Cancer Research Methods Workshop

Mayur M. Desai, PhD, MPH

Mwanza, Tanzania

September 18-19, 2012

Yale SCHOOL OF PUBLIC HEALTH

Bias

Internal vs. external validity

- Internal validity
  - Did the study measure what it set out to measure?
- External validity (generalizability)
  - Can the study results be extrapolated to a wider population?
- Our focus will be on threats to internal validity – "bias" undermines the internal validity of a study

Bias

- Defined as the result of ***systematic error*** in the design or conduct of a study
- Bias results from flaws in either the method of selection of study participants [***selection bias***] or the procedures for gathering exposure and/or outcome information [***information bias***]

Process vs. outcome

- Bias relates to the "process" of the study, not the "outcome" of any particular study, thus:
  - sound study design, methods, and procedures = unbiased (valid) study → on average, correct result
  - faulty study design, methods, and procedures = biased (invalid) study → on average, incorrect result
  - (note: even a biased study can yield the "correct result" by chance)
- Because "the truth" is unknown and an infinite number of studies aren't conducted, bias has to be assessed in the context of a careful evaluation of the present study

Selection bias

- Systematic error in the ascertainment of study subjects, resulting in a tendency toward distorting the exposure-outcome association
- Present when individuals have different probabilities of being included in the study sample according to the exposure and the outcome of interest

Selection bias is a major potential problem in case-control studies

Recall basic principles:

- Study cases should be representative of all cases that arise in the source population with respect to the exposure of interest
- Controls should be representative of the source population with respect to the exposure of interest

Biased cases in case-control studies

- The situation when ascertainment of cases is influenced by exposure status, resulting in either:

$$OR = \frac{ad}{bc}$$

  - overrepresentation of exposed cases (→ this would tend to overestimate OR)

	D+	D-
E+	**a**	
E-	c	

  - overrepresentation of unexposed cases (→ this would tend to underestimate OR)

	D+	D-
E+	a	
E-	**c**	

Biased cases in case-control studies (cont.)

- Problems may arise when, for example:
  - there is differential participation according to exposure status
  - exposure leads to increased likelihood of diagnosis (also called detection bias or medical surveillance bias or, in the case of hospital-based cases, Berkson's bias)
  - prevalent (or surviving) cases are used instead of incident cases
- Can we think of some examples?

Biased controls in case-control studies

- The situation when ascertainment of controls is influenced by exposure status, resulting in either:

$$OR = \frac{ad}{bc}$$

  - overrepresentation of exposed controls (→ this would tend to underestimate OR)

	D+	D-
E+		**b**
E-		d

  - overrepresentation of unexposed controls (→ this would tend to overestimate OR)

	D+	D-
E+		b
E-		**d**

Biased controls in case-control studies (cont.)

- Problems may arise when, for example:
  - there is differential participation according to exposure status
  - there are inappropriate criteria for selection of controls (e.g., hospital-based control diagnoses are not independent of exposure status)
- Can we think of some examples?

Assessment and prevention of selection bias

- There is no statistical test to determine whether selection bias has occurred – must be assessed through critical review of the study
- Generally speaking, there is no statistical procedure to "fix" selection bias – once it's occurred, you're pretty much @#$%&!
- In general, how to avoid selection bias (easier said than done!):
  - ensure that the study design is appropriate
  - establish appropriate selection criteria
  - minimize nonparticipation and loss to follow-up

Information bias
(misclassification bias)

Imperfect definitions of study variables or flawed data collection procedures, resulting in misclassification of exposure and/or outcome status

- Nondifferential misclassification bias
  - degree of misclassification is the same in the groups being compared (i.e., of exposure in a case-control study or of outcome in a cohort study or RCT)
- Differential misclassification bias
  - degree of misclassification differs between the groups being compared (i.e., of exposure in a case-control study or of outcome in a cohort study or RCT)

Effects of information bias

- Nondifferential misclassification bias:
  - biases an association toward the null, assuming a 2x2 table (situation is more complicated if >2 levels)
- Differential misclassification bias:
  - unpredictable! – can dilute or strengthen an association, or even produce a spurious one

Potential for differential misclassification of exposure

A major concern in case-control studies (less so in cohort studies)

- Recall bias/reporting bias
  - cases may be more (or less) motivated to recall or report past exposures, more aware of family history, etc. – that is, presence of disease can influence accuracy or completeness of recall
- Observation (observer) bias
  - possible if data collector is aware of disease status (e.g., may result in greater probing of cases)

Potential for differential misclassification of outcome

A major concern in cohort studies and RCTs
(detection bias or medical surveillance bias)

- Respondent bias
  - exposed individuals may be more (or less) likely to report outcomes, seek health services, etc.
- Observation (observer) bias
  - possible if data collector is aware of exposure status (e.g., may result in greater probing of exposed persons); may be particularly problematic for more subjective, “soft” outcomes

Prevention of information bias

- Some strategies for minimizing information bias, where applicable:
  - clearly and appropriately define study variables
  - use of more objective sources of information (e.g., physical exam, medical records, lab results, death certificates, etc.) vs. self-report
  - initial and ongoing training of research staff
  - blinding (at least to study hypothesis)
  - in case-control study, use of “diseased” controls to help “equalize” recall bias

Bias: take-home messages

- Recognizing and understanding the effects of potential biases are critical when evaluating a study
- No fancy statistical techniques can salvage a study that is poorly designed and carried out