Study types & Measures of association

Oscar Millán Iturbe R1MI
Epidemiological methods map

- Study types
- Measures of frequency
- Population data sources
- Measures of association
- Confounding / effect modification
- Selection bias
- Measurement error
- Systematic reviews
- Causal inference

CRITICAL APPRAISAL

Screening and test evaluation
Infectious disorder outbreak
Measures of association

How big, or how strong is the association between the study factor (exposure) and the outcome factor (disease)?

- relative risk
- risk difference
- odds ratio
- attributable fraction
- population attributable fraction
STUDY TYPES
# Asking questions

<table>
<thead>
<tr>
<th>1. Australian government needs to provide figures to WHO on the incidence of HIV in Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Anna’s mum is worried about Anna using her mobile phone so much – she’s heard they’re not safe</td>
</tr>
<tr>
<td>3. Mrs Smith’s GP is wondering whether acupuncture might help Mrs Smith’s shoulder pain</td>
</tr>
<tr>
<td>4. Medical Services Advisory Board is considering whether to offer a Medicare rebate for Magnetic Resonance Imaging for investigation of joint problems</td>
</tr>
</tbody>
</table>
What is the research question?

What is the main issue?

- How common is it? [Frequency]
- What caused it? [Aetiology]
- Does it work? [Intervention]
- How accurate is this test? [Diagnosis or Test evaluation]
What is the research question?

What is the main issue?

- How common is it? [Frequency]...HIV
- What caused it? [Aetiology]...Mobile phone
- Does it work? [Intervention]...Acupuncture
- How accurate is this test? [Diagnosis or Test evaluation]...Magnetic resonance imaging
<table>
<thead>
<tr>
<th>Question/issue</th>
<th>Study designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Randomised controlled trial</td>
</tr>
<tr>
<td></td>
<td>Cohort study (non-randomised trial)</td>
</tr>
<tr>
<td></td>
<td>Case-control study</td>
</tr>
<tr>
<td></td>
<td>Case series</td>
</tr>
<tr>
<td></td>
<td>Ecological study</td>
</tr>
<tr>
<td></td>
<td>Before and after study</td>
</tr>
<tr>
<td>Aetiology</td>
<td>Randomised controlled trial</td>
</tr>
<tr>
<td></td>
<td>Cohort study</td>
</tr>
<tr>
<td></td>
<td>Case-control study</td>
</tr>
<tr>
<td></td>
<td>Case series</td>
</tr>
<tr>
<td></td>
<td>Ecological study</td>
</tr>
<tr>
<td></td>
<td>Before and after study</td>
</tr>
<tr>
<td>Diagnosis (one test against reference standard)</td>
<td>Cross-sectional analysis</td>
</tr>
<tr>
<td>Diagnosis (comparison of two tests)</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td></td>
<td>Cohort study</td>
</tr>
<tr>
<td></td>
<td>Case-control study</td>
</tr>
<tr>
<td>Frequency of occurrence</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>- cross sectional (eg survey)</td>
</tr>
<tr>
<td></td>
<td>- cohort</td>
</tr>
</tbody>
</table>
How frequently does it occur?

Descriptive study: cross-sectional study / survey
Descriptive Studies

Population of Interest

Representative Sample

Measure Characteristic(s) of Interest
Descriptive Studies

Used for answering questions about frequency

Usually cross-sectional studies
May be cohort studies

Examples
• Prevalence survey of TB
• Prevalence of domestic violence
• Incidence of playground injury in a school population over one year
Causal relationships

Study factor (exposure) → Outcome factor
Analytic studies

Analytic studies are used to work out the answers to questions about cause and effect:
- “What caused this?”
- “Does this intervention work?”
- “Is this test better than another test?”

