Information Bias

Serena Houghton, PhD, MPH
EVMS
April 5, 2021

1

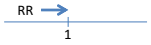
Opener

For a cohort study examining fruit and vegetable intake (exposed: >5 servings/wk, unexposed: ≤5 servings/wk) and the risk of hip fracture, what do you think would happen to the RR if 15% of ALL unexposed are misclassified as exposed AND 10% of ALL exposed are misclassified as unexposed?

A) Away from the null, more inverse

B) Away from the null, more positive

C) Towards the null



2

Recap: Random vs. Systematic Error

- Random error affects precision (repeatability)
- Systematic error affects validity (bias)

Precise, Valid

Not Precise, Valid

Random Error
-affects confidence interval

3

Recap: Random vs. Systematic Error

- Random error affects precision (repeatability)
- Systematic error affects validity (bias)

Precise, Valid

Systematic Error
-affects association measure

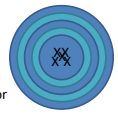
Precise, Not Valid

4

Recap: Random vs. Systematic Error

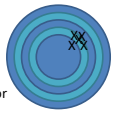
- Random error affects precision (repeatability)
- Systematic error affects validity (bias)

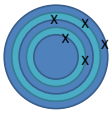
Precise, Valid
Low random error,
Low systematic error



Not Precise, Valid
High random error,
Low systematic error

Precise, Not Valid
Low random error,
High systematic error





Not Precise, Not Valid
High random error,
High systematic error

5

Recap: Bias

- 2 major categories of biases related to study design/procedures
  - Selection bias
  - Information bias
    - Participants erroneously placed (misclassified) in different exposure/ outcome category than they one they should be in

6

Misclassification of Exposure

- How accurately can these exposures be measured?
  - Age
  - Race/ ethnicity
  - Physical activity
  - Caffeine intake
  - Height and weight
  - Pollutants
  - Stress

7

Misclassification of Exposure

- Imprecise measurement
  - e.g., scale not calibrated and everyone is 5lbs heavier than true weight
- Subject self-report
  - e.g., issues around recall, stigma, etc.
- Incorrect coding of exposure data
  - e.g., systematic errors in data entry/ scanning

8

Non-differential Misclassification of Exposure

- Non-differential: probability of individuals being misclassified is <u>equal</u> across all groups in the study
  - Misclassification of exposure is unrelated to disease

	Outcome	No Outcome
Exposed		
Not Exposed	10%	10%

9

Non-differential Misclassification of Exposure

- Non-differential: probability of individuals being misclassified is equal across all groups in the study
  - Misclassification of exposure is unrelated to disease
- Effect:
  - 2 exposure categories: usually biases towards the null
  - 3+ exposure categories:
    - Overall exposure-outcome trend usually biased towards the null
    - Intermediate exposure groups may be biased away from null

10

Differential Misclassification of Exposure

- Differential: probability of being misclassified differs between groups in a study
  - Misclassification of exposure is related to disease

	Outcome	No Outcome
Exposed		
Not Exposed	20%	10%

11

Differential Misclassification of Exposure

- Differential: probability of being misclassified differs between groups in a study
  - Misclassification of exposure is related to disease
- Effect: bias can go in either direction, either towards or away from null; inflate or attenuate effect estimates

12

Recall/Reporting Bias

- Cases/ controls have different accuracies in recalling exposures
- To prevent/ reduce:
  - Verify exposure information (e.g., medical records)
  - Objective measures (e.g., blood levels, clinician measures)
  - Use diseased control group unlikely to be related to exposure
  - Well designed cohort study

13

Interviewer Bias

- Interviewer knows case or control status of participant and asks questions differently/ probes more
- To prevent:
  - Trained interviewers blinded to outcome status
  - Verify exposure information (e.g., medical records)
  - Objective measures (e.g., blood levels, clinician measures)
  - Well designed cohort study

14

ACTIVITY

15

Misclassification of Outcome

- Incorrect diagnosis
  - Limited knowledge
  - Diagnostic process complex
  - Laboratory error
  - Subclinical disease
  - Detection bias
- Subject Self-Report
  - Incorrect recall
  - Reluctant to be truthful
- Records incorrectly coded in database

16

Non-differential Misclassification of Outcome

- Non-differential: probability of individuals being misclassified is <u>equal</u> across all groups in the study
  - Misclassification of outcome is unrelated to exposure

	Outcome	No Outcome
Exposed		10%
Not Exposed		10%

17

Differential Misclassification of Outcome

- Differential: probability of being misclassified <u>differs</u> between groups in a study
  - Misclassification of exposure is related to disease

	Outcome	No Outcome
Exposed		20%
Not Exposed		10%

18

Information Bias - Cohort Studies

- Misclassification of exposure at baseline
  - Likely non-differential
- Changes in exposure over time
- Misclassification of outcome
  - Detection bias, diagnostic suspicion bias

19

Information Bias - Case-Control Studies

- General poor recall vs. recall bias
- Interviewer bias

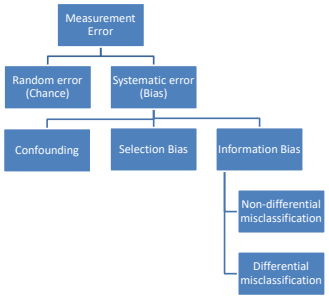
20

Summary

- Exposure and outcome can be misclassified
- Misclassification can be:
  - Non-differential (equal probability between groups compared)
    - Usually biases towards null
  - Differential (different probability between groups compared)
    - Can bias towards or away from the null

21

Bigger Picture



22