Proctored Exam 1

● Graded

Student

Valerie Stacey

Total Points

31.5 / 35.5 pts

Question 1

Honor code

0 / 0 pts

+ 0 pts Correct

✔ **+ 0 pts** Incorrect

+ 0 pts Answer key:
https://docs.google.com/document/d/1RCoa7qyU9oRHt7T2aDR9NENgjQVaJ4KV/edit?usp=sharing&ouid=112531867312124573521&rtpof=true&sd=true

Question 2

(no title)

1 / 1 pt

✔ **+ 1 pt** Something happened in this time period (i.e., the COVID-19 lockdown) where preterm birth decreased in pregnancies globally. This is not an example of cohort effects or age effects because the decrease occurred irrespective of age or year of birth.

+ 1 pt Correct

+ 1 pt Credit provided for *period and cohort* effect. It is possible that these preterm infants are a birth cohort and experiencing a different rate of preterm relative to other birth cohorts.

+ 0 pts Incorrect

+ 1 pt Credit provided for cohort effect.

Question 3

(no title) **13.5** / 15.5 pts

3.1 **(no title)** **2** / 2 pts

✔ **+ 1 pt** Correct mathematical expression. Accepted notation include E[Y1-Y0], P(Y1-Y0), E[Y1] /= E[Y0], P(Y1) /= P(Y0)

✔ **+ 1 pt** Correct interpretation: This equation represents the counterfactual (i.e. causal) difference in preterm birth if the whole population of pregnant people were exposed to stress during pregnancy compared to if the whole population of pregnant people were not exposed to stress during pregnancy.

+ 0.5 pts Partial credit: Reference of "population-level" in interpretation

+ 0.5 pts Partial credit: Reference of "counterfactual" in interpretation

+ 0 pts Incorrect equation

+ 0 pts Incorrect interpretation

+ 0.5 pts Correct equation, missed one value

3.2 **(no title)** **1** / 1 pt

✔ **+ 1 pt** The time scale is calendar time and new pregnancies can be added to the study cohort over time.

+ 0 pts Incorrect

+ 1 pt Correct

3.3 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Including pregnancies in 2020 violates the assumption that there are no secular trends.

+ 0 pts Incorrect

+ 1 pt Correct

3.4 **(no title)** **0** / 1 pt

+ 1 pt The effect of stress would appear stronger in your study population because more people likely have bacteria in the uterus, resulting in more people completing the sufficient cause. The strength of an effect for one variable will depend on how common the rest of the pie is in that population (how often it gets completed).

✔ **+ 0 pts** Incorrect

+ 1 pt Correct

3.5 **(no title)** **2** / 2 pts

✔ **+ 2 pts** Intrauterine infection and preterm birth are downstream of a weakened immune system.

+ 0 pts Incorrect. No partial credit is provided for only identifying one of the descendants.

+ 2 pts Correct

+ 0 pts Incorrect

3.6 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Structural causal models outline the quantitative relationships between variables.

+ 0 pts Incorrect

+ 1 pt Correct

3.7 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Addresses assumptions of simple cumulative method (e.g. closed population, competing risks, withdrawals, follow-up time).

+ 0 pts Incorrect reason

3.8 **(no title)** **Resolved** **2.5** / 2.5 pts

✔ **+ 2.5 pts** All values are correct. See answer key.

+ 0.25 pts A = 3

+ 2.5 pts Correct

✔ **+ 0 pts** Incorrect

+ 0.25 pts B = 4

+ 0.25 pts C = 4. In this scenario, it is NOT 5 because there is a withdrawal at time = 2. According to the slides, this withdrawal would count towards time = 2, but we wouldn't count it towards time = 3.

+ 0.25 pts D = 3

+ 0.25 pts E = 1

+ 0.25 pts F = 1

+ 0.25 pts G = 0.250

+ 0.25 pts H = 0.333

+ 0.25 pts I = 0.750

+ 0.25 pts J = 0.667

Regrade Request **Submitted on: Oct 03**

I looked through the answer key, and all my answers appear to match up, but I received 0 points out of 2.5 for this question.

must have been an error on the the auto-grader, I'll fix that now

Reviewed on: Oct 09

3.9 **(no title)** **2** / 2 pts

✔ **+ 1 pt** CI(t0, tj) = 1 - (0.750*0.667) = 1- 0.500 = 0.500

✔ **+ 1 pt** Full credit was provided to students who had the correct formula but used incorrect values carried over from the previous question.

