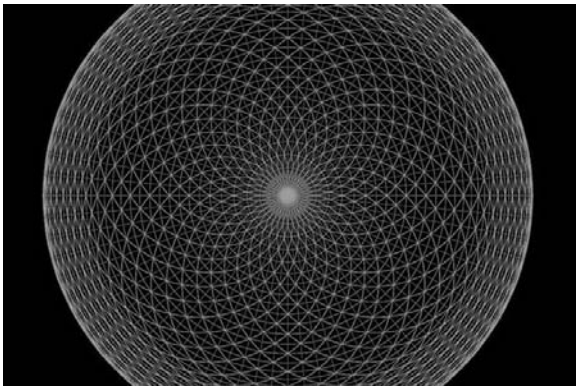
ERRORS IN EPIDEMIOLOGICAL STUDIES



Coffee and Cancer of the Pancreas

Brian MacMahon, M.D., Stella Yen, M.D., Dimitrios Trichopoulos, M.D., Kenneth Warren, M.D., and George Nardi, M.D.

N Engl J Med 1981; 304:630-633 March 1981

Coffee drinking ⟹ Pancreatic cancer

OR =2.7

With three or more cups per day was 2.7 (1.6 to 4.7).

Association

a 'statistical dependence between two or more events, characteristics, or other variables'.

Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.

The presence of an association does not necessarily imply a **causal** relationship.

Explanation for the observed difference

1. Chance (Random error)
2. Bias (Systematic error)
   - Selection
   - Information
   - Confounding
3. Effect of exposure

FRAMEWORK FOR THE INTERPRETATION OF AN EPIDEMIOLOGIC STUDY

IS THERE A VALID STATISTICAL ASSOCIATION?
- Is the association likely to be due chance?
- Is the association likely to be due bias?
- Is the association likely to be due confounding?

CAN THIS VALID STATISTICAL ASSOCIATION BE JUDGED AS CAUSE AND EFFECT?
- Is there a strong association?
- Is there biologic credibility to the hypothesis?
- Is there consistency with other studies?
- Is the time sequence compatible?
- Is there evidence of a dose-response relationship?

An important goal of

epidemiological studies is to measure accurately the occurrence of exposure/risk factors and disease outcome.

ERROR

Is defined as a false or mistaken result obtained in a study or experiment.

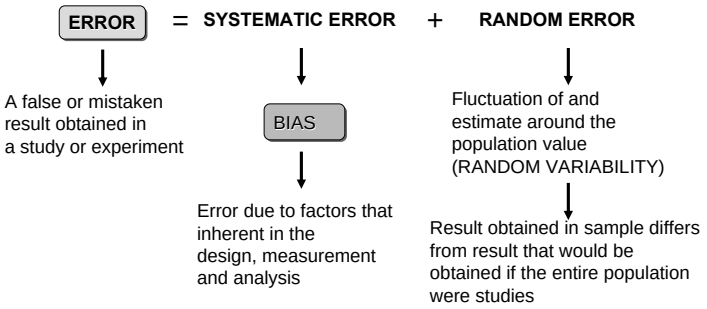
Discrepancy between measured and true effect.

ERROR

Consists of 2 components

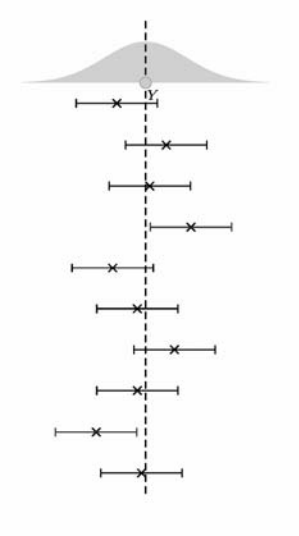
Systematic error

Random error



RANDOM ERROR

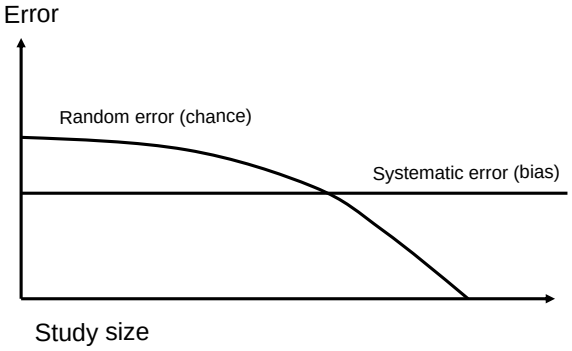
Refers to fluctuations around a true value because of Sampling variability



Random error

- An observed value that deviates from the true population value due to chance alone
- The unpredictable & uncontrollable element of an event or occurrence
- As a result → lack of precision in the measurement of an association

Errors in epidemiological studies



Source: Rothman, 2002

SYSTEMATIC ERROR

Any difference between the true value and that actually obtained that is the result of all causes other than Sampling variability.

Bias

- Bias occurs when an estimated association (RR, OR, difference in means etc.) deviates from the true measure of association
- Consequence of bias → systematic error in RR, OR etc.
- Bias may be introduced at design, implementation or analysis phase of a study

SYSTEMATIC ERROR :

- SELECTION BIAS
- INFORMATION BIAS
- CONFOUNDING

Classifying types of bias

- Selection bias – differential access to the study population
- Information bias – inaccuracy in measurement or classification
- Confounding bias – unfair comparison

VALIDITY :

A study is valid if its results corresponds to the truth, no systematic error or should be as small as possible

VALIDITY

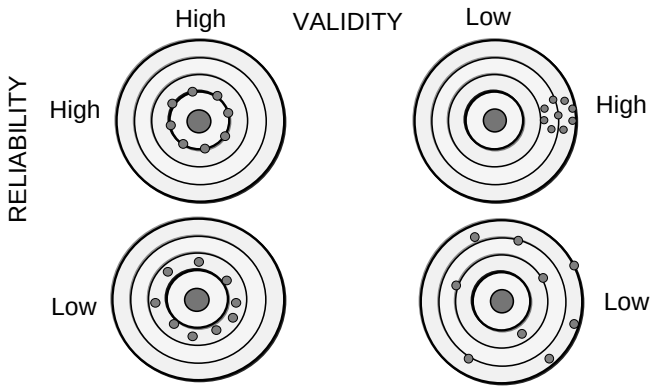
Is the expression of the degree to which a test is capable of measuring what it is intended to measure

A study is valid if its results corresponds to the truth, no systematic error and random error should be as small as possible

A high reliability means that in repeated measurements the results fall very close to each other; conversely,

A low reliability means that they are scattered.

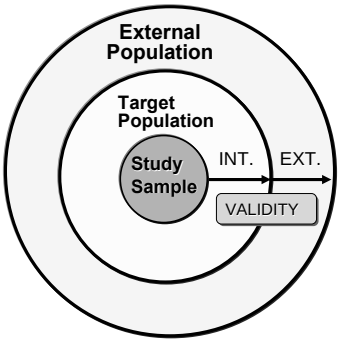
Different combinations of high and low reliability and validity



Internal validity versus external validity

- ***Internal validity***: whether the study provides an unbiased estimate of what it claims to estimate
- ***External validity***: whether the results from the study can be generalized to some other population

Internal and External Validity



SELECTION BIAS

is a distorsion in the estimate of effect resulting from the manner in which subject are selected for the study population

MAJOR SOUREC OF SELECTION BIAS

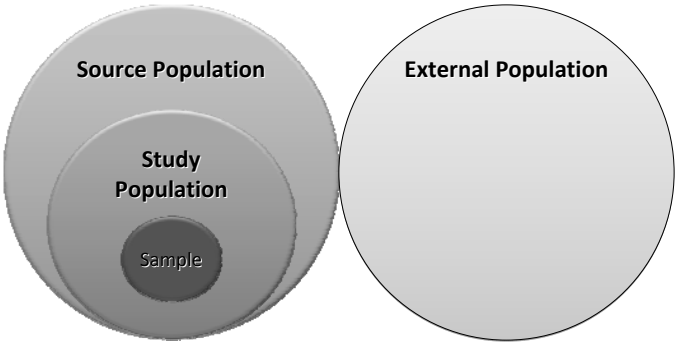
1) flaws in the choice of groups to be compared
2) choice of sampling frame
3) loss to follow up or nonresponse during data collection
4) selective survival

