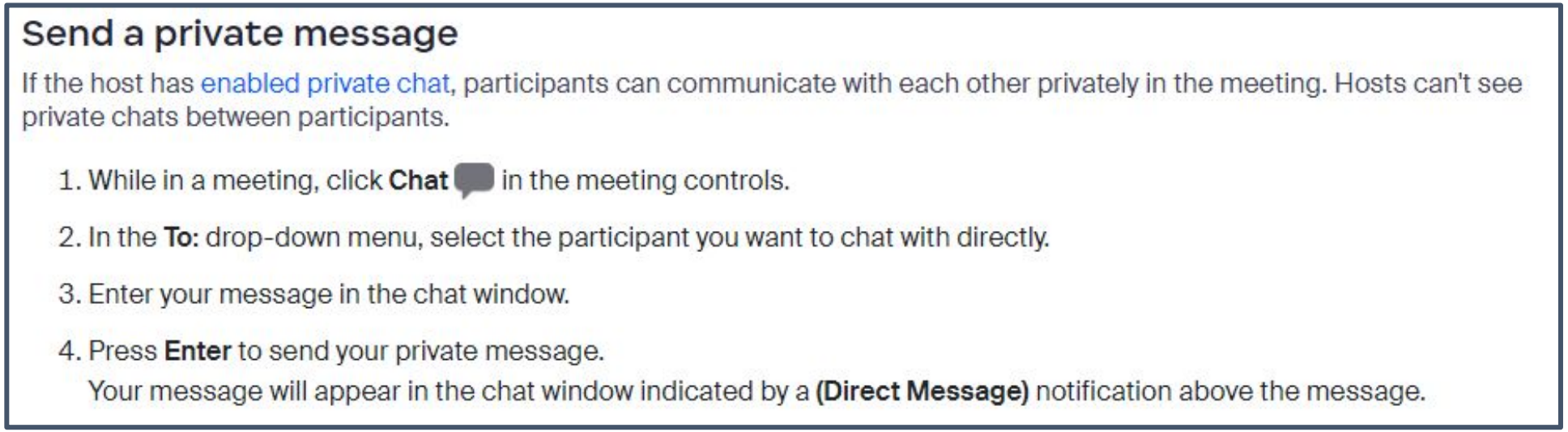
Recording sessions

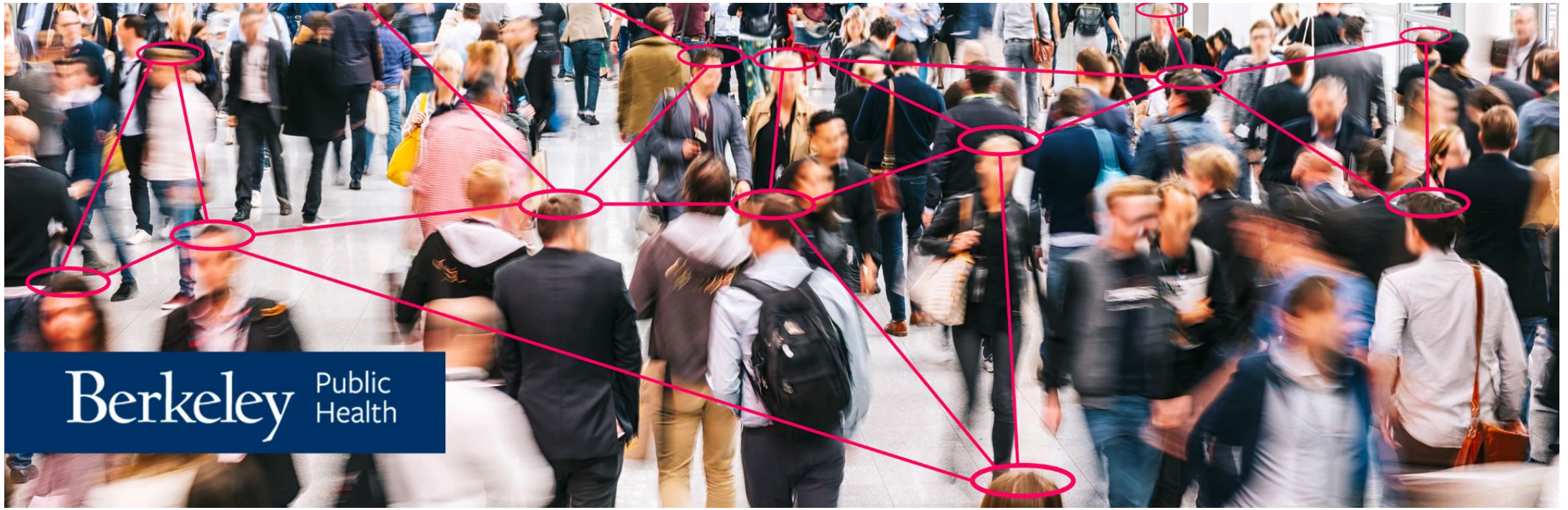
- All live sessions will be recorded and available on bCourses.
- If you want remain anonymous during the session, please send a private message to me on Zoom.



