How fragile are critical care trials?

Marion K Campbell
Trialist and Methodologist
Prof of Health Services Research

@MarionKCampbell

Health Services Research Unit
University of Aberdeen
Background

- Clinical trials are expensive to design and run
- **Average cost of clinical trial participant in US** – $16,500
- **£5.3bn per annum** spent on clinical trials in UK alone
- Want their results to change practice
The “importance” of a “significant finding”

Source: de Winter & Debou, 2015, PeerJ 3:e733
The scourge of the p-value

\[ p < 0.05 \]

\[ p > 0.05 \]
The Fragility Index

- Developed by Walsh et al (*J Clin Epi* 2014;67:622-28)
- Quantifies the “fragility” of trial results
- Calculates how many events would be required to change a significant result to a non-significant result
The Fragility Index

- Designed for dichotomous outcomes
- The *Fragility Index* is the number of events, \( f \), that would be required to obtain a p-value, \( p \geq 0.05 \)

**Example:** Fragility Index = 1; only one event needed to have been different to overturn the result
**Fragility Index**

<table>
<thead>
<tr>
<th>Control Events</th>
<th>Intervention Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Total</td>
<td>Intervention Total</td>
</tr>
</tbody>
</table>

### What is the Fragility Index (FI)?

A measure of how fragile a trial's result is. It describes the number of non-events that need to become events in order to render a trial result non-significant (i.e. p>0.05). P-values are calculated with Fisher's Exact Test.

For example, in a study with 100 patients in the placebo arm of which 20 die, and 100 in the intervention arm of which 12 die, we ask what might be the new p-value if the death rates are changed to 20/100 and 13/100. If the p-value remains significant (<0.05), we repeat the process, adding one death with each iteration until p>0.05. The number added is reported as the Fragility Index.

In other words, a lower Fragility Index indicates less statistically robust results.

### What kind of trials can FI be applied to?

FI can be applied to any trial reporting a binary outcome of any kind. This can be death, MI, intubation or anything for which the question can be asked "did this event occur in my patient?" and a yes or no answer be given.

A trial of this type will often report the number of outcomes (events) in the control/placebo group, and then compare this to the number of outcomes (events) in the intervention group.

### How to use the calculator

Enter the number of events that occurred in your trial's control group, the number of events that occurred in your trial's intervention group, and the total number of patients in each arm of the trial. Hit submit, and the calculator will give you the trial's FI.
Example

The Fragility Index of this trial was found to be 2 event(s).
This would increase the p-value to 0.06797574456022.

<table>
<thead>
<tr>
<th>Events</th>
<th>Non-Events</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>69</td>
<td>0.03444781</td>
</tr>
<tr>
<td>35</td>
<td>68</td>
<td>0.04888921</td>
</tr>
<tr>
<td>36</td>
<td>67</td>
<td>0.06797574</td>
</tr>
</tbody>
</table>

If Fragility Index < number missing, results likely to be particularly vulnerable
Are trials fragile?

- Walsh et al reviewed 399 trials in major journals from 2004-2010:
  - median Fragility Index = 8
  - 25th percentile of the Index was 3
  - for 53% of trials - more participants were lost to follow up than would be required to make a result non-significant
  - results held over different sample sizes and numbers of events
More recent evidence

- Ridgeon et al, 2016
- Reviewed multicentre, critical care trials looking at mortality outcomes
- Reviewed 20 most important critical care journals + 9 other general journals
- 862 critical care trials reviewed; 56 had significant mortality outcomes
- Median fragility index = 2 (IQR 1-3.5)
- 40% had index of ≤ 1

Crit Care Med, Mar 2016, Epub ahead of print; DOI:10.1097/CCM0000000000001670
Assessing a cohort of trials

- NIHR HTA programme – largest funders of trials in UK
- All studies published via journals library so should have complete picture
- Includes large/small trials – multiple disease areas
- Reviewed 5 year (2011-2015) inclusive cohort of published outputs
334 outputs

92 trials

45 trials with at least one dichotomous outcome

20 trials with >1 significant dichotomous outcome

13 trials had significant primary binary outcome
Fragility index for cohort of trials

- 13 trials eligible for Fragility Index assessment
- 3/13 had FI of zero
- median FI = 4
- 8/13 trials had FI of five or less
- Remaining 5 trials reasonable FI (ranged from 19 to 167)

Cohort suggests approx 10% of all trials might have a problematic Fragility Index
How should we use this in critical care trials?

