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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

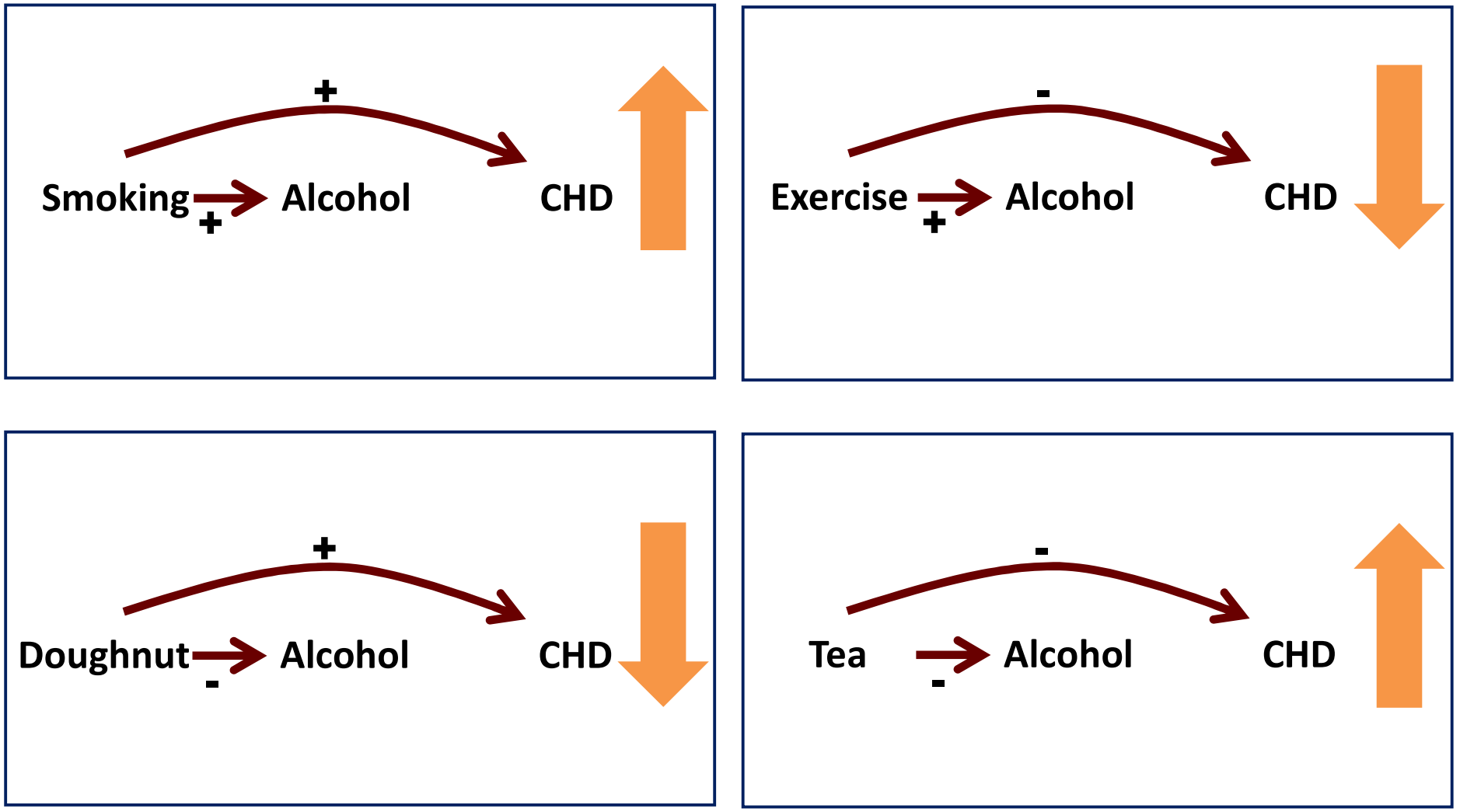
Week 5: Discussion Topics

1. Magnitude of bias due to uncontrolled confounding
   - Impact on testing and estimation
   - Partitioning the crude association
   - Potential maximum bias due to uncontrolled confounding
   - E-value
2. Misclassification
   - Terminology
   - Indices of accuracy – Sensitivity/specificty
   - Appropriate, unnecessary and overmatching
   - Misclassification correction
   - Impact of a misclassified exposure
   - Impact of a misclassified confounder

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.