Send a private message

If the host has enabled private chat, participants can communicate with each other privately in the meeting. Hosts can't see private chats between participants.

1. While in a meeting, click **Chat** in the meeting controls.
2. In the **To:** drop-down menu, select the participant you want to chat with directly.
3. Enter your message in the chat window.
4. Press **Enter** to send your private message.
   Your message will appear in the chat window indicated by a **(Direct Message)** notification above the message.



Epidemiologic Methods II

PHW250B

Week 9: Bias

Agenda

- Exam 2 Logistics
- Bias overview
- Practice problems
- Q & A

Exam 2

- **Materials from Weeks 5-9**
  - Cohort studies, case-control studies, trials, cross-sectional studies, ecological studies, bias (selection and information)
  - Cumulative exam, so material from Weeks 1-5 will be relevant (but not the focus)

- **Online Proctored Exam**
  - Open for self procter 10/29-11/2
  - Live proctored session: Sunday November 2nd 12pm-3pm PST

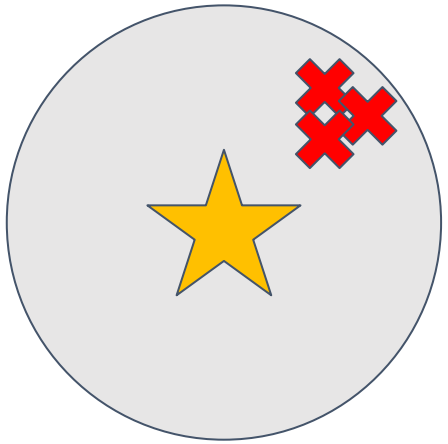
Bias Review

Types of Error:

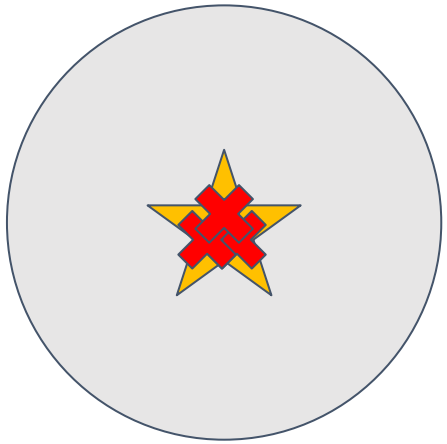
- Systematic error: Affects the validity of our estimates
- Random error: Affects the precision of our estimates