+ 1 pt Correct interpretation: The probability of preterm birth over the study period was 0.500.

+ 0 pts Incorrect interpretation or interpretation did not specify the time period for the cumulative incidence.

+ 0 pts Incorrect calculation

+ 0.75 pts Rounding error

+ 0.5 pts Interpretation did not include value

+ 0.5 pts Time period not t=6

3.10 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

3.11 **(no title)** **0** / 1 pt

+ 1 pt The density method assumes that disease is rare (and the prevalence of preterm birth is greater than 0.1 in our sample).

+ 0.5 pts Partial credit: Density method requires rare disease

+ 0.5 pts Partial credit: Identifying that disease (preterm birth) was not rare

✔ **+ 0 pts** Incorrect

The density method assumes that disease is rare (and the prevalence of preterm birth is greater than 0.1 in our sample).

Question 4

(no title) **2** / 2 pts

✔ **+ 2 pts** Correct

+ 0 pts Incorrect

+ 1 pt One correct out of two

Question 5

(no title) **5.25** / 7 pts

5.1 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct. OR = (70/130) / (5/20) = (20 x 70) / (130 x 5) = 2.15

+ 0.75 pts Rounding error

+ 0.5 pts Minor mathematical error

+ 0 pts Incorrect, see above.

5.2 **(no title)** **0.75** / 1 pt

+ 1 pt Correct. Adults living in neighborhoods with low to moderate levels of greenness had 2.15 times the odds (not risk) of depressive symptoms compared to adults living in neighborhoods with high levels of greenness, in this study population, over the study period.

+ 0.75 pts Interpretations correct except used "higher" or "greater" in describing odds

+ 0.75 pts Used risk instead of odds (or didn't specify odds)

+ 0.5 pts Partial credit

+ 0 pts Incorrect, see above.

✔ **+ 0.75 pts** Did not specify "in this study population over the study period"

5.3 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

5.4 **(no title)** **1.5** / 2 pts

✔ **+ 1 pt** Correct calculation:
CI low-mod - CI high = 0.55-0.30 = 0.25

+ 1 pt Correct interpretation. The risk of depressive symptoms among adults living in neighborhoods with low to moderate levels of greenness was 25 percentage points higher compared to adults living in neighborhoods with high levels of greenness, in this study population, over the five year study period.

+ 0 pts Incorrect, see above.

+ 0.75 pts Interpretation is missing "in this study population, over the study period"

+ 0.75 pts Partial credit for interpretation, did not state both exposure groups

+ 1 pt Correct interpretation but wrong number

✔ **+ 0.5 pts** Used "times" instead of increased/higher

Make sure to add "in study population over study period"

5.5 **(no title)** **0** / 1 pt

+ 1 pt Correct

✔ **+ 0 pts** Incorrect

5.6 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

Question 6

(no title) **2** / 2 pts

✔ **+ 1 pt** Describes how with perfect randomization we could assume that the population in the control arm and the treatment arm are exchangeable, meaning we would expect the same result regardless of treatment/placebo assignment.

✔ **+ 1 pt** Describes how with perfect randomization we can simulate the counterfactual by comparing the populations with and without the treatment as if they were the same population that first received the treatment, then went back in time and did not receive the treatment and were compared.

+ 0.5 pts Partial credit

+ 0 pts Incorrect, see above.

+ 0.5 pts Did not go into detail about the counterfactual aspect

Question 7

(no title) **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

Question 8

(no title) **3.75** / 4 pts

8.1 **(no title)** **0.75** / 1 pt

+ 1 pt Correct. The individual-level counterfactual outcome would be the person's hypertension status if they were not exposed to coffee consumption while holding all other factors constant.

+ 0.5 pts Stated population-level instead of individual-level counterfactual

+ 0.5 pts Stated counterfactual but also stated an outcome value (had hypertension or didn't). This is incorrect because we don't know whether they would've developed hypertension. The counterfactual outcome is their outcome **status** had they not drank coffee, though we don't know what the status actually is because we can't observe this scenario

+ 0.5 pts Stated counterfactual but did not specify hypertension status had the individual not consumed coffee.

+ 0 pts Incorrect, see above.