Direction of Confounding



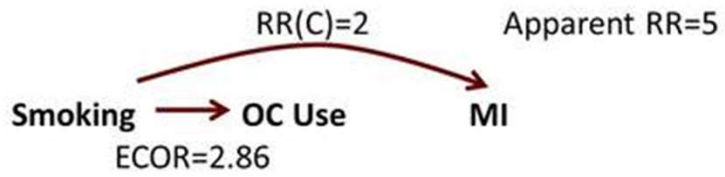
Assume exposures and outcomes are binary and there are no other sources of confounding or bias

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).

Based on the information shown, the crude RR is:



- Too low
- Unbiased
- Too high
- Do not know

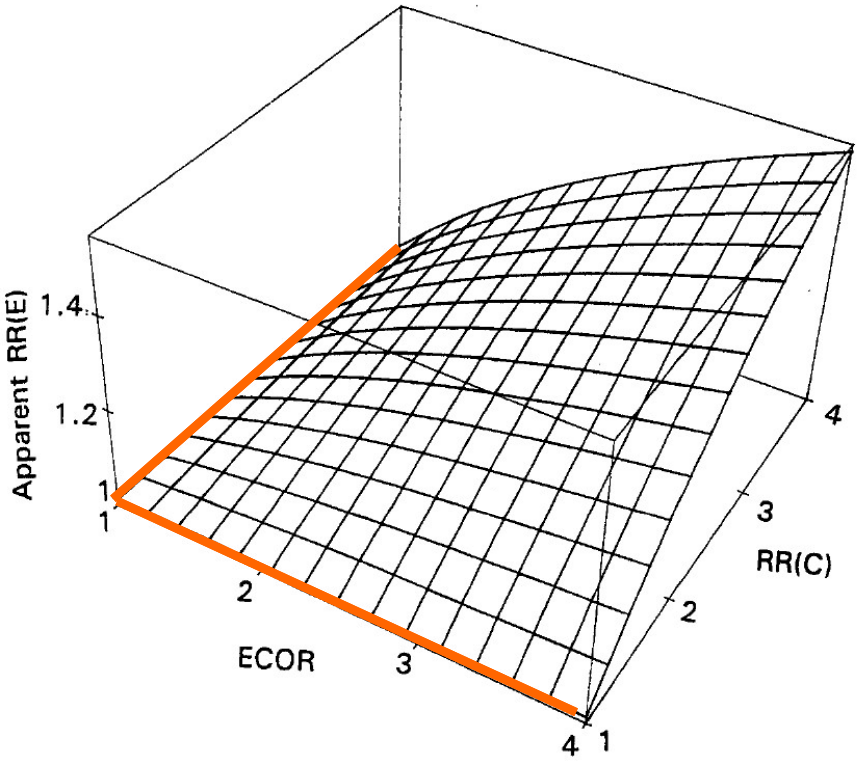
Total Results: 0

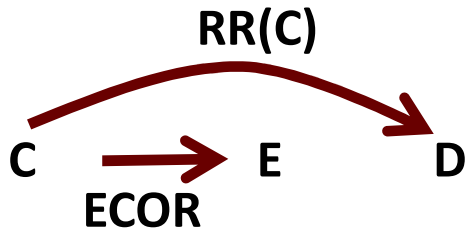
Magnitude of Confounding

- Determinants of the magnitude of bias due to uncontrolled confounding
  - Association between the confounder and exposure
  - Association between the confounder and outcome
  - Prevalence of the confounder in the study population

Walker's Bounds of Confounding (1)