= TRUTH

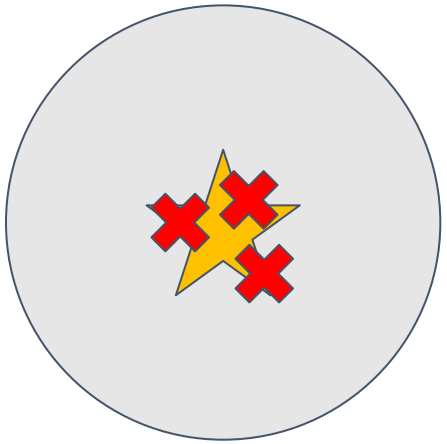
= ESTIMATE



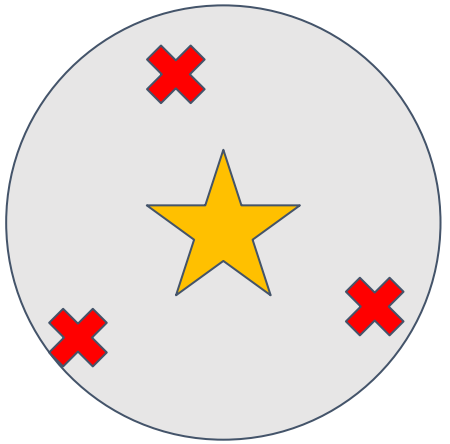
Precise, not valid



Precise and valid



Imprecise and valid



Imprecise and not valid

Systematic error describes how close or far away we are from the truth

Bias Review

Selection bias: Distortion of estimates that are related to who gets to participate in the study and who gets to stay in the study (loss-to-follow-up).

Information bias (misclassification): Distortion of estimates that are related to how the exposure/outcome gets measured or collected.

Types of exposure misclassification:

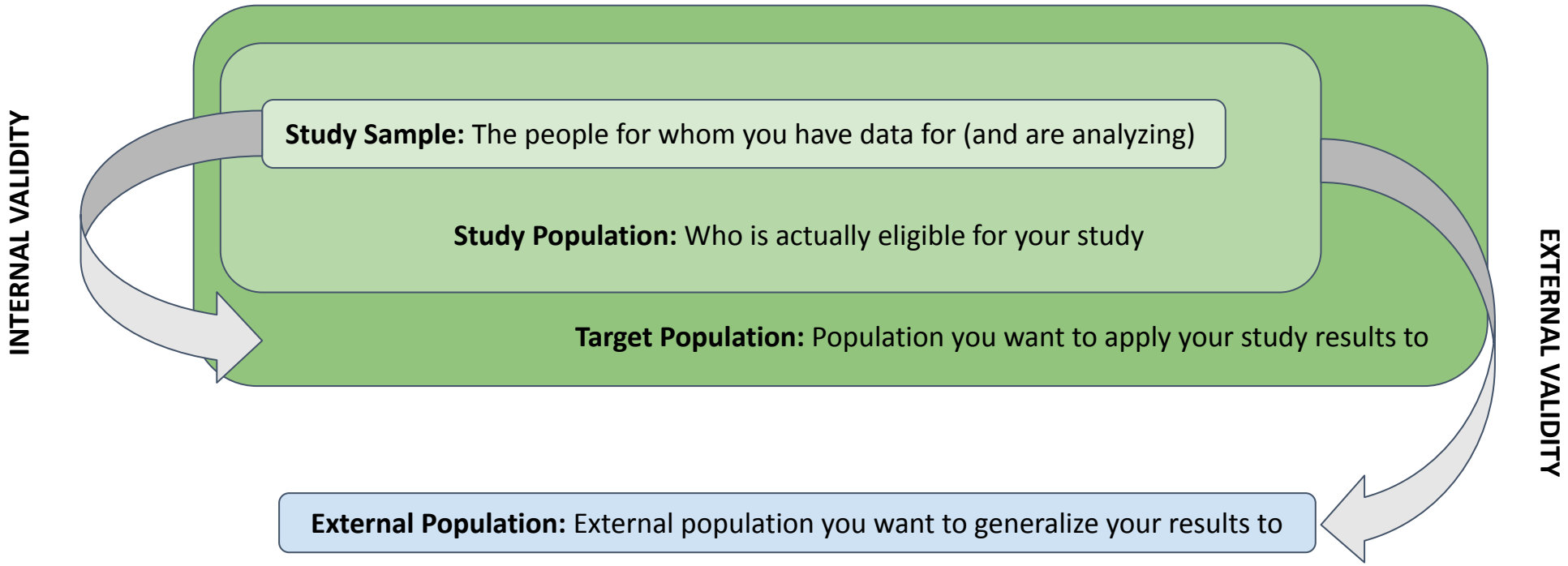
- Non-differential exposure misclassification
- Differential exposure misclassification by disease status
  - Level of misclassification depends on outcome

Types of outcome misclassification:

- Non-differential outcome misclassification
- Differential outcome misclassification by exposure status
  - Level of misclassification depends on exposure

Sensitivity & Specificity Review

Generalizability Review



Tab 1, Problem 1

An investigator decides to conduct a study examining the relationship between Facebook usage and anxiety among undergraduate students. She wants to enroll 75 participants from the largest undergraduate dorm on campus. She will measure their "Facebooking" behavior over a 2-week period and then test them for signs of anxiety. Two hundred students respond to her recruitment flyers. Fifteen are excluded because they do not use Facebook or are not undergrads. She randomly selects 75 students from those who responded and were eligible.

a. Who comprises the study, actual, and target populations?

