ISPE 2012 Mid-Year Meeting
Miami, FL

Introduction to Pharmacoepidemiology: Confounding and Bias

Soko Setoguchi, MD DrPH

Duke Clinical Research Institute
Duke University School of Medicine

April 22, 2012

Outline

- Safety concerns to formal epidemiologic studies
  - Challenges
- Bias (systematic error) vs. chance (random error)
- Types of Biases
  - Confounding
  - Information bias
  - Selection bias
- Summary

Safety Studies

Goal and Steps of Analytic Studies for Drug Safety

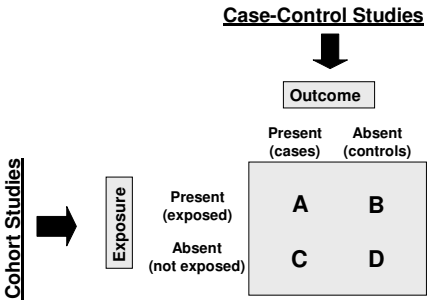
Goal

- Assess the causal association between an exposure drug and outcome of interest (adverse event)
  - Risk ratio (relative risk) or risk differences

Steps

- Identify exposed group (users of drug A) and non-exposed group (non-users of drug A or users of comparison drug or therapy)
- Follow up both groups to identify outcomes
- Calculate the risk (frequency of adverse event or incidence rate of adverse event) in two groups (exposed and unexposed groups)
- Compare the risks in two groups
  - Risk ratio (relative risk) or risk differences

The Landscape: All of Epidemiology is the Study of 2x2 Tables



From Strom BL. Medical databases in post-marketing drug surveillance. *Trends in Pharmaceutical Sciences* 1986; 7: 377.

Challenges in Observational Studies

Type of Study	Purpose	Validity
RCT (explanatory) 'Gold standard'; selected study population, unusual settings	**Efficacy** - Does it work in an ideal situation?	**Internal (no bias)**
RCT (pragmatic/large simple) Randomized ; usual setting of care; non-selected study population	**Effectiveness** - Does it work in the 'real world'?	**External (generalizability)**
Observational Analytic Studies Cohort Study, Case-Control Study, Case-Cohort Study, Case Cross Over Study		

Study design choices

Explanatory RCT
'Gold standard'; selected study population, unusual settings

Large Simple or Pragmatic Trial
Randomized ; usual setting of care; non-selected study population

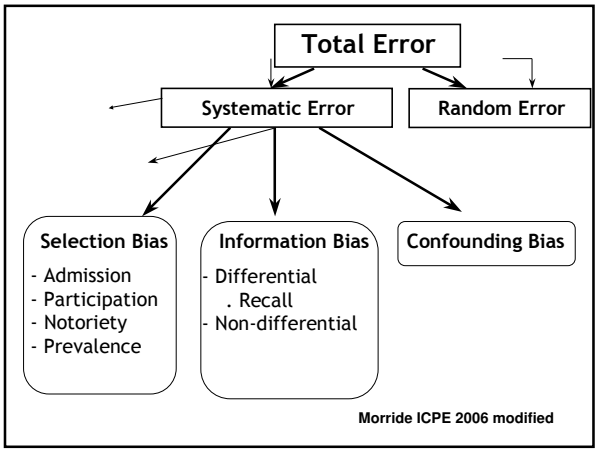
Observational Studies
Cohort Study
Case-Control Study
Case-Cohort Study
Case Cross Over Study

Efficacy
- Does it work in an ideal situation?

Effectiveness
- Does it work in the 'real world'?

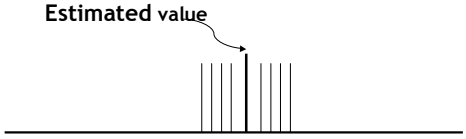
Internal ⟷ External

Bias (Systematic errors) vs. Chance (Random Error)



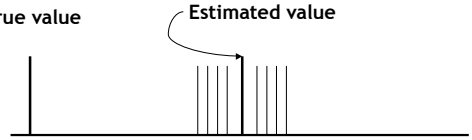
Morride ICPE 2006 modified

Random Errors (Precision)



