Exam practice

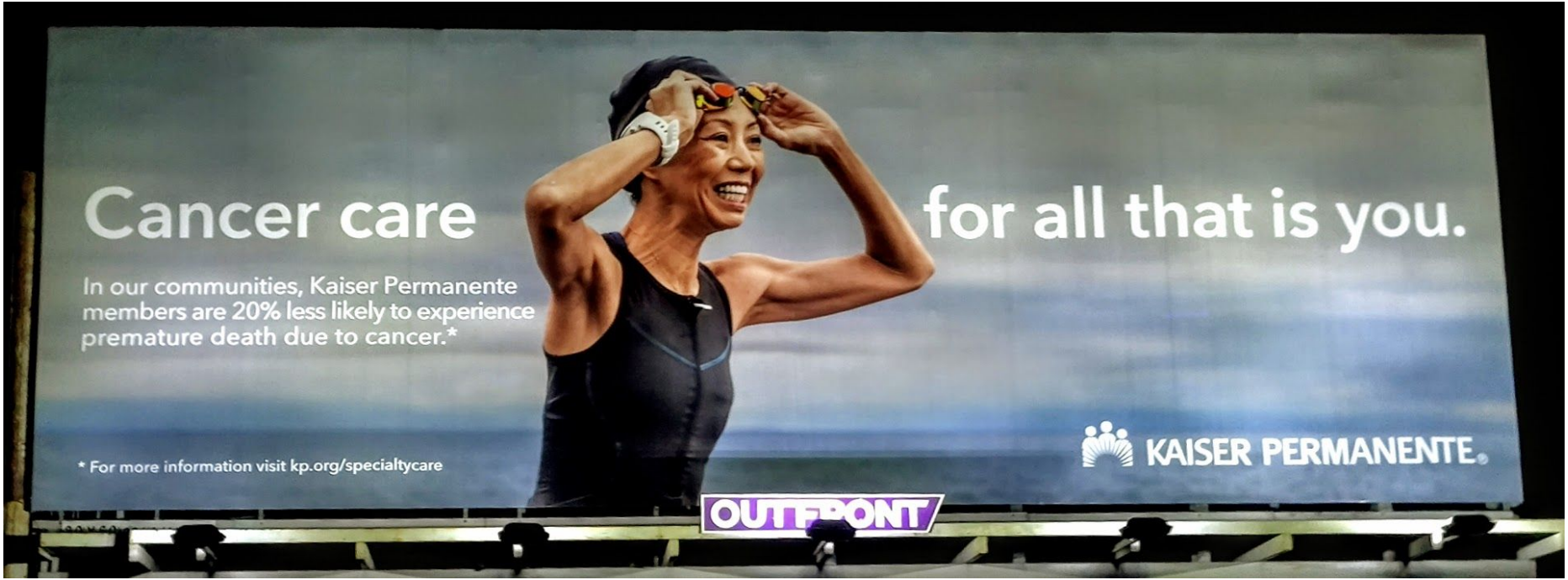
https://pollev.com/jonathanold608

Today is all about Midterm exam practice

- Your questions
- A very detailed mock exam question that talks through:
  - The fundamental problem of causal identification
  - Regression table interpretation
  - Omitted variable bias
  - The potential of RCTs

Mock exam question

Mock exam



Seen this week on the I-80!

Question 1.a: Conceptual

You see the following billboard on
way back from San Francisco. Based on the billboard, a friend of you advises you to switch to Kaiser Permanente insurance, as this will decrease your risk of premature death due to cancer by 20%.

- **Why is your friend likely wrong?**
- **What would have to hold for your friend's statement to be correct?**
- Bonus question: Is there anything else that's weird about the comparison on the billboard?

Question 1a: Answer

- My friend is wrong because a simple comparison of means does not deliver a causal estimate of the impact of the insurance. **People who are members of KP are likely fundamentally different from people that are not, in ways that are correlated with their risk to die of cancer**. They may, for example, be richer, younger, healthier, or more health-conscious. Thus, the two groups are not valid counterfactuals for each other.
- For my friend's statement two hold, KP-members and non-KP-members would have to be valid counterfactuals for each other. **That is, KP-members' expected cancer death risk without KP insurance would have to be the same as that of non-KP-members.** We call this the 'identifying assumption' of your friend's research design.
- Other weird things: The billboard does not specify the comparison group (all people? All people with insurance?). We also don't know whether this is a conditional or unconditional probability.

Question 1b: Regression tables

- You were able to obtain the data that KP used for their analysis, and you run a regression of whether an individual died from cancer, on a dummy variable whether they were members of KP. Interpret the regression output below.
- Test whether the coefficient on KP member is significant.

Outcome: Died from cancer	Model 1
(Intercept)	**0.020** (0.0010)
KP member	**-0.004** (0.0009)
N	10,324

Robust standard errors in parentheses.

Question 1b: Answer

Outcome: Died from cancer	Model 1
(Intercept)	**0.020** (0.0010)
KP member	**-0.004** (0.0009)
N	10,324

- The intercept indicates that, in this sample, 2% of individuals that were not members of KP died from cancer. The intercept is significantly different from zero, as |0.02/0.001| > 2
- The slope coefficient on KP member indicates that individuals that were members of KP were 0.4 percentage points less likely to die from cancer. As -0.004/0.02=-0.2, this means they were 20% less likely to die from cancer.
- The t-statistic for the slope coefficient is -0.004/0.0009=-4.44. This is less than -2, so the difference is statistically significant from zero.

