Is The Fragility Index A Clinical Guide In Our Use of Guidelines?

By Chuck Dinerstein — November 7, 2018

Randomized clinical trials remain the “gold-standard” evidence in making standards, that is the creation of clinical guidelines. They are a form of best practice creating a “floor” for the care patients receive. But as I have said before, it is difficult for a physician to always apply guidelines for populations to the patient sitting in front of me. Is my patient similar enough to the population studied to use the guidelines; or put another way how much do I have to push to get that square peg in the round hole?

How can clinician’s evaluate evidence?

We all know at this point that a study’s p-value may or may not be credible for a variety of reasons. We also know that even when a study has a genuinely significant statistical finding that it may have no clinical consequence and can be safely ignored. So how can I figure out how much squareness in the peg is allowable – how applicable is the recommendation outside the population that was studied? A study in JAMA Surgery offers clinicians a simple metric, the Fragility Index or Quotient. (FI)

The Fragility Index is a variation of sensitivity analysis, where you vary the values of your assumptions and see whether your decision changes. For example, you might look at the decision to operate on someone you assume will leave for the next ten years and find that the “answer” is yes; but when you change that assumption to one year, the answer is no – sensitivity analysis shows you how fragile or robust your decision is to your assumptions. FI looks only at the overall results and determines “the minimum number of patients whose status would have to change” from one outcome to another, “to make the study lose statistical significance.” In one of the
studies, the author cited it only required two patients to flip from significant to insignificant. The Fragility Quotient is nothing more than the same calculation divided by the number of study participants – a way of adjusting for populations; as with FI, FQ numbers that are smaller are more fragile.

The authors go on to discuss the use of this “scoring” in trauma guidelines, finding that many of the surgical guidelines were fragile, not widely applicable. In part this reflected small sample sizes a more significant problem for surgical studies than cardiac diseases, FI values were more robust in those studies, presumably in part because enrollment was easier.

As with any measure, FI has its limitations. It does not serve as well as sensitivity analysis does for observational studies; it requires a dichotomous result and applies only to primary outcomes. Moreover, no clearly defined FI value says a study is too fragile to use.

I think the Fragility Index is worthwhile to share with the recipients of guidelines, clinicians, who like myself, know enough statistics to be more dangerous than knowledgeable. And for those pushing meta-analysis, we could undoubtedly calculate an aggregate of FI to share along with those p-values.

Perhaps if the FI was large I, along with my colleagues might make a bit more effort to squeeze that square patient into that round guideline.

Source: The Fragility Index in Randomized Clinical Trials as a Means of Optimizing Patient Care JAMA Surgery DOI: 10.1001/jamasurg.2018.4318