

# Understanding the fragility index in experimental clinical studies: An example using the meta-analysis of compression-only v. conventional CPR in out-of-hospital cardiac arrest

Anthony M.-H. Ho, MD\*<sup>‡</sup>; Patrick Norman, MSc<sup>†</sup>; Song Wan, MD, PhD<sup>‡</sup>; Tarit Saha, MD\*<sup>‡</sup>; Glenio B. Mizubuti , MD, MSc\*

A randomized controlled trial comparing two treatments normally defines a null hypothesis of no outcome difference. Researchers calculate the  $p$ -value and the 95%-confidence interval (CI) of the treatment effect. A  $p$ -value of, say, 0.049 means that if there is no difference between the treatments, a difference at least as extreme as the one observed due to chance is  $< 5\%$ . A threshold of 0.05 is a tradition that dates back over a century. Since this chance is small, we reject the null hypothesis within  $\leq 5\%$  level of uncertainty and conclude that there is a difference. The 95%-CI gives an upper-lower bound, which we expect the true effect estimate to lie within a 95% certainty. Another interpretation of the 95%-CI is, if the study is repeated on the same population 1,000 times, 95% of the resulting CIs should include the true difference. A 95%-CI that does not include the null effect size corresponds to rejecting the null hypothesis at a  $p = 0.05$  level. A meta-analysis combines the results from comparable randomized controlled trials to derive a pooled point estimate and the 95%-CI of the treatment effect. Statistical significance is reached if the 95%-CI of the pooled treatment effect does not overlap the null effect.

Instead of the probability of the outcome difference if there was no difference (frequentist approach), Bayesian analysis involves calculating, based on prior beliefs and observed trial results, the probability of an outcome difference being true.

Use of a  $p$ -value of 0.05 to dichotomize statistical “significance” and “nonsignificance” has numerous shortcomings.<sup>1</sup> Nevertheless, a  $p$ -value  $< 0.05$  and a 95%-CI not crossing the null value remain a popular standard. For this reason, it is reasonable to ask just how robust those results are, and what would be required for a  $p$ -value  $< 0.05$  to cross the threshold to become  $> 0.05$  and the 95%-CI to retreat towards the null. Trial results are affected by protocol violations, loss to follow-up, attrition, outliers, random and/or systematic errors, or fraud. For meta-analysis, results could be further influenced by methodology, publication bias, inclusion/exclusion of non-English reports, and the heterogeneity of treatment effect estimates across trials. The fragility index (FI) is a metric introduced to address that concern.<sup>2</sup> The FI of a randomized controlled trial with a statistically significant result is the minimum number of nonevents that need to be changed to events in one arm to switch the result to statistically insignificant.<sup>2</sup> The FI in randomized controlled trials is applicable to studies with 1-to-1 randomization and dichotomous outcomes, and is calculated using Fisher’s exact test.<sup>2</sup> Walsh et al. found that in 53% of 399 randomized controlled trials, which had significant results, the FI is less than the patient attrition rates.<sup>2</sup> The FI in meta-analysis is the minimum number of patients from  $\geq 1$  included trials for which a modification in the event status would change the statistical significance of the pooled

From the \*Department of Anesthesiology and Perioperative Medicine, Queen’s University, Kingston, ON; <sup>†</sup>Kingston General Health Research Institute, Kingston Health Sciences Center, Kingston, ON; and the <sup>‡</sup>Division of Cardiac Surgery, The Chinese University of Hong Kong, Hong Kong, SAR China.