✔ **+ 0.75 pts** Did not specify holding all other factors constant

8.2 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

8.3 **(no title)** **2** / 2 pts

✔ **+ 1 pt** Correct Calculation.
E[Y1-Y0]
= [-1 + 1 + 0 + 1 + -1 + 0 + 1 + 0 + 0 + 1] / 10
= 0.2

✔ **+ 1 pt** Correct Interpretation. Since we can assume a causal interpretation, 0.2 of the hypertension risk in this population is due to coffee consumption.

+ 0 pts Incorrect, see above.

+ 0.5 pts Correct number but wrong direction (negative vs positive)

+ 0.75 pts Rounding or decimal error

+ 1 pt Correct interpretation but wrong value

+ 0.75 pts Did not refer to population, everyone, causal, or counterfactual

0.2 or 20%
Be sure to include the value in your interpretation

Question 9

(no title) **3** / 3 pts

9.1 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

9.2 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

9.3 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

Q1 Honor code
0 Points

For calculation-based questions, please include your calculations in the free-text boxes that appear after the answer box for full credit.

Please round answers to three decimal places, unless otherwise stated.

This exam is closed-book. The materials you are permitted to use consist of your study guide, the PHW250 formula sheet that has been provided, a calculator (not a cell phone calculator), and scratch paper/writing utensils.

As a member of the OOMPH community at UC Berkeley, I act with honesty and integrity to uphold the high academic standards in the code of student conduct. On my honor, I confirm that:

- I have neither given nor received any unauthorized aid on this exam.
- I have not plagiarized any original works.
- I have not collaborated with any other students during the exam.
- I have not made use of any inappropriate online sources for collaboration during the exam, such as direct messaging sites/apps, social media, Zoom, emails -- except to contact the teaching team when necessary.
- I will not speak to other students about exam questions until the end of the exam period.

Please type in your name as acknowledgment:

Val Stacey

Q2
1 Point

There was a reported decrease in the rate of preterm birth during the first three months of the COVID-19 pandemic lockdown (March, April, and May 2020) across multiple countries. This is a most likely an example of:

- (●) Period effects
- ( ) Cohort effects
- ( ) Age effects
- ( ) Period and cohort effects

Q3
15.5 Points

You are interested in studying the relationship between stress during pregnancy and preterm birth. In an ideal scenario, you would like to conduct a counterfactual experiment to estimate the causal effect of stress during pregnancy on preterm birth.

Q3.1
2 Points

Write a mathematical expression to represent the population-level causal effect of stress during pregnancy on preterm birth. Interpret this expression in one sentence max. (1 point for mathematical expression; 1 point for interpretation)

E[Y1_stressed - Y0_no_stress]

The expression represents the difference in preterm births if the whole population of pregnant people were stressed compared to (if we could wind back the clock keeping all else equal) the preterm births if the whole population was not stressed.

Q3.2
1 Point

You are unable to conduct a counterfactual experiment, so you and your colleagues design a study that looks at pregnancies in a large health care system from 2010 through 2019.

True or False: This is an example of an open population.

- (●) True
- ( ) False

Q3.3
1 Point

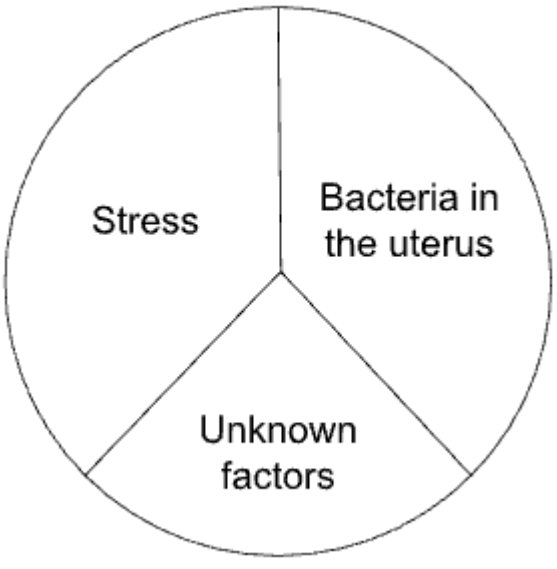
When designing your study, you decided to exclude pregnancies in 2020 because of the change in rate of preterm birth during the COVID-19 lockdown.

