Reviews of conditions with low event rates

Matthew Grainge and Jo Leonardi-Bee
Statistical Editors for CSG
Introduction

- CSG reviews frequently have common outcomes (>10%)
- Statisticians recommend using risk ratio
  - Odds ratio over estimates treatment effect
But what about rare outcomes?

- Rare outcomes (<5%)
  - Not a problem if sample size is large enough (>30 participants in total)

- BUT, if sample size is small and rare outcomes, then results can be misleading
What about no outcomes?

- Mathematically, can’t estimate Risk Ratio
  - Review Manager cheats by adding 0.5 people to each group with and without the outcomes

<table>
<thead>
<tr>
<th></th>
<th>Outcome</th>
<th>No outcome</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>5</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>5</td>
<td>16</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Outcome</th>
<th>No outcome</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>0.5</td>
<td>8.5</td>
<td>9</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>5.5</td>
<td>8.5</td>
<td>14</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>6</td>
<td>17</td>
<td>23</td>
</tr>
</tbody>
</table>
What about confidence intervals?

- Misleading confidence intervals
  - Assumes data is based on a large sample size
What does the handbook say?

- **Section 16.9 – Rare events (including zero frequencies)**

  “Many methods of meta-analysis are based on large sample approximations, and are unsuitable when events are rare”

  “The fixed correction (for zero frequencies) ... avoids computational errors [but] it has the undesirable effect of biasing study estimates towards no difference and overestimating variances of study estimates...”
Mycosis fungoides

- Most common type of cutaneous T-cell lymphoma
  - Rare malignant, chronic disease (not fungal infection)
  - Uncontrolled growth of blood cells within the skin
  - Several therapies used to induce clinical remission

- Weberschock et al. Interventions for mycosis fungoides. 2012; Issue 9
  - 14 RCTs (675 participants)
  - Majority of results are based on data from one study per treatment
Intervention: Combination therapy

- Active transfer factor with nitrogen mustard
  - Intramuscular injections

- Nitrogen Mustard
  - Topically applied chemotherapy agent

- Comparison: TNM with active transfer factor versus TNM with inactivated transfer factor

- Outcome: Clearance
“… differences were seen between the active transfer factor group and the inactivated transfer group for clearance: RR 0.09, 95% CI 0.01 to 1.41, 16 participants, Analysis 1.1”

N.B: Both these statistics assume large sample sizes
“P values for the subtotal RR were calculated with the Fisher test in order to avoid spurious (non-) significance in studies with small sample sizes or low numbers of events”

- Data: 0/8 v. 5/8
- P=0.03 (from Fisher’s Exact test)
Is the treatment effective?

- **P values**
  - From RevMan <0.00001
  - From Fisher’s Exact test = 0.03
  - Both significant

- **Measure of effect and confidence intervals**
  - Risk Ratio = 0.09
  - 95% Confidence Interval 0.01 to 1.41
  - Not significant
### Included study

<table>
<thead>
<tr>
<th>Outcome of interest</th>
<th>Event rate</th>
<th>RR (95% CI)</th>
<th>Fisher’s p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuVic (2001) Rash</td>
<td>6/29 vs. 1/35</td>
<td>7.24 (0.92, 56.8)</td>
<td>0.04</td>
</tr>
<tr>
<td>Wolf (1985) Fever</td>
<td>5/9 v. 0/9</td>
<td>11.0 (0.70,173.7)</td>
<td>0.03</td>
</tr>
<tr>
<td>Stadler (1998) Adverse events</td>
<td>9/42 v. 2/40</td>
<td>4.29 (0.99,18.6)</td>
<td>0.049</td>
</tr>
<tr>
<td>Child (2004) Improvement</td>
<td>0/8 v. 7/8</td>
<td>0.07 (0.00,1.00)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