Study sample: The 75 students who satisfied the inclusion criteria and actually participated in the study.

Study population: The 185 students who responded to the flyers and satisfied the inclusion criteria. (This includes everyone who was eligible and willing to participate in the study if chosen.)

Target population: Students who live in the selected undergraduate dorm. This definition of the study's target population could be different – remember that the target population may depend on the investigator's aims!

Tab 1, Problem 1

An investigator decides to conduct a study examining the relationship between Facebook usage and anxiety among undergraduate students. She wants to enroll 75 participants from the largest undergraduate dorm on campus. She will measure their “Facebooking” behavior over a 2-week period and then test them for signs of anxiety. Two hundred students respond to her recruitment flyers. Fifteen are excluded because they do not use Facebook or are not undergrads. She randomly selects 75 students from those who responded and were eligible.

b. What are two examples of potential external populations to whom the investigator may wish to generalize her results?

Possibilities: All Berkeley undergrads, all 18-22-year-olds in California, all 18-22-year-olds in the U.S. This would really depend on the investigator’s goals and the generalizability of the study population.

Tab 1, Problem 1

An investigator decides to conduct a study examining the relationship between Facebook usage and anxiety among undergraduate students. She wants to enroll 75 participants from the largest undergraduate dorm on campus. She will measure their “Facebooking” behavior over a 2-week period and then test them for signs of anxiety. Two hundred students respond to her recruitment flyers. Fifteen are excluded because they do not use Facebook or are not undergrads. She randomly selects 75 students from those who responded and were eligible.

c. The investigator realizes after the study is over that the study participants were students who had, on average, come from much wealthier families than the dorm residents as a whole. What sort of validity (internal or external) should she be worried about?

The internal validity of her study. The study population was not representative of the target population (selection bias).
Additionally, she should be worried about external validity because internal validity is a prerequisite for external validity.

Tab 1, Problem 1

An investigator decides to conduct a study examining the relationship between Facebook usage and anxiety among undergraduate students. She wants to enroll 75 participants from the largest undergraduate dorm on campus. She will measure their “Facebooking” behavior over a 2-week period and then test them for signs of anxiety. Two hundred students respond to her recruitment flyers. Fifteen are excluded because they do not use Facebook or are not undergrads. She randomly selects 75 students from those who responded and were eligible.

d. The investigator presents results that are very imprecise (the confidence interval is very wide). What kind of error affects the precision of a study?

Random error

Tab 1, Problem 7

This is case-control study. Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The table gives the data for the association between current exposure to biomass smoke and TB:

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

A. In this study, exposure assessment was made by interviews. If tuberculosis patients provide more accurate histories about past exposure to biomass smoke compared to controls, what bias can occur? Would it be differential or non-differential?

This bias is called recall bias (a form of information bias). To be considered recall bias, assessment of past exposure must affect cases and controls differently – i.e., differential.

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The table gives the data for the association between current exposure to biomass smoke and TB:

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

B. If tuberculosis patients changed their cooking fuel to cleaner fuels like gas after developing the disease, how could it affect the odds ratio for current use of a biomass stove?

This is misclassification error (a form of information bias) and in this example it is differential by disease status (only affects cases).
In this example, it will bias the OR toward the null because some cases are reported as *not* exposed but were actually exposed prior to developing the disease.

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The table gives the data for the association between current exposure to biomass smoke and TB:

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

C. In this study, patients with ear, nose and throat (ENT) conditions were chosen as controls. If some ENT conditions are associated with biomass smoke (like allergic rhinitis), could this affect the odds ratio? What would this type of bias represent?

In this case, the controls are also likely to have high levels of biomass exposure. This would bias the OR towards the null. The study includes more exposed (potential) controls than unexposed (potential) controls relative to the target population. This is a form of selection bias—specifically, Berkson's bias—and it would be differential by disease status.

