

PHS2000B Exam 2

Econometrics, Missing Data & Meta-Analysis

May 11, 2021

1 Instrumental Variables (12 points)

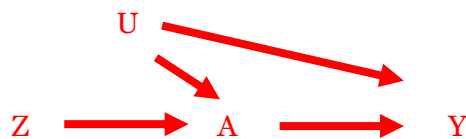
Imagine it is 1995 and you are part of a research team that has just won a grant to study the effect of hospital length of stay (LOS) following delivery on risk of hospital readmission for newborns in the three months after delivery. Lengths of postpartum stay in the United States had decreased from an average over 5 days pre-1970 to a low of 1.1 days on average in 1995 and the medical community is concerned about potential negative effects of this trend.

Your team decides to perform a small randomized controlled trial to answer this question by recruiting a cohort of pregnant individuals who are randomized to either 1 day ($Z = 1$) or 3 day ($Z = 0$) hospital lengths of stay after a healthy vaginal delivery. (In 1996, Congress passed the Newborns' and Mothers' Health Protection Act, which required health insurance providers to extend hospital stay benefits to at least 48 hours after vaginal delivery.)

Questions 1-3 are based on this randomized controlled trial design and the following results.

Randomized to 1 day				Randomized to 3 days			
		Postpartum LOS				Postpartum LOS	
Rehospitalization		1 day	3 days	Rehospitalization		1 day	3 days
	Yes	8	0		Yes	0	2
	No	22	0		No	0	28

1. Draw a DAG depicting this study design including the randomized treatment assignment. Remember to define any variables you use. (1 point)



Z is randomization. A is binary postpartum length of stay. Y is binary rehospitalization. U is an unmeasured confounder. (Don't have to draw U for full credit)

2. Based on the data in Table 1, calculate and interpret the intention-to-treat (ITT) effect. (2 points)

$$ITT = \Pr(Y = 1|Z = 1) - \Pr(Y = 1|Z = 0) = \frac{8}{30} - \frac{2}{30} = 0.2$$

Those who were randomized to 1 day postpartum length of stay had 0.2 greater risk of rehospitalization compared to those who were randomized to 3 days postpartum length of stay.

- What are two requirements that must be true for the ITT you calculated in question 2 to be equal to the average treatment effect (ATE) for the effect of postpartum hospital length of stay on risk of rehospitalization? (1 point)

May select any two of the following 3 options (or others at grader's discretion):

Perfect compliance.

No direct effect of randomization on rehospitalization.

No chance imbalance in characteristics due to small sample size.

Your team is concerned that the randomized controlled trial consists of a very small sample selected from a single hospital and will fail to generalize to a wider population. To address this issue, you decide to perform an observational study for the effect of postpartum hospital length of stay on the risk of hospital readmission for newborns in three months post-delivery using medical records data. Your team is worried about a host of unmeasured confounders of the association that are not captured in administrative data. To allay these concerns, you suggest an instrumental variable analysis using day of week of delivery as an instrument for postpartum hospital length of stay with the idea being that deliveries at the end of the week (e.g., Friday/Saturday) will have longer length of stay compared to earlier in the week (e.g., Monday/Tuesday) because hospital billing offices close on Fridays at 5 PM and do not reopen until Monday mornings.

Questions 4-8 are based on this instrumental variable design and the following results.

Monday/Tuesday				Friday/Saturday			
		Postpartum LOS				Postpartum LOS	
Rehospitalization		≤2 days	>2 days	Rehospitalization		≤2 days	>2 days
	Yes	1,575	1,565		Yes	1,129	945
	No	14,354	26,949		No	9,956	30,124

- Your colleague runs a first stage linear model (identity link) in which postpartum LOS is regressed against day of delivery and obtains the following output. Is the relevance assumption satisfied? Is day of delivery a strong instrument? What is the concern with weak instruments in IV estimation? (2 points)

(Low_LOS = 1 for ≤2 days, Low_LOS = 0 for >2 days; MT = 1 for Monday/Tuesday delivery, MT = 0 for Friday/Saturday delivery)

```

Call:
lm(formula = Low_LOS ~ MT, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.3584 -0.3584 -0.2630  0.6416  0.7370

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.262964   0.002245  117.16  <2e-16 ***
MT           0.095450   0.003133   30.46  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4608 on 86595 degrees of freedom
Multiple R-squared:  0.0106,    Adjusted R-squared:  0.01059
F-statistic: 928.1 on 1 and 86595 DF,  p-value: < 2.2e-16

```

Yes, the relevance assumption is satisfied because there is a positive association between day of week of delivery and hospital LOS. The F-statistic of 928.1 is greater than the traditional cutoff of 10 meaning that the day of delivery instrument is strong.

