How small studies mislead us

Joanna Chikwe, MD, a,b and Shinobu Itagaki, MS, MD a

What is the problem with small studies? Our specialty evolved through change and innovation published as case reports and series, and although these have been augmented recently by increasingly large registries and randomized trials, most cardiothoracic research still consists of comparisons of a few hundred patients.

For example, in this issue of the Journal, Patel and colleagues1 report 0.5% mortality after coronary artery bypass grafting and 0.5% mortality after a hybrid procedure in 207 propensity-matched patients pairs (P = 1.0), concluding these procedures are associated with equivalent outcomes. The problems with this kind of study have recently been highlighted in critiques (Figure 1) suggesting that most conclusions drawn from such research are unreliable for reasons including low statistical power.2,3

Imagine a manufacturer making thousands of cars at 2 factories. The manufacturer discovers that 0.5% of the cars made at 1 factory have a lethal construction fault. They stop production and recall all cars made at that factory. How many cars made at the second factory should they recall, and how many faults should they find, to prove both factories have the same problem?

Recalling and examining every car made by the second factory is the only way to be 100% sure. An alternative is recalling several thousand cars: the same 0.5% fault rate in a sample of that size indicates the factories very likely have the same problem, whereas forming this conclusion after finding one fault in a 200-car sample from the second factory is the equivalent of the study by Patel and colleagues1. A false negative is reduced (and the power of the study to identify a true difference is increased) if the sample is very large, or the event rate being compared is very high.

Therefore, there is a real possibility that hybrid revascularization could result in much better or much worse outcomes than conventional surgery—we simply cannot tell because the number of events in the study by Patel and colleagues1 is too small for reliable comparison. Either the study needs more patients (thousands), a more frequent primary endpoint (such as a composite of long-term mortality, repeat revascularization, stroke, and myocardial infarction), or both.

This is why the randomized Hybrid Coronary Revascularization Trial is designed to recruit 2354 patients, comparing a composite end point (5-year all-cause mortality, myocardial infarction, stroke, and unplanned revascularization).4 The trial is designed to have sufficient patients and “power” to reduce the risk of a false negative is reduced (and the power of the study to identify a true difference is increased) if the sample is very large, or the event rate being compared is very high.

Central Message

Most conclusions drawn from small comparative studies are unreliable, primarily because low statistical power increases the risk of false-negative results.

See Article page 1799.
false-negative result. Low statistical power is first in a catalogue of reasons why small studies frequently mislead us.

References