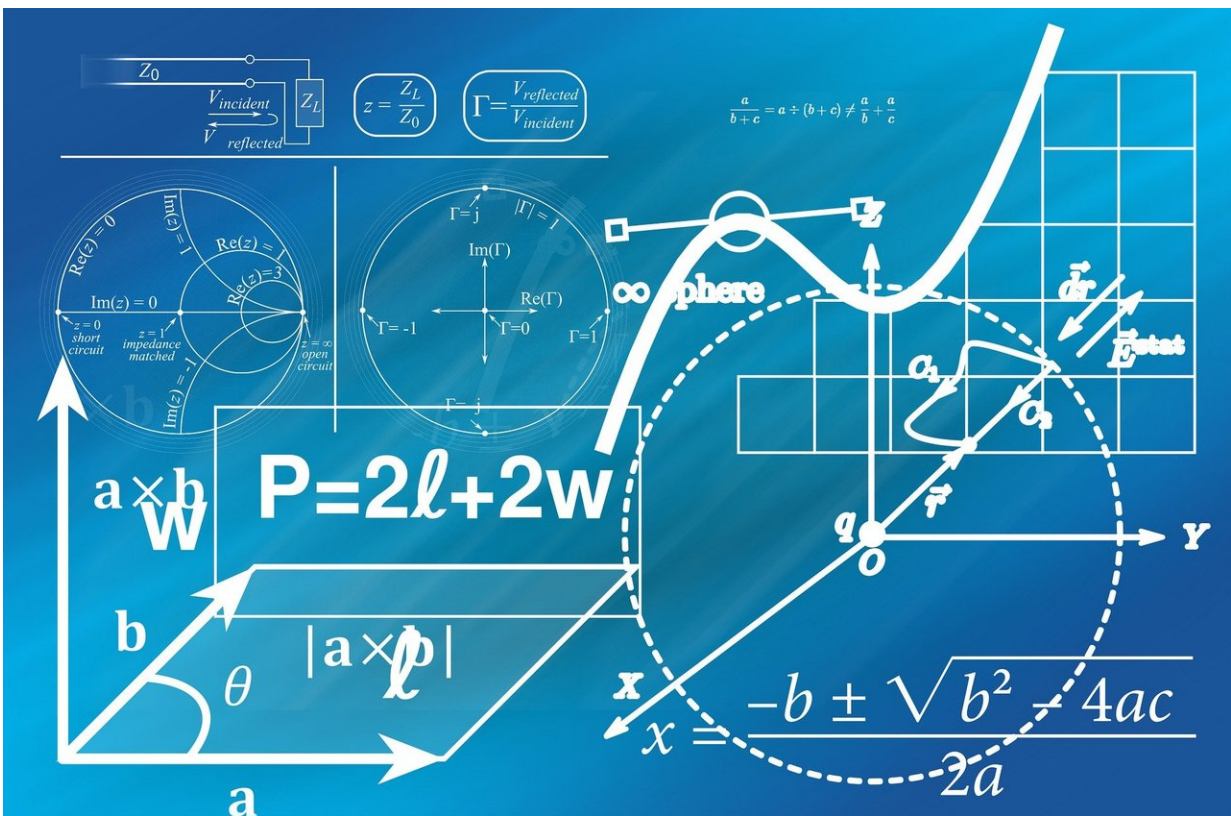


Alternative statistical method could improve clinical trials

December 21 2021



Credit: CC0 Public Domain

An alternative statistical method honed and advanced by Cornell researchers can make clinical trials more reliable and trustworthy while also helping to remedy what has been called a "replicability crisis" in the

scientific community.

In a new paper published this month in the *Proceedings of the National Academy of Sciences*, Cornell researchers further the "[fragility index](#)," a method gaining traction in the [medical community](#) as a supplement to the p-value, a probability measurement applied across science since the 1920s and cited, sometimes recklessly, as evidence of sound results.

"Clinicians trust that the procedures and protocols they carry out are informed by sound, [clinical trials](#). Anything less makes surgeons nervous, and rightly so," said Martin Wells, the Charles A. Alexander Professor of Statistical Sciences and a paper co-author. "We're discovering that many of these consequential [trials](#) that showed promising results and that were published in top journals are fragile. That was a disconcerting surprise that came out of this research."

The paper, written by statisticians from Cornell and doctors from Weill Cornell Medicine and the University of Toronto, proposes a new statistical toolkit using the fragility index as an alternative method to help researchers better determine if their trial results are, in fact, strong and reliable or merely a product of chance.

"When you tell the world a treatment should or shouldn't be used, you want that decision to be based on reliable results, not on results that can swing one way or another based on the outcomes of one or two patients," said Benjamin Baer, Ph.D. '21, a paper co-author and currently a postdoctoral researcher at the University of Rochester. "Such results can be considered fragile."

Randomized, clinical trials to test effectiveness are essential for surgical procedures and medical treatments. To interpret the statistical significance of trial results, researchers for decades have turned to an often-misunderstood measure, the p-value, to determine whether results

have merit or are just a chance occurrence.

But skepticism surrounding the p-value's reliability, when used on its own and without supporting methods, has grown in the last 15 years, particularly as past trial results initially deemed strong couldn't be replicated in follow-up trials. In a 2014 study using the fragility index, researchers analyzed 400 randomized clinical trials and found that 1 in 4 trials with "statistically significant" p-values in fact had alarmingly low fragility scores, indicating less reliable results.

"One can see why there is a replication crisis in science. Researchers find good results, but they don't hold up," Wells said. "These are serious, large trials studying cutting-edge issues, with findings published in top journals. And yet, some of these big trials have low fragility indices, which raises the question of the results' reliability."

With their latest research, Cornell scholars offer a solution by honing the fragility index, which investigates what number of patient outcomes could tip a trial either successful or unsuccessful. The lower the fragility number, the more fragile and unreliable the results. For example, a trial with 1,000 participants that flips either statistically significant or insignificant based on the results of a few patient outcomes has an extremely low fragility index.

Since it emerged in the 1990s, the fragility index has been criticized for rigidity—it's only applicable for data with two study groups, treatment and control, and a binary, event-or-not outcome. This latest research offers a more flexible fragility index that can be applied to any type of outcome and to any number of explanatory variables.

The team's method also gives researchers across science the ability to calculate the [fragility index](#) based on the likelihood of particular outcomes.

"The traditional framing of statistical significance in terms of yes-no is overly simplistic, and the problems we're investigating aren't," said Dr. Mary Charlson, the William Foley Distinguished Professor of Medicine at Weill Cornell Medical College and a paper coauthor. "With each clinical situation, there are different contexts you're dealing with. This method allows us a way to test assumptions and consider implications of a much narrower range of outcomes."

More information: Benjamin R. Baer et al, Fragility indices for only sufficiently likely modifications, *Proceedings of the National Academy of Sciences* (2021). [DOI: 10.1073/pnas.2105254118](https://doi.org/10.1073/pnas.2105254118)

Michael Walsh et al, The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index, *Journal of Clinical Epidemiology* (2014). [DOI: 10.1016/j.jclinepi.2013.10.019](https://doi.org/10.1016/j.jclinepi.2013.10.019)

Provided by Cornell University

Citation: Alternative statistical method could improve clinical trials (2021, December 21) retrieved 26 April 2024 from <https://phys.org/news/2021-12-alternative-statistical-method-clinical-trials.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.