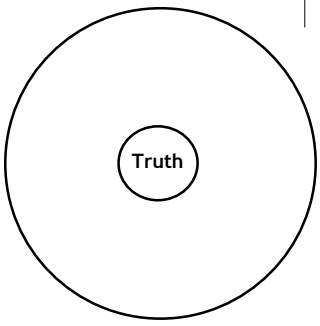
- Expressed by
  - p value
  - Confidence interval
- Precision is improved with increased sample size

Systematic Errors (Validity, internal)

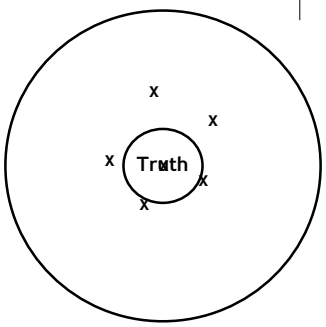


- Difference between what one estimates and the true value in the population
- *Methodological problems:*
  - Choice of participants/patients
  - Quality of collected information
  - Confounding

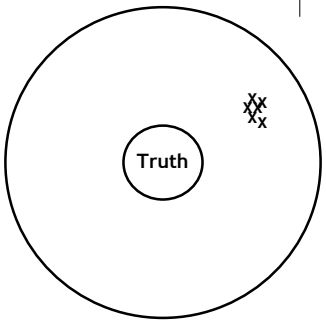
Precision vs. Validity



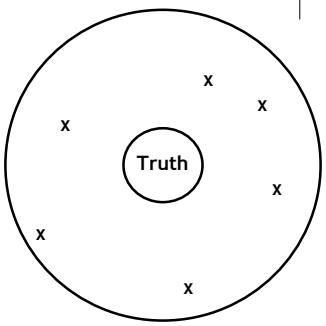
Valid but not precise



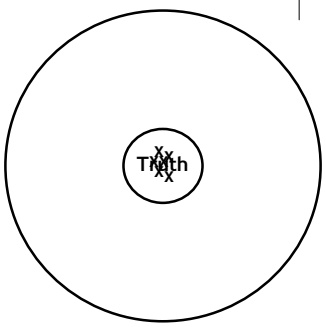
Precise but not valid



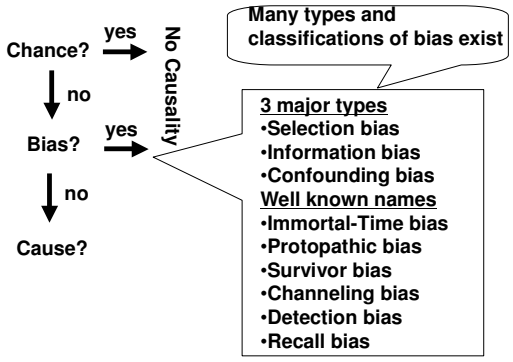
Invalid and imprecise



Valid and precise



Risk Ratio < or >1 or Risk Difference < or > 0 and Causality

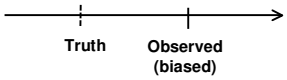


Types of Bias

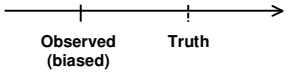
- **Confounding bias**
  - **Due to a third factor that distort association between exposure and outcome**
- **Information bias**
  - **Due to poor quality of information on exposure, outcome, confounding**
- **Selection bias**
  - **Due to choice of study population**

Describing Direction of Bias

- **Positive bias** – observed (biased) value is higher than the true value

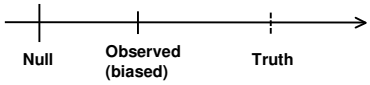


- **Negative bias** – observed (biased) value is lower than the true value

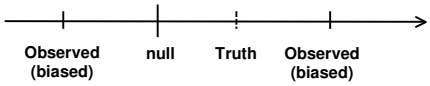


Describing Direction of Bias

- **Bias towards the null** – observed (biased) value is closer to the null than is the true value



- **Bias away from the null** – observed (biased) value is farther from the null than is the true value



- *Null: 1 for risk ratio or odds ratio, 0 for risk differences*

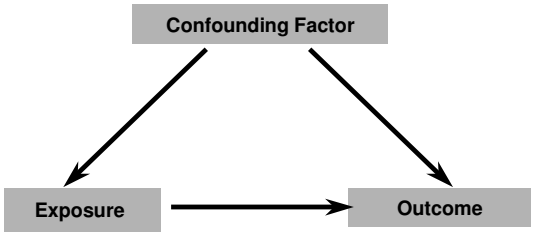
Confounding Bias

Confounding Bias

The quantitative association between exposure and outcome is distorted by a third factor with the following characteristics

1. Is a risk factor for the outcome of interest
2. Is a predictor of the exposure of interest
3. Is not an intermediate factor on the causal pathway between exposure and outcome

Confounding Bias- 'Famous' Triangle in Epidemiology



Dream vs. Real Epidemiologic Study

Dream Epidemiology = Time Machine Epidemiology

Does strontium (osteoporosis drug) cause deep vein thrombosis (DVT)?

