Published in final edited form as:

Stat Med. 2012 December 20; 31(29): 4140–4141. doi:10.1002/sim.5578.

Power, Ethics and Obligation

Jonathan Adam Gelfond,

UT Health Science Center San Antonio, Department of Epidemiology and Biostatistics, Mail Code 7933, 7703 Floyd Curl Drive, San Antonio, Texas 78229-3900, U.S.A.

Elizabeth Heitman,

Vanderbilt University Medical Center, Center for Biomedical Ethics and Society, 2525 West End Avenue, Suite 400, Nashville, Tennessee 37203, U.S.A.

Brad H. Pollock, and

UT Health Science Center San Antonio, Department of Epidemiology and Biostatistics, San Antonio, Texas 78229-3900, U.S.A.

Craig H. Klugman

UT Health Science Center San Antonio, Department of Medicine, San Antonio, Texas 78229, U.S.A.

Consideration of statistical power is critically important when assessing the ethicality of a study. The Nuremberg Code's point #2 (the experiment should give meaningful results) and point #6 (the humanitarian importance of the problem must justify the participant's risk) [1] offer strong support of this connection between power and ethics. The underlying principle is that for any given outlay of human risk or resources there is an obligation to maximize the power and efficiency of the experimental design. For example, of two experiments of equal cost with similar scientific and clinical importance the one with higher power is preferable because it minimizes risk and maximizes benefit. The determination and definition of power and efficiency, however, are not without controversy.

Bacchetti, et al., state that “underpowered” studies are not necessarily unethical. Their argument focuses on sample size as the determinant of power, and their argument hinges on the definition of the term underpowered, which they have interpreted to mean that the power of a study is less than some arbitrary value such as 80%. There are several straw men that are brought forward by Bacchetti, et al. First, our article does not state that underpowered studies are “completely worthless” or “necessarily unethical”, nor does it state that a study “must be definitive” as Bacchetti et al., incorrectly suggest. We even argued that a sample size of six could be too large [2]. Second, we did not mention or support the arbitrary cutoffs that Bacchetti, et al., rail against. Third, the article does not specify their definition of underpowered. Bacchetti, et al., offer a quantitative measure of cost efficiency related to ethicality [3, 4]. Any such formulaic conception of ethicality should be scrutinized, and notably, several other authors have significant critiques [5–7] to their formulation of ethicality that go beyond mere “misconceptions.” As these critiques are previously expounded upon at length, we do not repeat them here, but we do quote Simon (3) who writes that for Phase III trials “there is usually strong interest in obtaining a definitive result,” and we disagree that there is no lower limit for sample size that would result in a

study being “underpowered” by some measure. Bacchetti, et al., also repeat previous arguments about recovering meaningful conclusions from low precision studies through possible future meta-analyses. This meta-analysis strategy unduly depends on future studies actually being performed, ignores complications of cross-study heterogeneity that would hinder combinability of these studies, and is seemingly not time- or cost-efficient for testing a particular hypothesis. Lastly, Bacchetti, et al., mention the manipulation of power calculations (referred to as “sample size samba”). We agree that this manipulation is problematic, and we find that it violates several of our principles [8] including accuracy (3.6), objectivity (3.3), and openness and transparency (3.4).

Nevertheless, no matter what the definition of study efficiency, power is determined by the complete study design that includes many factors other than sample size, and one could define underpowered designs as having less power than the optimal feasible design, where the optimal design is determined by some efficiency criterion. Given this definition of underpowered, we could revise our statement in the article (Edits in *italics*) to “Underpowered studies are *less* likely to yield results with practical translational value; they *may both* put subjects at unnecessary risk and waste resources.” Finally, we reiterate that for any given outlay of human risk or resources there is an obligation to maximize the power and efficiency of the experimental design.

References

1. Shuster E. The Nuremberg Code: Hippocratic ethics and human rights. *Lancet*. 1998; 351(9107): 974–977. [PubMed: 9734958]
2. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, et al. Cytokine storm in a phase I trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med*. 2006; 355(10):1018–1028. [PubMed: 16908486]
3. Bacchetti P, McCulloch CE, Segal MR. Simple, defensible sample sizes based on cost efficiency. *Biometrics*. 2008; 64(2):577–585. [PubMed: 18482055]
4. Bacchetti P, Wolf LE, Segal MR, McCulloch CE. Ethics and sample size. *Am J Epidemiol*. 2005; 161(2):105–110. [PubMed: 15632258]
5. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *Jama-Journal of the American Medical Association*. 2002; 288(3):358–362.
6. Halpern SD, Karlawish JHT, Berlin JA. Re: “Ethics and sample size”. *Am J Epidemiol*. 2005; 162(2):195-U1. [PubMed: 15972938]
7. Prentice R. Invited commentary: Ethics and sample size - Another view. *Am J Epidemiol*. 2005; 161(2):111–112. [PubMed: 15632259]
8. Gelfond JAL, Heitman E, Pollock BH, Klugman CM. Principles for the ethical analysis of clinical and translational research. *Statistics in Medicine*. 2011; 30(23):2785–2792. [PubMed: 21751225]