P(E) = 0.2
P(C) = 0.2



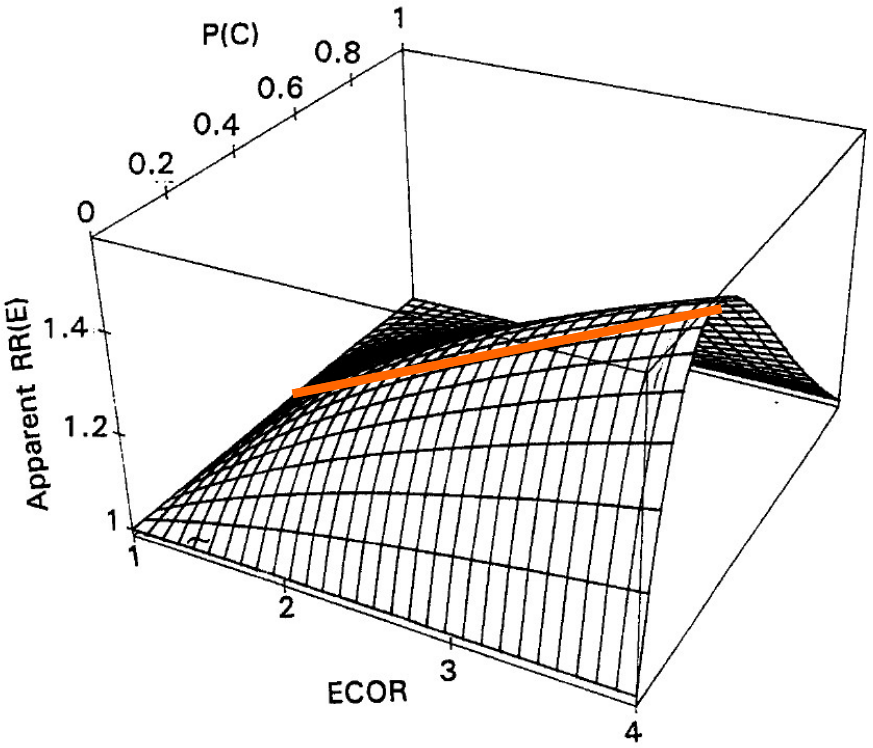


ECOR	Exposure-confounder OR
P(C)	Prevalence of confounder
RR(C)	Confounder-outcome RR
P(E)	Prevalence of exposure

Alexander Muir Walker. Observation and Inference: An Introduction to the Methods of Epidemiology. Epidemiology Resources, June 1991

Walker's Bounds of Confounding (2)