•Of 1000 women who took strontium (osteoporosis drug), 30 developed deep vein thromboembolism (DVT) after taking strontium.

•Exposure and outcome measurements were perfect.

•All women took strontium as prescribed and there is not skipping or discontinuing the drug

•No women were lost to follow-up

•We got on Time Machine and were sent back to the past.

- This time the same 1000 women did not take strontium and everything else stayed the same.

•Only 10 developed DVT after all 1000 women went back to the past and did not take strontium

•Strontium is causally associated with 3 fold increase in the risk of DVT

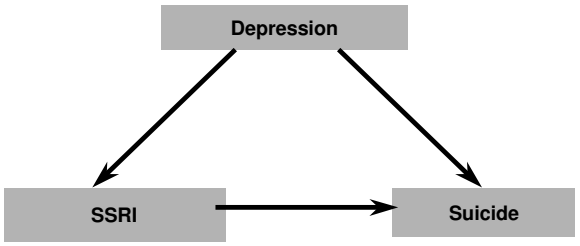
Reality in Drug Safety Studies

- No Time Machine
  - We cannot observe the same population with and without exposure
- The closest we can do is to compare two groups randomly assigned to exposure
  - They are not the same population but similar in characteristics on average as a results of randomization
- However, RCT cannot answer every possible questions on drug safety
  - Ethical concerns
  - Scarce resources
- But we can do our best to conduct high quality observational analytic studies
  - Using real-world complex data
  - Employing clever way of designing and analyzing data
  - Considering and combating all potential biases

Confounding by Indication or Severity

- Indication for an exposure drug or severity of the disease predict the use of the exposure drug
- The indication or severity is also associated with the risk of outcome of interest

Confounding by Indication



Confounding by Indication

Are ACEIs effective preventing MI in patients with hypertension?

ACEIs ← Hypertension

vs. vs.

No ACEIs ← No hypertension

Confounding by Severity

Are ACEIs effective preventing MI in patients with hypertension?

Everybody has hypertension

ACEIs ← Hypertension with Diabetes and proteinuria

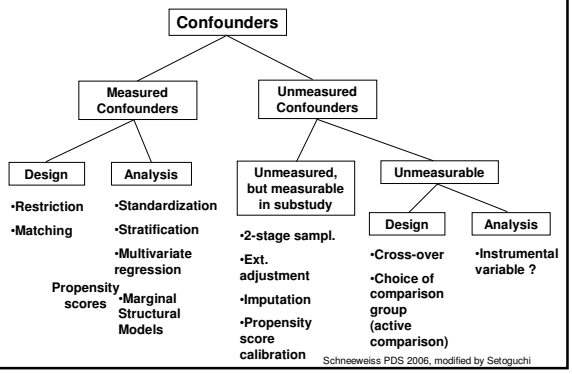
vs. vs.

No ACEIs ← Uncomplicated Hypertension

Dream Epidemiology

- Time-machine Epidemiology
  - No confounding as the exposed and non-exposed are the same persons
- Randomized trials
  - Randomization (if appropriately operationalized and the sample are large) assures the balance in the baseline characteristics of the patients

Combating Confounding in Observational Studies



Schneeweiss PDS 2006, modified by Setoguchi

Information Bias

Information or Misclassification Bias

- Measurement error or classification error on
  - Exposure status
  - Outcome status
  - Confounding

  can cause bias on the effect estimate

Non-Differential Misclassification Bias

- Degree of (or presence of) misclassification is not affected by exposure or outcome status
- Example of non-differential outcome misclassification
  - Some patients were coded as having PE because of typos in coding of diagnosis
  - Likely not associated with the use of strontium
- Example of non-differential exposure misclassification
  - Some patients forgot to take strontium during holiday seasons as they were busy with taking care of or distracted by grandchildren. (some exposure period were misclassified as a result)
  - Likely not related to the risk of developing PE
- *The bias will underestimate the risk or benefit*