**So was this a fluke?**

**Is this phenomena seen in other CSG reviews?**
Method 1: Survey of Skin Group full reviews

- Random sample of 20 skin group reviews published in the Cochrane Library

- Research question: What percent of analyses (Forest plots) are based on?
  - A single study
  - Denominator < 30
  - No. of events < 10
## Survey of Skin Group Reviews

<table>
<thead>
<tr>
<th></th>
<th>No. of Forest plots</th>
<th>No. based on 1 study</th>
<th>Denominator*</th>
<th>No. with &lt;10 events</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All reviews</strong></td>
<td>742 (0–298)</td>
<td>603 (81.3)</td>
<td>56 (0–12)</td>
<td>88 (0–22)</td>
</tr>
<tr>
<td><strong>All reviews</strong></td>
<td>444 (0–123)</td>
<td>336 (75.7)</td>
<td>56 (0–12)</td>
<td>66 (0–18)</td>
</tr>
</tbody>
</table>

* Number of randomised patients in both groups combined
So have small no. comparisons influenced overall findings?

- 20 reviews, 14 contained comparisons with small numbers

**Results reported in abstract:** 4 reviews reported data based on a single comparison with small numbers

**Implications for practice:** 2 occasions where implications for practice were influenced by small no. comparisons
“My query relates to the inconsistency in approaches used to generate the results, where the author is reporting both an exact (Fisher’s) p value and asymptotic based 95% confidence intervals for risk ratios from Rev Man. The issue is the discrepancy in the interpretation between the p value (which is less than 0.05) and the confidence intervals (which crosses 1)”
Responses

- Suggested formulae for confidence intervals which work well with small numbers
  - Robert Newcombe’s Excel worksheet
  - Miettinen–Nurminen formula

- Philosophical suggestions
  - Should use odds ratios instead of risk ratios if concordance between confidence intervals and p-values is required
  - Warned against over interpreting borderline results (i.e. where P is close to 0.05)
Comparison of two formulae for Mycosis fungoides review

<table>
<thead>
<tr>
<th>Study</th>
<th>Event rates</th>
<th>P</th>
<th>Mantel–Haenszel</th>
<th>Miettinen–Nurminen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thestrup-Pedersen (1982)</td>
<td>0/8 v. 5/8</td>
<td>0.03</td>
<td>0.09 (0.01, 1.41)</td>
<td>0.00 (0.00, 0.61)</td>
</tr>
<tr>
<td>Duvic (2001)</td>
<td>6/29 v. 1/35</td>
<td>0.04</td>
<td>7.24 (0.92, 56.8)</td>
<td>7.24 (1.22, 45.1)</td>
</tr>
<tr>
<td>Wolf (1985)</td>
<td>5/9 v. 0/9</td>
<td>0.03</td>
<td>11.0 (0.70, 173.7)</td>
<td>NE (1.59, NE)</td>
</tr>
<tr>
<td>Stadler (1998)</td>
<td>9/42 v. 2/40</td>
<td>0.049</td>
<td>4.29 (0.99, 18.6)</td>
<td>4.29 (1.13, 17.1)</td>
</tr>
<tr>
<td>Child (2004)</td>
<td>0/8 v. 7/8</td>
<td>0.002</td>
<td>0.07 (0.00, 1.00)</td>
<td>0.00 (0.00, 0.40)</td>
</tr>
</tbody>
</table>

NE – Not estimable
Solution

- Measures of effect and 95% CIs should be used ideally to inform strength of association.
- Where there are results from individual studies based on small numbers of included participants, individual studies are presented narratively and a two-sided Fisher’s exact p-value is used to determine statistical significance.
Student project
- How common is this phenomena in other Cochrane reviews?
- What methods are used to overcome the phenomena?

Standard sentence to be added to protocol:
- “Where results are estimated for individual studies with low numbers of outcomes (<10 in total) or where the total sample size is less than 30 participants, we will report the proportion of outcomes in each treatment group together with a p value from a Fisher’s Exact test.”

Meta-analyses which contain 2 or more small studies are beyond the scope of this presentation
Questions