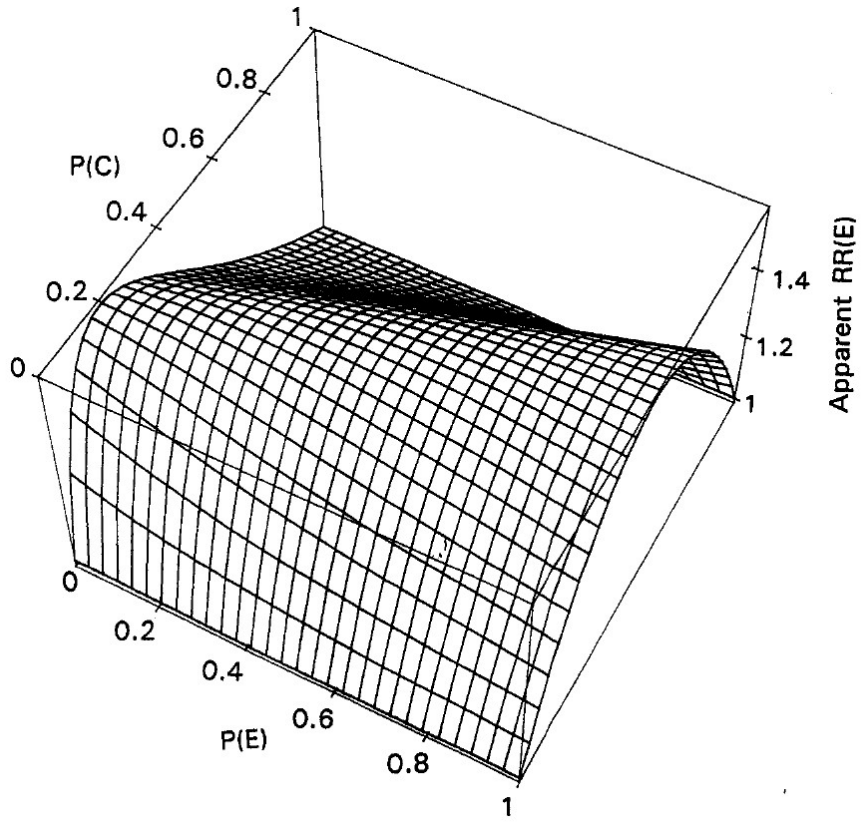
P(E) = 0.2
RR(C) = 4



Alexander Muir Walker. Observation and Inference: An Introduction to the Methods of Epidemiology. Epidemiology Resources, June 1991

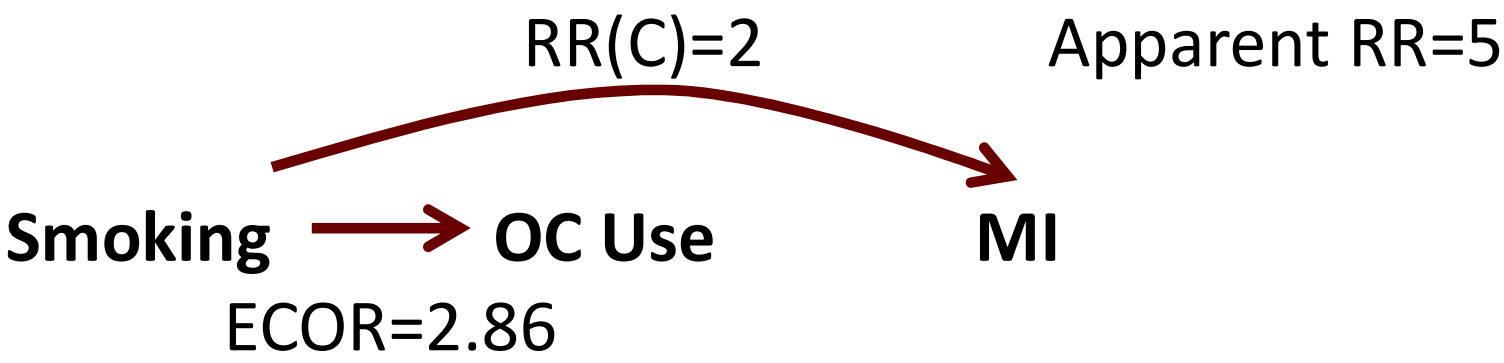
Walker's Bounds of Confounding (3)

ECOR = 4
RR(C) = 4



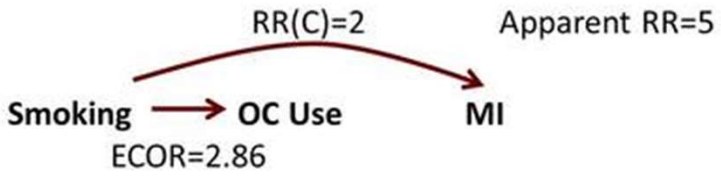
Alexander Muir Walker. Observation and Inference: An Introduction to the Methods of Epidemiology. Epidemiology Resources, June 1991

Maximum RR Due to Confounding
Example



- What is the maximum relative risk that could be attributed to smoking?
- If there are no other confounders of the observed association, then some of the observed elevation in relative risk from OC use must represent the real effect of OC use on MI incidence, regardless of the prevalence of the confounder.