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The table gives the data for the association between current exposure to biomass smoke and TB:

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

D. If cases and controls had similar problems in remembering the cooking fuels they used in the past, what type of misclassification could have occurred? How will that affect the odds ratio?

Since both cases and controls have similar problems in recall, the misclassification will be non-differential exposure misclassification. On average, it would bias the OR towards the null. This is NOT a form of recall bias, which would require that recall be differential between cases and controls. Contrast this with 4.A (above).

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The table gives the data for the association between current exposure to biomass smoke and TB:

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

E. If the interviewers who did the exposure assessment had known the disease status of the cases and controls, how could this have affected the odds ratio? What type of bias would this be?

If the interviewers knew the disease status, they could probe for better exposure data from the TB patients and this could bias the OR away from the null, assuming that this caused TB patients to report more exposure than they would have otherwise. This bias is called interviewer bias (a form of information bias) and it would be differential by disease status.

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

F. Biomass smoke is known to contain many toxic chemicals. If exposure to biomass smoke is one cause of tuberculosis, and if it leads to rapid progression of disease and death among these patients, how would it affect the OR in a case-control study examining biomass smoke as a risk factor for TB?

- Assuming a cumulative case-control design, if exposure leads to rapid death among TB patients, then the cases who end up being recruited into the study may have contracted TB due to other exposures, since they are still alive and well enough to take part in a study.
- This would result in a form of selection bias, as the exposure distribution among the cases does not reflect that in the target population – those exposed to biomass smoke are less likely to have been selected.
- The OR would likely be biased towards the null because the measured effect would likely be smaller than the true effect. People would be missing from the A cell. Remember, selection bias can be towards or away from the null, depending on what's happening!

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

G. The authors could have improved the measurement of exposures in this study by actually quantifying indoor air pollution levels due to biomass smoke using instruments. If they had done that, how would it have affected the precision of their OR estimates?

More refined measurement of the exposure status would increase the validity of those measurements (and therefore the OR), but it would not necessarily have any effect on the precision of the OR.

Tab 1, Problem 7

Cases were 288 patients with an active smear or culture positive tuberculosis, and controls were 545 patients with ear, nose and throat ailments from the same hospital, seen at the same time. Past or present exposure to biomass smoke was obtained by interview. The authors found a significant association between indoor smoke and TB (OR = 5.2). After adjusting for several confounders, a significant association remained (adjusted OR = 2.2).

	TB		
Biomass Smoke Exposure	Cases	Controls	Total
Exposed	50	21	71
Unexposed	238	524	762
Total	288	545	833

H. Suppose instead of interviews, biomass smoke exposure was assessed using air monitors worn by individuals in the study. You classify exposure to biomass smoke as high, medium and low using the monitor results. After you finish your study, the monitor company tells you their monitors have been acting strangely. A random set of monitors in your study malfunctioned such that some over-estimated smoke exposure while others under-estimated smoke exposure (malfunctions affected cases and controls equally). What, if anything, can you say about the odds ratios you calculated in your study?

We cannot be sure if our OR is biased toward or away from the null. Non-differential misclassification when there are more than two categories of the exposure or disease that is misclassified does not necessarily result in bias towards the null. We have three levels of biomass smoke exposure.

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	A	B	A + B
	No	C	D	C + D
	Total	A + C	B + D	

1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

a. For some of cases, aspirin use followed the onset of the earliest manifestation of Reye's syndrome rather than preceded it.

OR is biased **away from the null** because cell A is erroneously inflated and cell C is deflated.

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	A	B	A + B
	No	C	D	C + D
	Total	A + C	B + D	

1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

b. Parents tended to incorrectly report Tylenol use as aspirin use, non-differentially with respect to case or control status.

OR is biased **towards the null** because non-differential misclassification of the exposure will bias results towards the null.

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	A	B	A + B
	No	C	D	C + D
	Total	A + C	B + D	

1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

c. The retrospective nature of the case-control study led to some problems with recall. Parents of both cases and controls under-reported aspirin use by 10%.

OR is biased **towards the null** because non-differential misclassification of the exposure will bias results towards the null.