The concern with weak instruments is bias due to inflation of the effect estimate because the instrument-exposure relationship is in the denominator of the LATE estimator.

5. Explain the exclusion assumption in the context of this study and provide one plausible example of how the exclusion assumption could be violated in this study. (1 point)

In the context of this study, a violation of the exclusion assumption would mean that delivery day would have a direct or indirect effect on rehospitalization not through hospital LOS. This could occur if the hospital staff is different during different times of the week and it has an effect on the health of the newborn. For example, if the weekend hospital staff is less experienced or less attentive than the workweek staff, then it may have a direct effect on the health of the newborn and subsequent risk of rehospitalization.

6. Explain the independence assumption in the context of this study and provide one plausible example of how the independence assumption could be violated in this study. (1 point)

In the context of this study, a violation of the independence assumption would mean that delivery day would be correlated with another variable with an effect on rehospitalization. This could occur if air pollution is different during different times of the week and it has an effect on the health of the newborn. For example, ozone concentrations in the atmosphere are higher on Sundays compared to other days, which may affect the health of the newborn and subsequent risk of rehospitalization.

7. Based on the data in Table 2, calculate an IV estimate of the effect of hospital LOS on rehospitalization and provide a causal interpretation. (2 points)

$$LATE = \frac{\Pr(Rehosp|MT) - \Pr(Rehosp|FS)}{\Pr(\leq 2 \text{ days}|MT) - \Pr(\leq 2 \text{ days}|FS)} = \frac{(3,140/44,443) - (2,074/42,154)}{(15,929/44,443) - (11,085/42,154)} = 0.22$$

Among the compliers, being delivered with ≤ 2 days postpartum length of stay increases risk of hospital readmission by 0.22 compared to being delivered with > 2 days postpartum length of stay.

8. What assumption is needed to interpret your effect estimate from Question 7 as a LATE? Describe an individual who would violate this assumption in this study. (2 points)

We need the monotonicity assumption or the assumption of no defiers. In the context of this study, a defier would be an individual who would have had < 2 days postpartum length of stay if they delivered on a Friday or Saturday and would have had ≥ 2 days postpartum length of stay if they delivered on a Monday or Tuesday.

2. Regression Discontinuity Design (12 points)

In the United States, minimum legal drinking ages varied widely before 1975, when several states began to raise their drinking ages in response to drunk driving fatalities. In 1984, the National Minimum Drinking Age Act required states to raise their age of possession to 21, or else lose 10% of federal highway funds. Most states responded by raising both the age of possession and the minimum legal drinking age to 21.

While the 21-year minimum legal drinking age remains [highly popular](<https://news.gallup.com/poll/174077/lowering-drinking-age.aspx>), it is not without critics. The age of majority in most states is 18, while the minimum legal drinking age is 21, [causing some critics to observe](<https://www.nytimes.com/2020/01/18/us/usa-legal-age.html>) that young adults can vote, drive, and enlist in the military before they are viewed by the law as responsible enough to drink a beer. Others argue that the law is widely disregarded, with many young adults drinking before they are old enough to do so legally. As a [former college president wrote in the New York Times](<https://www.nytimes.com/roomfordebate/2015/02/10/you-must-be-21-to-drink/raising-the-drinking-age-to-21-has-been-a-disasterous-30-year-experiment>), “the overwhelming response of young people has been, not compliance, but contempt for the law.” The effects of minimum legal drinking age laws are therefore of significant interest, both for public health researchers, economists, and policy-makers and for the public at large. So far, the evidence suggests that the minimum legal drinking age of 21, despite the criticisms mentioned above, [saves lives and protects health](<https://www.cdc.gov/alcohol/fact-sheets/minimum-legal-drinking-age.htm>). Motor vehicle crashes saw significant declines when minimum ages were raised to 21, reducing alcohol-related mortality in young drivers. And while young adults do drink before turning 21, the overall effect of minimum legal drinking age laws has been a reduction in drinking, even among individuals \emph{above} the minimum age. Because the minimum drinking age of 21 is a widespread policy of significant interest, econometric studies have exploited a discontinuity in drinking habits just above and just below the 21-year-old threshold. The following questions are based on work by [Carpenter and Bobkin (2009)](https://hollis.harvard.edu/primo-explore/fulldisplay?docid=TN_cdi_proquest_miscellaneous_1835551179&context=PC&vid=HVD2&search_scope=everything&tab=everything&lang=en_US) which uses a regression discontinuity design to explore the causal effect of alcohol consumption on mortality at the threshold of the minimum drinking age.

