6.

ID	A	Y	Y_1	Y_0
1	0	0	0	0
2	0	0	0	0
3	0	1	1	1
4	0	1	1	1
5	0	1	1	1
6	0	1	1	1
7	1	0	0	0
8	1	0	0	0
9	1	0	0	0
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1

Table 6: Data under causal null hypothesis.

(b) It is not possible that the causal risk difference is equal to 1. This would only happen if $\Pr(Y_1 = 1) = 1$ and $\Pr(Y_0 = 1) = 0$, i.e. if there are no subjects for which $Y_0 = 1$. But we know from the first three columns in the table that Y_0 is not 0 for everyone, since there are subjects for which $A = 0$ and $Y = 1$. When $A = 0$, we have that $Y_0 = Y$, so for these subjects we have that $Y_0 = 1$.

(c) The causal risk difference is maximal when $\Pr(Y_1 = 1)$ is maximal and $\Pr(Y_0 = 1)$ is minimal. The maximal value for $\Pr(Y_1 = 1)$ occurs when all unexposed (those with $A = 0$) have $Y_1 = 1$. The minimal value for $\Pr(Y_0 = 1)$ occurs when all exposed (those with $A = 1$) have $Y_0 = 0$. See Table 7. We then have that $\Pr(Y_1 = 1) = 9/12$ and $\Pr(Y_0 = 1) = 4/12$, so that the causal risk difference is equal to $5/12$. The causal risk difference is minimal

ID	A	Y	Y_1	Y_0
1	0	0	1	0
2	0	0	1	0
3	0	1	1	1
4	0	1	1	1
5	0	1	1	1
6	0	1	1	1
7	1	0	0	0
8	1	0	0	0
9	1	0	0	0
10	1	1	1	0
11	1	1	1	0
12	1	1	1	0

Table 7: Data that gives maximal value for $\Pr(Y_1 = 1)$ and minimal value for $\Pr(Y_0 = 1)$.

when $\Pr(Y_1 = 1)$ is minimal and $\Pr(Y_0 = 1)$ is maximal. The minimal value for $\Pr(Y_1 = 1)$ occurs when all unexposed (those with $A = 0$) have $Y_1 = 0$. The maximal value for $\Pr(Y_0 = 1)$ occurs when all exposed (those with $A = 1$) have $Y_0 = 1$. See Table 8. We then have that $\Pr(Y_1 = 1) = 3/12$ and $\Pr(Y_0 = 1) = 10/12$, so that the causal risk difference is equal to $-7/12$. With no data, the maximal and minimal values for the causal risk difference are 1 and -1, respectively.

3. We first compute $\Pr(Y = 1|A = 1, L = 1) = 108/(108 + 252) = 0.3$, $\Pr(Y = 1|A = 0, L = 1) = 24/(24 + 16) = 0.6$, $\Pr(Y = 1|A = 1, L = 0) = 20/(20+30) = 0.4$, $\Pr(Y = 1|A = 0, L = 0) = 40/(40+10) = 0.8$.

ID	A	Y	Y_1	Y_0
1	0	0	0	0
2	0	0	0	0
3	0	1	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	0	1
7	1	0	0	1
8	1	0	0	1
9	1	0	0	1
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1

Table 8: Data that gives minimal value for $\Pr(Y_1 = 1)$ and maximal value for $\Pr(Y_0 = 1)$.

We have that

(a)

$$\begin{aligned} \Pr(Y_a = 1|L) &= \Pr(Y_a = 1|A = a, L) \\ &= \Pr(Y = 1|A = a, L), \end{aligned}$$

where the second equality follows from the assumption of conditional exchangeability, given L. The conditional causal risk difference, given $L = 1$, is equal to

$$\Pr(Y_1 = 1|L = 1) - \Pr(Y_0 = 1|L = 1) = 0.3 - 0.6 = -0.3.$$

The conditional causal risk difference, given $L = 1$, is equal to

$$\Pr(Y_1 = 1|L = 0) - \Pr(Y_0 = 1|L = 0) = 0.4 - 0.8 = -0.4.$$

(b) We first compute $\Pr(L = 1) = (108 + 252 + 24 + 16)/(108 + 252 + 24+16+20+30+40+10) = 0.8$, $\Pr(L = 0) = 1-\Pr(L = 1) = 0.2$. Under the assumption of conditional exchangeability, given L, the