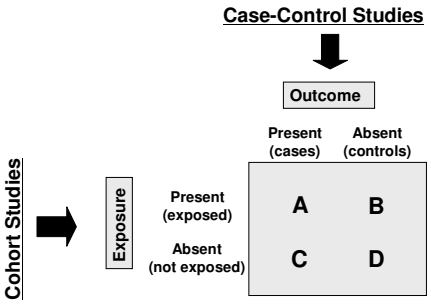
Numerical Examples of Bias Due to Non-differential Outcome Misclassification

Hypothesis: Exposure to strontium will increase the risk of DVT

Cohort Study: We followed 10,000 osteoporosis patients who were started on strontium and 10,000 similar osteoporosis patients who did not take strontium and assessed the occurrence of new DVTs. Assume that 1) all patients were followed for 5 years without lost-to-follow-up, 2) the compliance to the medication was 100%, and 3) there was not cross-over.

	Strontium		
	Exposed	Unexposed	
DVT	80	40	120
No DVT	9920	9960	19880
	10000	10000	20000

The Landscape: All of Epidemiology is the Study of 2x2 Tables



From Strom BL. Medical databases in post-marketing drug surveillance. *Trends in Pharmaceutical Sciences* 1986; 7: 377.

Analysis of Cohort Study

Outcome	Exposure: Yes	Exposure: No	
Yes	A	C	A+C
No	B	D	B+D
	A+B	C+D	N=A+B+C+D

Risk Ratio = $Risk_{exposed}/Risk_{unexposed}$ = A/(A+B) ÷ C/(C+D)

Risk Difference = $Risk_{exposed}$ - $Risk_{unexposed}$ = A/(A+B) - C/(C+D)

Non-differential Outcome Misclassification Bias

No outcome misclassification

Outcome	Exposure: Yes	Exposure: No	
Yes	80	40	120
No	9920	9960	19880
	10000	10000	20000

Risk Ratio =80/10000 / 40/10000 = 2

Non-differential Outcome Misclassification Bias

No outcome misclassification

Outcome	Exposure: Yes	Exposure: No	
Yes	80	40	120
No	9920	9960	19880
	10000	10000	20000

Risk Ratio =80/10000 / 40/10000 = 2

100% sensitivity and 50% specificity for outcome definition

	Exposed	Unexposed	
Yes	5040	5020	10060
No	4960	4980	9940
	10000	10000	20000

Risk Ratio =5040/10000 / 5020/10000 = 1.004

Non-differential Outcome Misclassification Bias

No outcome misclassification

Point 1

Non-differential misclassification of outcomes generally cause bias toward the null (direction of the bias is expected toward the null)

Risk Ratio =80/10000 / 40/10000 = 2

100% sensitivity and 50% specificity for outcome definition

	Exposed	Unexposed	
Yes	5040	5020	10060
No	4960	4980	9940
	10000	10000	20000

Risk Ratio =5040/10000 / 5020/10000 = 1.004

Non-differential Outcome Misclassification Bias

No outcome misclassification

Outcome	Exposure: Yes	Exposure: No	
Yes	80	40	120
No	9920	9960	19880
	10000	10000	20000

Risk Ratio =80/10000 / 40/10000 = 2

50% sensitivity and 50% specificity for outcome definition

	Exposed	Unexposed	
Yes	5000	5000	10000
No	5000	5000	10000
	10000	10000	20000

Risk Ratio =5000/10000 / 5000/10000 = 1

Non-differential Outcome Misclassification Bias

No misclassification (100% specificity and 100% sensitivity)

		Exposed	Unexposed	
Outcome	Yes	80	40	120
	No	9920	9960	19880
		10000	10000	20000

Risk ratio=80/10000 / 40/10000 = 2

50% sensitivity and 100% specificity

		Exposed	Unexposed	
Outcome	Yes	40	20	60
	No	9960	9980	19940
		10000	10000	20000

Risk ratio =80/10000 / 40/10000 = 2

Non-differential Outcome Misclassification Bias

No misclassification (100% specificity and 100% sensitivity)

Point 2

Incomplete sensitivity in outcome definition does not cause bias in risk ratio if specificity is 100%

