**A few thoughts on Confidence Levels for Pilot Studies**

William V. Harper, 2/18/2014

To perform a sound statistical study many numerous issues involving experimental design should be addressed prior to data collection. One aspect of this is what sample size is needed to obtain the desired controls on Type I and Type II errors. Without going into details here, the probability of a Type I error = α and the probability of a Type II error = β (or 1 – power). All of this gets very technical but are key components of serious research to (for example) provide the FDA sufficient information to assess a new drug application. Other components such as the number of factors to be assessed, the variability for the type of data, the type of data itself (e.g., continuous data, proportions) are part of what is commonly called a Power analysis in the biostatistics field.

Many studies do not have the resources or potential patient basis to obtain sample sizes rigorously. Often the researcher does the study without evaluation of such concerns. This is the norm in small academic studies at Otterbein that I have aided over the last 15+ years.

To address this it is my practice to treat the project as a pilot study. In this sense the work is not aimed at journal publication but rather to do a preliminary assessment of what factors show promise for a potential more in-depth study later.

In doing such a pilot study, the α level is relaxed from its often world-wide default of 0.05 to something larger. In my engineering studies and student based work I am willing to relax it to α = 0.20. In doing so it is possible to find more statistically significant findings at the danger of an increased Type I error. What is a Type I error? A Type I error is the rejection of the null hypothesis Ho when it is true. In a similar vein a Type II error is not rejecting Ho when it is false. As α goes up, the P(Type II error = β) goes down. These two (α, β) always have this inverse relationship.

A quick web search this morning (2/18/2014) found 3 biostatistics references that in some sense deal with this issue. The web reference is given below along with a small subset of the material provided.

**1st Source**

<http://aje.oxfordjournals.org/content/138/11/923.abstract>

[**American Journal of Epidemiology**](http://aje.oxfordjournals.org/), [Volume 138, Issue 11](http://aje.oxfordjournals.org/content/138/11.toc), Pp. 923-936

© 1993 by The Johns Hopkins University School of Hygiene and Public Health

Simulation Study of Confounder-Selection Strategies, George Maldonado and Sander Greenland

Last line of abstract: The significance test strategies performed best when the alpha level was set to much higher than conventional levels (0.20).

**2nd Source**

**http://www.uth.tmc.edu/uth\_orgs/educ\_dev/oser/L2\_2.HTM**

**Biostatistics for the Clinician**



**University of Texas-Houston
Health Science Center**

**Lesson 2.2**

**Hypothesis Testing**

#### Factors Affecting the Choice of Alpha and Power

That is, your experiment is designed in a way that you have a 10% chance of a false positive. It doesn't matter if you misapply this drug. It's not going to hurt anybody. So, on the one hand, you may want to have a .01 a point .001 alpha level if the drug has nasty side effects. Or, you may want to have a .10 alpha level if you are doing a pilot study.

Note by Bill: Here they are talking about potentially dangerous drugs. In the small academic studies (Allied Health, Nursing) often done at Otterbein University, there is no danger involved. In the spirit of a screening study, I let α be 0.20 for very small pilot studies. Otherwise, due to the very small sample size, it is virtually impossible to find statistical significance.

**3rd Source**

<http://www.ats.ucla.edu/stat/seminars/Intro_power/>

### Statistical Computing Seminars Introduction to Power Analysis

Pilot studies:  There are lots of good reasons to do a pilot study prior to conducting the actual study.  From a power analysis prospective, a pilot study can give you a rough estimate of the effect size, as well as a rough estimate of the variability in your measures.  You can also get some idea about where missing data might occur, and as we will discuss later, how you handle missing data can greatly affect your power.  Other benefits of a pilot study include allowing you to identify coding problems, setting up the data base, and inputting the data for a practice analysis.  This will allow you to determine if the data are input in the correct shape, etc. (please listen to our podcast # for more information on this).

Of course, there are some limitations to the information that you can get from a pilot study.  (Many of these limitations apply to small samples in general.)  First of all, when estimating effect sizes based on nonsignificant results, the effect size estimate will necessarily have an increased error; in other words, the standard error of the effect size estimate will be larger than when the result is significant.  The effect size estimate that you obtain may be unduly influenced by some peculiarity of the small sample.  Also, you often cannot get a good idea of the degree of missingness and attrition that will be seen in the real study.  Despite these limitations, we strongly encourage researchers to conduct a pilot study.  The opportunity to identify and correct "bugs" before collecting the real data is often invaluable.  Also, because of the number of values that need to be guestimated in a power analysis, the precision of any one of these values is not that important.  If you can estimate the effect size to within 10% or 20% of the true value, that is probably sufficient for you to conduct a meaningful power analysis, and such fluctuations can be taken into account during the power analysis.

Alpha level:  One obvious way to increase your power is to increase your alpha (from .05 to say, .1).  Whereas this might be an advisable strategy when doing a pilot study, increasing your alpha usually is not a viable option.  We should point out here that many researchers are starting to prefer to use .01 as an alpha level instead of .05 as a crude attempt to assure results are clinically relevant; this alpha reduction reduces power.

Note by Bill: see earlier note by Bill.