1. To overcome endogeneity, researchers can randomize exposure in experimental studies or attempt to adjust for confounding variables in an observational study. Briefly explain why these approaches are not appropriate in this context, or why they would not necessary yield an unbiased causal estimate. \points{2}

An RCT would not be ethical, as alcohol consumption is known to be harmful. The study, further, would offer no potential benefit to patients, yet may involve significant risks. Thus, an RCT would not be appropriate in this context.

In an observational study, may attempt to overcome endogeneity by adjusting for known confounders, but it is likely that we may not have data on all possible common causes. For example, preferences around risk may be difficult to observe; more risk-adverse individuals might avoid drinking and be more cautious in avoiding accidents and life-threatening actions. Omitting this would cause omitted variable bias, resulting in an overestimation of the causal effect of alcohol consumption and mortality.

2. As one of the critics of the minimum drinking age law who is quoted in the prompt argues, many people under the age of 21 drink, despite it being illegal to do. If the critic is correct in saying that disregard for the minimum legal drinking age is the “overwhelming” response of individuals under 21, 1) what would this imply about the discontinuity of drinking behaviors at the threshold and 2) how might this cause a challenge to fuzzy RD estimation in this study?

If the minimum legal drinking age is widely disregarded, we may be concerned that our assumption of relevance is not met. Relevance in this context means that the probability of alcohol consumption is discontinuous at 21, such that those above the cut-off are more likely to drink alcohol than those below it. If the minimum drinking age law has no effect on drinking patterns, we would expect drinking to be continuous at the threshold, such that there is no discontinuity at all. As a result, our first stage will be weak, and our fuzzy RD estimate will likely be biased away from the true causal effect.

3. Recall that in a fuzzy RD, the estimand will yield the complier average causal effect (or the CACE) under the assumption of monotonicity, or no defiers at the threshold. In one or two sentences, briefly describe who a defier would be in the context of this study.

Monotonicity implies that the threshold affects the probability of receiving the exposure in the same direction for all individuals at and near the threshold. In this case, we can interpret it as the assumption that everyone’s probability of alcohol consumption would be either unchanged or greater if they were to be above the legal drinking age than if they were to be below it. A defier, therefore, is someone who’s probability is lesser, rather than greater, under these conditions. A defier in this study would be someone who would be more likely to consume alcohol **if** they were under 21 **and** less likely to consume alcohol **if** they were 21 and older.

4. Econometricians using regression discontinuity designs may be concerned with manipulation of the forcing variable at the threshold, as well as measurement error if people’s exposure is self-reported. Consider the following:
 - The minimum legal drinking age is well publicized
 - There can be significant disciplinary, social, and legal consequences for individuals who are caught drinking under the age of 21.
 - Data for this study were obtained from the National Health Interview Survey, which is a general longitudinal survey of health, and from the National Center for Health Statistic’s mortality data files, which contains a registry of all deaths in the U.S. from 1997 to 2004. The mortality data include the decedent’s date of birth and date of death.

