**Question 1**

Which of the following is a causal question?

1. How many ventilators are there in Texas?
2. What is the average age of people living in Florida?
3. Can we prevent lung cancer if we prevent smoking?
4. Who is most likely to get heart disease?

A and B are descriptive questions. D is a predictive question. It does ask why people get heart disease or if we can change their risk of heart disease. It simply asks who we can expect to see popping up on the radar with heart disease. Only C implies a causal question -- Does smoking cause lung disease?

**Question 2**

What is a statistical association (The Hernan videos gave two informal definitions)?

The distribution of one variable differs, on average, within levels of another variable. Or, knowing something about one variable gives us information about another variable.

**Question 3**

Why do things happen?

The answers that are most useful to us typically include some form of thing 2 (effect) happened because thing 1 (cause) happened. Had thing 1 not happened, thing 2 would not have happened (counterfactual).

**Question 4**

Based on Rothman and Greenland (2005), Hill’s criteria are a sufficient method for inferring causation.

1. True
2. False

They are useful for historical reference, but there are many examples of non-causal associations that meet the criteria, and many examples of causal associations that don't meet the criteria.

**Question 5**

How can we say that smoking causes cancer if everyone who smokes doesn't get cancer?

Think about Rothman's sufficient component cause model. Smoking is neither necessary nor sufficient. Think about causes in individuals vs. causes in populations.

**Question 6**

In terms of disease prevention, it is necessary to understand causal mechanisms between risk factors and outcomes in their entirety.

1. True
2. False

Knocking out even one component cause prevents all sufficient causes with that component cause.

**Question 7**

Factor A, B, C, D can individually cause a certain disease without the other three factors but only when followed by exposure to factor X. Exposure to factor X alone is not followed by the disease, but the disease never occurs in the absence of exposure to factor X.

According to the causal model described above, factor C is:

1. A necessary and sufficient cause
2. A necessary but not sufficient cause
3. A sufficient but not necessary cause
4. Neither necessary nor sufficient
5. None of these

If factor C were necessary, then it would be part of every set of sufficient causes. But, A|X, B|X, and D|X are all sufficient cause sets that do not include C. Likewise, if factor C were sufficient, then factor X wouldn't be necessary. However, the passage above says that factor C can cause the disease "only when followed by exposure to factor X." Therefore, factor C, by itself, is neither necessary nor sufficient to cause the disease.

**Question 8**

Factor A, B, C, D can individually cause a certain disease without the other three factors but only when followed by exposure to factor X. Exposure to factor X alone is not followed by the disease, but the disease never occurs in the absence of exposure to factor X.

According to the causal model described above, factor X is:

1. A necessary and sufficient cause
2. A necessary but not sufficient cause
3. A sufficient but not necessary cause
4. Neither necessary nor sufficient
5. None of these

Factor X is necessary because it is a component cause of every set of sufficient causes. However, Factor X alone cannot cause the disease. Therefore, factor X is not a sufficient cause all on its own.

**Question 9**

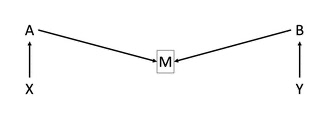
Randomization always leads to an unbiased estimate of the causal effect.

1. True
2. False

In theory, ideal randomized experiments (i.e., no loss to follow- up, full adherence to the assigned treatment over the duration of the study, a single version of the treatment, and double-blind assignment) with a large enough sample size always lead to estimates of association that can be interpreted as causal. Unfortunately, randomized experiments often do not meet this standard. Additionally, randomized experiments are often unethical, impractical, or untimely.

**Question 10**

Given this causal DAG, we expect X to have a causal effect on Y.

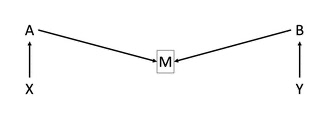


1. True
2. False

The lack of an arrow between X and Y encodes our belief that X has no causal effect on Y

**Question 11**

Given this causal DAG, we expect X and Y to be statistically associated in our data.



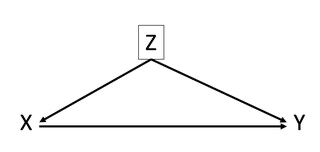
1. True
2. False

A

Conditioning on the collider, M, opens the path X -> A -> M <-B <- Y. Therefore, we would expect to find a non-causal statistical association between X and Y in our data.

**Question 12**

Given this DAG, we expect X to have a causal effect on Y.

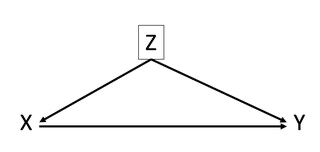


1. True
2. False

The arrow between X and Y encodes our belief that X has a causal effect on Y.

**Question 13**

Given this DAG, we expect our estimate of the causal effect of X on Y to be unbiased.



1. True
2. False

Conditioning on Z, as indicated by the box around Z, blocks the non-causal statistical association that would have flowed through the path x <- Z -> Y. Because that path is blocked, any remaining statistical association must arise from the causal effect of X on Y.