Analytic studies allow you to analyse the relationship between two factors:
- exposure and outcome
- treatment and outcome
- test and disease
Analytic studies

Used for answering questions of aetiology or the effect of intervention (causality):

- Randomised controlled trial
- Cohort study
- Case control study
- Cross sectional analytic study
- Before and after study
- Ecological study
Analytic studies - RCT

Chance alone determines exposure to study factor

Outcome is measured later

MAJOR STRENGTH:
There is no difference between the exposed and unexposed group EXCEPT for the study factor
Randomised controlled trial

Study population

Randomly allocated to exposure

Randomly allocated to no exposure

Measure outcome
Analytic studies - RCT

Randomised control trial of efficacy of midwife-managed care

Lancet 1996;348:213-218

1299 pregnant women

randomised

648 midwife managed care

651 shared care

Measured obstetric outcomes and satisfaction with care
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Shared care (n=635)</th>
<th>Midwife care (n=643)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>25.5 yrs</td>
<td>25.8 yrs</td>
</tr>
<tr>
<td>Current smokers</td>
<td>38.6%</td>
<td>37.9%</td>
</tr>
<tr>
<td>Married</td>
<td>54.8%</td>
<td>53.6%</td>
</tr>
<tr>
<td>1st baby</td>
<td>53.5%</td>
<td>54.7%</td>
</tr>
<tr>
<td>Subsequent baby</td>
<td>46.5%</td>
<td>45.3%</td>
</tr>
<tr>
<td>Star sign = Gemini</td>
<td>8.1%</td>
<td>8.3%</td>
</tr>
</tbody>
</table>
Analytic studies - cohort

Exposure to study factor determined by subjects

Investigators measure the extent of exposure

Outcome is measured LATER
Cohort study

Study population

- Chose / allocated to exposure
  - Measure outcome

- Chose / allocated to no exposure
  - Measure outcome
Analytic studies - cohort

NURSES’ HEALTH STUDY

US NURSES AGED
34 - 59 YEARS

n = 89,538

MEASURE DIETARY FAT INTAKE

very low
low
median
high
very high

4 years follow-up

MEASURE OUTCOME - BREAST CANCER
Analytic studies - case control

Subjects selected on presence [cases] or absence [controls] of the outcome factor

Then exposure factor[s] is measured in cases and controls [ie after outcome is known].

Relative frequency of the exposure in cases and controls is compared.
Case-control study

Study base

Case

Measure exposure

Control

Measure exposure
What is a study base?

- Defined place
- Defined time
- Defined group

- E.g. All children under 15 living in Brisbane between 1 Jan 1990 and 31 Dec 2003
Is the study base well defined?

Is the study base poorly defined?
Analytic studies - Case Control

STUDY BASE
WOMEN AGED 20 - 70 YEARS
LIVING IN SOUTH AUSTRALIA 1982 - 1984

CASES
WOMEN WITH NEWLY DIAGOSED BREAST CANCER
N = 451

CONTROLS
RANDOM SAMPLE OF WOMEN WITHOUT BREAST CANCER
N = 451

EXPOSURE TO DIETARY FAT
Analytic Studies: cross sectional analytic

Subjects selected because they were present at the time of the study.

Selection is NOT on the basis of either exposure or outcome.

Example:
• Does being overweight cause arthritis?

This could be examined in a study in which both weight and arthritis symptoms are measured at the same time
Cross-sectional study

Study population

Measure exposure and outcome

Measure exposure and outcome
Ecological studies

The group is the unit of analysis

Individual exposure and outcome are not considered
Dental caries per 100 children by Fluoride content of water supply (parts per million)
Pub Health Rep 57;1155-1179, 1942

Fluoride content of public water (parts per million)
Case-control versus cohort studies

In a COHORT STUDY of smoking (exposure) and lung cancer (outcome):

• begin with a group of smokers (exposed to study factor) and a group of non-smokers (not exposed to study factor; “controls”)

• follow them forward in time to see who develops lung cancer (outcome factor)
Case-control versus cohort studies

In a CASE-CONTROL STUDY of smoking and lung cancer

- begin with a group in whom outcome is known: eg lung cancer patients (cases) and a group of people without lung cancer (controls)

- assess their past history of smoking (exposure factor)
Case-control versus cohort studies

Cohort study:
• “Control” = person without the exposure

Case-control study:
• “Control” = person without the outcome

• “Control” = “Referent”
Analytic studies – test evaluation

For answering questions of diagnosis (test evaluation):

Evaluating one test against a reference standard:
• Cross sectional analytic study

Comparing two tests:
• Randomised controlled trial
• Cohort study
• Case control study
Questions about diagnosis:
How accurate is the test?