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	A	B ↓	A + B
	No	C	D	C + D
	Total	A + C	B + D	

1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

d. Parents of children with Reye's syndrome recalled aspirin use with greater accuracy than the parents of children without the syndrome, who tended to underreport aspirin use.

OR is biased **away from the null** because B is deflated.

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	↑ A	B	A + B
	No	C	D	C + D
	Total	A + C	B + D	

1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

e. Due to the increased publicity about possible consequences associated with aspirin, parents of cases were more likely to report Tylenol use as aspirin use.

OR is biased **away from the null** because cell A is inflated (differential misclassification).

Tab 1, Problem 9

The association between Reye's syndrome and aspirin is being investigated. Reye's syndrome occurs in children due to consumption of aspirin medication. It leads to a condition called hepatic encephalopathy where the liver and brain are affected. If the true odds ratio (OR) for Reye's syndrome and aspirin is **significantly greater than one**, what will be the effect of the following scenarios? In each instance, indicate if the odds ratio would be:

		Reye's Syndrome		Total
		Yes	No	
Aspirin Use	Yes	↑ A	B	A + B
	No	C	D	C + D
	Total	A + C	B + D	

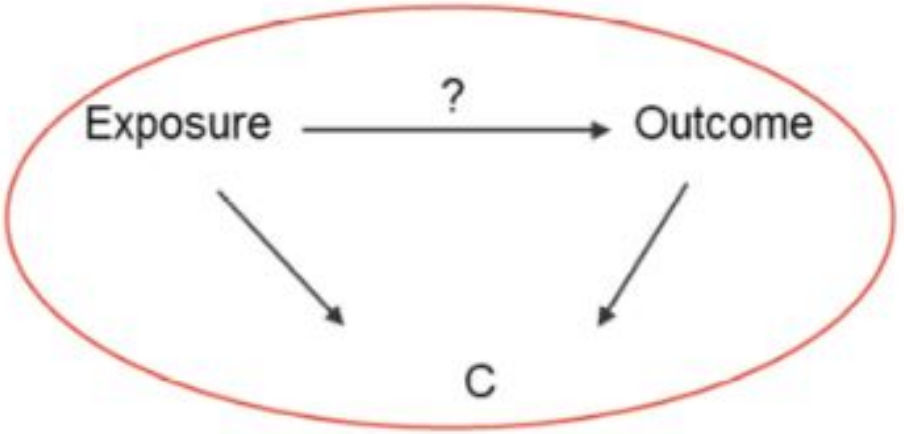
1. Biased away from null (overestimated)
2. Biased towards null (underestimated)
3. Unbiased

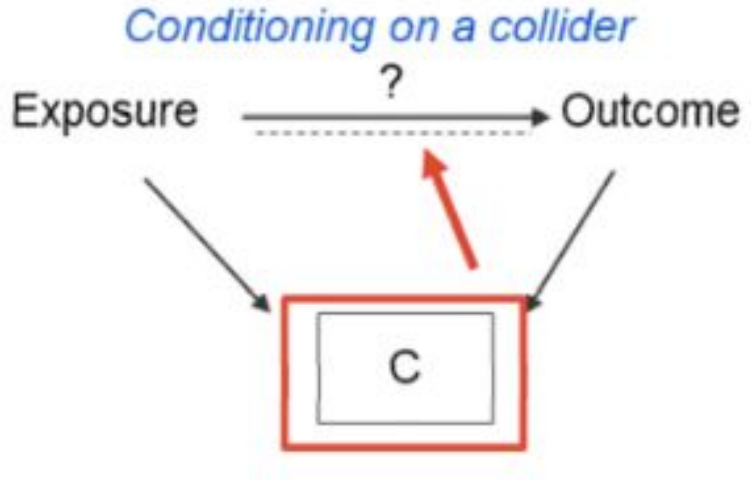
f. Cases were selected from physician reports, and physicians were more likely to diagnose Reye's syndrome if the child had used aspirin.

OR is biased **away from the null** because cell A is inflated (differential misclassification).

