RISK: a statistician’s viewpoint

Graeme L. Hickey
Department of Biostatistics

University of Liverpool
What is risk?

**Definition**: [noun] exposure to the chance of injury or loss; a hazard or dangerous change

Lots of terminology used

Understanding these terms is crucial to *interpreting* risk

* http://www.dictionary.com/browse/risk
Reporting differences

**Absolute difference**

*The Independent*

*Vitamin D can prevent asthma attacks, study finds*

Taking oral vitamin D supplements can reduce the risk of severe asthma attacks by 3 per cent, according to research.

*The Guardian*

*Vitamin D supplements could halve risk of serious asthma attacks*

Major research review suggests that taking vitamin D might have fewer attacks compared to placebo.

**Relative difference**

*The Independent*

*Vitamin D can prevent asthma attacks, study finds*

Taking oral vitamin D tablets can reduce the likelihood of bouts requiring hospital admission or emergency department attendance from 6 per cent to 3 per cent, analysts at medical research group Cochrane concluded.

*The Guardian*

*Vitamin D supplements could halve risk of serious asthma attacks*

Vitamin D pills can halve the risk of serious asthma attacks according to a major review of research into the impact the supplements have on the condition.

---

http://www.independent.co.uk/news/science/vitamin-d-asthma-attacks-prevent-study-cochrane-a7226756.html

https://www.theguardian.com/society/2016/sep/05/vitamin-d-supplements-could-halve-risk-of-serious-asthma-attacks
Randomization
\[ N = 200 \]

Treatment
\[ n = 100 \]

Dead at 30-days
\[ n = 30 \]
Alive at 30-days
\[ n = 70 \]

Control
\[ n = 100 \]

Dead at 30-days
\[ n = 40 \]
Alive at 30-days
\[ n = 60 \]
Example

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died within 30-days</td>
<td>30</td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td>Alive at 30-days</td>
<td>70</td>
<td>60</td>
<td>130</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>N = 200</td>
</tr>
</tbody>
</table>

A 2x2 contingency table + marginal totals
### Example

A 2x2 contingency table + marginal totals

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died within 30-days</td>
<td>$a$</td>
<td>$b$</td>
<td>$a + b$</td>
</tr>
<tr>
<td>Alive at 30-days</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>$N = a + b + c + d$</td>
</tr>
</tbody>
</table>
Absolute risk reduction

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute risk in treatment group ( \text{AR}_{\text{treat}} ) =</td>
<td>( \frac{a}{a + c} )</td>
<td>( \frac{30}{100} = 0.3 )</td>
</tr>
<tr>
<td>Absolute risk in control group ( \text{AR}_{\text{control}} ) =</td>
<td>( \frac{b}{b + d} )</td>
<td>( \frac{40}{100} = 0.4 )</td>
</tr>
<tr>
<td>Absolute risk reduction ( \text{ARR} ) =</td>
<td>( \text{AR}<em>{\text{control}} - \text{AR}</em>{\text{treat}} )</td>
<td>( 0.4 - 0.3 = 0.1 )</td>
</tr>
</tbody>
</table>
## Absolute risk reduction

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
</table>
| Absolute risk in treatment group \( \text{AR}_t \) = | \[
\frac{a}{a + c}
\] | \[
\frac{30}{100} = 30\%
\] |
| Absolute risk in control group \( \text{AR}_c \) = | \[
\frac{b}{b + d}
\] | \[
\frac{40}{100} = 40\%
\] |
| Absolute risk reduction (ARR) =              | \( \text{AR}_c - \text{AR}_t \) | 40\% - 30\% = 10\% |
Number needed to treat

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number needed to treat (NNT) =</td>
<td>[\frac{1}{\text{ARR}}]</td>
<td>[\frac{1}{0.1} = 10]</td>
</tr>
</tbody>
</table>

Equivalent to the average number of patients who need to be treated to prevent one additional event
## Relative risk (reduction)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk (RR)</td>
<td>$\frac{AR_{treat}}{AR_{control}}$</td>
<td>$\frac{0.3}{0.4} = 0.75$</td>
</tr>
<tr>
<td>Relative risk reduction (RRR)</td>
<td>$1 - \text{RR}$</td>
<td>$1 - 0.75 = 0.25$</td>
</tr>
</tbody>
</table>
Results from 3 hypothetical RCTs of the same treatment

**High risk**
- ARR = 0.1 (or 10%)
- NNT = 10
- RRR = 0.25 (or 25%)

**Intermediate risk**
- ARR = 0.05 (or 5%)
- RRR = 0.25 (or 25%)

**Low risk**
- ARR = 0.01 (or 1%)
- RRR = 0.25 (or 25%)
Relative vs. absolute differences

• Ratios more robust to baseline risk
  • better for meta-analysis
  • not as useful for patients and clinicians
  • inappropriate to extrapolate ARR from one population to another if baseline risk is different
Odds

