Making Sense of Biostatistics: Randomization

By Carleton Southworth

A typical clinical study is an experiment designed to test a hypothesis of no difference between groups, with an alternative hypothesis that a difference exists.¹ When a clinical study compares one treatment to a placebo or to a different treatment, it is essential to minimize possible factors that might confound the result. For this reason, most clinical studies randomize the assignment of study subjects across study arms. The alternative — enrolling matched, essentially identical, subjects across arms — is limited, usually impractical, and prone to selection bias.

There are three types of variables to consider when designing a clinical study:

- The independent variable controlled by the experiment (two or more treatments)
- A dependent variable (the outcome of interest, e.g., the presence or absence of a disease)
- Potentially confounding variables (other factors that could impact the dependent variable but are not the focus of the study)

Collectively, these potentially confounding variables are referred to as statistical “noise.” Making sure that statistical noise is distributed evenly between treatments is essential to measure the relationship between a treatment difference and a result of interest. The primary tool for achieving this even distribution is randomization. True randomization means that any given study subject has an equal chance of receiving either treatment prior to treatment assignment. Randomization can be augmented by other methods to help ensure that statistical noise is distributed evenly between treatments and that the sample for each arm is truly representative of the target population.

We will start by discussing a clinical study that compares two treatments across two groups of the same size. The same principles discussed here also apply to study designs with more than two treatment groups and where the study design stipulates that treatment groups are not the same size.

Treatments affect outcomes, but so do a host of other factors, such as the subject’s age, gender, genetic predisposition, and nutrition. The treatment variable (experimental or control) is referred to as the “independent” variable, and the outcome of interest is measured by a “dependent” variable. The experimenter controls the “independent” variable and is interested in the effect of this variable on the “dependent” variable. The trick is to isolate the effect of the treatment difference — the independent variable — against the background of “noise” caused by the other factors, some of which are unknown.

With careful planning, the cause-and-effect relationship between the independent and dependent variable can be measured accurately. Rather than trying to create two matched samples by manipulating all these noise factors, we assign subjects randomly. The assumption is that, with enough study subjects, any differences between the two treatment arms prior to the introduction of the treatment difference will be inconsequential. An additional benefit of randomization is that important variables affecting outcome that have gone unnoticed will also be distributed evenly between treatment groups in samples of adequate size. Isolating the treatment difference adequately creates a “strong inference” that rules out alternative explanations for the treatment difference observed in the dependent variable. In other words, the only possible reasons for the difference are the treatment difference and pure chance.² (Pure chance in this context refers to the likelihood of finding a difference at least as large as the observed difference.) It is only when strong

Subscribe free at www.firstclinical.com
© 2012 First Clinical Research and the Author(s)
inference is present that p-values and other statistical analytics are truly bona fide and unbiased.

Randomization Over Time

In addition to variations across study subjects, variation can be introduced by the simple passage of time. Recruiting subjects for a clinical study may take a year or longer. During this time, standard of care may change, personnel may change, and the weather may change. To maintain strong inference, it may be critical to keep the number of subjects in each treatment arm approximately even at any given time. Consider a drug designed to reduce inflammation. Airborne pollen could increase inflammatory responses in the spring season. If one treatment group has a higher count of subjects enrolled in the spring than the other, statistical bias would be introduced and the principle of strong inference would be violated. An unknown seasonal factor like humidity could also confound the results. The more factors that exist, the more likely it is that one of them could cause problems, based on chance alone.

Blocking

Blocking is a way to exert some control over randomization. For example, a block of four subjects would assign two subjects to each arm randomly. In any blocking scheme, the assignments even out exactly when a block is full. Further, if the block is filled at about the same time, then variations over time will be represented evenly in each study arm. Blocking is especially important if the study might end before planned enrollment is completed, e.g., based on early stopping rules.

In some blinded studies, the investigator can sometimes guess the particular study arm to which the next subject will be assigned. In such cases, the investigator can bias the randomization process, especially for the last subject in a block. It is important, therefore, to hide block size and even to vary their sizes randomly within a study, e.g. 6 then 4 then 8 then 6.

If the study plan calls for each research site to enroll a large number of subjects, then each site can work from its own blocks. This ensures an even subject distribution within each study arm and within each site. However, if the plan is for each site to enroll, say, only five subjects, then sites can share blocks (assuming the randomization technology cooperates). In this case, it is good practice to share blocks between similar sites. Stout has created a SAS program that produces random blocks of 2, 4 and 6 subjects. This program is available through the website referenced in the citation and can be used to generate random block sizes automatically.

Stratification

Sample stratification is a method to ensure that subjects are evenly spread demographically or across sites. Strata are groups of subjects that are more or less homogeneous. In many instances, it is important for the study sample to reflect the proportion of patients falling within certain more-or-less homogeneous groups. For example, if it is known that the intended population includes 40% women and 60% men, then a representative sample would include 40% women and 60% men. Thus, the sample would be stratified by gender and enrollment would be stopped for a given stratum when the desired proportion for that stratum is met. Note that variables for stratification do not include variables under the control of the experiment, e.g., the treatment variable. When samples are stratified, it is typical for them to first be stratified and then blocked within each stratum.

A common variable for stratification is study site. Outcomes might vary by site because of differences in the patient populations they serve, available equipment, personnel training, investigator skill, and a host of other factors. In this instance, stratification would entail
randomizing each site separately, as if each site were a separate experiment. Stratification by site, combined with blocking within each site, maintains balance between treatment groups within each site, helping to ensure that strong inference is maintained. When this is done, any difference between treatment group outcomes cannot be attributed to an over-representation of one or the other treatment at any particular site.

Other variables where stratification may need to be considered include initial severity of disease, gender, initial concomitant conditions, and others. For example, if hypertension could affect outcome, then it might be important to stratify by hypertensive status (present or not) and then use randomized blocks within each of these categories separately to ensure that an even number of subjects are represented in each study arm within each strata (hypertensive or non-hypertensive). The benefit of stratifying by a particular variable is offset by an administrative cost and may require extra effort to find qualified subjects. Therefore it is important to make sure that the cost:benefit ratio of stratification is favorable.

Conclusion
When designing a randomization scheme, sample blocking and stratification should be considered to ensure that any difference observed between treatment groups is attributable to the difference in treatment and to no other cause, known or unknown. It is important to recognize that the guiding principle in any experimental design is the maintenance of strong inference — the elimination of all potential explanations for a difference between treatment groups except for the treatment difference under investigation.

References

Acknowledgement
The author thanks Elisabeth Southworth for editorial assistance.

Author
Carleton Southworth, AB MS RAC, is Director of Biostatistics at American Research Partners. Contact him at 1.574.367.8076 or csouthworth@americanresearchpartners.com.