Statistics for the terrified

Amanda Burls
Evidence-Based Teachers and Developers Conference, Taormina, Sicily
October 2013
Talk to your neighbour about what you want to get out of this session.
Hypothermia vs. control

In severe head injury
Mortality or incapacity (n=158)

Clifton 1993
Clifton 1992
Hirayama 1994
Marion 1997
Total (95%CI)

RR 0.63 (0.46, 0.87)

1 minute to discuss with your neighbour
Then write down what you think this graphic tells you
Learning objectives

• By the end of this session you will
  – Know how measures of effect are reported
  – Be able to interpret p-values
  – Be able to interpret confidence intervals
  – Be able to calculate relative risks (RR, OR)
  – Be able to explain the difference between statistical significance and clinical significance
  – Like to use blobbograms and be able to interpret them with ease

• Have fun!
Statistics without fear
Before we start, let’s remind ourselves

What are the important things to think about when we are using research evidence to help inform your decisions?
Validity for an intervention study?

• Randomised controlled trial

End
Validity for an RCT – Getting similar groups and keeping them similar

- Randomized
- Concealment of allocation
- Similar baseline characteristics
- Blinding
- Treating groups the same
- Minimal losses to follow up
- Intention to treat analysis
Appraisal of any study must consider

- **Validity**
  - Can the results be trusted? $\leftarrow$ **Systematic error (Bias)**

- **Results**
  - What are the results
  - How are they (or can they be) expressed
  - Could they have occurred by chance $\leftarrow$ **Random error**

- **Relevance**
  - Do these results apply to the local context or to me or to my patient?
Warning!

Everything I say from now onwards assumes that the results being considered come from an unbiased study!

(assumes NO systematic errors)
How are results summarised?

• Most useful studies compare at least two alternatives.
• How can the results of such comparisons be expressed?
Well-conducted RCT (no bias)
Expressing results:

What did the study show?

• Patients with backache:
  – 10 randomised to receive Potters
  – 10 randomised to receive placebo

• After 3 months:
  – 2 get better on Potters
  – 1 get better on placebo

• Summarise this result to your neighbour in at least three different ways
Summarise

• 2 out of 10 (20%) better on Potters
• 1 out of 10 (10%) better on placebo

• Twice as likely to get better on Potters
• An extra 10% of people get better on Potters
• For every 10 people with back pain given Potters, one case of back pain is improved
Line of no difference
Relative Risk

• How much more likely one group is to recover than the other

• Twice as many recovered on Potters means the relative risk is 2, or $RR = 2.0$
Line of no difference
Risk difference

- The difference in the proportions recovering – the proportion of patients benefiting from treatment
- 20% improved on Potters, but 10% improved on placebo, so the risk difference is 10%
Line of no difference
Number needed to treat (NNT)

- The number of patients to whom the new intervention needs to be given to produce one extra patient who is helped
- NNT = 1/risk difference
- Why?
How were the results summarised?

Two basic ways to summarise results of studies that compare groups:

1. Difference (take them away)
2. Ratio (divide)
Do you think this study proves that Potters works?
I think I've discovered something.
“It could have happened by chance!”
“It could have happened by chance!”

• If there had been 1000 people in the trial
  • 200 got better with Potters
  • 100 got better on placebo
• Would you believe Potters works?
So how many would you want before you believe the results?

- 10 in each arm?
- 20 in each arm?
- 100?
- 1000?
What is the minimum number you would want in each arm to believe the trial?

Assume similar effect size:
10% better with placebo
20% with Potters

- Write on a piece of paper your estimate
- Fold your paper in half and half again
- Swap it with your neighbour
- Swap the paper again with someone else
- Keep swapping until you don’t know who’s paper you have
Scores

- 0-20
- 21-40
- 41-60
- 61-99
- 100
- 101-200
- >200
"I think you should be more explicit here in step two."
Quantifying uncertainty due to chance

p-value
The Null Hypothesis

... is the assumption of no difference between treatments being compared

Umm... It could be due to chance!
So what does $p = 0.5$ mean?

So what does $p = 0.05$ mean?