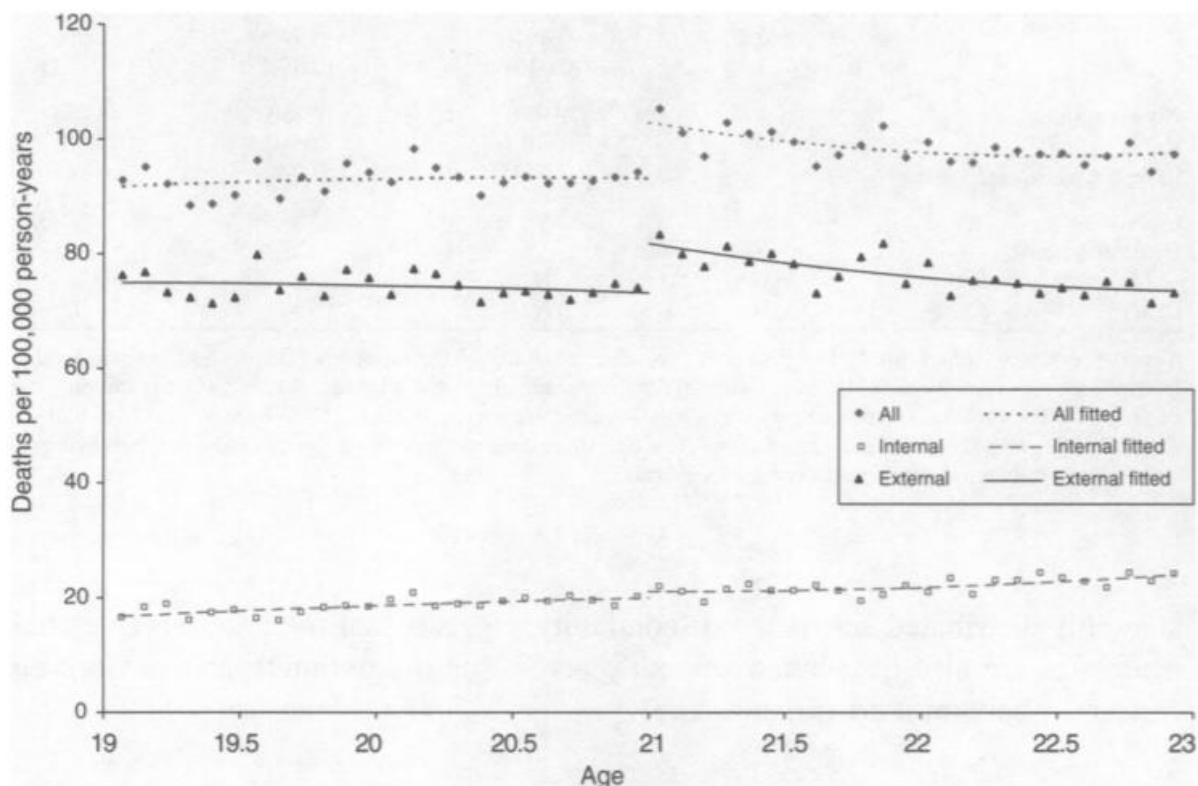
In two to three sentences, briefly discuss what manipulation of the forcing variable would mean in the context of this study and whether the facts above suggest it is likely to be present in this study (Note: be careful in distinguishing manipulation of the forcing variable from other problems for RD studies.)

While the threshold is well known and carries strong incentives against underage drinking, manipulation of the forcing variable (age) is unlikely given the source of the data. The NHIS collect

broad data on many health topics, so manipulation of age for the specific context of alcohol consumption would be unlikely. Furthermore, date of birth in vital registries like the NCHS mortality file are likely subject to verification and cross-referencing, and manipulation by officials is highly unlikely.

Note that it is possible individuals under 21 may underreport their alcohol consumption due to social desirability bias. This would introduce measurement error in our measurement of the exposure, but would not constitute manipulation of the forcing variable, which is age, not alcohol consumption.

The authors find that individuals just over age 21 report drinking on ~21 percent more days than individuals just under 21 years of age. They find a similar discontinuity in heavy drinking days. Satisfied that the alcohol consumption is discontinuous at the threshold, the researchers use the discontinuity to explore three outcomes of interest: all-cause mortality, mortality due to external causes (homicides, suicides, motor vehicle accidents, etc.), and mortality due to internal causes. They present the following graph, which shows lines fitted above and below the threshold for these three outcomes.



5. Based on the information in the prompt and in the graph, what conclusions do you draw from this graph and what might they imply regarding the impacts of alcohol on mortality? \

The graph shows that all-cause mortality is higher for individuals just above 21 compared to those just under 21. We fitted lines show that this increase is due to a jump in external causes of death, while mortality due to internal causes remains continuous at the threshold.