standardization formula gives that

$$\Pr(Y_a = 1) = \sum_L \Pr(Y = 1|A = a, L)\Pr(L).$$

For $a = 1$ we have that

$$\begin{aligned}\Pr(Y_1 = 1) &= \overbrace{0.3}^{\Pr(Y=1|A=1,L=1)} \times \overbrace{0.8}^{\Pr(L=1)} + \overbrace{0.4}^{\Pr(Y=1|A=1,L=0)} \times \overbrace{0.2}^{\Pr(L=0)} \\ &= 0.32.\end{aligned}$$

For $a = 0$ we have that

$$\begin{aligned}\Pr(Y_0 = 1) &= \overbrace{0.6}^{\Pr(Y=1|A=0,L=1)} \times \overbrace{0.8}^{\Pr(L=1)} + \overbrace{0.8}^{\Pr(Y=1|A=0,L=0)} \times \overbrace{0.2}^{\Pr(L=0)} \\ &= 0.64.\end{aligned}$$

The causal risk difference is equal to

$$\Pr(Y_1 = 1) - \Pr(Y_0 = 1) = 0.32 - 0.64 = -0.32.$$

(c) The causal risk difference for the treated is defined as

$$\Pr(Y_1 = 1|A = 1) - \Pr(Y_0 = 1|A = 1).$$

Form the definition of Y_1 we have that

$$\begin{aligned}\Pr(Y_1 = 1|A = 1) &= \Pr(Y = 1|A = 1) \\ &= (20 + 108)/(20 + 108 + 30 + 252) \\ &= 0.312.\end{aligned}$$

We first compute $\Pr(L = 1|A = 1) = (108+252)/(108+252+20+30) = 0.878$ and $\Pr(L = 0|A = 1) = 1-\Pr(L = 1|A = 1) = 1.122$. We then compute $\Pr(Y_0 = 1|A = 1)$ as

$$\begin{aligned}\Pr(Y_0 = 1|A = 1) &= \sum_L \Pr(Y_0 = 1|A = 1, L)\Pr(L|A = 1) \\ &= \sum_L \Pr(Y_0 = 1|A = 0, L)\Pr(L|A = 1) \\ &= \sum_L \Pr(Y = 1|A = 0, L)\Pr(L|A = 1) \\ &= \overbrace{0.6}^{\Pr(Y=1|A=0,L=1)} \times \overbrace{0.878}^{\Pr(L=1|A=1)} + \overbrace{0.8}^{\Pr(Y=1|A=0,L=0)} \times \overbrace{0.122}^{\Pr(L=0|A=1)} \\ &= 0.624,\end{aligned}$$

where the second equality follows from the assumption of conditional exchangeability, given L. The causal risk difference for the treated is equal to

$$\Pr(Y_1 = 1|A = 1) - \Pr(Y_0 = 1|A = 1) = 0.312 - 0.624 = -0.312.$$

4. Discuss in class.

2.3 Directed Acyclic Graphs

1. Discuss in class.

2. (a) See Figure 3. Here, U represents all (usually unmeasured) factors that affect both treatment taken and the outcome. For instance, subjects with a bad health may be more eager to take the treatment, and also more likely to have an unfavorable outcome.

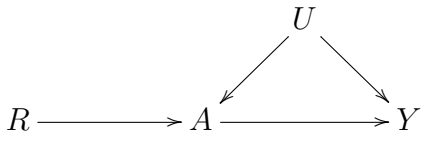

Figure 3: DAG for a randomized experiment with non-compliance.

   (b) An association between A and Y may be explained by the non-causal path $A \leftarrow U \rightarrow Y$.

   (c) If the arrow from A to Y is missing, then there would be no association between R and Y, since the path $R \rightarrow A \leftarrow U \rightarrow Y$ is blocked at A. Hence, if we observe an association between R and Y, then we can say that the arrow from A to Y exists.