Impossible

Absolutely certain
Bag of 20 sweets

1 Blue
19 Green
Bag of 20 sweets

10 Blue
10 Green
Bag of 30 sweets

20 Blue
10 Green
“Statistical significance”

• When a similar result would happen by chance on fewer than 1 in 20 occasions

• $p<0.05$
<table>
<thead>
<tr>
<th>Potters</th>
<th>Placebo</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/10</td>
<td>1/10</td>
<td>P = 0.531</td>
</tr>
<tr>
<td>4/20</td>
<td>2/20</td>
<td>P = 0.376</td>
</tr>
<tr>
<td>6/30</td>
<td>3/30</td>
<td>P = 0.278</td>
</tr>
<tr>
<td>8/40</td>
<td>4/40</td>
<td>P = 0.210</td>
</tr>
<tr>
<td>10/50</td>
<td>5/50</td>
<td>P = 0.161</td>
</tr>
<tr>
<td>12/60</td>
<td>6/60</td>
<td>P = 0.125</td>
</tr>
<tr>
<td>14/70</td>
<td>7/70</td>
<td>P = 0.097</td>
</tr>
<tr>
<td>16/80</td>
<td>8/80</td>
<td>P = 0.076</td>
</tr>
<tr>
<td>18/90</td>
<td>9/90</td>
<td>P = 0.060</td>
</tr>
<tr>
<td>20/100</td>
<td>10/100</td>
<td>P = 0.048</td>
</tr>
<tr>
<td>100/500</td>
<td>50/500</td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td>200/1000</td>
<td>100/1000</td>
<td>P &lt; 0.0001</td>
</tr>
</tbody>
</table>
Why \( p<0.05 \) as the cut-off?

- Convention!
- There is no magic cut-off between “statistically significant” and not
- Although many behave as if there were!
Toss a coin 8 times in a row and record the number of heads.

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

$P < 0.016$
“Odds ratio”

Pre and Post Workshop Scores

Percentage
Do you think this is likely to have happened by chance?

1. Yes
2. Don’t know
3. No
Do you think this is likely to have happened by chance?

1. Yes
2. Don’t know (~1000)
3. No
P<0.00001
Do you think this is likely to have happened by chance?

1. Yes
2. No
P<0.00001
Statistical significance does not imply clinical significance!
Limitations of the p-value

Any genuine difference between two groups, *no matter how small*, can be made to be “statistically significant” - at *any* level of significance - by taking a sufficiently large sample.
We need a better way to express uncertainty due to chance….. [?]
Introduction to confidence intervals

- CIs are a way of showing the uncertainty surrounding a point estimate.
How many **Red** sweets did I pick?

P < 0.000001
Hypothermia vs. control

In severe head injury
Mortality or incapacity (n=158)

Marion 1997

Clifton 1993
Clifton 1992
Hirayama 1994
Marion 1997

Total (95% CI)

RR 0.63 (0.46, 0.87)
Hypothermia vs. control

In severe head injury
Mortality or incapacity (n=158)

Clifton 1993
Clifton 1992
Hirayama 1994
Marion 1997

Total (95%CI)

RR 0.63 (0.46, 0.87)
When making health care decisions we often want to know

- The chance that people with a specific disease have a certain risk factor
- The chance that people with a specific trait have or are likely to develop a particular outcome
- Or the chance that a patient given a treatment will get better (or be harmed)
How are these things summarised?

• Write down as many measures of comparison and association as you can
On the next slide

- There is a field
- Write down how you could summarise quantitatively what you see to someone who cannot see the field
- Talk to your neighbour if you want
Possible summaries

• There are sixteen animals in the field
• There are sixteen sheep and goats in the field
• There are sixteen animals in the field, of which 10 are sheep and 6 are goats
Calculate (you have 1 minute)

• The risk of being a sheep
  (note that in epidemiology we use the word “risk” to mean “chance” – it doesn’t necessarily mean something unwanted)
• The odds of being a sheep

End
Answers

• Risk
  – 10/16
  – 0.625
• Odds
  – 10:6
  – 1.66
Odds

• Odds are when you separate the sheep from the goats!
**Odds** are a way of describing how many people in a population have a disease, risk factor or other outcome of interest.

e.g. The **odds** could be the number of people with the disease compared to the number without the disease:

\[
\frac{\text{number with disease}}{\text{number without disease}}
\]
Comparing two or more groups…

• The difference between the groups
• The ratio between the groups (e.g. experimental/control, exposed/not expose)
Mountain side

- Calculate the odds of being a sheep at the foot of the hill
- Calculate the odds and risk of being a sheep on the mountain side
- Using these calculate comparative measures
Summarise the relative risk of being in the field compared to the mountain in terms of odds and risks

- Odds of being a sheep at the foot = 7:3 = 2.3
- Risk of being a sheep at the foot = 7/10 = 0.7 of the hill
- Odds of being a sheep on the mountain = 3:7 = 0.43
- Risk of being a sheep on the mountain = 3/10 = 0.3
- Odds ratio (OR) = 5.4
- Risk ratio (RR) = 2.3
Is it reasonable to conclude that there is an association between being in the field and being a sheep?

1. Yes
2. No
3. Don’t know
P-value

- > 0.07

- [http://www.vassarstats.net/odds2x2.html](http://www.vassarstats.net/odds2x2.html)
How likely is it that we would get a result as big (or as small) as the one observed if there is nothing going on?