What assumption for the calculation of incidence density would the inclusion of the 2020 pregnancies violate?

- ○ Independence of censoring and survival
- ◉ No secular trends
- ○ Constant risk over interval period
- ○ Closed cohort

Q3.4
1 Point

Based on prior literature, you outline one potential mechanism that links stress to preterm birth. All component causes occur during pregnancy.



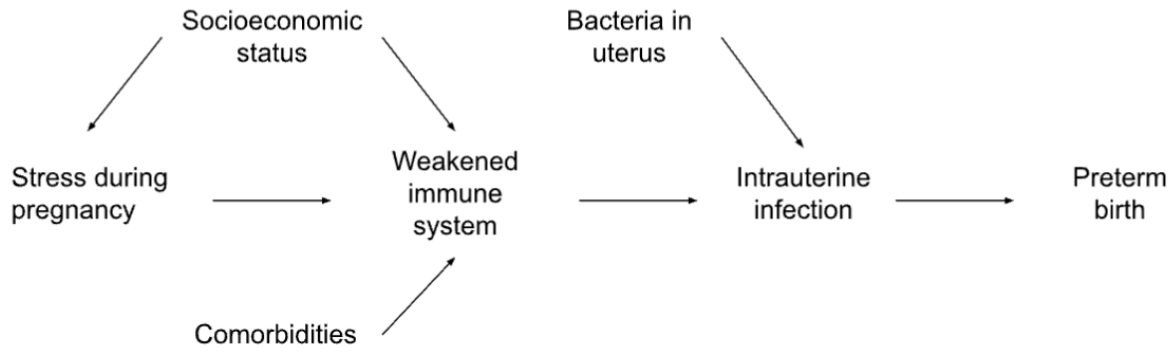
You look at the medical histories of your study population and notice that they have a higher prevalence of bacterial vaginosis compared to the general population of pregnant people. Bacterial vaginosis increases the likelihood of bacteria in the uterus.

If you only consider the causal mechanism above, you expect that the effect of stress in your study population may appear:

- ○ Stronger compared to the general population of pregnant people
- ◉ Weaker compared to the general population of pregnant people
- ○ The same as the general population of pregnant people

Q3.5
2 Points

You use a directed acyclic graph to further develop your conceptual model.



Which of the following variables are descendants of a weakened immune system? Select all that apply.

- [ ] Stress during pregnancy
- [x] Intrauterine infection
- [x] Preterm birth
- [ ] Bacteria in uterus
- [ ] Comorbidities

Q3.6
1 Point

True or False: Directed acyclic graphs are better at encoding quantitative information than structural causal models.

- ○ True
- ◉ False

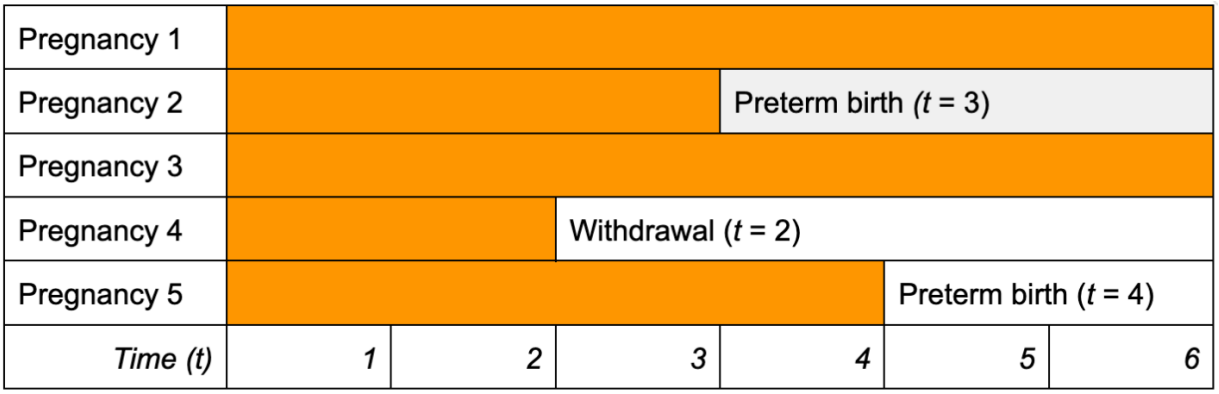
Q3.7
1 Point

One colleague proposes that you calculate cumulative incidence using the simple cumulative method. List one reason why you disagree with that decision (1 sentence max):

This method is best for shorter time frames, and not appropriate to use when there are withdrawals.