3. (a) See Figure 4.

   (b) If the arrow from A to Y is missing, then there would be no association between Z and Y, since the path $Z \rightarrow A \leftarrow U \rightarrow Y$ is blocked at A. Hence, if we observe an association between Z and Y, then we can say that the arrow from A to Y exists.

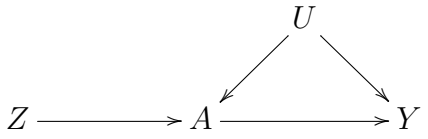

Figure 4: DAG for an instrumental variable setting.

(c) Discuss in class.

4. (a) See Figure 5. The arrow from Y to Y^* reflects the fact that true weight loss (hopefully) influences reported weight loss. U is the set of all factors that affect both Y and Y^*. For instance, young people may loss more weight when put on a high protein diet, but may also tend to overestimate (or underestimate!) their true weight loss, as compared to old people. The arrow from A to Y^* represents the influence of diet on reported weight loss. For instance, people who have followed the high protein diet may feel more confident that they have lost weight than people who have followed the ordinary diet (a placebo effect).

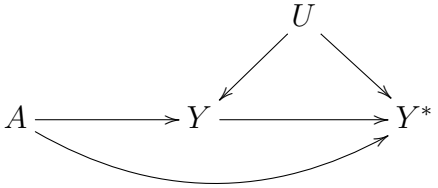

Figure 5: DAG for a randomized study with measurement error in the outcome.

(b) No; an association between A and Y^* could be explained by the path $A \rightarrow Y^*$.

(c) If the arrow from A to Y^* is absent, then an association between A and Y^* proves that the arrow from A to Y exists. However, the association between A and Y^* is probably weaker then the association between A and Y. Thus, even though we can test for

a causal effect, we cannot estimate its mganitude. Also, measurement errors are likely to reduce the power of the study.

5. (a) See Figure 6. Here, U is the set of all factors (e.g. genetics and lifestyle) that may influence the child's birth weight and the child's risk of death.

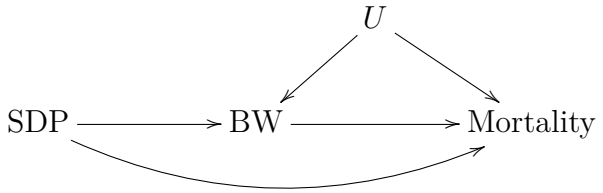

Figure 6: DAG for the causal relations between smoking during pregnancy (SDP), birth weight (BW), and infant mortality.

   (b) The 'birth weight paradox' may be explained by the non-causal path SDP $\rightarrow$ BW $\leftarrow U \rightarrow$ Mortality; this path becomes open when stratifying on birth weight.

6. Discuss in class.

7. The path $A \rightarrow B \leftarrow C$ is blocked at B, so $A \amalg C$. The path $A \rightarrow B \leftarrow C$ is blocked at B, if we adjust for D, so $A \amalg C|D$. The path $A \rightarrow B \leftarrow C \rightarrow D$ is blocked at B, so $A \amalg D$. The path $A \rightarrow B \leftarrow C \rightarrow D$ is blocked at B, if we adjust for C, so $A \amalg D|C$. The path $A \rightarrow B \leftarrow C \rightarrow D$ is blocked at B, if we adjust for (B, C), so $A \amalg D|(B, C)$.The path $B \leftarrow C \rightarrow D$ is blocked at C, if we adjust for C, so $B \amalg D|C$. The path $B \leftarrow C \rightarrow D$ is blocked at C, if we adjust for (A, C), so $B \amalg D|(A, C)$.

8. (a) See Figure 7.
   (b) See Figure 8.
   (c) See Figure 9.

2.4 Multiple exposures

1. (a) By adjusting for A^0 and L, we block all paths between A^1 and Y, except the path $A^1 \rightarrow Y$. Hence, if there is a conditional

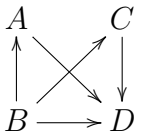

Figure 7: DAG that implies $A \amalg C|B$ and no other (conditional) independencies.