The answer is given as a p-value

- Write down the letter of the studies you would believe, if:
  - A. $P = 0.24$
  - B. $P < 0.5$
  - C. $P = 0.05$
  - D. $P = 0.049$
  - E. $P < 0.01$
  - F. $P > 0.001$
How likely is it that we would get a result as big (or as small) as the one observed if there is nothing going on?

<table>
<thead>
<tr>
<th>Option</th>
<th>P</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>P = 0.24</td>
<td>✗</td>
</tr>
<tr>
<td>B</td>
<td>P &lt; 0.5</td>
<td>✗</td>
</tr>
<tr>
<td>C</td>
<td>P = 0.05</td>
<td>✗</td>
</tr>
<tr>
<td>D</td>
<td>P = 0.049</td>
<td>✓</td>
</tr>
<tr>
<td>E</td>
<td>P &lt; 0.01</td>
<td>✓</td>
</tr>
<tr>
<td>F</td>
<td>P &gt; 0.001</td>
<td>✗</td>
</tr>
</tbody>
</table>
Meta-analysis
More fun!
Fixed effects
• Each bag of sweets have been drawn from the same barrel

Random effects
• Each bag of sweets is drawn from a random barrel and all we know about the proportions of sweets in the barrels in the world is what we can deduce from our samples
Review: Hypothermia for traumatic head injury
Comparison: 1 Immediate hypothermia versus normothermia
Outcome: 1 Death at final follow-up

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Hypothermia n/N</th>
<th>Control n/N</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adelson 2005 HYP01</td>
<td>4/25</td>
<td>4/25</td>
<td>1.00 [0.22, 4.54]</td>
<td></td>
</tr>
<tr>
<td>Adelson 2005 HYP02</td>
<td>3/14</td>
<td>3/13</td>
<td>0.91 [0.15, 5.58]</td>
<td></td>
</tr>
<tr>
<td>Abiki 2000</td>
<td>1/15</td>
<td>3/11</td>
<td>0.19 [0.02, 2.15]</td>
<td></td>
</tr>
<tr>
<td>Biswas 2002</td>
<td>3/8</td>
<td>0/6</td>
<td>8.27 [0.35, 197.61]</td>
<td></td>
</tr>
<tr>
<td>Clifton 1992</td>
<td>1/5</td>
<td>1/5</td>
<td>1.00 [0.05, 22.18]</td>
<td></td>
</tr>
<tr>
<td>Clifton 1993</td>
<td>8/23</td>
<td>8/22</td>
<td>0.93 [0.28, 3.16]</td>
<td></td>
</tr>
<tr>
<td>Clifton 2001</td>
<td>53/190</td>
<td>48/178</td>
<td>1.05 [0.66, 1.66]</td>
<td></td>
</tr>
<tr>
<td>Harris 2009</td>
<td>6/12</td>
<td>4/13</td>
<td>2.25 [0.44, 11.52]</td>
<td></td>
</tr>
<tr>
<td>Hashiguchi 2003</td>
<td>1/9</td>
<td>0/8</td>
<td>3.00 [0.11, 84.56]</td>
<td></td>
</tr>
<tr>
<td>Hirayama 1994</td>
<td>4/12</td>
<td>5/10</td>
<td>0.50 [0.09, 2.81]</td>
<td></td>
</tr>
<tr>
<td>Hutchison 2008</td>
<td>23/102</td>
<td>14/103</td>
<td>1.85 [0.89, 3.84]</td>
<td></td>
</tr>
<tr>
<td>Jiang 2000</td>
<td>11/43</td>
<td>20/44</td>
<td>0.41 [0.17, 1.02]</td>
<td></td>
</tr>
<tr>
<td>Marion 1997</td>
<td>9/39</td>
<td>10/42</td>
<td>0.96 [0.34, 2.69]</td>
<td></td>
</tr>
<tr>
<td>Meissner 2003b</td>
<td>3/15</td>
<td>3/13</td>
<td>0.83 [0.14, 5.08]</td>
<td></td>
</tr>
<tr>
<td>Qiu 2007</td>
<td>9/40</td>
<td>13/40</td>
<td>0.60 [0.22, 1.63]</td>
<td></td>
</tr>
<tr>
<td>Shiozaki 1993</td>
<td>8/16</td>
<td>14/17</td>
<td>0.21 [0.04, 1.05]</td>
<td></td>
</tr>
<tr>
<td>Shiozaki 1999</td>
<td>0/8</td>
<td>0/8</td>
<td>0.00 [0.00, 0.00]</td>
<td></td>
</tr>
<tr>
<td>Shiozaki 2001</td>
<td>8/45</td>
<td>6/46</td>
<td>1.44 [0.46, 4.55]</td>
<td></td>
</tr>
<tr>
<td>Smrcka 2005</td>
<td>5/35</td>
<td>11/37</td>
<td>0.39 [0.12, 1.28]</td>
<td></td>
</tr>
<tr>
<td>Yan 2001</td>
<td>13/24</td>
<td>16/20</td>
<td>0.30 [0.08, 1.15]</td>
<td></td>
</tr>
<tr>
<td>Zhang 2000</td>
<td>41/123</td>
<td>50/123</td>
<td>0.73 [0.43, 1.23]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI) 803 / 784 = 0.85 [0.68, 1.06]