<table>
<thead>
<tr>
<th></th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease present</strong></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>True positive</td>
</tr>
<tr>
<td>Test negative</td>
<td>False negative</td>
</tr>
<tr>
<td><strong>Disease absent</strong></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>False positive</td>
</tr>
<tr>
<td>Test negative</td>
<td>True negative</td>
</tr>
</tbody>
</table>
### Questions about diagnosis: How accurate is the test?

<table>
<thead>
<tr>
<th>Blood alcohol concentration (=&quot;Truth&quot;)</th>
<th>&gt;0.1 mg/dl</th>
<th>&lt;0.1 mg/dl</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intoxicated</td>
<td>444</td>
<td>85</td>
<td>529</td>
</tr>
<tr>
<td>Sober</td>
<td>45</td>
<td>762</td>
<td>807</td>
</tr>
<tr>
<td>Total</td>
<td>489</td>
<td>847</td>
<td>1,336</td>
</tr>
</tbody>
</table>
Systematic reviews and meta-analysis

Examines results from different studies

Meta-analysis is a statistical method of combining results from multiple studies
MEASURES OF ASSOCIATION
Incidence

CUMULATIVE INCIDENCE

\[ I = \frac{\text{# of individuals experiencing a NEW event during a time period}}{\text{# of susceptible individuals at the beginning of the time period}} \]

Incidence is a measure of events (event rate)

Incidence is a measure of risk
Relative Risk
(risk ratio)

Incidence in the exposed group
Incidence in the control group
How large is the effect: relative risk
How large is the effect: relative risk

- Exposed
- Unexposed
Pregnant women

Exposed
1,000

Not exposed
4,000

OUTCOME: Birth defects
20 / 1,000
25 / 4,000
Cohort study of effect of pesticide on risk of birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>20</td>
<td></td>
<td>1,000</td>
</tr>
<tr>
<td>Unexposed</td>
<td>25</td>
<td></td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td></td>
<td>5,000</td>
</tr>
</tbody>
</table>
Cohort study of effect of pesticide on risk of birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>20</td>
<td></td>
<td>1,000</td>
</tr>
<tr>
<td>Unexposed</td>
<td>25</td>
<td></td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td></td>
<td>5,000</td>
</tr>
</tbody>
</table>

Incidence in exposed group: \( \frac{20}{1,000} = 0.02 = 2\% \)

Incidence in unexposed group: \( \frac{25}{4,000} = 0.00625 = 0.625\% \)

Relative risk = \( \frac{2\%}{0.625\%} \)

= 3.2
Trial of folate to reduce birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>343</td>
<td></td>
<td>10,335</td>
</tr>
<tr>
<td>Unexposed</td>
<td>398</td>
<td></td>
<td>10,320</td>
</tr>
<tr>
<td>Total</td>
<td>741</td>
<td></td>
<td>20,655</td>
</tr>
</tbody>
</table>

Incidence in exposed group: \( \frac{343}{10,335} = 0.033 = 3.32\% \)

Incidence in unexposed group: \( \frac{398}{10,320} = 0.039 = 3.86\% \)

Relative risk \( = \frac{3.3\%}{3.9\%} = 0.86 \)
Trial of folate to reduce birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>343</td>
<td></td>
<td>10,335</td>
</tr>
<tr>
<td>Unexposed</td>
<td>398</td>
<td></td>
<td>10,320</td>
</tr>
<tr>
<td>Total</td>
<td>741</td>
<td></td>
<td>20,655</td>
</tr>
</tbody>
</table>

