Fragility Index: Beyond the P Value

PJ Devereaux, MD, PhD
McMaster University, HHS, PHRI

2 hypothetical RCTs

- Evaluating new treatments to prevent MI
- Placebo controlled
- Outcome MI
- Identical methodology
  - concealed, blinded, complete f/u, ITT
    - high quality

Trial 1	Tx A (n=100)	Placebo (n=100)	P value
MI	1	9	0.02

Trial 2	Tx B (n=4000)	Placebo (n=4000)	P value
MI	200	250	0.02

Question

- What best characterizes your believe in whether there is a real effect of TxA and TxB?
- 1. Similar for both trials
- 2. Substantially more likely for Tx A
- 3. Modestly more likely for Tx A
- 4. Substantially more likely for Tx B
- 5. Modestly more likely for Tx B

Trial 1	Tx A (n=100)	Placebo (n=100)	RRR (95% CI)
MI	1	9	90% (23-100)

Trial 2	Tx B (n=4000)	Placebo (n=4000)	RRR 95% CI
MI	200	250	20% (5-37)

Question

- What best characterizes your believe in whether there is a real effect of TxA and TxB?
- 1. Similar for both trials
- 2. Substantially more likely for Tx A
- 3. Modestly more likely for Tx A
- 4. Substantially more likely for Tx B
- 5. Modestly more likely for Tx B

Small variation in hypothetical trials

- Add 2 events to treatment groups
  - what would happen to p values?
    - 1st trial of 200 patients P value 0.13 (i.e., fragile p value)
    - 2nd trial of 8000 patients P value remains 0.02
- Fragility Index (FI)
  - minimum number of patients required to switch from non-event to event in one group to reverse statistical significance
    - 1st trial FI – 1
    - 2nd trial FI - 9

Trial 1	Tx A (n=100)	Placebo (n=100)	P value	RRR (95% CI)	FI
MI	1	9	0.02	90% (23-100)	1

Trial 1	Tx B (n=4000)	Placebo (n=4000)	P value	RRR (95% CI)	FI
MI	200	250	0.02	20% (5-37)	9

Fragility of trial results

- Highly cited studies in leading journals
  - not uncommonly contradicted (16%)
  - demonstrated to have exaggerated effects (16%)
  - only identified factor explaining these findings
    - outcome in initial trial had a small sample size

What is goal of randomization?

- Achieve balance of prognosis between treatment groups outside of investigational interventions

Balance of prognosis

- Given 9 independent RF associated with MI
  - prevalence of RF varies from 18-65%
  - RF have substantially larger associates with MI
    - (e.g., current smoker OR 2.87) than realistic drug effects
  - not difficult to understand how effect seen in initial 1st trial
    - may be due to imbalance in RFs despite randomization
  - whereas size of 2nd trial minimizes likelihood of meaningful imbalance in RFs that could explain result

Recent experience

- Beta-blocker trial
  - randomized 112 patients
  - 11 deaths
    - 2 beta-blocker group, 9 control group
      - P = 0.02
- POISE Trial
  - randomized 8351 patients
  - 226 deaths
    - 129 beta-blocker group, 97 placebo group
      - P = 0.03

Fragility index in high impact RCTs

- JCE 2014 – 399 RCTs published in high impact journals with statistically significant result
- Median sample size n= 682 (range: 15 - 112,604)
- Median of 112 events (range: 8 - 5,142)
- 53% reported a p-value <0.01
- Median FI 8 (range: 0 - 109)
- 25% had FI $\leq$3
- FI was < number of patients lost to f/u in 53% of trials

Conclusion

- Zeal to turn folks onto RCTs may have resulted in overconfidence in trial results
- Need large sample sizes to achieve balance of prognosis
  - what defines large is focus of ongoing work
- “Positive” trial results frequently hinge on few events
- Fragility Index may improve trial interpretation