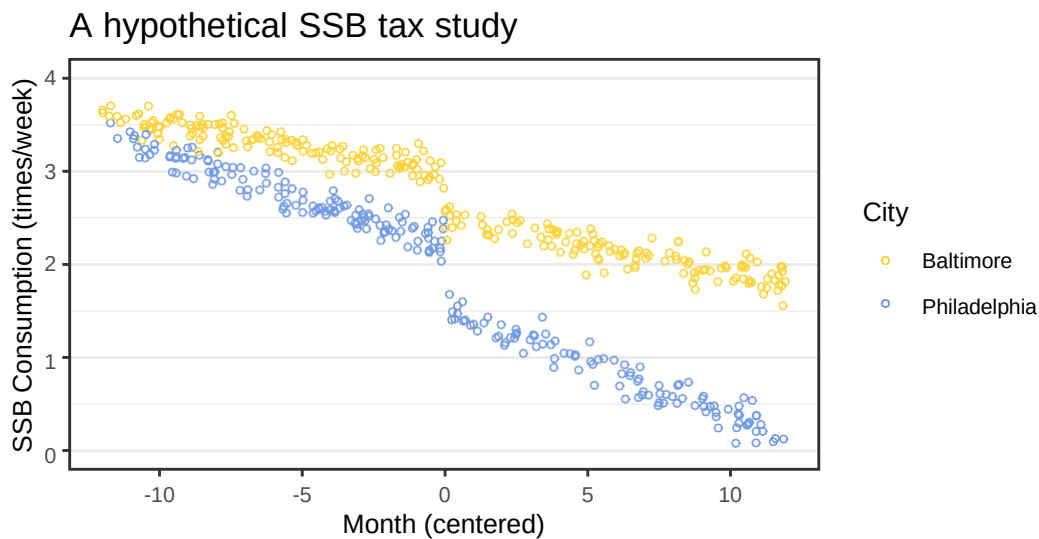
With the information from the prompt, it appears likely that the difference in mortality is due to alcohol consumption, as many of the external causes (including motor vehicle accidents) may be made more likely due to impairment from alcohol. The continuous trend for internal mortality at

the threshold—which we would expect to be unaffected by alcohol consumption, at least in the short term—also provides indirect evidence to support this conclusion.

3 Interrupted Time Series/Difference in Differences (12 points)

Sugar-sweetened beverages (SSBs) are the main source of added sugars in the American diet: in 2013-2014, 60.7% of children and 50.0% of adults consumed SSBs on a given day according to dietary recall data from the National Health and Nutrition Examination Survey (NHANES).¹ To reduce consumption, some cities in the US have enacted taxes on sugar-sweetened beverages to raise prices and discourage purchasing. You are a researcher at the Philadelphia Department of Public Health charged with **evaluating the effect of an SSB tax that was implemented on January 1, 2017**.

As part of your work, you've collected weekly consumption data (number of SSBs consumed per week) on 500 individuals - 250 from Philadelphia, PA (*tax, i.e. treated*) and 250 from Baltimore, MD (*no tax, i.e. control*) - for 12 months before and after the tax was implemented. **These data are plotted below (with a centered version of your month variable):**



¹See Bleich et al. (2017)

1. You plan on implementing an **controlled interrupted time series (CITS)** approach to estimate the effect of the Philadelphia SSB tax on SSB consumption frequency. You have variables for week of measurement ($week$), an indicator for whether this measurement was before or after 01/01/2017 ($post$), an indicator for whether an individual lived in Philadelphia (trt), and their weekly consumption frequency (ssb). Write out the CITS model you would fit to this data that includes terms for a tax-attributable level change and a tax-attributable slope change in SSB consumption frequency. (2 points)

CITS model is given by:

$$\begin{aligned} ssb_{it} = & \lambda_0 + \lambda_1(week_t) + \lambda_2(post_i) + \lambda_3(week_t \times post_i) \\ & + \alpha(trt_i) + \beta(week_t \times trt_i) \\ & + \delta(post_i \times trt_i) + \gamma(week_t \times post_i \times trt_i) + \varepsilon_{it} \end{aligned}$$

2. Write down the regression parameters from your CITS model in question 1 that correspond to the tax-attributable level change in SSB consumption frequency, and the tax-attributable slope change in SSB consumption frequency. (1 point) What two key assumptions are needed to identify these as causal effect? Describe them within the context of the study. (2 points)

δ and γ denote the causal effects of interest. Two assumptions needed to identify these from the data:

- Timing of SSB tax implementation is as good as random (all ITS assumption)
- Deviations from weekly SSB consumption trend in Philadelphia prior to Jan 1, 2017 would be the same as the observed deviation in Baltimore prior to Jan 1, 2017. (CITS assumption)