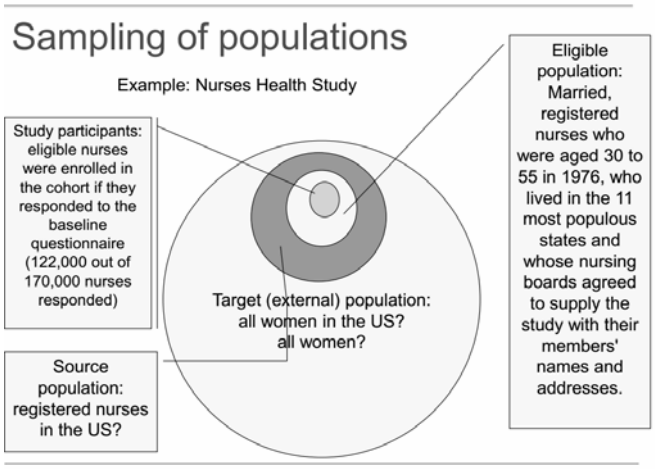
Selection Bias

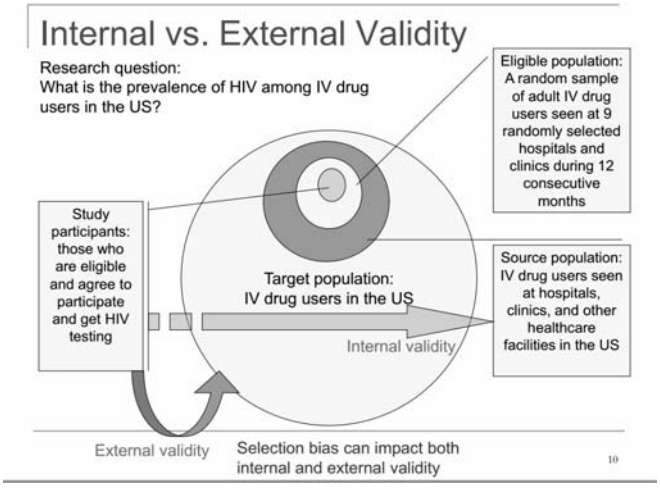
- Systematic error resulting from manner in which subjects are selected or retained in the study
- Can occur when:
  - Characteristics of subjects selected for study differ systematically from those in the target population
  - Study and comparison groups are selected from different populations

Selection Bias

- Distortions that arise from
  - Procedures used to select subjects
  - Factors that influence study participation
  - Factors that influence participant attrition
- **Systematic error** in identifying or selecting subjects
  - Examples are…







Coffee and Cancer of the Pancreas

Brian MacMahon, M.D., Stella Yen, M.D., Dimitrios Trichopoulos, M.D., Kenneth Warren, M.D., and George Nardi, M.D.

N Engl J Med 1981; 304:630-633 March 1981

Coffee drinking ⟹ Pancreatic cancer

OR =2.7

With three or more cups per day was 2.7 (1.6 to 4.7).

Selection bias in case-control studies

630 THE NEW ENGLAND JOURNAL OF MEDICINE March 12, 1981

COFFEE AND CANCER OF THE PANCREAS

BRIAN MACMAHON, M.D., STELLA YEN, M.D., DIMITRIOS TRICHOPOULOS, M.D., KENNETH WARREN, M.D., AND GEORGE NARDI, M.D.

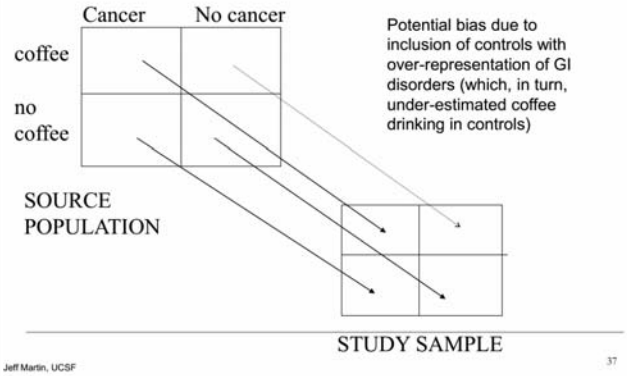
Abstract We questioned 369 patients with histologically proved cancer of the pancreas and 644 control patients about their use of tobacco, alcohol, tea, and coffee. There was a weak positive association between pancreatic cancer and cigarette smoking, but we found no association with use of cigars, pipe tobacco, alcoholic beverages, or tea. A strong association between coffee consumption and pancreatic cancer was evident in both sexes. The association was not affected by controlling for cigarette use. For the sexes combined, there was a significant dose-response relation ($P \sim 0.001$); after adjustment for cigarette smoking, the relative risk associated with drinking up to two cups of coffee per day was 1.8 (95 per cent confidence limits, 1.0 to 3.0), and that with three or more cups per day was 2.7 (1.6 to 4.7). This association should be evaluated with other data; if it reflects a causal relation between coffee drinking and pancreatic cancer, coffee use might account for a substantial proportion of the cases of this disease in the United States. (N Engl J Med. 1981; 304:630-3.)

Controls in this study were selected from a group of patients hospitalized by the same physicians who had diagnosed and hospitalized the cases' disease. The idea was to make the selection process of cases and controls similar. It was also logistically easier to get controls using this method. However, as the exposure factor was coffee drinking, it turned out that patients seen by the physicians who diagnosed pancreatic cancer often had gastrointestinal disorders and were thus advised not to drink coffee (or had chosen to reduce coffee drinking by themselves). So, this led to the selection of controls with higher prevalence of gastrointestinal disorders, and these controls had an unusually low odds of exposure (coffee intake). These in turn may have led to a spurious positive association between coffee intake and pancreatic cancer that could not be subsequently confirmed.

MacMahon et al. N Engl J Med. 1981 Mar 12;304(11):630-3

36

Case-control Study of Coffee and Pancreatic Cancer: Selection Bias



Jeff Martin, UCSF

37

Coffee and cancer of the pancreas: Use of population-based controls

•Gold et al. *Cancer* 1985

	Case	Control
Coffee: ≥1 cup day	84	82
No coffee	10	14

OR= (84/10) / (82/14) = 1.4 (95% CI, 0.55 - 3.8)

So, when population-based controls were used, there was no strong association between coffee and pancreatic cancer

Jeff Martin, UCSF

39

Bias due to selection of hospital controls

- Example:
- In a case-control study of smoking and chronic obstructive pulmonary disease (COPD), controls were selected from the same hospital with other lung diseases (e.g. tuberculosis, lung cancer, occupational lung diseases).
- The authors found a weak association between smoking and COPD
- What is the problem with this study??
  - Smoking causes many diseases resulting in higher hospitalization rate of smokers
  - Hospital controls do not represent the prevalence of exposure (smoking) in the source population from which cases of COPD arose
  - Also, hospitalized people tend to have multiple diseases, and this can result in the distortion of the exposure frequencies in hospitalized controls (Berkson's bias)

Selection Bias

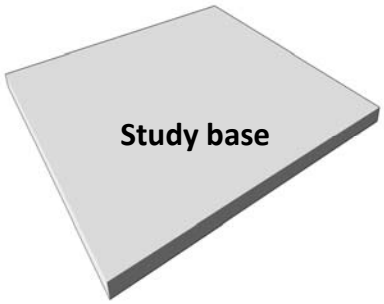
Example:

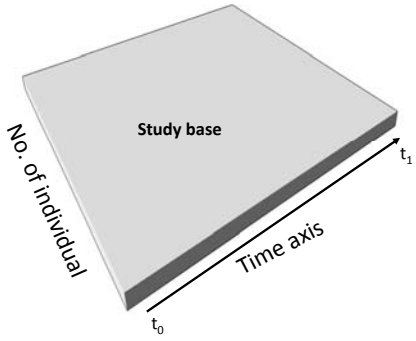
- If cases & controls or exposed & non-exposed individuals were selected in such a way that an association is observed even though exposure & disease are not associated

