World Health Organization simplified system in emergency department work.

Fred Bernardes Filho, MD
Dermatology Division
Department of Medical Clinics
Ribeirão Preto Medical School
University of São Paulo
Ribeirão Preto, São Paulo, Brazil

Andreia d.O.A. Bernardes, MD
Emergency Department
Hospital Imaculada Conceição da Sociedade Portuguesa de Beneficência
Ribeirão Preto, São Paulo, Brazil

Valéria S. Ferreira, MD
Department of Epidemiological Surveillance
Boa Vista, Roraima, Brazil

Loan Towersey, MD, PhD
AIDS Division
Carlos Tortelly Municipal Hospital
Ministry of Health
Niterói, Rio de Janeiro, Brazil

https://doi.org/10.1016/j.annemergmed.2019.01.022

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

1. Solomon AW, Peeling RW, Foster A, et al. Diagnosis and assessment of trachoma. *Clin Microbiol Rev*. 2004;17:982-1011.
2. Azari AA, Barney NP. Conjunctivitis: a systematic review of diagnosis and treatment. *JAMA*. 2013;310:1721-1729.
3. Luna EJ, Lopes MF, Medina NH, et al. Prevalence of trachoma in schoolchildren in Brazil. *Ophthalmic Epidemiol*. 2016;23:360-365.
4. Talero SL, Resnikoff S, Saboyá-Díaz MI, et al. Giant papillae versus lymphoid follicles of the conjunctiva. *Ann Emerg Med*. 2019;73:694-695.

Fragility Measures: More Limitations Considered

To the Editor:

We commend the work by Brown et al[1] that, using the fragility index and fragility quotient, assesses the fragility of randomized controlled trials in the emergency medicine literature. Although they provide a good overview of the limitations of the 2 metrics, we would like to further question the utility of fragility measures.

In their critique of fragility measures, Carter et al[2] suggested that fragility index is a "P-value in sheep's clothing." In their article, they demonstrated the strong relationship between *P* values, fragility index, and sample size. The results of their study were based on 60,000 simulated clinical trials using a combination of 10 sample sizes (100 to 1,000), with relative risks of 1.0, 0.67, and 0.33 for intervention versus control.[2] They showed that *P* values decrease as sample size increases when the effect size is nonzero; the inverse relationship is observed when fragility index and sample size are compared. Stated another way, if the effect size is held constant and the sample size increases, the fragility index likewise increases.[2]

In their study, Brown et al noted that the overall sample size for included trials was a median of 140 (interquartile range 69.5 to 286), with the overall fragility index for primary outcomes of the 74 included studies equal to 5 (interquartile range 2 to 11.75) and a fragility quotient of 0.039 (interquartile range 0.015 to 0.081). The small median sample size in randomized controlled trials in this study could be one reason for the low fragility index, as mentioned above. Furthermore, as Carter et al[2] observed, randomized controlled trials "are designed to balance the sample size with expected efficacy. In doing so, the [fragility index] is also minimized and results will necessarily hinge on fewer events. This is unavoidable, particularly in the context of clinical equipoise and finite resources."

Although certainly fragility quotient (fragility index normalized to study size) addresses some of these concerns, relative measures of fragility have not proven to be a reliable indicator of study quality, as the authors suggested (for example, the Collaborative Study Group captopril study, which established the use of angiotensin-converting enzyme inhibitors for the prevention of worsening diabetic nephropathy).[3] The primary endpoint of Collaborative Study Group was a doubling of the baseline serum creatinine concentration. Using the methods of Brown et al, we calculated the fragility index of Collaborative Study Group's primary endpoint to be 4, with a fragility quotient of 0.009, or 1 event per 100 patients. Or consider the West of Scotland Coronary Prevention trial, the first major study to demonstrate the efficacy of statins in the primary prevention of nonfatal myocardial infarction or death from coronary artery disease in men with hypercholesterolemia.[4] The fragility index of this study was 34, with a fragility quotient of 0.005. The fragility index and fragility quotient of these studies suggest these trials are quite fragile, yet have proved clinically meaningful in daily practice.