Q3.8
2.5 Points

You decide to go with the Kaplan-Meier approach because you have access to the date and time of preterm birth in your cohort. Data from a subset of your cohort is shown below.



Complete the table using the data provided.

Time j	**Population at risk** N'_j	**Incident disease** I_j	**Interval risk** R_j	**Interval survival** S_j
A	C	E	G	I
B	D	F	H	J

A:

3

B:

4

C:

4

D:

3

E:

1

F:

1

G:

.250

H:

.333

I:

0.750

J:

0.667

Q3.9
2 Points

Calculate the cumulative incidence using the table in Part 3.8. Interpret your result in one sentence. (1 point for estimate; 1 point for interpretation)

1 - (0.750 X 0.667) = 0.500
The risk of having a pre-term birth among pregnant people in the study over the study's time period was near 0.50, or 50%.

Q3.10
1 Point

Another colleague on your research team calculated cumulative incidence using the actuarial method using the same data. You found that their estimate was the same as your estimate! Why?

- ○ There were no withdrawals.
- ○ Withdrawals occurred in the same interval as the outcome.
- ◉ Withdrawals did not occur in the same interval as the outcome.
- ○ The actuarial method does not account for withdrawals.
- ○ None of the above.

Q3.11
1 Point

Another colleague on your research team calculated cumulative incidence using the density method using the same data. You found that their estimate was different from your estimate. In one sentence explain why you believe that their estimate was different from yours.

The density method accounts for the person who withdrew from the study (and importantly here - it did not occur in the same period as the preterm birth), so the estimate will be different.

Q4
2 Points

Which of the following scenarios can the odds ratio likely be substituted for relative risk? Select all that apply.

- [x] A case-control study examining a genetic disease that has a global prevalence of 0.2%.
- [ ] A cross-sectional study examining a chronic condition that has a US prevalence of 11%.
- [x] A case-cohort study that samples controls from the study population at baseline.
- [ ] All of the above
- [ ] None of the above

Q5
7 Points

Researchers conducted a cross-sectional survey of 225 adults to examine if living close to green space (e.g., forests and parks) was associated with depressive symptoms. Of the 75 adults who lived in neighborhoods with high levels of greenness, 5 reported depressive symptoms. Of the 150 adults who lived with low to moderate levels of greenness, 20 reported depressive symptoms.

	Depressive symptoms	No depressive symptoms
Low to moderate levels of greenness	20	130
High levels of greenness	5	70
Total	25	200

Q5.1
1 Point

Calculate the odds ratio for those living in neighborhoods with low to moderate levels of greenness compared to those living in neighborhoods with high levels of greenness. Round to two decimal places.

OR = (20*70) / (130*5) = 1400/650 = 2.15

Q5.2
1 Point

Interpret the odds ratio.

The odds of depressive symptoms in people living in neighborhoods with low to moderate levels of greenness is 2.15 times the odds of depressive symptoms in people living in neighborhoods with high levels of greenness.

Q5.3
1 Point

In this cross-sectional survey, the exposure was measured at the same time as the outcome. Which Bradford Hill criteria does this violate?

- ○ Consistency
- ○ Specificity
- ◉ Temporality
- ○ Biological gradient

Q5.4
2 Points

Another research group was interested in your research question and decided to conduct a follow-up study. They recruited a cohort of adults and followed them for five years. Over the follow-up period, they estimated that the cumulative incidence of depressive symptoms among adults living in neighborhoods with high levels of greenness was 0.30 and the cumulative incidence of depressive symptoms among adults living in neighborhoods with low to moderate levels of greenness was 0.55.

Calculate and interpret the appropriate absolute measure of association using the information from the new study. Compare low to moderate levels of greenness to high levels of greenness (1 point for calculation; 1 point for interpretation).

Cumulative incidence difference = 0.55 - 0.30 = 0.25

The cumulative incidence of depressive symptoms in low to moderate levels of greenness neighborhoods is 0.25 times the cumulative incidence of depressive symptoms in high levels of greenness neighborhoods.

Q5.5
1 Point

True or False: The absolute measure of association is often considered a better measure of causal impact because it excludes causal subtypes of doomed and immune.