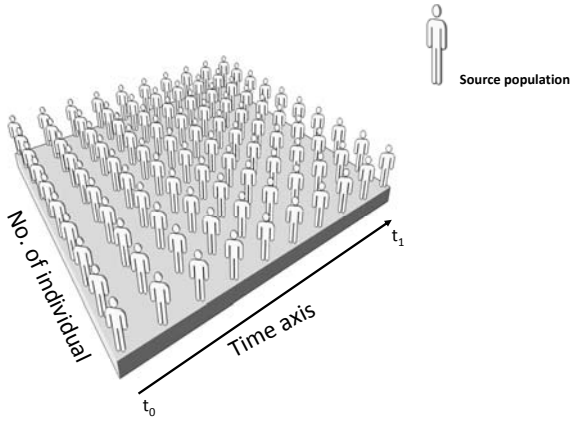
- May result from withdrawal or losses to follow-up of study subjects

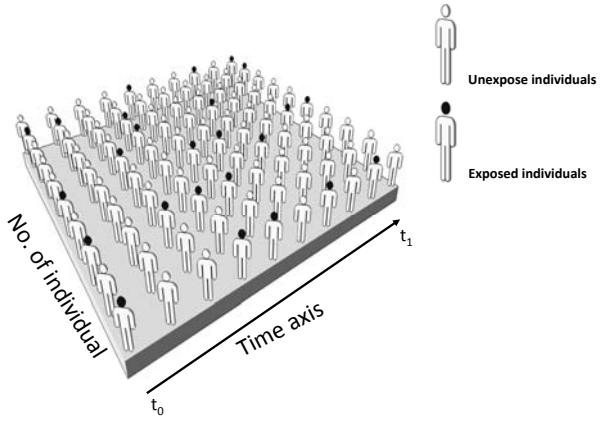
Case-Control Study

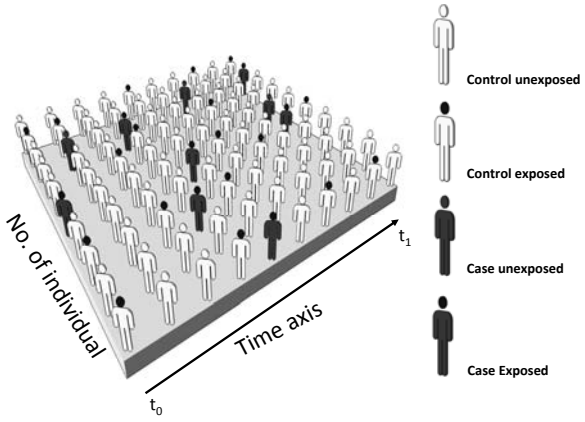
Case-control studies are prone to selection bias attributable to flawed sampling of base populations.



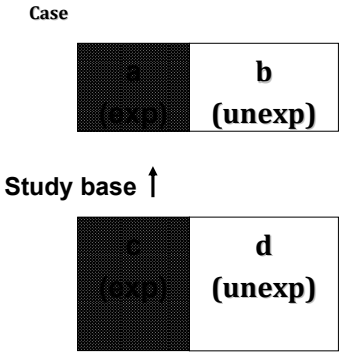




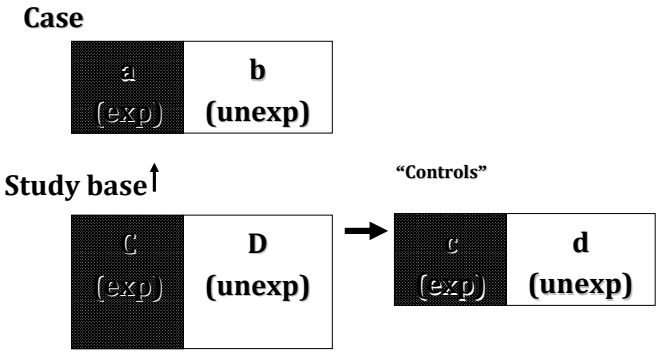


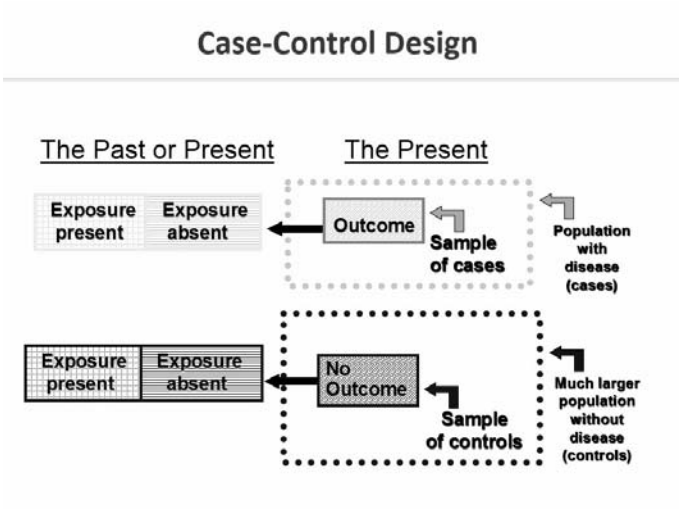


COHORT STUDY



CASE-CONTROL STUDY





"Cases and Controls should be representative of the same base experience"

The identification of the appropriate

study base (source population) from which to select controls is the primary challenge in the design of case-control studies

Case-control studies

- Case-control studies are highly vulnerable to selection bias, particularly in the control group.
- The purpose of the control group is to estimate exposure in the base population.
- Selection bias results if control selection is not neutral with respect to exposure.

Types of Selection Bias

- ***Response Bias* – those who agree to be in a study may be in some way different from those who refuse to participate**
  - **Volunteers may be different from those who are enlisted**

Selection Bias

- Problematic
  - Can result in over- or under- estimation of the true magnitude of the relationship between an exposure and an outcome
  - May produce an apparent association when none exists
    - OR/RR may be incorrect estimates ⇒ Invalid inferences about association of exposure & disease
  - May conceal a real association

Selection Bias

- To avoid it, ensure that:
  - Subjects are representative of target population
  - Study and comparison groups are similar except for variables being investigated
  - Subject losses are kept to a minimum

INFORMATION BIAS is a distortion in the measurement error or misclassification of subject on one or more variables

MAJOR SOURCES OF INFORMATION BIAS

1) invalid measurement
2) incorrect diagnostic criteria
3) omissions and imprecisions
4) other inadequacies in previously recorded data

NO MEASUREMENT IS PERFECT

The quantity intended to be measured and the measurement result inevitably differ because of measurement error.

Information bias

- **Systematic error in the measurements of information on exposure or outcome**
- Result in:
- **Differences in accuracy** of:
  - exposure data between cases and controls
  - outcome data between different exposure groups

Information bias

- Sources of information bias include:
- - Defects in the measurement instruments
- - Deficiencies in the questionnaires
- - Inaccurate diagnostic procedures
- - Ambigious definition of exposure
- - Poorly defined diagnostic criteria of disease
- - Incomplete or unreliable data sources

Information Bias

- Cause:
- Information bias arises when study variables (exposure, disease, or confounders) are inaccurately measured or classified resulting in Misclassification

Types of Information Bias

- ***Interviewer Bias* – an interviewer's knowledge may influence the structure of questions and the manner of presentation, which may influence responses**
- ***Recall Bias* – those with a particular outcome or exposure may remember events more clearly or amplify their recollections**

Types of Information Bias

- ***Observer Bias* – observers may have preconceived expectations of what they should find in an examination**
- ***Loss to follow-up* – those that are lost to follow-up or who withdraw from the study may be different from those who are followed for the entire study**

Information Bias (cont.)

