LETTERS TO THE EDITORS

Hypothesis testing complexity in the name of ethics: Response to commentary

The issue of test sidedness in hypothesis testing for clinical trial analyses has been the subject of debate in the medical and statistical literature [1–4], dividing clinical investigators’ deep-seated beliefs in therapy effectiveness from their obligatory prime concern for patient welfare. J.A. Knottnerus and L.M. Bouter, in their commentary “The ethics of sample size: two sided testing and one-sided thinking” [5], advance the thesis that for individual trials, especially those evaluating new interventions not previously studied, a one-sided hypothesis test seems sensible from the perspective of both ethics and cost (sample size) efficiency. However, we argue here that one-tailed testing should be avoided in health care research, especially in randomized trials in which the investigator controls the intervention. Rather than reflect the investigators’ a priori intuition or suggestive information from prior studies, the type I error should reflect the uncertainty of the research effort’s future conclusions.

The two-sided test shines bright, direct light on the possible production of harm by the intervention, a critical illumination in a field in which health care practitioners and health care researchers inadvertently do harm to their patients. As clinical researchers, we do not like to harbor the notion that the interventions we have developed for the benefit of our patients can produce harm. Nevertheless, because harm is often the result of our best intentions, health care practitioners and researchers must be ever vigilant for its occurrence. The more strongly we believe in (but remain uncertain about) the benefit of a therapy, the more observant we must become for the unsuspected occurrence of harm. We argue that it is an ethical obligation to know that harm has not been caused in a research endeavor since this information can directly impact clinical or public health action. Furthermore, only a design that accepts the possibility of harm also allows prospective monitoring and early stopping rules for harm if indeed it occurs.

An argument raised in defense of one-sided testing is sample size efficiency. However, this reduction in sample size in a clinical experiment produced by carrying out a one-sided hypothesis test comes at the price of being unable to draw appropriate conclusions in the population if the investigators are wrong, and the study demonstrates harm in the sample. Thus, although less efficient, the experiment designed for a two-tailed hypothesis test is more effective by removing the necessity of repetition (with its attendant ethical dilemma) when the findings of harm and not benefit are produced. Furthermore, because medications and other interventions even when novel often acquire additional indications, the detection of harm for one end point when studies of interventions do not yield expected benefits for another end point provides useful information. Referring to Knottnerus’ example, this additional useful information is obtained in a single two-sided experiment using only 63% of the total sample size used if a second one-sided test had to be conducted: 199 per group for the two sided versus 157 per group × 2 for two-one-sided tests. Knottnerus is correct that more patients may be exposed to the control therapy and perhaps receive the inferior treatment in a two-sided test, but this criticism is blunted by the investigators’ use of prospectively designed monitoring rules that can help terminate the study prematurely in light of early strong evidence of benefit as well as for harm.

The intelligent application of the two-tailed test requires deliberate, overt effort to consider the possibility of patient harm during the design phase of any experiment. This concern, expressed early and formally in the trial’s design, can be very naturally translated into effective steps to ensure patient safety during the course of the experiment. In circumstances where clinical intuition is overwhelmingly in favor of a finding of benefit, the investigators should exert the required discipline to provide adequate ability to determine if the intervention produces harm. It is fine to hope for the best, as long as we prepare for the worst. Although the two-sided hypothesis test can complicate experimental design, complexity in the name of ethics is no vice.

Lemuel A. Moyé
Alan T. N. Tita
University of Texas
School of Public Health
Houston, TX 77030

References