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



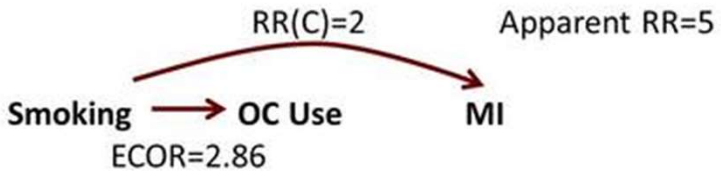
Yes

No

Do not know

Total Results: 0

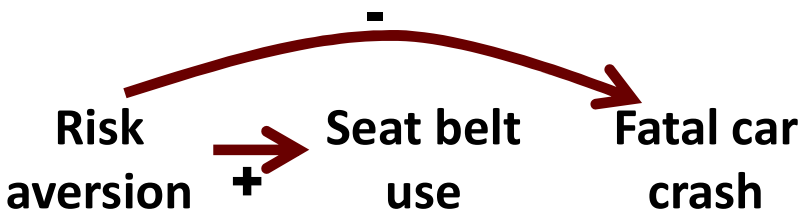
If the null were true, what is the theoretical maximum RR that might be observed due to confounding by smoking?



- < 2.0 ✓
- 2.0 - 2.86
- 2.86 - 5.0
- > 5.0

Total Results: 0

E-Value



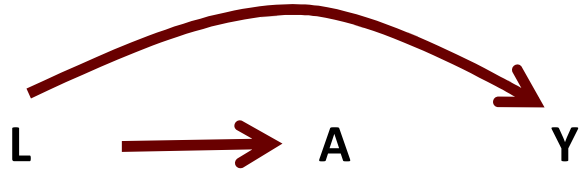
- The E-value is the smallest value for the association between an unmeasured confounder and both exposure and outcome that could fully explain the observed association.

$$\text{E}-\text{value} = RR + \sqrt{RR(RR-1)}$$

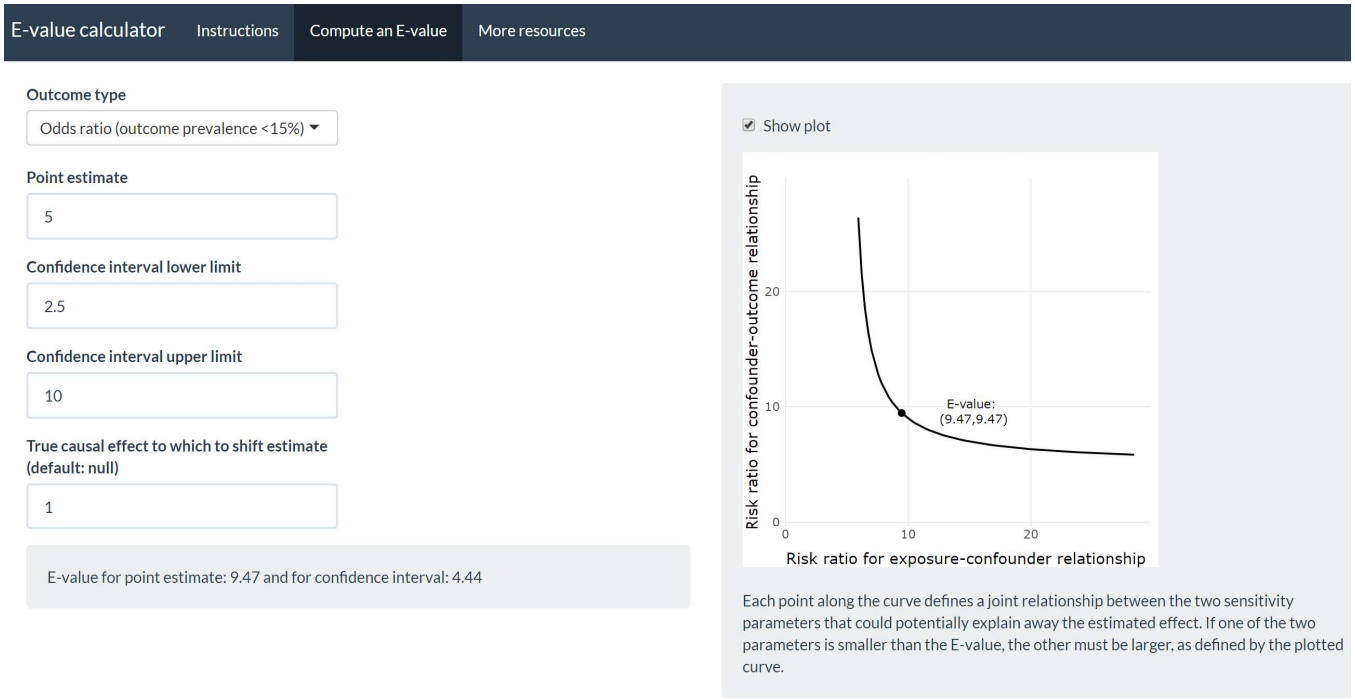
- If RR < 1, take the reciprocal before computing the E-value