- ***Hawthorne effect* – an effect first documented at a Hawthorne manufacturing plant; people act differently if they know they are being watched**
- ***Surveillance bias* – the group with the known exposure or outcome may be followed more closely or longer than the comparison group**

MISCLASSIFICATION BIAS

Definition:

the erroneous classification of an individual, a value, or an attribute into a category other than that to which it should be assigned

- often results from an improper "cutoff point" in disease diagnosis or exposure classification
- **Hence errors are made in classifying to either disease or exposure status**

MISCLASSIFICATION BIAS

- **Types of misclassification bias**
  - **Non differential (random)**
  - **Differential (systematic)**

Nondifferential Misclassification Bias

- Occurs when there is equal misclassification of exposure between diseased and non-diseased study subjects
- OR
- When there is equal misclassification of disease between exposed and non-exposed study subjects

Non-differential Misclassification Bias

- If exposure or disease is dichotomous,

  then,

  Non-differential misclassification causes a bias of the RR or OR towards the null

Nondifferential Misclassification Bias

True Classification

	Cases	Controls	Total
Exposed	**100**	**50**	**150**
Nonexposed	**50**	**50**	**100**
	150	**100**	**250**

OR = ad/bc = 2.0; RR = a/(a+b)/c/(c+d) = 1.3

Nondifferential misclassification → Overestimate exposure in 10 cases, 10 controls → bias towards null

	Cases	Controls	Total
Exposed	**110**	**60**	**170**
Nonexposed	**40**	**40**	**80**
	150	**100**	**250**

OR = ad/bc = 1.8; RR = a/(a+b)/c/(c+d) = 1.3

Non-Differential Misclassification

"True Situation"

	Cases	Controls	Total
Exp.	85	40	125
Not Exp.	15	60	75
Total	100	100	200

OR= 8.5

50% of exposed misclassified as unexposed

	Cases	Controls
Exp.	43	20
Not Exp.	15 + 42	60 + 20
Total	100	100

OR=3.0

Bias towards the null (1.0)

Differential Misclassification

- Occurs when misclassification of exposure is not equal between diseased and non-diseased study subjects
- OR
- When misclassification of disease is not equal between exposed and non-exposed study subjects

Differential Misclassification

- Causes a bias in the RR or OR
  - either towards or away from the null,
  - depending on the proportions of study subjects misclassified

Differential Misclassification

- Direction of bias is **towards the null** if
  - fewer cases are considered to be exposed or
  - fewer exposed are considered to be diseased
- Direction of bias is **away from the null** if
  - more cases are considered to be exposed or
  - more exposed are considered to be diseased

Differential Misclassification

– **Example**
- **Case-Control study**:
- If more cases are *mistakenly* classified as being exposed than controls → overestimation of OR
- **Cohort study**:
- If exposed group is more likely to be *mistakenly* classified as having developed the outcome than the unexposed group → overestimation of RR

– Leads to **over- or under- estimation** of the true magnitude of the measure of association

EXAMPLES OF MISCLASSIFICATION BIAS

- people who have disease (cases) classified as controls
  – due to inadequate description or criteria for what constitutes disease
- EXAMPLE:
  – GOAL: retrospective analysis of hypertension and stroke
  – MISCLASSIFICATION SOURCE: hypertension diagnosis
  – BIAS: in the 1960's and 1970's medical guidelines diagnosed hypertension only when diastolic pressure exceeded 100 therefore many individuals who, by today's standards were hypertensive, were "misclassified" into control groups

Differential Misclassification Bias

True Classification

	Cases	Controls	Total
Exposed	100	50	150
Nonexposed	50	50	100
	150	100	250

OR = ad/bc = 2.0; RR = a/(a+b)/c/(c+d) = 1.3

Differential misclassification - Overestimate exposure for 10 cases, inflate rates

	Cases	Controls	Total
Exposed	110	50	160
Nonexposed	40	50	90
	150	100	250

OR = ad/bc = 2.8; RR = a/(a+b)/c/(c+d) = 1.6

Differential Misclassification Bias

True Classification

	Cases	Controls	Total
Exposed	100	50	150
Nonexposed	50	50	100
	150	100	250

OR = ad/bc = 2.0; RR = a/(a+b)/c/(c+d) = 1.3

Differential misclassification - Underestimate exposure for 10 cases, deflate rates

	Cases	Controls	Total
Exposed	90	50	140
Nonexposed	60	50	110
	150	100	250

OR = ad/bc = 1.5; RR = a/(a+b)/c/(c+d) = 1.2

Differential Misclassification Bias

True Classification

	Cases	Controls	Total
Exposed	**100**	**50**	**150**
Nonexposed	**50**	**50**	**100**
	150	**100**	**250**

OR = ad/bc = 2.0; RR = a/(a+b)/c/(c+d) = 1.3

Differential misclassification - Underestimate exposure for 10 controls, inflate rates

	Cases	Controls	Total
Exposed	**100**	**40**	**140**
Nonexposed	**50**	**60**	**110**
	150	**100**	**250**

OR = ad/bc = 3.0; RR = a/(a+b)/c/(c+d) = 1.6

Differential Misclassification Bias

True Classification

	Cases	Controls	Total
Exposed	**100**	**50**	**150**
Nonexposed	**50**	**50**	**100**
	150	**100**	**250**

OR = ad/bc = 2.0; RR = a/(a+b)/c/(c+d) = 1.3

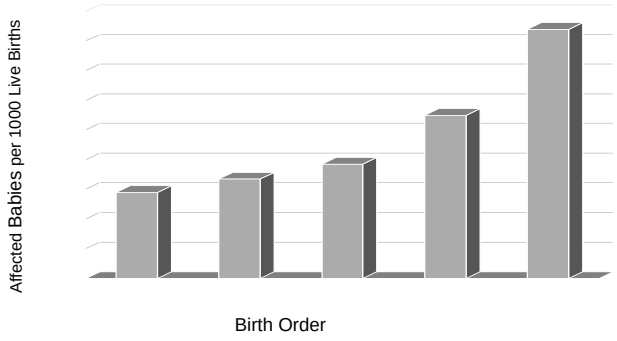
Differential misclassification - Overestimate exposure for 10 controls, deflate rates

	Cases	Controls	Total
Exposed	**100**	**60**	**160**
Nonexposed	**50**	**40**	**90**
	150	**100**	**250**

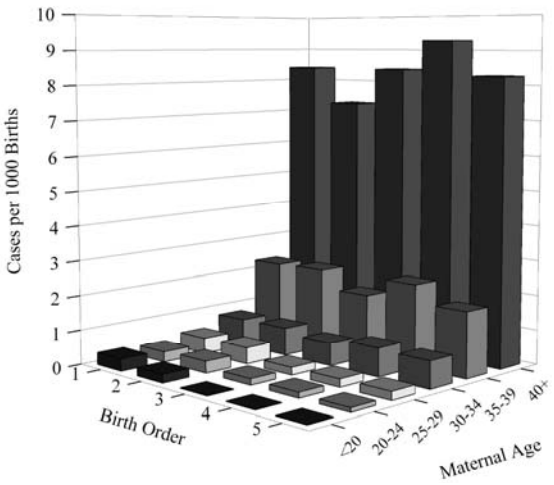
OR = ad/bc = 1.3; RR = a/(a+b)/c/(c+d) = 1.1

Confounding

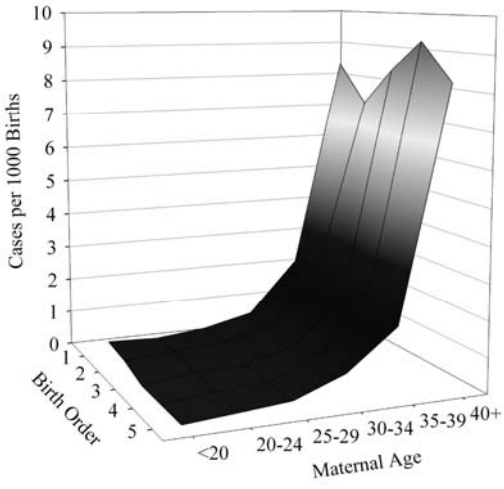
Prevalence of Down syndrome at birth by birth order



Prevalence of Down syndrome at birth by birth order and mother's age



Prevalence of Down syndrome at birth by birth order and mother's age



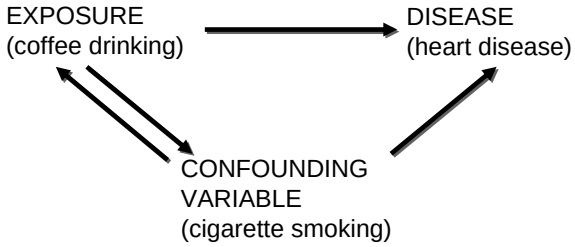
CONFOUNDING

MIXING OF EFFECTS

The estimate of the effect of the exposure of interest is distorted because it is mixed With the effect of an extraneous factor