Despite the more recent use of the fragility index and fragility quotient in the literature, we recommend caution in relation to their use until research establishes their utility in clinical context. Until then, their clinical utility for randomized controlled trials remains uninterpretable.

Joshua D. Niforatos, MTS
Cleveland Clinic Lerner College of Medicine of Case Western Reserve University
Cleveland, OH

Alexander R. Zheutlin, BS
University of Michigan School of Medicine
Ann Arbor, MI

Richard M. Pescatore, DO
Department of Emergency Medicine
Crozer-Keystone Health System
Upland, PA

https://doi.org/10.1016/j.annemergmed.2019.01.035

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

1. Brown J, Lane A, Cooper C, et al. The results of randomized controlled trials in emergency medicine are frequently fragile. *Ann Emerg Med*. 2019;73:565-576.
2. Carter RE, McKie PM, Storlie CB. The fragility index: a P-value in sheep's clothing? *Eur Heart J*. 2017;38:346-348.
3. Lewis EJ, Hunsicker LG, Bain RP, et al. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. Collaborative Study Group. *N Engl J Med*. 1993;329:1456-1462.
4. Shepherd J, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*. 1995;333:1301-1308.

In reply:

We thank Niforatos et al for their contribution to the discussion in regard to the limitations of the fragility index and fragility quotient. In their letter, they expand on some of the limitations to the application of the fragility index and fragility quotient discussed in our original work and ultimately question their utility. Although they highlight some important considerations, we believe that the measures have clinical utility. Here, we explore further some key points about their application and provide insight into how clinicians might use these tools.

The fragility index is a metric that provides the absolute number of events necessary to reverse statistical significance. We note that because its calculation relies on *P* values, the fragility index is subject to the same limitations inherent to *P* values themselves. Carter et al[1] expanded on this notion and illustrated how *P* values are repackaged into the calculation of the fragility index to generate an inverse relationship between the two. Thus, as is true with *P* values, the fragility index should not be misinterpreted as a measure of clinical effect. Narayan et al[2] noted that "just as a *P* value of 0.001 compared with 0.04 does not signal a greater importance of the findings, a fragility index of 10 should not be interpreted to imply greater clinical effect than a fragility index of 1; rather, it simply illustrates the strength of the statistical significance itself." Clinicians must therefore recognize that the fragility index represents an easily interpretable measure of the stability of statistically significant findings and not a measure of clinical effect.

Because the fragility index provides the absolute number of events required to reverse significance, it has particular utility in interpreting study findings in the context of loss to follow-up. When the number of patients lost to follow-up exceeds the fragility index, results should be viewed with particular skepticism because factoring in the unknown outcomes of those patients might reverse significance.[3] Moreover, the fragility index should not be interpreted in isolation. Rather, it complements the *P* value and confidence interval.[4] When interpreted in conjunction with the aforementioned and paired with other statistical measures such as the number needed to treat, the fragility index may offer further insight into the reliability of statistically significant findings. Carter et al[1] recommended yet a broader approach to study interpretation that encompasses estimated efficacy, study design, and mitigation of biases. This may be best accomplished by abandoning the frequentist approach altogether in favor of alternative methodology such as Bayesian analysis. However, as long as researchers insist on a frequentist framework, added metrics such as the fragility index may provide further insight into the stability of statistically significant results.

Although the fragility index has limitations, it also has clear utility as an intuitive measure of the reliability of study findings when interpreted in conjunction with the *P* value and confidence interval. We maintain that the fragility index should be normalized to sample size with the fragility quotient. Few studies have reported on the fragility quotient, however, and we are unaware of any that have investigated its reliability. Thus, we encourage further research on the concept of fragility as a whole.