Measures of Frequency and Association

Frances Boreland
Broken Hill UDRH
June 2005

Key Questions

• How frequent is the event?
• How strongly associated with suspected cause?
• Likely to be real or due to chance?

How frequent?

• Main measures of frequency
  – Number of existing cases (prevalence)
  – Number of new cases (incidence)

Prevalence

The proportion of a defined population with an attribute of interest (e.g., disease).

\[
\text{Prevalence} = \frac{\text{Persons with an attribute at a given time}}{\text{Total population at a given time}}
\]

E.g., Number of adults with arthritis
Number of adults resident in Broken Hill

Incidence

The number of new cases of disease in a population, over a given period of time.

\[
\text{Cumulative Incidence} = \frac{\text{New cases over specified time period}}{\text{Number susceptible individuals at the beginning of time period}}
\]

Calculating incidence

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>HEART DISEASE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>1,980</td>
</tr>
<tr>
<td>No</td>
<td>32</td>
<td>1,968</td>
</tr>
<tr>
<td>TOTAL</td>
<td>52</td>
<td>3,948</td>
</tr>
</tbody>
</table>
Did the treatment work?

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>HEART DISEASE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>1,980</td>
</tr>
<tr>
<td>No</td>
<td>32</td>
<td>1,968</td>
</tr>
<tr>
<td>TOTAL</td>
<td>52</td>
<td>3,948</td>
</tr>
</tbody>
</table>

Incidence yes = 20 / 2,000 = 10/1,000
Incidence no = 32 / 2,000 = 16/1,000

How strong?

Relative risk
- Derived from estimates of incidence so can only be directly estimated from RCT & cohort studies.
- $RR = \frac{\text{Incidence in exposed group}}{\text{Incidence in unexposed group}}$

Incidence yes = 10/1,000
Incidence no = 16/1,000

Interpreting the strength

- $RR > 1$ Incidence higher in treatment group
- $RR = 1$ No difference
- $RR < 1$ Incidence lower in treatment group

- Relative Risk Reduction (RRR)
  - In trials of therapy or prevention, the relative risk for a successful intervention is below 1. This is often described using the Relative Risk Reduction.
  - Relative Risk Reduction = $1 - \text{Relative Risk}$
  - Relative Risk = 0.625
  - RRR = $1 - 0.625 = 0.375$

- Risk Difference (RD)
  - Measures the difference in incidence between the treatment and control groups in a clinical trial.
  - $RD = \frac{\text{Incidence in control group}}{\text{Incidence in intervention group}}$
  - Incidence in control group = 16 / 1,000
  - Incidence in intervention group = 10 / 1,000
  - $RD = 16 / 1,000 - 10 / 1,000 = 0.006$
Interpreting the strength

- Number needed to treat (NNT)
  - This concept has arisen from the literature to describe the outcome of randomised trials.
  - \( NNT = \frac{1}{Risk\ Difference} \)
  - \( Risk\ Difference = \frac{6}{1,000} = 0.006 \)
  - \( NNT = \frac{1}{0.006} = 167 \)

Case-control studies

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>HEART DISEASE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>1,980</td>
</tr>
<tr>
<td>No</td>
<td>32</td>
<td>1,968</td>
</tr>
<tr>
<td>TOTAL</td>
<td>52</td>
<td>3,948</td>
</tr>
</tbody>
</table>

Case-control studies

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>HEART DISEASE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>104</td>
</tr>
<tr>
<td>No</td>
<td>32</td>
<td>104</td>
</tr>
<tr>
<td>TOTAL</td>
<td>52</td>
<td>208</td>
</tr>
</tbody>
</table>

Case-control studies

- Incidence yes = \( \frac{20}{124} = \frac{1}{6} \) (not \( \frac{10}{1,000} \))
- Incidence no = \( \frac{32}{104} = \frac{1}{3} \) (not \( \frac{16}{1,000} \))

Case-control studies

- Odds Ratio (OR)
  - Based on the ratio of the odds of exposure in the diseased, divided by the odds of exposure in the non-diseased.
  - \( OR = \frac{Odds\ of\ exposure\ in\ diseased}{Odds\ of\ exposure\ in\ non-diseased} \)
  - \( OR = \frac{20}{32} = 0.625 \)
  - \( \frac{104}{104} \)
Interpreting the strength

- OR > 1 Exposure higher among cases
- OR = 1 No difference
- OR < 1 Exposure lower among cases

Likelihood it’s real?

From population to sample

- Population
- Representative sample
- Study results

From population to sample

Risk of heart disease
From population to sample

How do we decide whether the difference reflects a real difference between the treatments, or just the usual variation we would expect between samples drawn from the same population?

We can never know for certain..
But we can infer, based on three pieces of information:
1. The likelihood the two groups are really different.
2. The size of the difference.
3. The precision with which the size of the difference is known.

Likely to be a real difference?

<table>
<thead>
<tr>
<th>P</th>
<th>little or no evidence</th>
<th>weak evidence</th>
<th>evidence</th>
<th>Statistically significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
<td>0.1</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.001</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

How large is the difference?

- Statistical significance does not necessarily mean practical importance
  - How large is the difference?
  - Is it large enough to be of practical or clinical importance? This is not a question that can be answered by statistics!

You can never know for certain......

\[ P = 0.001 \]

999 times in 1,000 we will correctly conclude there is a difference between the groups

BUT

1 time in 1,000 we will think there is a difference when there really isn't

And we can never know which is the 1 time!
Might we have missed an important difference?

- If no evidence of difference was found, was the sample size large enough to detect a clinically important difference if one was present?
  - Was there truly no difference?
  - Were the results inconclusive?

How precisely is the size of the difference known?

- Confidence Intervals (CI)
  - Indicate how precisely the size of the difference between groups is known
  - Provide an estimate of the range of values with which the data are consistent
  - Based on sample size and degree of variability within sample

How precisely is the size of the difference known?

Eg, reduction in blood pressure on new treatment: 15 mm Hg +/- 5
95% likely that true value is in range 10-20 mm Hg reduction

Eg, RR of experiencing clinically significant reduction in blood pressure on new treatment:
RR 1.6, +/- 0.2
95% likelihood that true benefit is within range of 1.4 – 1.8

Accuracy vs precision

- Precise but not accurate
- Accurate but not very precise

<table>
<thead>
<tr>
<th>Large difference</th>
<th>Important</th>
<th>Inconclusive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important difference</td>
<td>Not important</td>
<td>True negative</td>
</tr>
<tr>
<td>No difference</td>
<td>Statistically significant (P &lt;=0.05)</td>
<td>Not statistically significant (P&gt;0.05)</td>
</tr>
</tbody>
</table>