VanderWeele TJ, & Ding P. (2017). Sensitivity analysis in observational research: introducing the E-value. Annals of Internal Medicine, 167(4), 268-274.

E-Value Web Calculator



OR= 5.0, 95% CI (2.5,10.0)



https://www.evalue-calculator.com/

Mathur MB, Ding P, Riddell CA, VanderWeele TJ. (2018). Website and R package for computing E-values. Epidemiology, 29(5), e45-e47.

Misclassification of a Dichotomous Exposure or Outcome

	Non-Differential	Differential
Exposure	▪ Misclassification of exposure is similar for cases and non-cases ▪ Generally leads to a bias towards the null*	▪ Misclassification of exposure is different for cases and non-cases ▪ Lead to a bias in either direction, depending on situation
Outcome	▪ Misclassification of outcome is similar for exposed and unexposed ▪ Generally leads to a bias towards the null*	▪ Misclassification of outcome is different for exposed and unexposed ▪ Lead to a bias in either direction, depending on situation

***This holds when exposure and outcome are binary**

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

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

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

Indices of Measurement Accuracy (2)

		Truth E+	Truth E-	
Observed	T+	True Positive TP	False Positive FP	TP+FP
	T-	False Negative FN	True Negative TN	FN+TN
		TP+FN	FP+TN	

Sensitivity	Pr[T+|E+]	TP / (TP + FN)
Specificity	Pr[T-|E-]	TN / (FP + TN)
False Negative Rate (1-Sensitivity)	Pr[T-|E+]	FN / (TP + FN)
False Positive Rate (1-Specificity)	Pr[T+|E-]	FP / (FP + TN)
Positive Predictive Value	Pr[E+|T+]	TP / (TP + FP)
Negative Predictive Value	Pr[E-|T-]	TN / (TN + FN)

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

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

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

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

Yes

No

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

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

Exposure Misclassification Impact:
Observed → Truth
Notation

Based on the formulae above:

Observed			
	E+	E-	
D+	A	B	m_1
D-	C	D	m_0

Truth			
	E+	E-	
D+	a	b	m_1
D-	c	d	m_0

Cases	a:	$= \frac{(\text{specificity})*m_1 - B}{\text{sensitivity} + \text{specificity} - 1}$
	b:	$= \frac{(\text{sensitivity})*m_1 - A}{\text{sensitivity} + \text{specificity} - 1}$
Non-cases	c:	$= \frac{(\text{specificity})*m_0 - D}{\text{sensitivity} + \text{specificity} - 1}$
	d:	$= \frac{(\text{sensitivity})*m_0 - C}{\text{sensitivity} + \text{specificity} - 1}$

Note: Since we are assuming no misclassification of disease,

$m_1 = A+B = a + b$

and

$m_0 = C + D = c + d$

HAPPY THANKSGIVING!