**Correspondence to:** Dr. Glenio Bitencourt Mizubuti, Department of Anesthesiology and Perioperative Medicine, Kingston General Hospital, 76 Stuart Street, Kingston, ON K7L 2V7; Email: [gleniomizubuti@hotmail.com](mailto:gleniomizubuti@hotmail.com).

treatment effect.<sup>3</sup> They found in 906 meta-analyses that the statistical significance of 33% of them depended on the event status of  $\leq 5$  participants from  $\geq 1$  trial.<sup>3</sup> The smaller the number of event change required to flip the  $p$ -value from  $< 0.05$  to  $> 0.05$ , or change the 95%-CI to include the null value, the smaller the FI and the more fragile the result. The value of FI is related to sample size, observed effect size, event rate, and meta-analytical methodology used (e.g., random v. fixed effects, study weighting methods). This makes it difficult to provide a rule-of-thumb for what sample size or magnitude of effect will result in a high FI. In general, studies with a large sample size, high event rate, and/or large observed effect size tend to produce larger FIs. We herein illustrate FI by using a meta-analysis<sup>4</sup> of three randomized controlled trials<sup>5-7</sup> on compression-only CPR by untrained bystanders in out-of-hospital cardiac arrest.

In 2000, Hallstrom et al. published a randomized controlled trial comparing compression-only CPR and standard CPR in out-of-hospital cardiac arrest performed by untrained bystanders receiving instructions over the phone.<sup>5</sup> An intention-to-treat analysis of 241 patients receiving compression-only CPR and 279 patients receiving standard CPR showed that survival to hospital discharge was 14.6% v. 10.4%, respectively ( $p = 0.18$ ).<sup>5</sup> In 2010, two similar randomized controlled trials were published showing that compression-only CPR by untrained bystanders is possibly superior, but again, the differences were not statistically significant.<sup>6,7</sup> A meta-analysis<sup>4</sup> of the randomized controlled trials,<sup>5-7</sup> using intention-to-treat data, showed a risk ratio of 1.22 (95% CI: 1.01–1.46;  $p = 0.040$ ) in favor of compression-only CPR as measured by survival to hospital discharge.<sup>4</sup> The effect estimate, presented as a risk ratio, means that the cumulative “risk” of survival to hospital discharge in the pooled compression-only-CPR cohort is  $1.22 \times$  that of the same “risk” in the pooled conventional-CPR cohort.

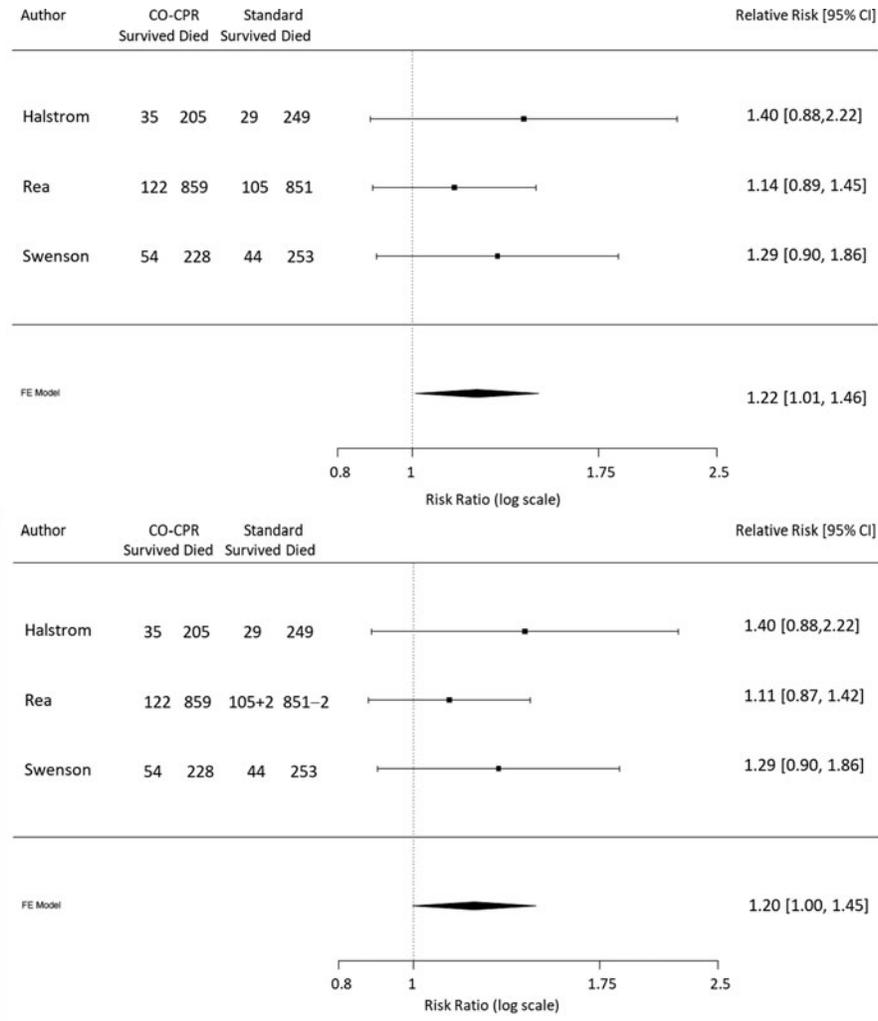
To calculate the FI of meta-analyses, Atal et al. recommended changing the event status of individual trials included.<sup>3</sup> Figure 1 shows that, by small changes in the event status in Rea’s randomized controlled trial, the 95%-CI of the pooled analysis in Hüpfl et al.’s meta-analysis changed to insignificant, yielding an FI of 2.