3. A data analyst runs your model in R, and they email you an **incomplete** snippet of the output (shown below). What are your estimate(s) of the causal effect you wrote down in question 2? Interpret within the context of the study. (3 points)

```
## Coefficients (**abbreviated!):**
##           Estimate Std. Error t value Pr(>|t|)
## ...           ...           ...      ...
## week:post    -0.006095   0.003686  -1.654   0.0989 .
## week:trt     -0.047896   0.003756 -12.750 < 2e-16 ***
## post:trt     -0.194403   0.037083  -5.242 2.36e-07 ***
## week:post:trt 0.000000   0.005446  -1.623  0.8999
## ...           ...           ...      ...
```

$\hat{\delta} = -0.194$, and $\hat{\gamma} = 0$. Therefore, an SSB tax implemented in Philadelphia resulted in a 0.194 reduction in the mean weekly SSB consumption frequency (times/week) at the time of enactment, above and beyond the change that occurred in Baltimore. However,

there was no change in the rate of SSB consumption frequency over time in Philadelphia before versus after the tax above and beyond the change in the rate of SSB consumption frequency in Baltimore over the same period.

4. A colleague suggests that people may consume SSBs more frequently during the summer months when it's warmer, and less frequently during the winter months. What challenge does this present to the analysis, and how could you address this problem? (2 points)

This describes seasonality of outcomes. Lack of adjustment may lead to biased estimates since the SSB tax was implemented in January, when consumption is probably lower. To address this problem in the analysis, could include season dummies or de-trend the time series before analyzing the data.

5. Yet another colleague suggests you implement a **difference-in-differences** approach, since this is what has been done in some research studies of SSB taxes previously. Describe what the **common shocks assumption** and **parallel pre-trends** mean in the context of the study. (2 points)

For Diff-in-Diff approach:

- Parallel pre-trends: trend in SSB consumption frequency is the same in Philadelphia and Baltimore prior to implementation of the tax (Jan 1, 2017)
- Common shocks assumption: Any event that may affect SSB consumption frequency in the post-tax period affects those in Philadelphia and Baltimore equally.

4 Missing Data and Meta-Analysis (14 points)

You want to assess the association between ambient noise pollution and sleep quality in Boston, Massachusetts. Luckily, you had unlimited funding, so you provided personal noise dosimeters (a device used to measure noise exposure) to all study participants. Participants were requested to carry their dosimeters with them for 24 hours prior to filling out a questionnaire. The questionnaire collected data about sleep quality, personal characteristics such as age, lifestyle factors, and other medical conditions. Two months later, participants repeated this process and filled out a second questionnaire.

1. In the context of this study, define unit non-response and item non-response. (2 points)

Item non-response: if an individual neglected to fill out a question about occupation but answered all other questions.

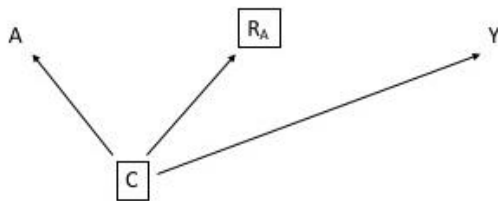
Unit non-response: if an individual did not fill out the second questionnaire in its entirety (no data collected for this individual at the second time point).

2. Suppose that age is a predictor of noise exposure and sleep quality. You also realize that older study participants were more likely to forget to carry around their dosimeter compared to younger study participants, and thus data on ambient noise exposure is more likely to be missing for older study participants. Assume that there are no other predictors of missing dosimeter data and that there are no other confounders of the exposure-outcome association.

- a. Identify the missingness mechanism for ambient noise exposure. (1 points)

The missingness mechanism is missing at random (MAR). This is because, conditional on age, the missingness in exposure is random.

- b. Based on the description, **draw a DAG** to represent a Complete Case Analysis (CCA) in this scenario. Use R_A to represent the missingness indicator for exposure. (1 point)



- c. Under what conditions would a complete case analysis yield an unbiased estimate of the effect of ambient noise on sleep quality? (2 points)

If we condition on age, then we would be able to obtain an unbiased estimate because all paths between A and Y would then be blocked.

3. Upon further examination, you see that about 20% of the data is missing. Your colleague wants to use CCA to address the missingness scenario presented in question 2. What is a reason that you might argue in favor of using multiple imputation instead? (1 point)

CCA leads to a reduction in sample size which in turn may compromise on the efficiency. MI addresses this issue because we are imputing missing values rather than omitting observations with missing data.