- ○ True
- ◉ False

Q5.6
1 Point

The city planning department is interested in how many cases of depressive symptoms could be avoided among people living in neighborhoods with low to moderate levels of greenness if they built more parks and types of green spaces. Which type of measure would you estimate to inform the city planning department?

- ○ Risk difference
- ○ Relative risk
- ◉ Attributable proportion among the exposed (APe%)
- ○ Attributable proportion among the total population (APt%)

Q6
2 Points

Imagine you are conducting a randomized control trial (RCT) investigating a novel COVID-19 treatment with perfect randomization between the control and treatment arms. Briefly explain (2 sentences max) if you could make causal claims in this study using the concepts of exchangeability and counterfactuals.

Yes, we could make causal claims when randomization was "perfect," because what that means is that those receiving treatment and those receiving the placebo are "exchangeable", i.e. all confounding variables end up balancing out across the two groups. This set up replicates a counterfactual scenario quite well, as if we could swap treatment and placebos and wind back the clock, and would expect no difference in the occurrence of covid between them.

Q7
1 Point

Imagine you are working at a state health department that is interested in comparing lung cancer mortality between your state and the neighboring state. You have the number of lung cancer deaths (counts) in each state and the lung cancer mortality rate for the entire US population, stratified by age. Because you know age is a confounding factor for lung cancer, you decide to standardize the two state populations by age and calculate the standardized mortality ratio (SMR) for each population. Given this information, which standardization method is most appropriate and how could you interpret/compare the findings?

- ○ Direct standardization, could compare the SMRs from each state directly to each other
- ○ Direct standardization, could only compare the SMRs from each state to the reference population (i.e. US rates)
- ○ Indirect standardization, could compare the SMRs from each state directly to each other
- ◉ Indirect standardization, could only compare the SMRs from each state to the reference population (i.e. US rates)

Q8
4 Points

You are interested in exploring the relationship between coffee consumption (exposure) and hypertension (outcome). Many coffee drinkers have hypertension, but this may be due to the fact that those who consume coffee have other risk factors associated with hypertension (stress, diet, tobacco/alcohol consumption etc.).

Q8.1
1 Point

For this research question, describe in one sentence the individual-level counterfactual outcome for a person who drinks coffee and has hypertension.

The counterfactual outcome for this person would be if that same person did not drink coffee and either still had hypertension or did not have hypertension.

Given: Person drinks coffee -- has hypertension

Possible hypothetical counterfactual outcomes to think about:

Person does not drink coffee -- still has hypertension
Person does not drink coffee -- does not have hypertension

Q8.2
1 Point

The counterfactual information for each individual in the study is given below, where a is the exposure (coffee consumption) and Y is the outcome of interest (hypertension):

Participant	a	Y_0	a	Y_1	Individual-level causal effect
1	0	1	1	0	
2	0	0	1	1	
3	0	0	1	0	
4	0	0	1	1	
5	0	1	1	0	
6	0	1	1	1	
7	0	0	1	1	
8	0	0	1	0	
9	0	1	1	1	
10	0	0	1	1	

What is the individual-level causal effect for participant #5 comparing coffee consumption (a = 1) to no coffee consumption (a = 0) ?

- (●) -1
- ( ) 0
- ( ) 1
- ( ) 0.4

Q8.3
2 Points

What is the population-level causal effect on coffee consumption and hypertension? Using the data in the table from question 7.2, calculate the effect and provide an interpretation. (1 point for effect; 1 point for interpretation).

Drinks coffee and has hypertension: 6/10
Does not drink coffee and has hypertension: 4/10

Effect: + 0.20
In this hypothetical counterfactual study where we could take the same people and only change exposure to coffee, we can say that, overall, drinking coffee increases hypertension in the population.

Q9
3 Points

Please indicate whether the following statements are true or false:

Q9.1
1 Point

According to the sufficient component-causal model, blocking the action of a necessary cause will prevent all cases of a disease by all of its causal mechanisms.

- (●) True
- ( ) False

Q9.2
1 Point

It is possible to calculate real life individual-level causal effects, which are considered the gold standard for epidemiologic research

- ( ) True
- (●) False

Q9.3
1 Point

Odds ratios are the only appropriate measure of association in case-control study designs

- ( ) True
- (●) False