- Need to be aware of the possibility of fragile results – especially in critical care

Uses:
- Interpreting newly completed studies
- Informing data monitoring decisions

Routinely consider how fragility of results may change/challenge our interpretation
Newly published trials

**Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension Compared with Usual Care**

Ben Andrews, MD; Matthew W. Semler, MD, MSc; Levy Muchenwa, MB; Douglas C. Heimbürger, MD, MS; Chileshe Mabula, MBChB; Mwango Bwisseur, MBChB

**IMPORTANCE** The effect of an early resuscitation protocol on in-hospital mortality among adults with sepsis and hypotension compared with usual care remains unknown.

**OBJECTIVE** To determine whether an early resuscitation protocol, consisting of intravenous fluids, vasopressors, and blood transfusion, decreases in-hospital mortality among adults with sepsis and hypotension compared with usual care.

**DESIGN, SETTING, AND PARTICIPANTS** Randomized clinical trial (suspected infection plus ≥2 systemic inflammatory response syndrome criteria plus hypotension [systolic blood pressure ≤90 mm Hg] or presenting to the emergency department at a 1500-bed referral center) involving 2207 patients, October 22, 2012, to November 11, 2013. Data collection was completed on November 11, 2013. A total of 1826 patients were randomized, with 913 patients in the protocol group and 913 in the usual-care group.

**CONCLUSIONS AND RELEVANCE** Among adults with sepsis and hypotension, the early resuscitation protocol reduced in-hospital mortality compared with usual care, particularly in patients with severe sepsis and septic shock. These findings may not be generalizable to settings with different levels of resources and care, and further studies are needed to confirm these results.

**Effect of Lung Recruitment and Titrated Positive End-Expiratory Pressure (PEEP) vs Low PEEP on Mortality in Patients With Acute Respiratory Distress Syndrome**

Writing Group for the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial (ART) Investigators

**IMPORTANCE** The effects of recruitment maneuvers and positive end-expiratory pressure (PEEP) titration on clinical outcomes in patients with acute respiratory distress syndrome (ARDS) remain uncertain.

**OBJECTIVE** To determine if lung recruitment associated with PEEP titration according to the best observed systemic compliance decreases 28-day mortality of patients with moderate to severe ARDS compared with a conventional low-PEEP strategy.

**DESIGN, SETTING, AND PARTICIPANTS** Multicenter, randomized trial conducted in 120 intensive care units (ICUs) from 9 countries from November 17, 2011, through April 25, 2017, enrolling adults with moderate to severe ARDS.

**INTERVENTIONS** An experimental strategy with a lung recruitment maneuver and PEEP titration according to the best observed systemic compliance (n = 504; experimental group) or a control strategy of low PEEP (n = 509). All patients received volume-assist control mode until weaning.

**CONCLUSIONS AND RELEVANCE** The experimental strategy was associated with lower mortality than the control strategy (17.3% vs 24.8%). The experimental strategy was also associated with lower incidence of nosocomial pneumonia and ventilator-associated events compared with the control strategy. These findings indicate that lung recruitment and PEEP titration associated with the best observed systemic compliance are associated with improved outcomes in patients with severe ARDS.
Use in DMCs

The scenario ...

- Emerging results of apparent benefit/harm
- Often incomplete data
- Picture fluctuates as data accumulates
- When is the signal strong enough to suggest termination?
- Decision WILL be criticised
Use in DMCs

- Fragility index – can give more confidence that the decision to stop/continue is not unduly fragile
- Would result be likely overturned by missing data?
- Add to the DMC toolkit
Conclusions

• Trial results often fragile (and sometimes VERY fragile)
• Cohort results suggest about 10% of all trials may be affected
• Critical care trials particularly susceptible
• Reporting of fragility index may guard against over-interpretation
• Fragility index can also help in critical care DMC deliberations
Thank you

If you have any further questions please contact:

m.k.campbell@abdn.ac.uk

@MarionKCampbell