4. You decide to use multiple imputation to address the missingness scenario in question 2. You include a set of variables in the imputation model that you were not planning on including in your main analysis. Your colleague is nervous that failing to include this additional set of variables in your main analysis will lead to biased estimates. Is this the case? Why or why not? (2 points)

Assuming that our main analysis model is correctly specified, there would not be bias as a result of omitting these variables. Our multiple imputation model should be as rich as possible and can include variables that are not included in main analyses. This is because we are trying our best to predict the missing values and using all tools at our disposal.

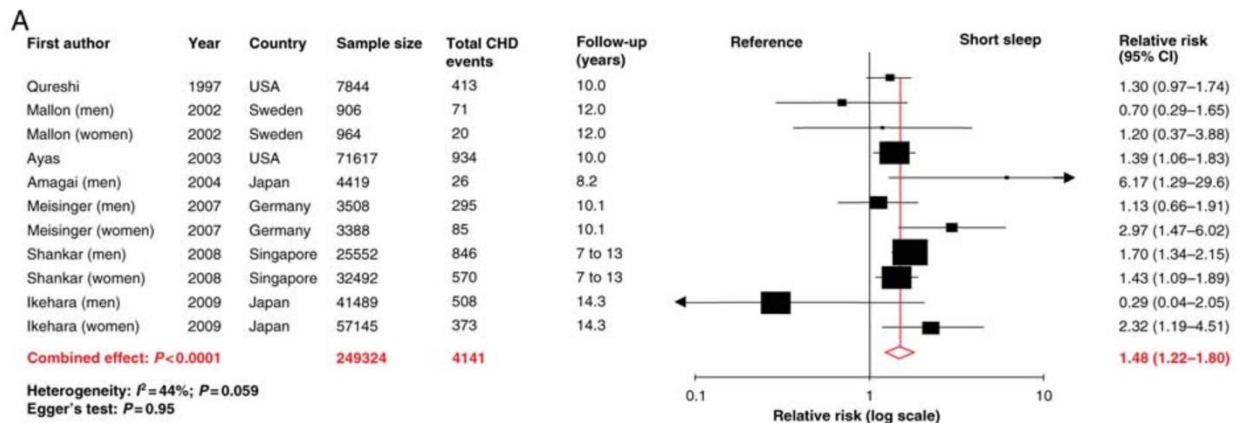
5. Suppose that occupation and number of hours worked are confounders of the association between noise exposure and sleep quality. You then discover that several study participants who work long hours (students, medical professionals, etc.) were less likely to respond to questions about occupation and number of hours worked. Identify the missingness mechanism at play in this scenario. (1 point)

The missingness mechanism for the occupation data is missing not at random (MNAR) because the missingness is dependent on the missing values themselves.

6. In the missingness scenario described in question 5, would using inverse probability weighting to address the missingness mechanism yield an unbiased estimate? Why or why not? (2 points)

No, because the missingness mechanism is MNAR. We would not be able to construct a sufficient model for the weights because data are missing for occupation and number of hours worked. Because the missingness is dependent on the missing values, we would get a biased estimate.

7. Your friends ask you why it is important to study factors that may influence sleep quality. In response, you show them a meta-analysis from 2011 that assessed the association between short duration of sleep and risk of coronary heart disease (CHD). Below is the forest plot of the risk of developing or dying of CHD associated with short duration of sleep compared with the reference group.



- a. Interpret the pooled relative risk value of 1.48. (1 point)

Shorter sleep duration is associated **on average across studies** with increased risk of developing/dying of CHD by $RR = 1.48$ (95% CI 1.22-1.80).

- b. In this meta-analysis, authors reported an I^2 statistic. What does the I^2 represent? What kind of meta-analysis is this? (2 points)

The I^2 statistic describes the proportion of the variability in estimates that is not due to sampling error. We can use the I^2 statistic to assess whether or not there is evidence of heterogeneity between studies. The presentation of an I^2 statistic indicates that this is a random effects meta-analysis.

- c. The I^2 statistic is 44% with a p-value of 0.059. Interpret the magnitude and significance of the I^2 in the context of this meta-analysis. (2 points)

We can see that about 44% of the variability in estimates is not due to sampling error. With a p-value of 0.059, at an alpha level of 0.05, we (barely) have insufficient evidence to reject the null hypothesis. Based on this hypothesis test alone, we may conclude that there is some evidence of heterogeneity because the p-value is so borderline. However, it should be noted that I^2 alone should not be used to conclude whether or not there is heterogeneity between studies.