Week 5 – Tuesday session

Magnitude of Confounding

Misclassifiaction

EPI202 – Epidemiologic Methods II
Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Health

Which of the following statements about confounding are true? Select all that apply.

Confounding arises from non-exchangeability between the exposed and non-exposed groups.

Confounding introduces bias into point estimates.

The presence of confounding is represented on a DAG by conditioning on a common consequence of exposure and outcome.

The presence of confounding invalidates the interpretation of statistical inferential procedures such as hypothesis tests and confidence interval estimation.

Answer

When the exposed and non-exposed groups are not exchangeable, a non-causal association will emerge. Hence the observed association will not be equal to the causal effect of exposure on outcome. This form of bias is called confounding. An unblocked backdoor path on a DAG indicates the presence of confounding. Paradigmatically, this occurs when there is a common cause of exposure and outcome. A variable that can be conditioned on to block a backdoor path is called a confounder. Hence, whether a particular variable is a confounder or not depends on which other variables are conditioned on in the analysis. When confounding is present, the point estimate for the association between exposure and outcome is biased and statistical inferential procedures are not valid.

Conditioning on a collider that is a consequence of exposure and outcome leads to selection bias.

Direction of bias due to confounding

A study found that current use of smokeless tobacco ("snuff" or "snus") appeared to increase the incidence rate of ischemic stroke mortality. Younger individuals have a lower incidence of ischemic stroke mortality than older people, and based on data from this study, younger people are more likely to be current users of smokeless tobacco than older people. If the investigators did not control for age in the analysis, then uncontrolled confounding by age will lead to a crude incidence rate ratio that is:

The direction of the bias can not be determined because, based on the information above, age must be an effect modifier of the association between current smokeless tobacco use and ischemic stroke mortality.

Larger in absolute magnitude than the unconfounded effect (biased upward).

Not biased because, based on the information above, age is not a confounder.

Smaller in absolute magnitude than the unconfounded effect (biased downward).

Answer

The crude association will be smaller in absolute magnitude than the unconfounded effect.

Younger age is positively associated with smokeless tobacco use and is associated with a lower rate of ischemic stroke (in the absence of exposure). The difference in age between the exposed and non-exposed causes non-exchangeability; younger age is a confounder. Furthermore, the direction of the bias due to confounding in the crude analysis is in the downward direction. Thus after adjusting for age, the point estimate will be larger (in absolute magnitude). Since smokeless tobacco was associated with an increased rate of ischemic stroke mortality in the crude analysis, after conditioning on age, smokeless tobacco will appear to be even more harmful.

Based on the information shown, the crude RR is:

RR(C)=2   Apparent RR=5

Smoking → **OC Use**   **MI**

ECOR=2.86
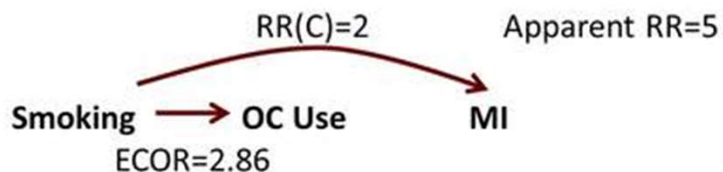
Too low

Unbiased

Too high

Do not know

Total Results: 0

Answer

C. Too high

Because smoking is positively associated with OC use in the study base and is positively associated with MI incidence, the bias will be in the upward direction and the crude estimate will be too high. Based on this, we expect that the adjusted RR will be less than 5.

Based on the information shown, is any of the association due to the effect of OC use on MI (assume no other bias)?

RR(C)=2   Apparent RR=5

Smoking → OC Use   MI

ECOR=2.86
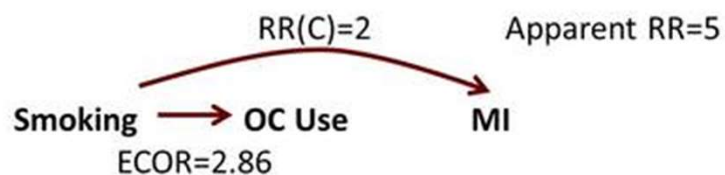
Yes

No

Do not know

Total Results: 0

Answer

A. Yes

Since the maximum bias due to confounding would lead to an association of an RR of 2.0 or smaller, some part of the crude association must be due to the true association between OC use and MI incidence. Of course, for this to represent a causal effect, we will need to make additional strong assumption including that there is no residual confounding by smoking, no confounding by other determinants, no selection or information bias.

If the null were true, what is the theoretical maximum RR that might be observed due to confounding by smoking?

RR(C)=2   Apparent RR=5

Smoking → OC Use   MI

ECOR=2.86
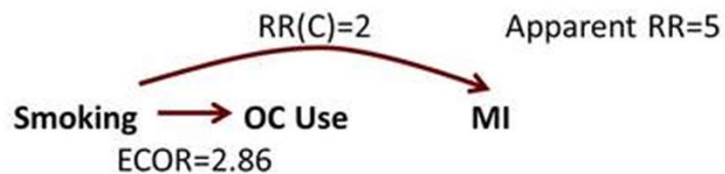
< 2.0

2.0 - 2.86

2.86 - 5.0

> 5.0

Total Results: 0

Answer

A.   <2.0

The magnitude of the bias due to confounding will be limited by the strength of the association between the confounder and the exposure and the confounder and the outcome. In this example, the association between the confounder (smoking) and the outcome (MI) was the weaker of these association. Therefore, the theoretical maximum bias due to confounding under the null would be an RR of 2.0. However, this maximum could only be attained if the association between the confounder (smoking) and the exposure (OC use) was infinitely strong (i.e., a perfect 1:1 association). In this example the association was only moderately strong (ECOR = 2.86), and thus the realized bias due to confounding will be much less than the theoretical maximum.