CONFOUNDING

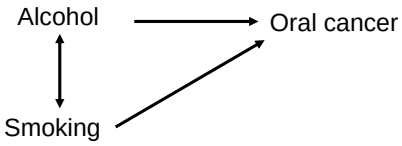
COFFEE DRINKING, CIGARETTE SMOKING AND CORONARY HEART DISEASE

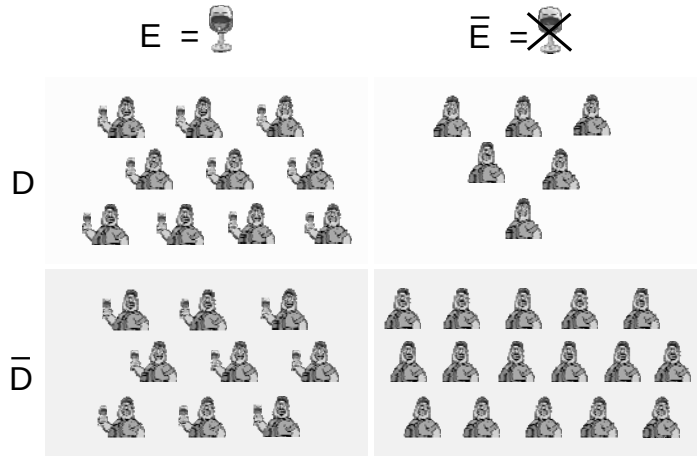


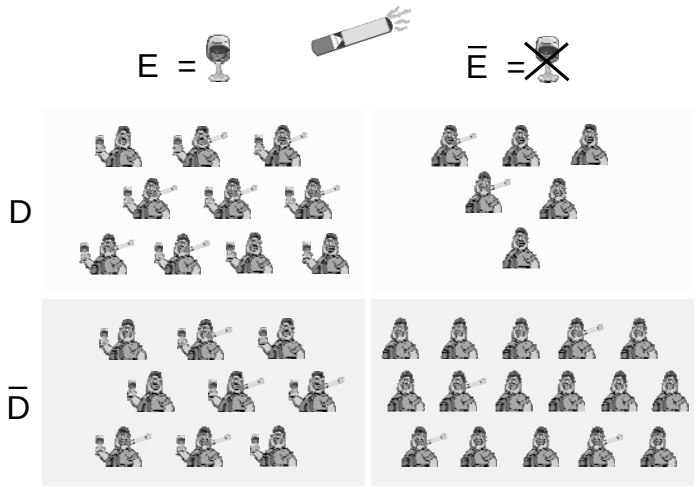
The distortion introduced by a confounding factor can lead to overestimation or under estimation of an effect depending on the direction of the association that the confounding factor has with exposure and disease.

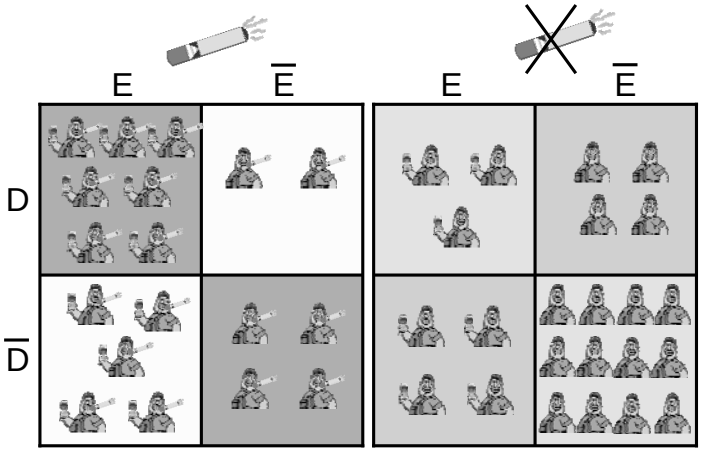
Confounding can even change the apparent direction of an effect.

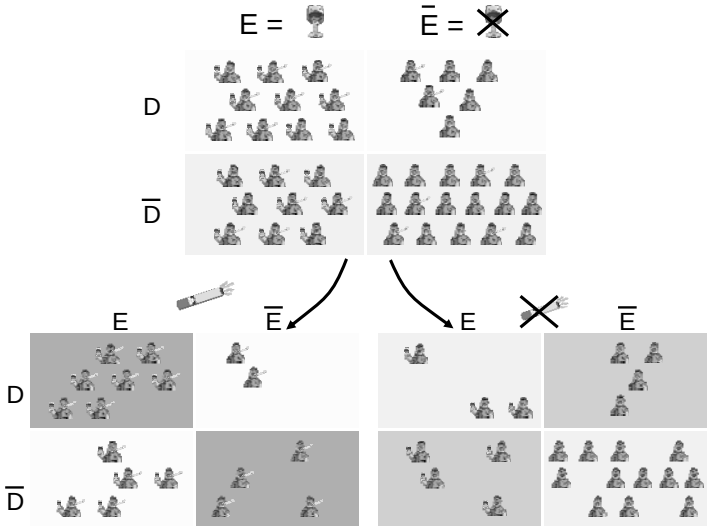
Example :



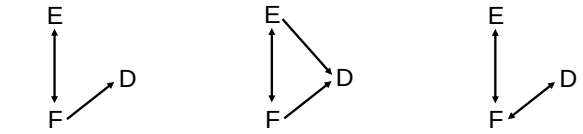




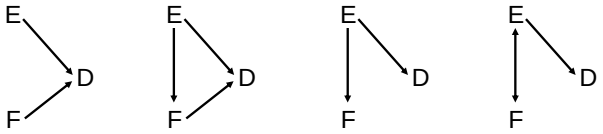




Situation in which F is a confounder for a D - E association.



Situation in which F *is not* a confounder for a D - E association.



To be confounding, the extraneous variable must have the following characteristics

A confounding variable must be a risk factor for the disease.

A confounding variable must be associated with the exposure under study (in the population from which the case derive).

A confounding variable must not be an intermediate step in the causal path between the exposure and the disease.

The data-based criterion for establishing the presence or absence of confounding involve the comparison of a <u>crude effect measure</u> with an <u>adjusted effect</u> measure that corrects for distortions due to extraneous variables.

Confounding is acknowledged to be present when the crude and adjusted effect measures differ in value.

CONTROL OF CONFOUNDING

- RESTRICTION
- MATCHING

} DESIGN

- STRATIFICATION
- MATHEMATICAL MODEL (Multivariate analysis)

} ANALYSIS

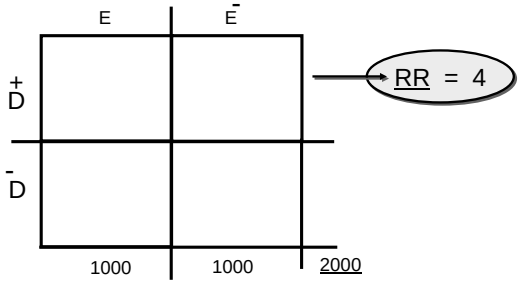
Relation of Confounder to Disease and Exposure

	DISEASE		EXPOSURE
AGE	*MI (%)	CONTROLS (%)	**OC USE (%)
25-29	3	16	29
30-34	9	14	10
35-36	16	20	8
40-44	30	21	4
45-49	42	18	3

