



Invited Commentary: Ethics and Sample Size—Another View

Ross Prentice

From the Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Received for publication August 5, 2004; accepted for publication September 15, 2004.

In their article entitled, “Ethics and Sample Size,” Bacchetti et al. (1) provide a spirited justification, based on ethical considerations, for the conduct of clinical trials that may have little potential to provide powerful tests of therapeutic or public health hypotheses. This perspective is somewhat surprising given the longstanding encouragement by clinical trialists and bioethicists in favor of large trials (2–4). Heretofore, the defenders of smaller trials have essentially argued only that small, underpowered trials need not be unethical if well conducted given their contribution to intervention effect estimation and their potential contribution to meta-analyses (5, 6). However, Bacchetti et al. evidently go further on the basis of certain risk-benefit considerations, and they conclude: “In general, ethics committees and others concerned with the protection of research subjects need not consider whether a study is too small.... Indeed, a more legitimate ethical issue regarding sample size is whether it is too large” (1, p. 108).

What rationale leads Bacchetti et al. to conclusions that seem so dramatically opposed to prevailing standards for high-quality clinical trial research? They begin with the well-established element of ethical clinical research that the potential benefits to individuals and knowledge gained for society must outweigh the risks (2). Like several previous writers, Bacchetti et al. then note that risks and benefits of research are not readily quantifiable on comparable scales, but they assert that they have a means of studying “the influence of sample size on the ethical balance in a way that does not depend on specific calculations or on the specific way in which burden or value is measured” (1, p. 106). To do so, they consider a “projected net burden” (1, p. 106) for study subjects that quantifies the difference between risks and any direct (nonsocietal) benefits to the individual. This quantity is regarded as a fixed number, independent of sample size, as seems a sensible assumption. To quantify the projected value to society, Bacchetti et al. consider the per participant contribution to study power and note that this tends to diminish with increasing sample size. They illustrate their risk-benefit framework by a figure (their figure 1) in which a horizontal line representing a specified projected net burden per partic-

ipant is imposed on the power per participant curve for a comparison of two normal means as a function of sample size per group. The intersection of the two lines, in this illustration, defines studies of sample sizes of about 130 or less per group as having a favorable benefit-versus-risk profile and hence “ethical sample size” (1, p. 106), whereas larger studies would evidently be labeled as unethical. One can note that the power at the largest “ethical” sample size of 130 per group is $260 \times 0.002 = 0.52$, so that in this scenario the authors’ arguments imply that only trials having power less than 52 percent are ethically defensible! In addition, if one thinks of a sequence of such trials that may be ultimately combined in a meta-analysis, a subsequent trial to a trial with 130 participants would presumably be “unethical” according to the framework of Bacchetti et al. because the (conditional) power per participant would fall below the corresponding net burden. These types of implications seem quite unappealing and counterintuitive and cause one to question their societal value formulation.

In fact, Bacchetti et al. raise this question themselves in writing that they adopted the “questionable premise that a study’s projected value is determined only by its power” (1, p. 106). This premise implies, for example, that the maximum societal benefit of a well-conducted trial is a fixed quantity that needs to be divided up among the study participants!

A contrasting view, more attractive to this writer, is that the value to a participant from his or her altruistic contribution to a definitive study of an important clinical or public health question is relatively independent of the number of trial participants. More generally, as a function of sample size, one might expect the projected value per participant to start low since there is modest benefit from a trial (in isolation) that is insufficient to affect medical or public health practice, then to be relatively constant over a range of sample sizes that have potential clinical impact, and eventually to decline beyond sample sizes where the research question will have been reliably answered (though even larger sample sizes may contribute valuable additional information, for example, about intervention effects in important subsets of the study population or concerning treatment contrasts for

Correspondence to Dr. Ross Prentice, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M3-A410, P.O. Box 19024, Seattle, WA 98109-1024 (e-mail: rprentic@whi.org).

rarer outcomes). This shape of a value per participant curve would tend to define an interval of preferred sample sizes, leaving out sample sizes that are too small unless there are specific well-developed plans for subsequent related trials or meta-analyses, and possibly leaving out trials that are unnecessarily large to answer key research questions if there are noteworthy risks or burdens to study participants.

In summary, it seems to me that Bacchetti et al. have provided a service by encouraging the reader to think of value and burden on a per participant basis, and by arguing that the projected net burden per participant may often be largely independent of sample size. However, the “questionable premise” they adopt to reach their conclusions, namely, that projected value per participant can be specified by study power divided by sample size, seems inherently flawed and leads to the misleading conclusion that small trials quite generally have favorable benefits compared with risks. Whether or not a projected value per participant curve can ever be quantified on a comparable scale to a net burden per participant line, I think that a value per participant curve that increases with sample size, at least up to samples sizes giving power in traditional study-planning ranges (80–95

percent), will better reflect the views of clinical trialists and biomedical researchers and will reinforce longstanding good advice in favor of sample sizes that are large enough to identify plausible intervention effects with high probability.

REFERENCES

1. Bacchetti P, Wolf LE, Segal MR, et al. Ethics and sample size. *Am J Epidemiol* 2005;161:105–110.
2. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA* 2000;283:2701–11.
3. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358–62.
4. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;48:23–40.
5. Edwards SJL, Lilford RJ, Brauholtz D, et al. Why “underpowered” trials are not necessarily unethical. *Lancet* 1997;350:804–7.
6. Lilford R, Stevens AJ. Underpowered studies. *Br J Surg* 2002; 89:129–31.