Misclassification can occur in which of the following types of studies?

Prospective cohort study

Retrospective cohort study

Retrospective case-control study

Randomized experiment

All of the above

Total Results: 0

Answer

E. All of the above

Misclassification can occur in any type of study.

In a study of habitual exercise and dementia, participants were asked to report whether they engaged in 30 minutes or more of moderate exercise at least 3 times per week. If this were a prospective cohort study of healthy participants interviewed about exercise at baseline and followed for 10 years to identify cases of dementia, the investigators would be most concerned about

nondifferential exposure misclassification

differential exposure misclassification

Total Results: 0

Answer

A. Nondifferential exposure misclassification

Since the participants were interviewed about exercise before the outcome occurred, it seems likely that any poor recall of usual exercise habits would be similar for individuals who subsequently do and do not develop dementia.

In a study of habitual exercise and dementia, participants were asked to report whether they engaged in 30 minutes or more of moderate exercise at least 3 times per week. If this were a retrospective case-control study of individuals interviewed at dementia clinics and controls from the underlying population and participants were interviewed about exercise at the time that they were recruited as a case of dementia or as a control, the investigators may be concerned about

nondifferential exposure misclassification

differential exposure misclassification

Total Results: 0

Answer

B. Differential exposure misclassification

Since the participants were interviewed about exercise after the outcome occurred, this may result in a bias in the upward or downward direction; if cases are more likely to under-report exercise exposure, this may lead to an overestimate of the protective benefits of exercise but if cases are more likely to over-report exercise exposure, this may lead to an underestimate of the protective benefits of exercise.

90% of the participants who had high exercise levels according to the accelerometer had reported high levels of exercise in the study questionnaire.

The sensitivity of self-report was 10%

The specificity of self-report was 10%

The sensitivity of self-report was 90%

The specificity of self-report was 90%

Total Results: 0

Answer

C. Sensitivity=90%

The sensitivity of the self-reported exercise data corresponds to the proportion of people who had high levels of exercise according to the accelerometer data who correctly reported high exercise levels in their self-report

P(T+|E+) = TP / (TP+FN) = 90%

25% of the participants who did not have high exercise levels according to the accelerometer had reported high levels of exercise in the study questionnaire

The sensitivity of self-report was 25%

The specificity of self-report was 25%

The sensitivity of self-report was 75%

The specificity of self-report was 75%

Total Results: 0

Answer

D. Specificity=75%

The specificity of the self-reported exercise data corresponds to the proportion of people who did not have high levels of exercise according to the accelerometer data who correctly reported low exercise levels in their self-report

P(T-|E-) = TN / (FP+TN) = 75%

Therefore, the proportion of people who were truly "unexposed" to high levels of exercise that incorrectly self-reported that they had high levels is equal to 1-specificity, or the False Positive Rate

P(T+|E-) = FP/ (FP+TN) = 25%

In Neyman-Pearson hypothesis testing:

Type 1 error (alpha) is defined as the probability of incorrectly rejecting the null hypothesis when in fact the null hypothesis is true

Type 2 error (beta) is defined as the probability of incorrectly not rejecting the null hypothesis when in fact the alternative is true

Power is the probability of rejecting the null hypothesis when in fact the alternative is true

All of the above

Answer

D. All of the above

If the null were true, would nondifferential exposure misclassification result in more false positive hypothesis tests?

Yes

No

Answer

B. No

Nondifferential misclassification generally causes bias towards the null and therefore alpha is preserved.

If there is a true association between exposure and outcome (the alternative hypothesis is true), then nondifferential misclassification of exposure will result in a bias towards the null. Therefore, compared to an identical study using a gold standard for exposure classification, you are less likely to reject the null hypothesis of an association in the study with the misclassified exposure.

True

False

Answer

A. True

Power is defined as the probability of rejecting the null hypothesis when an association truly exists. Since the misclassified data results in an estimate that is closer to the null compared with an identical study using the gold standard of exposure assessment, the observed association will be smaller and therefore there will be lower power.

In a study of physical activity and periodontitis, smokers are less likely to be physically active and they are more likely to develop periodontitis. The OR_{crude}=1.4. After adjusting for smoking with correctly classified pack years, the $OR_{adjusted}$ would be 3.5. The authors believed that there was independent nondifferential misclassification of smoking. Therefore, compared to the OR adjusted correctly, the OR adjusted for approximate pack-year will be

<1.4

Between 1.4 and 3.5

>3.5

Total Results: 0

Answer

B. Between 1.4 and 3.5

Residual confounding will result in an estimate in the direction of the bias. We expect the crude OR to be too low so a partially adjusted estimate is expected to be between 1.4 and 3.5

HAPPY THANKSGIVING!