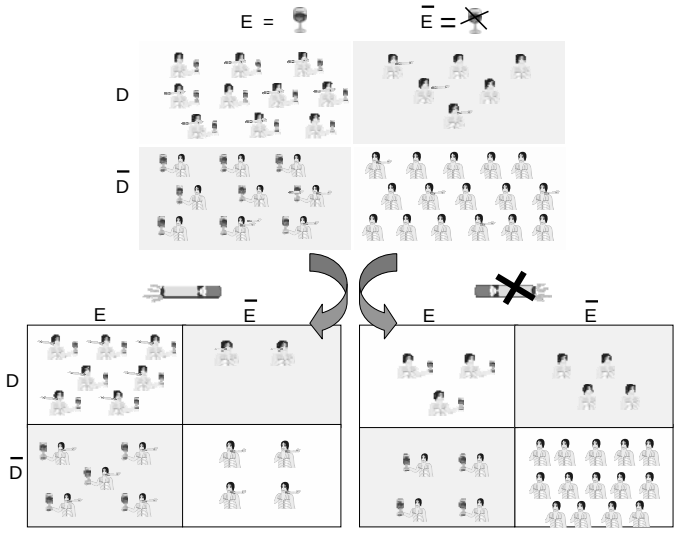
***MI : Myocardial Infarction**

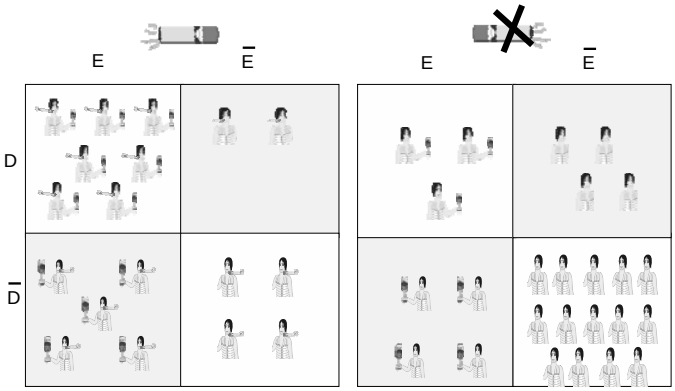
****OC : Oral Contraceptive**

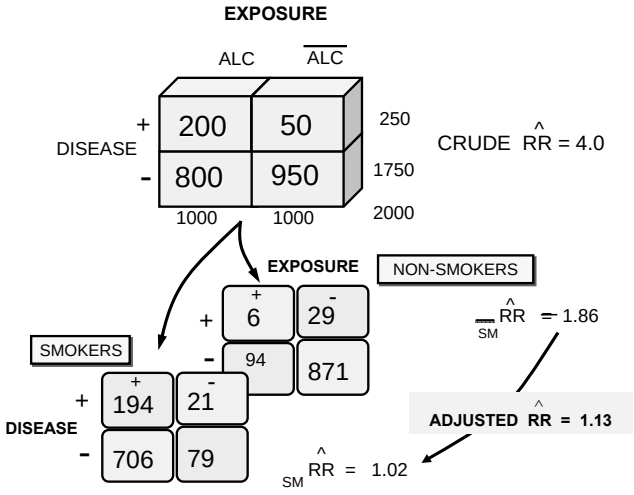
CRUDE RR

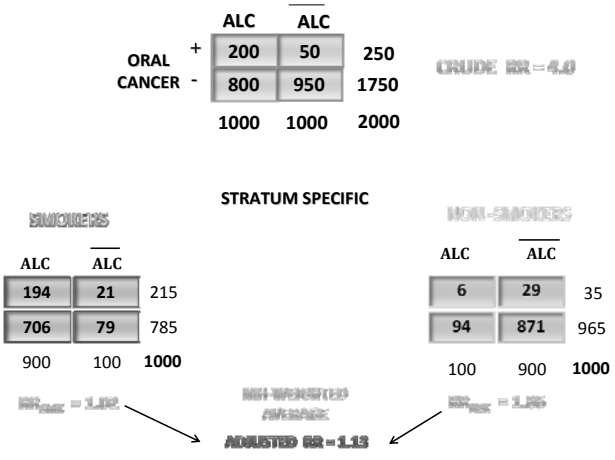


- Collapsed
- Collapsed in 1 table without separation into subgroup.









Mantel-Haenszel Estimators

Cohort Study

1. Cumulative Incidence (RISK) data

		Exposure +	Exposure −	
Disease	+	a	b	n_1
	−	c	d	n_0
Total		a + c	b+ d	N

2. Incidence Rate Data (Incidence density)

	Exposure +	Exposure −	
Disease	a	b	a + b
Person-time	L	L_1	L

Mantel-Haenszel Estimation

- Provides formula for estimating adjusted OR from case control studies
- Method generalized to estimate adjusted RR from cohort studies

Cohort Study

	E	$\bar{E}$	
D^+	a	b	m_1
D^-	c	d	m_0
	n_1	n_0	N

$$RR_{MH} = \frac{\sum_{i=1}^{k} \frac{a_i n_{0i}}{N_i}}{\sum_{i=1}^{k} \frac{b_i n_{1i}}{N_i}} = \frac{\text{Sum of disease and exposed x Non-exposed / Total}}{\text{Sum of disease and not exposed x Exposed / Total}}$$

Case-Control Study

	E	$\bar{E}$	
Case	a	b	m_1
Control	c	d	m_0
	n_1	n_0	N

$$OR_{MH} = \frac{\sum \frac{a_i d_i}{N_i}}{\sum \frac{b_i c_i}{N_i}}$$

Degree of Confounding

Measures the amount of confounding rather than mere presence or absence

degree of confounding = $\frac{\text{crude measure}}{\text{adjusted measure}}$

= $\frac{4.00}{1.13}$ = 3.53 ← over estimation

Crude = 1.68

Adjusted = 3.97

d.c. = $\frac{1.68}{3.97}$ = 0.42 ← under estimation

4 fold↗ risk of MI among recent of OC users as compared to non-users.

$\hat{a}OR(MH) = 3.97$

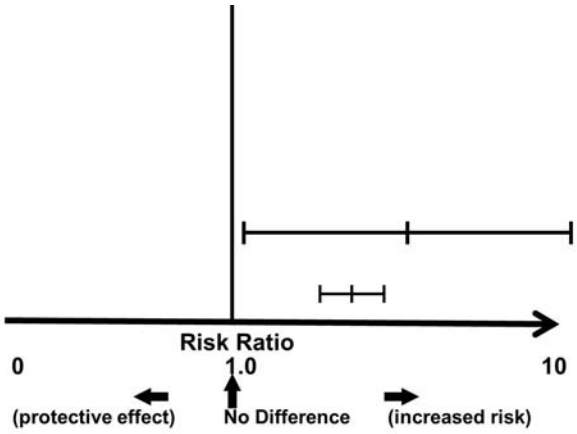
AGE	Recent Use of OC	MI	Controls	OR
25-29	Yes	4	62	7.2
	No	2	244	
30-34	Yes	9	33	8.9
	No	12	390	
35-39	Yes	4	26	1.5
	No	33	330	
40-44	Yes	6	9	3.7
	No	65	362	
45-49	Yes	6	5	3.9
	No	93	301	
TOTAL	Yes	29	135	1.7
	No	205	1607	

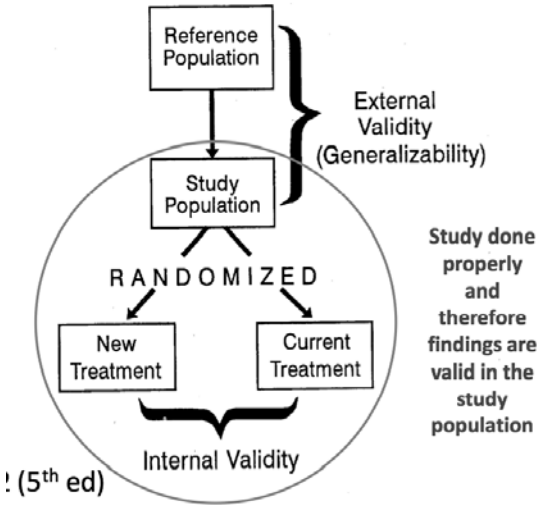
Hypothetical Examples of Unadjusted and Adjusted Relative Risks According to Type of confounding (Positive or Negative)

Example No.	Type of Confounding	Unadjusted Relative Risk	Adjusted Relative Risk
1	Positive	3.5	1.0
2	Positive	3.5	2.1
3	Positive	0.3	0.7
4	Negative	1.0	3.2
5	Negative	1.5	3.2
6	Negative	0.8	0.2
7	Qualitative	2.0	0.7
8	Qualitative	0.6	1.8