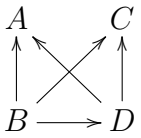

Figure 8: DAG that implies $A \amalg C|B, D$ and no other (conditional) independencies.

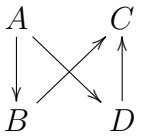

Figure 9: DAG that implies $A \amalg C|B, D$ and $B \amalg D|A$, and no other (conditional) independencies.

association between A^1 and Y, given A^0, L and A^1, then the arrow from A^1 to Y is present. We have that

$$\begin{aligned}
\Pr(Y = 1|A^0 = 0, L = 0, A^1 = 0) &= \frac{48}{48 + 192} = 0.2 \\
\Pr(Y = 1|A^0 = 0, L = 0, A^1 = 1) &= \frac{12}{12 + 48} = 0.2
\end{aligned}$$

$$\begin{aligned}
\Pr(Y = 1|A^0 = 0, L = 1, A^1 = 0) &= \frac{64}{64 + 96} = 0.4 \\
\Pr(Y = 1|A^0 = 0, L = 1, A^1 = 1) &= \frac{16}{64 + 24} = 0.4
\end{aligned}$$

$$\begin{aligned}
\Pr(Y = 1|A^0 = 1, L = 0, A^1 = 0) &= \frac{10}{10 + 10} = 0.5 \\
\Pr(Y = 1|A^0 = 1, L = 0, A^1 = 1) &= \frac{40}{40 + 40} = 0.5
\end{aligned}$$

$$\begin{aligned}
\Pr(Y = 1|A^0 = 1, L = 1, A^1 = 0) &= \frac{64}{64 + 16} = 0.8 \\
\Pr(Y = 1|A^0 = 1, L = 1, A^1 = 1) &= \frac{256}{256 + 64} = 0.8
\end{aligned}$$

We observe that $\Pr(Y = 1|A^0, L, A^1 = 0) = \Pr(Y = 1|A^0, L, A^1 = 1)$ for all L and A^1. Hence, $Y \amalg A^1|A^0, L$, which implies that the arrow from A^1 to Y is absent.

By adjusting for A^0 we block all paths between L and A^1, except the path $L \rightarrow A^1$. Hence, if there is a conditional association between L and A^1, given A^0, then the arrow from L to A^1 is present. We have that

$$\begin{aligned}
\Pr(A^1 = 1|A^0 = 0, L = 0) &= \frac{48 + 12}{48 + 12 + 192 + 48} = 0.2 \\
\Pr(A^1 = 1|A^0 = 0, L = 1) &= \frac{24 + 16}{24 + 16 + 96 + 64} = 0.2
\end{aligned}$$

$$\begin{aligned}\Pr(A^1=1|A^0=1,L=0) &= \frac{40+40}{40+40+10+10}=0.8\\ \Pr(A^1=1|A^0=1,L=1) &= \frac{64+256}{64+256+16+64}=0.8\end{aligned}$$

We observe that $\Pr(A^1=1|A^0,L=0)=\Pr(A^1=1|A^0,L=1)$ for all A^0. Hence, $A^1 \amalg L|A^0$, which implies that the arrow from L to A^1 is absent.

By adjusting for L we block all paths between A^0 and A^1, except the path $A^0 \to A^1$. Hence, if there is a conditional association between A^0 and A^1, given L, then the arrow from A^0 to A^1 is present. We observe that $\Pr(A^1=1|A^0=0,L) \neq \Pr(A^1=1|A^0=1,L)$ for all L. Hence, $A^1 \not\amalg A^0|L$, which implies that the arrow from A^0 to A^1 is present.

Without making any adjustments, all paths are blocked between A^0 and L, except the path $A^0 \to L$. Hence, if there is a marginal association between A^0 and L, then the arrow from A^0 to L is present. We have that

$$\begin{aligned}\Pr(L=1|A^0=0) &= \frac{96+64+24+16}{96+64+24+16+192+48+48+12}=0.4\\ \Pr(L=1|A^0=1) &= \frac{16+64+64+256}{16+64+64+256+10+10+40+40}=0.8\end{aligned}$$

We observe that $\Pr(L=1|A^0=0) \neq \Pr(L=1|A^0=1)$. Hence, $L \not\amalg A^0$, which implies that the arrow from A^0 to L is present.