50% sensitivity and 100% specificity

		Exposed	Unexposed	
Outcome	Yes	40	20	60
	No	9960	9980	19940
		10000	10000	20000

Risk ratio =80/10000 / 40/10000 = 2

Differential Misclassification

- **The degree (or presence) of misclassification differ by exposure or outcome status**

Famous example of differential exposure misclassification is 'recall bias'

- **e.g., cases tend to recall exposure status better than controls in case-control studies**
- **Direction of bias unknown (over or underestimation of the risk)**

Misclassification in Confounding Variables

- Adjustment with a binary nondifferentially misclassified confounder reduces the bias and produces a partially adjusted effect estimate that falls between the crude and true effect – residual confounding Greenland and Robins, AJE 1985
  - "Residual confounding" decreases with increasing sensitivity and specificity of the misclassified confounder Savitz and Baron, AJE 1986
  - Additional assumption – Effect of the confounder on the outcome is in the same direction among the treated and the untreated (ie, there is no qualitative interaction between the treatment and the confounder) Ogburn and VanderWeele, Epidemiology 2012
  - Assumption of no qualitative interaction between treatment and confounder will likely hold in most applications in epidemiology
- Polytomous confounding
  - Conflicting studies Fung and Howe, Int J Epi 1984
    Brenner, J Clin Epi 1993

Selection Bias

What is Selection Bias?

- **Definition**

  Distortions that result from procedures used to select subjects and from factors that influence study participation' (Modern Epidemiology)
- **Case control study**

  –Selection of cases and controls was affected by exposure status
- **Cohort study**

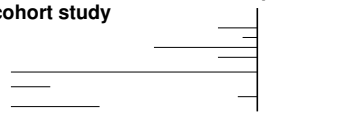
  –Selection of exposure and non-exposure group was affected by risks of outcome of interest

Potential Sources of Selection Bias

- Participant selection procedure, e.g., exposure affects case ascertainment ("detection bias") or control selection
- Differential participation due to death ('selective survival'), illness, migration, or refusal ('nonresponse bias')
- Loss to follow-up / attrition / missing data

Prevalent User Bias

When prevalent users were included in exposed group in a cohort study



Start of Follow-up

- Those who developed outcomes stopped taking the exposure drug (Depletion of susceptibles), leading to survivor bias
- Cannot identify outcomes occur soon after the initiation of the exposure drug
- Prevalent users tend to be healthy adherer, leading to compliance bias

Importance of New User Design

- All exposed persons should be new users or initiators
  - RCT is an experimental new user design
  - Follow-up starts after initiation
    - When defining exposure period, consider not only the actual use but also latency period and residual effect
- Naïve case-control study cannot control for prevalent user bias
  - Can employ new user design in nested case control studies

Summary

Combating Bias

Same as combating disease

- **Best remedy is 'prevention' rather than treatment**

- Information bias
  - Accurate, detailed drug data
  - Use validated definitions
  - Misclassification of confounding will lead to residual confounding
- Selection bias
  - Population-based data
  - New user design

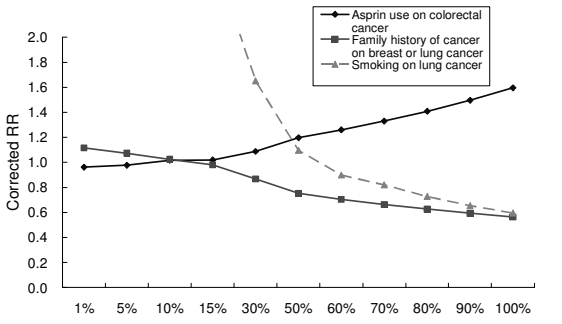
Combating Bias

- **Prevention!**
  - **Confounding bias**
    - **Consider all predictors of drug use**
      - **Utilization studies are important!**
    - **Consider all risk factors for the outcome**
      - **Literature search**
      - **Look at your data**
- **Bias is question- and data-specific**
  - **Know your data well**
  - **Ask questions in a way that you can minimize biases and are still clinically relevant and sound**

If you suspect bias while conducing your study

- **Sensitivity analysis**
- **Consider collecting additional data and/or studies**
- **Be honest about your assessment when presenting (truth vs. bias)**

Sensitivity Analysis to Assess the Impact of Unmeasured or Incompletely Measured Factors



Setoguchi et al. Circ 2007

Questions?