Using DAGs for Confounding

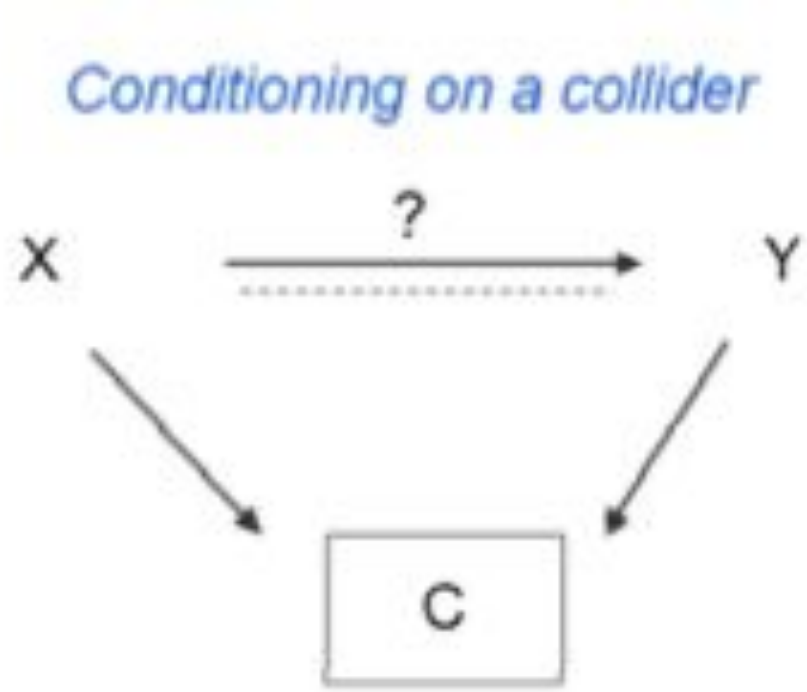
Conditioning on a collider creates bias!





Using DAGs for Confounding

Conditioning on a collider creates bias!

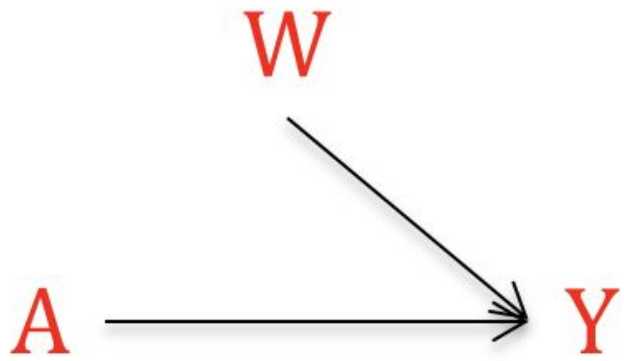


Tab 2, Problem 1

Researchers are conducting a placebo-controlled randomized trial to test whether anti-inflammatory medication can prevent preeclampsia (preeclampsia = gestational hypertension = high blood pressure during pregnancy). (5 points)

a. Draw a DAG to show the causal structure underlying the trial data; include the exposure, outcome, and pre-pregnancy blood pressure.

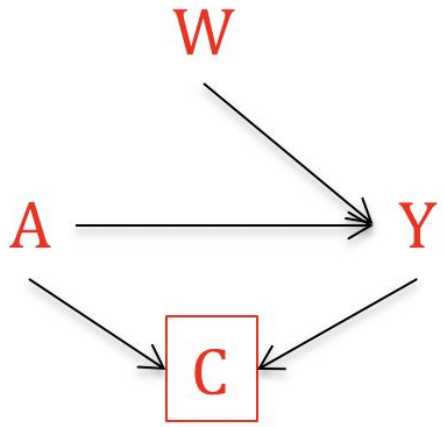
A is exposure, Y outcome, and W pre-pregnancy BP in the graph below.
There is no line between A and W because A is randomized.



Tab 2, Problem 1

Researchers are conducting a placebo-controlled randomized trial to test whether anti-inflammatory medication can prevent preeclampsia (preeclampsia = gestational hypertension = high blood pressure during pregnancy). (5 points)

b. During the trial, participants report minor side effects of the medication; these reports are more common in the treatment arm than the placebo arm. 80% of the cohort remains in the trial throughout while 20% withdraw. Researchers believe side effects contributed to retention in the study. The researchers also know that those showing early signs of preeclampsia were more likely to stay in the study. Update your DAG above or re-draw it to include retention in the study and how loss to follow up impacts the existing variables in the DAG.



Remember: daggity.net!