- Probability ($P$): a number between 0 (never happen) to 1 (definitely happen)
- Odds = $P / (1 - P)$
- Can invert this: $P = 1 / (\text{Odds} + 1)$

- E.g. odds of Native River to win Cheltenham Cup was 5-to-1
- $\text{Prob}[\text{Native River wins}] = 1 / (5 + 1) = 17\%$
<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
</table>
| Relative risk (RR)     | \[
\frac{a(b + d)}{b(a + c)}
\] = 0.75                    |
| Odds ratio (OR)        | \[
\frac{\text{odds}_{\text{treat}}}{\text{odds}_{\text{control}}} = \frac{a}{c}/\frac{b}{d}
\] \[
\frac{18}{28} = 0.64
\] |
Are they ever equivalent?

• OR ≈ RR for **low baseline risk**
  • High risk trial (event rate = 0.30; RR = 0.75 vs. OR = 0.64)
  • Intermediate risk trial (event rate = 0.15; RR = 0.75 vs. OR = 0.71)
  • Low risk trial (event rate = 0.04; RR = 0.75 vs. OR = 0.74)

• But ORs exaggerate reduction effect otherwise

• ORs are easily confused with RRs!


\[ RR = \frac{OR}{1 - AR_{control} + (1 - AR_{control})OR} \]
An interpretative disadvantage

- What if we think of ‘survival at 30-days’ as the outcome rather than ‘death at 30-days’?

\[
\text{OR}_{\text{survival}} = \frac{28}{18} = 1.56 = \frac{1}{\text{OR}_{\text{death}}}
\]

\[
\text{RR}_{\text{survival}} = \frac{0.7}{0.6} = 1.17 \neq \frac{1}{\text{RR}_{\text{death}}}
\]
Why are ORs reported so frequently?

- **Logistic regression**
  - \( \text{logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \)
  - \( P \) is the probability of the event
  - \( x_1, x_2, \ldots \) are covariates (sometimes called ‘risk factors’)
  - \( \exp(\beta_m) \) is the odds ratio for a unit increase (e.g. 1-year of age) in risk factor \( x_m \)
  - With estimates for all \( \beta \) coefficients (including intercept \( \beta_0 \)) we can estimate \( p \) for a patient → **absolute risk**

- **Case-control studies**
  - Can’t estimate outcome prevalence ⇒ **can’t estimate RR**
**Time-to-event outcomes**

Relative effect:  
HR = 0.55

- HR uses all data at each time point
- Not robust to departures from proportionality

Absolute effect:  
ARR(12-months) = 20.0%  
30.7% in the TAVI group  
50.7% in the standard therapy group  
NNT(12-months) = 5

What is hazard?

• The hazard function (or instantaneous rate) is

$$\lambda(t) = \lim_{dt \to 0} \frac{P[t \leq T < t + dt \mid T \geq t]}{dt}$$

• Cox proportional hazards regression models

• Example
  • Time-to-death of patients randomized to CABG or PCI
  • $x = 0$ if patient randomized to CABG (reference)
  • $x = 1$ if patient randomized to PCI
  • Univariable Cox model $\log[\lambda(t)] = \lambda_0(t) + \beta x$
Hazard ratio (HR)

- $\text{HR}_{\text{Tx}} = \lambda(t \mid x = 1) / \lambda(t \mid x = 0)$
- $\exp(\beta)$ is the HR for PCI (relative to CABG)
- HRs are not RRss (strictly speaking) or probabilities
  - E.g. HR = 1.2 $\Rightarrow$ 20% increase in hazard of death, not 20% more likely to die
- Two useful results$^1$
  1. $S_{\text{PCI}}(t) = S_{\text{CABG}}(t)^{\text{HR}}$
  2. Probability that a PCI patient survives longer than a CABG patient would be $1 / (1 + \text{HR})$

$^1$ https://understandinguncertainty.org/node/759
HR vs. absolute risk

As we observed earlier for RR, the HR does not convey treatment benefit alone.

Median survival times (years)

<table>
<thead>
<tr>
<th>Treatment arm</th>
<th>Panel 1 (HR = 3)</th>
<th>Panel 2 (HR = 1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0.9</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Multiple risks

Grant SW et al. Health Technol Assess. 2015;19(32)


Holistic view of risk is required
Conclusions

• Preferable to report **both** absolute and relative effects

<table>
<thead>
<tr>
<th>Outcomes and estimation</th>
<th>17a For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended</td>
</tr>
</tbody>
</table>

• Evidence suggests reporting choice affects interpretation*
• If presented with a relative risk, always ask ‘relative to what?’
• The uncertainty about effect sizes should always be reported

Questions?

Slides available from www.glhickey.com

THINGS GOT REALLY INTERESTING WHEN THE STATISTICIAN STARTED DOING WARD ROUNDS.