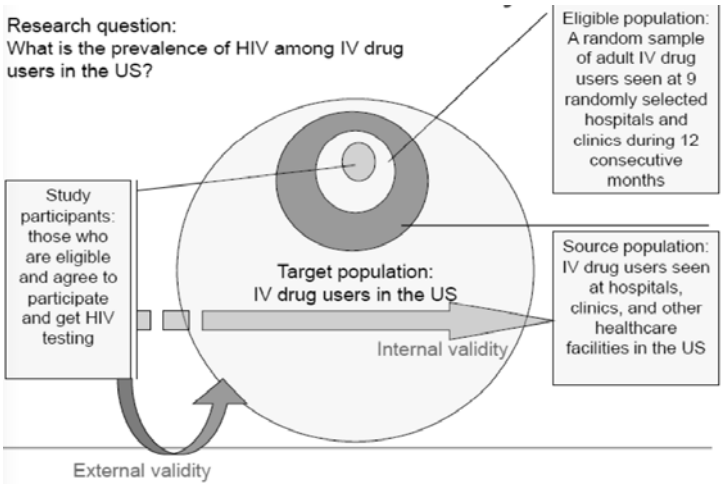
Detail from the Ramakian (Thai version of the Ramayana epic), Wat Phra Rattanasatsadaram

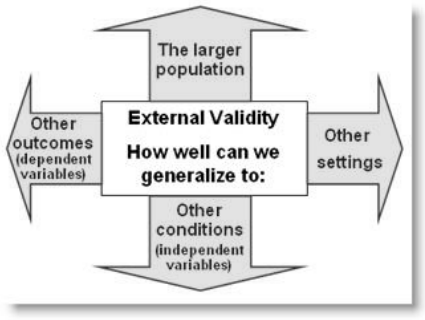
Image courtesy of Dr. C. L





(5th ed)





INTERACTION

Pratap Singhasivanon

Faculty of Tropical Medicine, Mahidol University

Interactions

The definitions

- a situation where the risk or rate of disease in the presence of 2 or more risk factors differs from the rate expected to result from their individual effects
- rate can be greater than expected
  - positive interaction or synergism
- rate can be less than expected
  - negative interaction or antagonism
- an interaction (or effect modification) is formed when a third variable modifies the relationship between an exposure and outcome

Interaction

When the incidence rate of disease in the presence of two or more risk factors differs from the incidence rate expected to result from their individual effects

Interaction

The effect can be greater than what we would expect (positive interaction) or less than we would expect (negative interaction)

Interaction (Effect Modification)

- Represents the phenomenon where the risk associated with the presence of two risk factors exceeds the risk we expect from the combination of the component risk

$$X \longrightarrow R_1$$
$$Y \longrightarrow R_2$$

X and Y > R_1 and R_2

Interaction (Miettinen 1974)

SAMPLE BASED

(Statistical Interaction)

POPULATION BASED

(Effect Modification)
(Biological Interaction)

Statistical Interaction

- Model Dependent
- Depends on deviation from statistical model (not biologic)

2 Models

Additive Model

Multiplicative Model

		Y: Absent	Y: Present
X	Absent	R_{00}	R_{01}
X	Present	R_{10}	R_{11}

$RR_{11} = R_{11} / R_{00}$
$RR_{01} = R_{01} / R_{00}$
$RR_{10} = R_{10} / R_{00}$

$R_{11} = \Pr(D/X^+Y^+)$ Probability of disease in the presence of factors **X and Y**

$R_{10} = \Pr(D/X^+Y^-)$ Probability of disease in the presence of factors **X only**

$R_{01} = \Pr(D/X^-Y^+)$ Probability of disease in the presence of factors **Y only**

$R_{00} = \Pr(D/X^-Y^-)$ Probability of disease in the **absence** of both **X and Y**

→ **Background RISK**

Additive Model

1. In term of excess **over "ONE"**

$$(RR_{11} - 1) = (RR_{\overline{10}}1) + (RR_{\overline{01}}1)$$

Stage of "No interaction" on **additive scale**

2. HOGANS

$$T = R_{11} - R_{10} - R_{01} + R_{00} = 0$$

Multiplicative Model

$$RR_{11} = RR_{10} \times RR_{01}$$

Stage of "No interaction" on **Multiplicative model**

Example :

smoke \ asbestos	+	-
+	50	10
-	5	1

ID/1000 PY

RR_{11} = 50/1 (smoking + asbestos)
RR_{10} = 10/1 (smoking alone)
RR_{10} = 5/1 (asbestos alone)

Additive Model :

(50-1) ≠ (10-1) + (5-1)

Presence of "**Interaction**" on **Additive model**

Multiplicative Model :

(50) = (10) * (5)

Absence of "Interaction" on **Multiplicative model**

Example :

	XY	X^+Y^-	X^-Y^+	X^-Y^-
D^+	**40**	**20**	**20**	**10**
D^-	**60**	**80**	**80**	**90**
	100	100	100	100

R_{11} = 40/100 = .4 **RR_{11} = .4/.1 = 4**
R_{10} = R_{01} = 20/100 = .2 **RR_{10} = .2/.1 = 2**
R_{00} = 10/100 = .1 **RR_{01} = .2/.1 = 2**

Multiplicative Model :

$RR_{11} = RR_{10} * RR_{01}$
$4 = 2 * 2$

No interaction on Multiplicative Model

Additive Model :

$(RR_{11}-1) \neq (RR_{10}-1)+(RR_{01}-1)$
$3 \neq 1 + 1$

$T = R_{11}-R_{10}-R_{01}+R_{00} = 0$
$= .40 - .20 - .20 + .10 = .10$

There is evidence of interaction on Additive Model

NOTES :

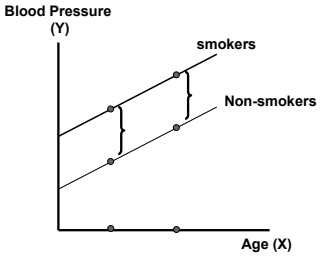
1. Neither model is right or wrong. They are simply devices for modeling data and may be more or less suitable for a particular application.
2. Most statistical techniques are based on multiplicative model.

The presence or absence of **interaction** pertains to whether or not a particular effect measure (RR, OR) varies in value over categories or strata based on level of some factor(s).

Equivalent to an assessment regarding interaction based on **multiplicative model**

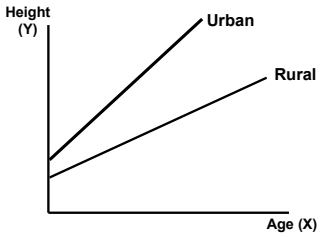
Which of the 2 models we should use :

1. For addressing public health concerns regarding disease frequency reduction, deviation from additivity appears to be most relevant
2. Contribution to the understanding of disease etiology → multiplicative model



Additive Model (No interaction)

Only change in intercepts no change in slope irrespective of the value of Xi which is being held constant



Interactive Model

There is change in both intercepts and slope as the level of Xi which is held constant and varied

$$\chi^2_{\text{TOTAL}} = \sum W_g \left[\ln(\text{OR})\right]^2 = 67.404$$

$$\chi^2_{\text{ASSO.}} = (\sum W_g)\hat{\theta}^2 = (49.290)(1.0898)^2 = 58.54$$

$$\hat{\theta}^2 = \frac{\sum W_g \ln(\text{OR})}{\sum W_g} = \frac{53.717}{49.290} = 1.0898$$

$$\chi^2_{\text{HOMO.}} = 67.404 - 58.54 = 8.86$$ **P < 0.005**

Conclude that the non uniformity of the observed OR's is unlikely to have occurred by chance; thus there is some evidence of interaction.

males

	VAC	$\overline{\text{VAC}}$
D	10	14
$\overline{\text{D}}$	191	15
	201	29

females

	VAC	$\overline{\text{VAC}}$
D	22	12
$\overline{\text{D}}$	155	17
	177	29

$\hat{P}$	0.05 (P_1) .483 (P_2)	.124 (P_1) .414 (P_2)
$\hat{RD}$	.483-0.05 = .433	.144-.124 = .290