Total events: 214 (Hypothermia), 233 (Control)
Heterogeneity: Chi^2 = 21.91, df = 19 (P = 0.29); I^2 = 13%
Test for overall effect: Z = 1.44 (P = 0.15)
## Review: Hypothermia for traumatic head injury

### Comparison: Immediate hypothermia versus normothermia

### Outcome: Death at final follow-up stratified by trial quality

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Hypothermia n/N</th>
<th>Control n/N</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Concealed allocation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adelson 2005 HYPO1</td>
<td>4/25</td>
<td>4/25</td>
<td>1.00 [0.22, 4.54]</td>
</tr>
<tr>
<td>Adelson 2005 HYPO2</td>
<td>3/14</td>
<td>3/13</td>
<td>0.91 [0.15, 5.58]</td>
</tr>
<tr>
<td>Clifton 1992</td>
<td>1/5</td>
<td>1/5</td>
<td>1.00 [0.05, 22.18]</td>
</tr>
<tr>
<td>Clifton 1993</td>
<td>8/23</td>
<td>8/22</td>
<td>0.93 [0.28, 3.16]</td>
</tr>
<tr>
<td>Clifton 2001</td>
<td>53/190</td>
<td>48/178</td>
<td>1.05 [0.66, 1.66]</td>
</tr>
<tr>
<td>Harris 2009</td>
<td>6/12</td>
<td>4/13</td>
<td>2.25 [0.44, 11.52]</td>
</tr>
<tr>
<td>Hutchison 2008</td>
<td>23/102</td>
<td>14/103</td>
<td>1.85 [0.89, 3.84]</td>
</tr>
<tr>
<td>Marion 1997</td>
<td>9/39</td>
<td>10/42</td>
<td>0.96 [0.34, 2.69]</td>
</tr>
<tr>
<td>Qiu 2007</td>
<td>9/40</td>
<td>13/40</td>
<td>0.60 [0.22, 1.63]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td><strong>450</strong></td>
<td><strong>441</strong></td>
<td><strong>1.11 [0.82, 1.51]</strong></td>
</tr>
</tbody>
</table>

Total events: 116 (Hypothermia), 105 (Control)

Heterogeneity: $\chi^2 = 4.33, df = 8 (P = 0.83); I^2 = 0.0$

Test for overall effect: $Z = 0.66 (P = 0.51)$

| 2 Non-concealed allocation | | | |
| Aibiki 2000 | 1/15 | 3/11 | 0.19 [0.02, 2.15] |
| Biswas 2002 | 3/8 | 0/6 | 8.27 [0.35, 197.61] |
| Hashiguchi 2003 | 1/9 | 0/8 | 3.00 [0.11, 84.56] |
| Hirayama 1994 | 4/12 | 5/10 | 0.50 [0.09, 2.81] |
| Jiang 2000 | 11/43 | 20/44 | 0.41 [0.17, 1.02] |
| Meissner 2003b | 3/15 | 3/13 | 0.83 [0.14, 5.08] |
| Shiozaki 1993 | 8/16 | 14/17 | 0.21 [0.04, 1.05] |
| Shiozaki 1999 | 0/8 | 0/8 | 0.00 [0.00, 0.00] |
| Shiozaki 2001 | 8/45 | 6/46 | 1.44 [0.46, 4.55] |
| Smrcka 2005 | 5/35 | 11/37 | 0.39 [0.12, 1.28] |
| Yan 2001 | 13/24 | 16/20 | 0.30 [0.08, 1.15] |
| Zhang 2000 | 41/123 | 50/123 | 0.73 [0.43, 1.23] |
| **Subtotal (95% CI)** | **353** | **343** | **0.62 [0.44, 0.86]** |

Total events: 98 (Hypothermia), 128 (Control)

Heterogeneity: $\chi^2 = 11.14, df = 10 (P = 0.35); I^2 = 10$

Test for overall effect: $Z = 2.83 (P = 0.0046)$

| **Total (95% CI)** | **803** | **784** | **0.85 [0.68, 1.06]** |

Total events: 214 (Hypothermia), 233 (Control)

Heterogeneity: $\chi^2 = 21.91, df = 19 (P = 0.29); I^2 = 13$

Test for overall effect: $Z = 1.44 (P = 0.15)$
“Hey, no problem!”
Statistics scary? – Nah, all bark and no bite