Question 1c: Omitted variable bias

Another friend suggests ‘controlling for income’.

- Explain intuitively what it means to control for a variable in a regression
- Using the omitted variable bias formula, explain what you would expect to happen to the coefficient on KP member.

Outcome: Died from cancer	Model 1	Model 2
(Intercept)	**0.020** (0.0010)	**0.030** (0.0011)
KP member	**-0.002** (0.0009)	
Annual income (in 10,000)		
N	10,324	8,925

Robust standard errors in parentheses.

Question 1c: Answer

Died from cancer	Model 1	Model 2
(Intercept)	**0.020** (0.0010)	**0.030** (0.0011)
KP member	**-0.002** (0.0009)	
Annual income (10,000)		
N	10,324	8,925

- Controlling for a variable removes omitted variable bias caused by the omission of that variable. Controlling for a variable is the econometric equivalent of “keeping a variable constant”. Mathematically, including a control variable removes systematic differences in both the other independent variables and the dependent variable that are due to the included control. This allows us to compare “apples to apples” by looking at the correlation of the independent and dependent variable **for given (fixed) values of the control variable**.
- We expect annual income to be positively correlated to being a KP member. We also expect annual income to be negatively related to the risk of death from cancer. Therefore, we expect the coefficient on KP member in the short regression to be downward biased. If we include annual income, the coefficient will be larger.

Question 1d: Omitted variable bias II

To confirm your suspicions about the OVB, you run an auxiliary regression of Annual income on the KP member dummy. You get a coefficient of 0.7. You then run the full (long) regression, but your computer has a bug. Can you calculate the coefficient on KP member?

Outcome: Died from cancer	Model 1	Model 2
(Intercept)	**0.020** (0.0010)	**0.030** (0.0011)
KP member	**-0.002** (0.0009)	**XXXXXXXXXXX**
Annual income (in 10,000)		**-0.003** (0.001)
N	10,324	8,925

Robust standard errors in parentheses.

Question 1d: Answer

To calculate this, use the OVB formula:

$$\beta_S = \beta + \pi \times \delta \rightarrow \beta = \beta_S - \pi \times \delta$$

In our case β_S=-0.02, π=0.7 and δ=-0.003.

Therefore, β=0.0001.

Outcome: Died from cancer	Model 1	Model 2
(Intercept)	**0.020** (0.0010)	**0.030** (0.0011)
KP member	**-0.002** (0.0009)	**0.0001**
Annual income (in 10,000)		**-0.003** (0.001)
N	10,324	8,925

Robust standard errors in parentheses.

Question 1e: Regression tables

- Interpret the output of the table for model 2
- Looking at the output, your friend claims: The coefficient on KP member is close to zero. Therefore, being a member of the insurance does have no effect on cancer risk. Give **three reasons** why your friend may be wrong.

Outcome: Died from cancer	Model 1	Model 2
(Intercept)	**0.020** (0.0010)	**0.030** (0.0011)
KP member	**-0.002** (0.0009)	**0.0001** (0.050)
Annual income (in 10,000)		**-0.003** (0.001)
N	10,324	8,925

Robust standard errors in parentheses.

Question 1e: Answer

Died from cancer	Model 2
(Intercept)	**0.030** (0.0011)
KP member	**-0.0001** (0.050)
Income (10,000)	**-0.003** (0.001)
N	8,925

- Intercept: For a hypothetical individual with 0 income that is not a member of KP, we expect their risk of dying from cancer to be 3%.
- Coefficient on KP member: For any given level of income, being a KP member is associated with a 0.01 percentage point lower risk of dying from cancer. This is not significantly different from zero.
- Coefficient on income: Holding KP membership constant, each additional $10,000 income is associated with a 0.3 percentage point lower risk of dying from cancer.
- Reasons why friend is wrong:
  - Coefficient is insignificant, but not zero (although it is very small)
  - Hypothesis testing: We fail to reject the null, but we do not accept it!
  - Controlling for one variable does not fully remove OVB and our estimate is still not causal.

Question 1f: RCT

You managed to get a summer internship with Kaiser Permanente, who are sympathetic to your concerns. They have a lot of money and suggest running an RCT in order to find out what their impact on the risk of dying from cancer actually is.

- How would an RCT solve the issues discussed before?
- **Bonus question**: We cannot simply randomly give insurances to people (they have to sign up, etc.). Therefore, your manager asks you to design an RCT that can help you to identify the causal effect of being a KP member. Give at least one idea how to design such an RCT.

Question 1f: Answer

- RCTs help us solve the fundamental problem of causal identification via random assignment of the treatment. Recall that our problem was that the two groups did not have the same expected outcome in the absence of a treatment. With an RCT, the two groups are, in expectation, not different on average, and so the mean difference or a regression just gives us the causal effect. Note: We need a large enough sample for this to work!
- This is a creative question with no right or wrong. The key is to look for a feasible randomized treatment that affects the probability of joining KP, but not other things. Some ideas:
    - Giving people without any insurance randomly vouchers to join KP for free or at a discount
    - Give people who just sign up for insurance (e.g. incoming international students) an incentive (such as a default option)
    - Randomly give people a cash incentive to change insurance - away from KP towards other insurances
    - Note: These three experiments all would give very different estimates and interpretations!