	Males	Females	Total
1. $\hat{R}D$	0.433	0.290	
2. $\text{Var}(\hat{R}D_{..})$ $(p_1q_1/n_1 + p_1q_2/n_2)$	0.08847	0.08979	
3. $W_g = (1/\text{Var}(\hat{R}D_g)$	113.03	111.37	224.4
4. $W_g(\hat{R}D_g)$	48.94	32.30	81.24
5. $W_g(RD_g)^2$	21.19	9.37	30.56

$\chi^2_{TOTAL} = 30.56$ $(\hat{\theta})\hat{R}D = 81.24/224.4 = .362$

$\chi^2_{ASS} = 224.4(.362)^2 = 29.41$

$\chi^2_{HOMO} = 30.56 - 29.41 = 1.15$

Relative risk of oral cancer according to presence or absence or two exposures : **smoking and alcohol consumption**

		smoking	
		No	**Yes**
alcohol	**No**	**1.00**	**1.53**
	Yes	**1.23**	**5.71**

Relative risk of liver cancer for persons exposed to **Aflatoxin** and/or **Chronic Hepatitis B infection** : An example of interaction

		Aflatoxin	
		Negative	**Positive**
HBs Ag	**Negative**	**1.00**	**3.4**
	Positive	**7.3**	**59.4**

Deaths from lung cancer (per 100,000) among individuals with and without exposure to **cigarette smoking** and **asbestos**

	Asbestos Exposure	
Cigarette smoking	No	Yes
No	**11.3**	**58.4**
Yes	**122.6**	**601.6**

Age-Adjusted Odds Ratios Estimated from Logistic Models with and without an Interaction between SMOKING and ORAL CONTRACEPTIVE USE

	No interaction Model		**Interaction Model**	
	OC Use		**OC use**	
Cig/day	**No**	**Yes**	**No**	**Yes**
None	1.0	3.3 (2.0, 5.5)	1.0	3.6 (1.2, 11.1)
1 - 24	3.1 (2.0, 4.6)	10.1 (5.2, 19.5)	3.3 (2.2, 5.1)	3.7 (1.04, 13.0)
≥ 25	8.5 (5.6, 12.8)	27.8 (14.4, 53.5)	8.0 (5.2, 12.4)	40.4 (19.4, 84.1)

Conceptual Framework of the definition of interaction based on comparing expected and observed joint effects

A. When there is no interaction, the joint effect of risk factors A and Z equals the sum of their independent effects :

A + **Z** =

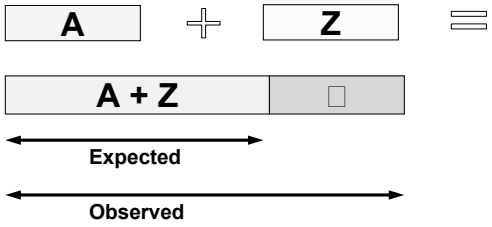
A + Z

Expected

Observed

Conceptual Framework of the definition of interaction based on comparing expected and observed joint effects

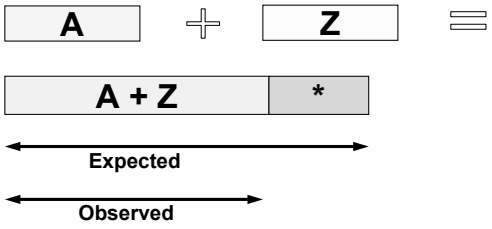
B. When there is positive interaction (synergism). The observed joint effect of risk factors A and Z is greater than that expected on the basis of summing the independent effects of A and Z :



□ Excess due to positive interaction

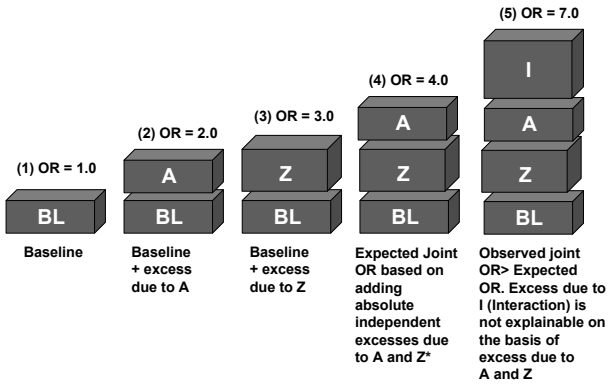
Conceptual Framework of the definition of interaction based on comparing expected and observed joint effects

C. When there is negative interaction (antagonism), the observed joint effect of risk factors A and Z is smaller than that expected on the basis of summing the independent effects of A and Z :

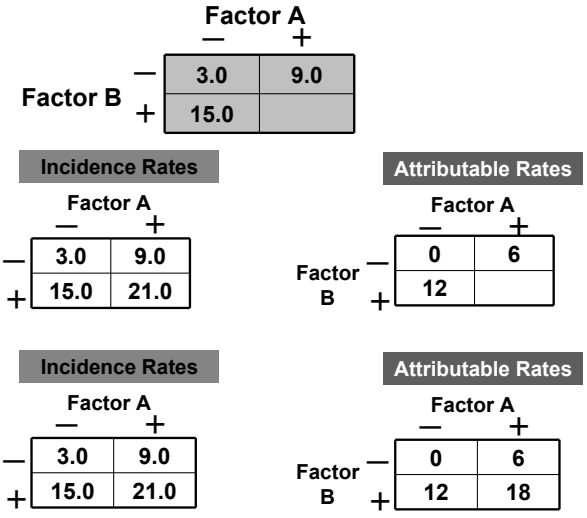


* "Deficit" due to negative interaction

Schematic representation of the meaning of the formula, Expected OR_{A+Z+}=Observed OR_{A+Z-}+Observed OR_{A-Z+}-1.0.



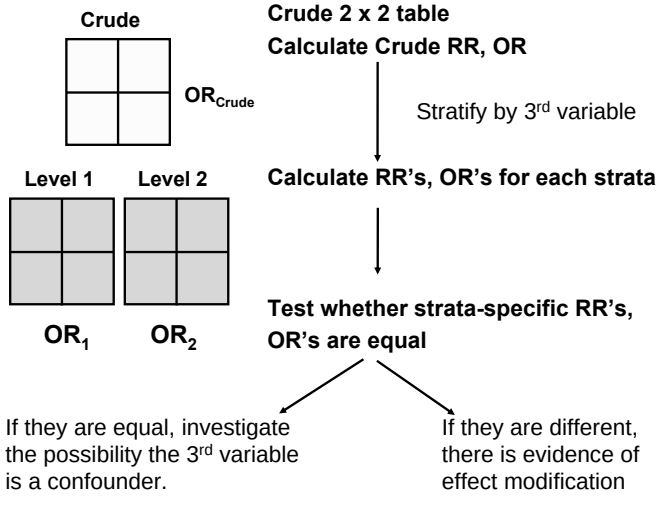
* Note that when the independent relative odds for A and Z are added, the baseline is added twice; thus, it is necessary to subtract 1.0 from the joint expected OR: that is, Expected OR_{A+Z+}=(Excess due to A + baseline) + (Excess due to Z + baseline) – baseline = OR_{A+Z-} + OR_{A-Z+} - 1.0.



INTERACTIONS

The definition…

- a situation where the risk or rate of disease in the presence of 2 or more risk factors differs from the rate expected to result from their individual effects
- rate can be greater than expected
  - positive interaction or synergism
- rate can be less than expected
  - negative interaction or antagonism
- an interaction (or effect modification) is formed when a third variable modifies the relationship between an exposure and outcome



IDENTIFYING AN INTERACTION - an example

1) Calculate crude measure of association .....

	MI	No MI	Total
Smokers	42	158	200
Non smokers	21	175	196
Total	63	333	396

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

OR = 2.22 (1.26, 3.91)

2. Calculate stratum-specific measures of association...

STRATUM 1: Dietary fat consumption **< 30%** of calories

	MI	No MI	Total
Smokers	42	158	200
Nonsmokers	21	175	196
Total	63	333	396

OR = 2.22 (1.26, 3.91)

STRATUM 2: Dietary fat consumption **> 30%** of calories

	MI	No MI	Total
Smokers	42	158	200
Nonsmokers	21	175	196
Total	63	333	396

OR = 6.29 (2.64, 14.75)

THIRD VARIABLE SUMMARY