Relative risk  = 0.86  
Relative risk reduction  = 0.14  
RRR = (1-RR)
Risk Difference (Absolute Risk Increase)

- Risk in exposed group minus risk in the unexposed group
- How many extra outcomes are due to the exposure?

\[ \text{Risk Difference} = \left| I_{\text{exposed}} - I_{\text{unexposed}} \right] \]
Cohort study of effect of pesticide on risk of birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>20</td>
<td>980</td>
<td>1,000</td>
</tr>
<tr>
<td>Unexposed</td>
<td>25</td>
<td>3,975</td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>4,955</td>
<td>5,000</td>
</tr>
</tbody>
</table>

Risk Difference = 2% - 0.625% = 1.375%
Trial of folate to reduce birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>343</td>
<td>9,992</td>
<td>10,335</td>
</tr>
<tr>
<td>Unexposed</td>
<td>398</td>
<td>9,922</td>
<td>10,320</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>741</strong></td>
<td><strong>19,914</strong></td>
<td><strong>20,655</strong></td>
</tr>
</tbody>
</table>

Risk Difference = $I_{UE} - I_E$

= 3.9% - 3.3%

= 0.6%
Number needed to treat

- How many people must be treated (exposed) to prevent one person from developing the outcome?

\[
\text{NNT} = \frac{1}{\text{Risk Difference (ARR)}}
\]
Trial of folate to reduce birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>343</td>
<td></td>
<td>10,335</td>
</tr>
<tr>
<td>Unexposed</td>
<td>398</td>
<td></td>
<td>10,320</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Risk Difference = 0.6% = 0.006 = 6/1000

Number needed to treat = 1 / 0.006 (or 1000 / 6)

= 167
How large is the effect?
How large is the effect: relative risk
Measures of effect

• Measures of effect based on INCIDENCE or EVENT RATES can only be calculated for studies that:
  – involve some form of follow-up
  – have known numbers of at-risk subjects
  – have high rates of follow-up ...

• RANDOMIZED TRIALS
• COHORT STUDIES
### Odds ratio for case-control studies

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Lung cancer</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

\[
\text{Odds Ratio (OR)} = \frac{a \times d}{b \times c}
\]
Odds ratio for case-control studies

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

Odds Ratio (OR) = \( \frac{a}{c} \) \quad \frac{a \times d}{b \times c}
Case control study of lung cancer

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Lung cancer</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>647</td>
<td>1,269</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>622</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>649</td>
<td>1,298</td>
</tr>
</tbody>
</table>

Odds Ratio (OR) = \frac{647 \times 27}{622 \times 2} = 14.0
Case control study of birth defects

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Birth defect</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>20</td>
<td>98</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>25</td>
<td>398</td>
<td>423</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>45</td>
<td>496</td>
<td>541</td>
</tr>
</tbody>
</table>

\[
\text{Odds Ratio (OR)} = \frac{20 \times 398}{98 \times 25} = 3.25
\]
Attributable risk

Incidence (unexposed) = Background rate (i.e., Incidence due to OTHER exposures)

Incidence (exposed) = Incidence (DUE TO exposure) PLUS Incidence (unexposed)

Therefore, incidence due to the exposure (attributable risk):

Incidence (exposed) - Incidence (unexposed)

So Risk Difference, Absolute Risk Increase and Attributable Risk are all words for the same measure of effect.
Incidence in unexposed population

Incidence due to other exposures or background rate
Incidence in exposed population

Incidence due to exposure

Background Incidence
Attributable risk
Attributable risk

\[ AR = I_{\text{exposed}} - I_{\text{unexposed}} \]
Attributable fraction
(Attributable risk %)

\[ AF = \left( \frac{I_{\text{exposed}} - I_{\text{unexposed}}}{I_{\text{exposed}}} \right) \times 100 \]