We conclude that the DAG can be simplified as in Figure 10.

(b) Let $Y_{a^0a^1}$ be the potential outcome under joint exposures $A^0=a^0$ and $A^1=a^1$. The G-formula gives that

$$\begin{aligned}\Pr(Y_{00}=1) &= \overbrace{0.2}^{\Pr(Y=1|A^0=0,L=0,A^1=0)} \times \overbrace{0.6}^{\Pr(L=0|A^0=0)}\\ &+ \underbrace{0.4}_{\Pr(Y=1|A^0=0,L=1,A^1=0)} \times \underbrace{0.4}_{\Pr(L=1|A^0=0)} = 0.28\end{aligned}$$

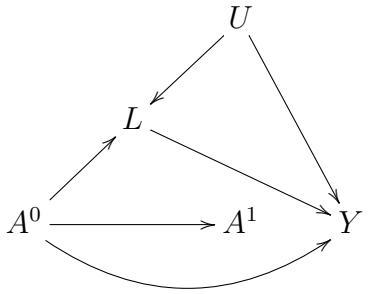

Figure 10:

$$\begin{aligned}
\Pr(Y_{10}=1) &= \overbrace{0.5}^{\Pr(Y=1|A^0=1,L=0,A^1=0)} \times \overbrace{0.2}^{\Pr(L=0|A^0=1)} \\
&+ \underbrace{0.8}_{\Pr(Y=1|A^0=1,L=1,A^1=0)} \times \underbrace{0.8}_{\Pr(L=1|A^0=1)} = 0.74
\end{aligned}$$

$$\begin{aligned}
\Pr(Y_{01}=1) &= \overbrace{0.2}^{\Pr(Y=1|A^0=0,L=0,A^1=1)} \times \overbrace{0.6}^{\Pr(L=0|A^0=0)} \\
&+ \underbrace{0.4}_{\Pr(Y=1|A^0=0,L=1,A^1=1)} \times \underbrace{0.4}_{\Pr(L=1|A^0=0)} = 0.28
\end{aligned}$$

$$\begin{aligned}
\Pr(Y_{11}=1) &= \overbrace{0.5}^{\Pr(Y=1|A^0=1,L=0,A^1=1)} \times \overbrace{0.2}^{\Pr(L=0|A^0=1)} \\
&+ \underbrace{0.8}_{\Pr(Y=1|A^0=1,L=1,A^1=1)} \times \underbrace{0.8}_{\Pr(L=1|A^0=1)} = 0.74
\end{aligned}$$

Hence, the direct effect, as a risk difference, is equal to $0.74 - 0.28 = 0.46$, at both $a^1 = 0$ and $a^1 = 1$.

From the DAG in Figure 10 we observe that there is no mediated (through A^1) effect of A^0 on Y. Thus, the total effect of A^0 on Y is equal to the direct effect. Let Y_{a^0} be the potential outcome

under exposure $A^0 = a^0$. We have that

$$\begin{aligned}
\Pr(Y_0 = 1) &= \Pr(Y = 1 | A^0 = 0) \\
&= \frac{48 + 12 + 64 + 16}{48 + 12 + 64 + 16 + 192 + 48 + 96 + 24} = 0.28 \\
\Pr(Y_1 = 1) &= \Pr(Y = 1 | A^0 = 1) \\
&= \frac{10 + 40 + 64 + 256}{10 + 40 + 64 + 256 + 10 + 40 + 16 + 64} = 0.74
\end{aligned}$$

We observe that $\Pr(Y_{a^0} = 1) = \Pr(Y_{a^0 a^1} = 1)$ for all a^0 and a^1. Hence, the total effect is indeed equal to the direct effect.

2.5 Regression models

1. (a) Discuss in class.

   (b) Discuss in class.

2. (a) Discuss in class.

   (b) Discuss in class.