

## Power and sample size in clinical studies

Todd M. Brown, MD, MSPH<sup>a</sup>

<sup>a</sup> Division of Cardiovascular Disease, University of Alabama at Birmingham, Birmingham, AL

Received Sep 5, 2015; accepted Sep 5, 2015  
doi:10.1007/s12350-015-0288-z

---

### See related article, pp. 1300–1313

---

In this issue, Chiuzan et al.<sup>1</sup> present the most commonly applied methods for the calculation of sample size in clinical studies. Sample size estimation is an important, and often, overlooked component of all study designs. Grant agencies and other funding sources as well as scientific journals expect that appropriate sample size calculations have been performed, and that they are described in grant applications or scientific manuscripts. This will frequently involve collaboration with a statistician, but for simple and straightforward study designs, investigators may be able to perform these themselves, as outlined by Chiuzan et al.<sup>1</sup> The ability to do this saves time and cost in the study design period. Regardless, even if an investigator plans to consult with a statistician, some basic understanding by the investigator of the requirements for appropriate sample size calculation will make the process considerably more efficient and less costly, saving time and money.

The most common scenario that investigators will face is the desire to perform a prospective study. In this scenario, the investigators have not yet collected the data but will have some idea of what they are going to collect and what they want to study. As emphasized by Chiuzan et al.,<sup>1</sup> the first step of any study design, prior to the calculation of estimated sample size, is the formulation of the research question. Once the research question is formulated and before sample size can be calculated, the investigators will need to determine or estimate the values that are needed to populate the sample size formulas.

First, the investigators will need to determine what the alpha and beta values will be for their study. Most commonly, the alpha level (type 1 error rate) will be set at 0.05 and the beta level (type 2 error rate) will be set somewhere between 0.1 and 0.2, giving the investigators 80–90% power in their study. Obviously, these decisions will affect the ultimate size of the study as reducing the type 1 error rate (alpha) or raising the power (1—beta) to lower the type 2 error rate will result in larger required sample sizes. In general, type 1 error is considered to be more egregious statistically, so alpha (type 1 error) rates are rarely greater than 0.05, while beta (type 2 error) rates are frequently as high as 0.2.

Second, the investigators will need to estimate the effect size and the expected variance of the measurement. These estimates can come from prior studies done in similar situations or may come from pilot data that are collected before a much larger and more expensive study is conducted. Larger effect sizes will result in a lower needed sample size as it is easier to observe differences between groups if the groups are further apart. Likewise, to detect relatively small effect sizes, one would need a very large study. Using the example of summed stress scores (SSS) by Chiuzan et al.,<sup>1</sup> the larger the expected difference in the mean SSS between groups (delta or effect size), the smaller the sample sizes will need to be to reach statistical significance. The variance is related to how wide the range of this variable, for example SSS, is expected to be in each group in the study. Variation can occur because of the natural variation of the variable in nature, such as expected fluctuations in SSS between participants. Importantly, however, variation can also come from the imprecision of the instrument used to measure the variable. Higher levels of expected variance will result in larger sample size calculations as it will be more difficult to separate out the groups to observe differences between the groups as large amounts of variance create more overlap between the groups. Using the example of the SSS, this explains why frequently clinical studies might want to utilize a centralized data center where all myocardial perfusion images (MPI) are interpreted by one person on

Reprint requests: Todd M. Brown, MD, MSPH, Division of Cardiovascular Disease, University of Alabama at Birmingham, LHRB 313, 1720 Second Avenue South, Birmingham, AL 35294; [mbrown@uab.edu](mailto:mbrown@uab.edu)

J Nucl Cardiol 2015;22:1314–5.  
1071-3581/\$34.00

Copyright © 2015 American Society of Nuclear Cardiology.

one software program so that the SSS does not vary from participant to participant because of variation in the instrument used to measure the SSS or inter-reader variability. By centralizing the measurement, the investigators only need to account for the natural variation of the SSS in nature and not the artificial variation of the SSS resulting from imprecision in measurement. This will reduce the variance and in turn reduce the needed sample size, often making the performance of the study less expensive with shorter follow-up times and more efficiency.

Although usually investigators are looking to calculate sample sizes for prospective studies, frequently investigators may be interested in collecting data on participants who are part of an existing clinical database. In this instance, the sample size is already determined based on how many participants are present in the database. However, before beginning data collection, which might be time-consuming or expensive, investigators might want to determine if they are likely to reach a meaningful conclusion at the end of the study given the sample size available. To do this, the principles in the manuscript by Chiuzan et al<sup>1</sup> can still be applied. As previously mentioned, there are essentially 5 components to these calculations: sample size, alpha level, beta level, effect size (delta), and variance. In this instance, instead of solving the equations for the sample size, one can calculate the effect size (delta) that can be observed given the sample size available, alpha level, beta level, and estimated variance.

For instance, if investigators wanted to study the risk of myocardial infarction in those with  $\geq 20\%$  ischemia by MPI as compared to those with  $< 20\%$  ischemia, they could go to a clinical database of individuals who underwent MPI, and then obtain information on myocardial infarction in these participants following completion of the MPI. Because the investigators are using an established database, they do not have any control over the sample size. However, based on the alpha level (usually 0.05), sample size of the database, and expected variance, they can calculate the expected hazard ratio that can be observed for any given level of beta. For instance, the investigators may

discover that they have 80% power to detect a hazard ratio of 1.5 but only 50% power to detect a hazard ratio of 1.2. This means that they will have a good amount of power to detect a relatively large increase in myocardial infarction risk (50% increase) but relatively low power to detect a relatively small increased risk (20% increase). Based on these data and what they expect the actual increased risk to be, the investigators can decide, before conducting the study, if they are likely to reach a meaningful conclusion. For instance, if the investigators expect that they are only likely to observe a 20% increase in myocardial infarction risk in those with  $\geq 20\%$  ischemia by MPI as compared to those with  $< 20\%$  ischemia, they will know that they have relatively low power, only a 50% chance, of observing a statistically significant result. This could very easily provide good information to the investigators about the feasibility of their study before time and money are spent obtaining the data and analyzing the results. These types of calculations are also frequently requested by funding agencies before grant money is committed to fund these types of studies.

In conclusion, the calculation of sample size and estimation of power in a study with a fixed sample size are important components of most if not all clinical studies. Frequently, investigators can perform these calculations themselves as outline by Chiuzan et al.<sup>1</sup> However, even in instances where investigators are going to collaborate with a statistician, understanding the components of these calculations will make the time spent with the statistician more efficient and productive.

## Disclosure

*None.*

## References

1. Chiuzan C, West EA, Duong J, Cheung KYK, Einstein AJ. Sample size considerations for clinical research studies in nuclear cardiology. *J Nucl Cardiol* 2015. doi:[10.1007/s12350-015-0256-7](https://doi.org/10.1007/s12350-015-0256-7).