Disclaimer: This column is based almost exclusively on a video, “Evidence-based Medicine or Evidence Bias” by Michael Greger, MD, FACLM, that was posted by NutritionFacts.org in 2014 and is available at https://nutritionfacts.org/video/evidence-based-medicine-or-evidence-biased/ (last accessed 21 April 2018).

Sometimes in life, bigger is not always better. In fact, there are times when “bigger” can border on unethical—I’m talking here about sample size in clinical trials. When a statistical difference is found between outcomes for groups A and B, both sampled from a larger population but randomized to different interventions, the likelihood that the difference was due to chance (that is, that there was no actual difference between groups) is less than 5 in 100 (assuming a $P$ value < .05). Since the introduction of evidence-based medicine in the early 1990s, medical practice has been increasingly shaped only by large ($N \geq 200$) randomized clinical trials in which the results were statistically significant but for which the effect size—that is, the size of change in outcome produced by the intervention—was rather small. But what about small studies in which the effect is very dramatic? Is it even ethical to proceed with a randomized clinical trial following such a study?

Rabies is an extreme example of this scenario. Since early times, rabies infection had been recognized as universally fatal and had been described as “a most awful death.” In 1885, 9-year-old Joseph Meiste, who was mauled by a rabid dog, was given a new vaccine and became the first human to survive the illness. Neither the efficacy nor the safety of the vaccine had been shown in a randomized clinical trial, but, after treatment of 1 additional patient (who also survived), the vaccine was accepted worldwide.

A slightly less extreme example involves a study from the 1980s with a small number of patients but a dramatic effect: the introduction of extracorporeal membrane oxygenation (ECMO) in neonatology. A highly invasive technique, ECMO has been used in neonates with severe, reversible cardiac and respiratory illness unresponsive to maximal medical therapy. Pioneers in the field noted that in very ill neonates (those with a mortality risk $\geq 80\%$), ECMO therapy upended the outcome, such that survival was now 80%. These results looked good—almost too good. Even though researchers felt strongly that they were “condemning” some children to death, they were also concerned that their colleagues would not accept this undeniably risky procedure without a randomized controlled trial. After the fourth death in the control group ($n = 10$), the trial was terminated. At that time, 9 of 9 neonates receiving ECMO had survived.

In 1995, Esselstyn et al. published a more controversial dietary intervention in which patients with angiographically documented, severe coronary artery disease were coached on a plant-based diet containing $<10\%$ fat. Of the 22 patients enrolled, 5 dropped out within 2 years, and 11 of the remaining 17 maintained the diet through a mean of 5 years. These 11 patients reduced their serum cholesterol levels and saw an angiographic reduction in or stabilization of their coronary artery lesions. None had a new cardiovascular event. In contrast, the 5 patients who resumed their pre-study diet had an additional 10 cardiovascular events. Extended follow-up through 10 years showed that the 6 patients who had continued the extremely low-fat diet had no further coronary events. This study demonstrated a dramatic response to dietary intervention but was roundly criticized for its small sample size and for not having been a blinded study.

We’ve grown accustomed to seeing larger randomized controlled trials that are rigorously blinded (not possible in most outpatient nutritional trials) but that often show only a modest effect compared with standard of care—yet these studies often change practice patterns. The Esselstyn et al. study...
cle described a 100% response rate with no safety issues. If a plant-based diet was a pill and the control therapy was the average American diet, scientists would argue whether it was ethical to even have a control group in future studies of the intervention.

Traditionally, phase 3 trials have committed sponsors to a large sample size with no possibility of early termination for benefit, harm, or futility. If a phase 2 trial suggests that an intervention is markedly beneficial, is it ethical to randomize half or a third of phase 3 patients to placebo before we are more certain of the interventional outcome? Or if early studies suggest that the intervention is no different from placebo or even harmful, can we argue that it is ethical to randomize patients to the intervention in order to confirm initial impressions?

Adaptive design allows for an interim analysis in which accumulating data are used to decide how to modify aspects of the trial as it continues, without undermining validity and integrity. “In such trials, changes are made ‘by design,’ and not on an ad hoc basis; therefore, adaptation is a design feature aimed to enhance the trial, not a remedy for inadequate planning,” wrote Dr Paul Gallo in a 2006 publication. Many adaptations are possible, including modification of dose, treatment, randomization, sample size estimation, early stopping rules, hypothesis, primary outcome variable, eligibility criteria, or statistical analysis plan. There are pitfalls associated with each adaptation, so adaptive trial design requires careful and extensive planning. Even with that planning, adaptive design still poses multiple concerns, including increasing the probability that the drug will be deemed effective when it is not, introducing unanticipated sources of bias, or misinterpreting results. Also, there are logistical and procedural issues associated with adaptive design.

Done right, interim analyses may help us straddle the gap between small studies with large effect and large studies with small effect. Ethical trials should include only as many patients as are needed to answer a specific clinical question. It is easy to get caught up in statistical significance; but statistical calculations should be used to identify a plan that will be effective in answering the clinical question—not the other way around.

Author contact: tballmd@gmail.com

References