While a high FI inspires confidence, a low FI should not necessarily mean the lack thereof. Indeed, a randomized controlled trial with a low FI may merely mean that the initial sample size prediction based on an estimated effect size required to achieve a false-positive

error of 5% was accurate. Patient well-being and cost considerations also often mean that a randomized controlled trial is terminated after a statistically significant treatment effect has been demonstrated. In other words, ideally, most randomized controlled trials should have a low FI, as should the meta-analysis they constitute. In practice, however, most studies have flaws. The utility of the FI, therefore, lies in how it compares to potential biases.

In all three randomized controlled trials,<sup>5-7</sup> bystanders were guided over the phone through the standard- or compression-only-CPR steps. In the standard cohort, a bystander begins CPR by performing head-tilt-chin-lift, pinching the nostrils, and delivering two mouth-to-mouth breaths. Laypeople need 16 seconds to perform two rescue breaths in a “sanitized” setting.<sup>8</sup> An anxious naïve layperson might conceivably have taken longer to be taught over the phone to perform those two rescue breaths on a dying stranger. In contrast, bystanders in the compression-only-CPR cohorts started chest compression immediately. Indeed, according to Rea et al.,<sup>6</sup> more bystanders failed to execute all standard-CPR instructions than those failing to execute all compression-only-CPR instructions because of the arrival of emergency personnel  $\sim 5$  minutes later. This suggests that many bystanders in the standard-CPR cohort were possibly still struggling with the initial two-rescue breaths stage after several minutes. Similarly, according to Hallstrom et al.,<sup>5</sup> the arrival of paramedics caused 20.8% of bystanders in the standard-CPR cohort and 7.9% of bystanders in the compression-only-CPR cohort to fail to complete the CPR instructions. In other words, in the standard-CPR group, some bystanders had not progressed to the chest compression stage. For other various additional reasons, the same group reported that 38% of laypeople in the standard-CPR cohort did not successfully receive all phone instructions v. 19% of those assigned to the compression-only-CPR cohort ( $p = 0.005$ ).<sup>5</sup>

There were other potential sources of bias. Svensson<sup>7</sup> reported that a “small” number of dispatchers, possibly believing that sicker patients should receive ventilation, instructed the bystander to perform mouth-to-mouth breathing and chest compression, despite the case being randomized to compression-only CPR. In Rea’s study, 4.8% of those bystanders randomized to standard CPR did not perform mouth-to-mouth breathing.<sup>6</sup> It is not clear whether they simply did not do mouth-to-mouth while waiting for the chest compression phase to begin. Likewise, in Hallstrom’s study, 7.2% of