The percentage of outcomes in the exposed group that are due to exposure
Cohort study of effect of pesticide on risk of birth defects

<table>
<thead>
<tr>
<th></th>
<th>Birth defect present</th>
<th>Birth defect absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>20</td>
<td>980</td>
<td>1,000</td>
</tr>
<tr>
<td>Unexposed</td>
<td>25</td>
<td>3,975</td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>4,955</td>
<td>5,000</td>
</tr>
</tbody>
</table>

Incidence in exposed group: \( \frac{20}{1,000} = 2\% \)
Incidence in unexposed group: \( \frac{25}{4,000} = 0.625\% \)

Attributable risk = \((2\% - 0.625\%) / 2\% \times 100\) = 1.375\% 

\[ AF = \frac{I_E - I_{UE}}{I_E} \times 100 \]

\[ = \frac{2\% - 0.625\%}{2\%} \times 100 \]

\[ = 69\% \]
The interpretation of Attributable Fraction

- If the AF = 69%,

- this means that 69% of the birth defects observed in the babies of women exposed to the pesticide are attributable to exposure to the pesticide.
Attributable Fraction

\[ AF = \frac{RR - 1}{RR} \times 100 \]
Attributable Fraction from Relative Risk

- $RR = 3.2$

- $AF = \frac{3.2 - 1}{3.2} \times 100$

  $= 69\%$
Case control study of birth defects

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Birth defect</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>20</td>
<td>98</td>
<td>118</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>25</td>
<td>398</td>
<td>423</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>45</td>
<td>496</td>
<td>541</td>
</tr>
</tbody>
</table>

Odds Ratio (OR) = \((20 \times 398) / (98 \times 25) = 3.25\)

AF \(= [(RR - 1) / RR] \times 100 = [(OR - 1) / OR] \times 100\)

\[= [(3.25 - 1) / 3.25] \times 100\]

\[= 69\%\]
Population Attributable Fraction

- Estimates the effect of the exposure in the whole population

- Includes both a measure of the size of the effect of the exposure and a measure of the frequency (prevalence of exposure)
Population Attributable Fraction

\[ \text{PAF} = \frac{F \times (RR - 1)}{1 + F \times (RR - 1)} \times 100 \]

where F is the prevalence of exposure in the population
Population attributable fraction

• Example:

Community A is using a pesticide called “Dead Weed”. In a cohort study, Dead Weed was found to increase the risk of birth deformity (RR = 3.2). It is estimated that 10% of women of child bearing age are exposed to Dead Weed.
Population attributable fraction

- \( \text{PAF} = \frac{0.1 \times (3.2-1)}{1 + 0.1 \times (3.2-1)} \times 100 \)
  
  \( = \frac{0.22 \times 100}{1.22} \)
  
  \( = 18.0\% \)

ie 18.0% of birth defects in the town are attributable to exposure to Dead Weed
Measures of effect

- Randomized trials
- Cohort studies
- Relative risk
- Risk difference (Absolute Risk Increase or Reduction, Attributable Risk)
- Number needed to treat
Measures of effect

- Case control studies
- Odds ratio
- RCTs, cohort studies and case control studies
- AF and PAF can be calculated from RCTs and COHORT studies, and estimated from the odds ratio in CASE-CONTROL studies.
How large is the treatment effect: relative and absolute risk

$$RR = \frac{0.4}{0.8} = 0.5 \quad RD = 0.8 - 0.4 = 0.4 \quad NNT = \frac{1}{0.4} = 2.5$$
How large is the treatment effect: relative and absolute risk

RR = 0.1 / 0.2 = 0.5  
RD = 0.2 – 0.1 = 0.1  
NNT = 1 / 0.1 = 10
How large is the treatment effect: relative and absolute risk

$RR = 0.00005 / 0.00010 = 0.5$

$RD = 0.0001 - 0.00005 = 0.00005$

$NNT = 1 / 0.00005 = 20,000$