**Figure 1.** Primary analysis of survival to hospital discharge in randomized trials as reported by Hüpfl et al.<sup>4</sup> (upper Forest plot); crossing the threshold to nonsignificance by event change of 2 in Rea et al.’s trial,<sup>6</sup> yielding a fragility index of 2 (lower Forest plot).

bystanders refused to perform standard CPR v. 2.9% of bystanders refused to perform compression-only CPR.<sup>5</sup>

Clearly, the percentages of protocol violations reported by all three randomized controlled trials – not to mention the executional problems in laypeople performing complex airway and breathing maneuvers based on phone instructions, as alluded to earlier – dwarf the FI of Hüpfl et al.’s meta-analysis. The purpose of this commentary is to illustrate the FI concept and not to contest the merits of compression-only CPR by untrained bystanders. Indeed, the reluctance/difficulties of performing airway + breathing by laypeople present true barriers. The intention-to-treat results of the meta-analysis reflect that reality and show that compression-only CPR is more effective in the real world. Furthermore, there is an abundance of animal and observational data

supporting compression-only CPR by laypeople. Such prior positive data could be combined with the data from the three randomized controlled trials/meta-analysis in a Bayesian analysis.

Putting the FI into context, a *p*-value of 0.05 and a 95% CI are conservative, and, while achieving statistical significance is impressive, a minor step over the artificial demarcation line should not mean that the study is unreliable. Nevertheless, authors of randomized controlled trials should discuss how the FI compares to potential biases, such as dropout and protocol violation rates, and how that may or may not be relevant. Authors of meta-analyses should likewise discuss the FI in the context of their methodology, quality of the included trials, possibility of publication bias, appropriateness of inclusion/exclusion of non-English papers, and so forth.

Authors should also consider a Bayesian interpretation of their results. Interested clinicians should seek out reviews to further understand the FI concept and related indices: the fragility quotient, the susceptibility index,<sup>9</sup> and solidity index.<sup>10</sup> Online calculators are available for FI calculation for randomized controlled trials (e.g., <https://clincalc.com/Stats/FragilityIndex.aspx>).

**Keywords:** Cardiac arrest, clinical decision rules, fragility index, prehospital/EMS, research methods

**Competing interests:** None declared.

## REFERENCES

1. Wasserstein R, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat* 2016;70:129–33.
2. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol* 2014;67:622–8.
3. Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. *J Clin Epidemiol* 2019;111:32–40.
4. Hüpfel M, Selig H, Nagele P. Chest-compression-only versus standard cardiopulmonary resuscitation: a meta-analysis. *Lancet* 2010;376:1552–7.
5. Hallstrom A, Cobb L, Johnson E, Copass M. Cardiopulmonary resuscitation by chest compression alone or with mouth-to-mouth ventilation. *N Engl J Med* 2000;342:1546–53.
6. Rea TD, Fahrenbruch C, Culley L, et al. CPR with chest compression alone or with rescue breathing. *N Engl J Med* 2010;363:423–33.
7. Svensson L, Bohm K, Castrèn M, et al. Compression-only CPR or standard CPR in out-of-hospital cardiac arrest. *N Engl J Med* 2010;363:434–42.
8. Assar D, Chamberlain D, Colquhoun M, et al. Randomised controlled trials of staged teaching for basic life support. Skill acquisition at bronze stage. *Resuscitation* 2000;45:7–15.
9. Majeed M, Agrwal R, Attar BM, et al. Fragility index: how fragile is the data that support the American College of Gastroenterology guidelines for the management of Crohn's disease? *Eur J Gastroenterol Hepatol* 2020;32:193–8.
10. Boyd O. The fragility of evidence in critical care research. Is it just a house of cards? *Crit Care Med* 2016;44:1423–4.