

**Good Practices in Clinical Research**  
**Lecture 8**  
**Dr. Alan Moses**

**Basic Concepts of Applied Statistics**

I am Dr. Alan Moses, senior vice president and chief medical officer of the Joslin Diabetes Center in Boston, Massachusetts. In this segment of the GCP we will consider basic concepts of applied statistics. We will review the role of statistics in clinical research. We are going to talk about basic statistical concepts. We are going to look at commonly used statistical tests and finally we are going to discuss some of the so-called "demons of data analysis."

What is the role of statistics in clinical research? What we are really doing is looking to make a differentiation between truth and chance. We want to compare group differences and we want to examine the effects of an intervention.

What then is the role of the biostatistician in clinical research? While we all would like to feel comfortable with data and know how to design a trial and analyze the data, we often have to rely on the expertise of a biostatistician to select the appropriate trial design and to calculate an appropriate sample size. We must all realize that sample size is really based on outcome measurements that include precision, accuracy, reproducibility and practicality. In addition, a biostatistician helps determine the analytical tools to utilize for data analysis. Methods need to be defined before the trial begins, and very importantly, analysis techniques are never changed after a trial is completed.

What is the role of statistics in analyzing a trial? It is very important to remember that statistical techniques are merely tools to assist in the interpretation of data generated in the context of an experiment. They are a tool and not the end result of a trial, and indeed, like any tools, statistical tools should be used carefully. A computer can produce numbers that may or may not have statistical significance, but only an investigator knows what tests should be utilized to perform the statistical analysis. The educated investigator is able to choose these tests rather easily. It is very important to remember that a powerful statistical software package can be a lethal weapon in unsophisticated hands.

One of the key concepts of bio statistics is that of validity. Validity is central to the critical analysis of data, the results of a trial and indeed the medical literature. There are two types of validity: internal and external (or generalizability). Internal validity means that within the confines of the trial the results are accurate. The methods and analysis employed are able to stand up to scrutiny, and the interpretation and conclusions of the investigators are supportable by the data and by a review of the relevant medical literature.

External validity or generalizability determines whether or not the trial design is adequate enough to allow the observations made and the conclusions drawn to be applied to an entire population. Indeed, it is the selection of the trial population that permits maximum generalization, a concept that we have touched upon in other sections of this program. If both males and females are chosen, different ethnic groups are included, as well as different age groups, there is a greater chance of applying the results of a clinical trial to a general population. On the other hand, subject selection can also define the limits of the population to which observations and conclusions apply. For example, if you choose children only between the ages of 5 and 10 years of age in a trial, that is the only group to which you can apply the results of that trial. If you choose Asian men over the age of 45, that is the group to which the trial results apply.

It is important in the concept of validity to recognize that sample size must be sufficient to support the conclusions drawn, and the appropriate control population has been selected with particular emphasis on the concept of randomized, double-blind comparative trials, the basis of scientific investigation in clinical research. Noted author Isaac Asimov recognized the importance of validity within the context of any kind of science including clinical trials when he quoted “One of the glories of scientific endeavor is that any scientific belief, however firmly established, is constantly being tested to see if it is truly, universally valid”.

We now are going to turn to how we look at some types of data. The first relates to relative risk and odds ratios. These are outcome measurements that are very helpful when comparing the effects of exposure on an outcome. Odds ratio is utilized primarily for case control studies. Relative risk is used primarily for cohort studies. These are two types of study designs that you have already addressed within the context of the GCP

seminars.

Let us first look at relative risk. This table indicates how one derives a relative risk factor. This table is broken up into two columns and two rows. In the first column we have disease. In the second column we have absence of disease and we look at whether the disease is a result of exposure to a risk factor or if there was no exposure to a risk factor. Disease in the presence of exposure to a risk factor is designated A. Absence of disease with exposure to the risk factor is designated B. Presence of disease without exposure to the risk factor is designated C and absence of disease without exposure to the risk factor is D. From this table we can derive the relative risk. The relative risk is the incidence of exposure over the incidence of no exposure. That translates into  $(A / (A+B))$  divided by  $(C / (C+D))$ . This is the relative risk.

Odds ratio is similar in terms of the construct of the table used to derive it, but different mathematically. Here again we have the independent variable of exposure to risk factor or negative exposure to risk factor vs. the presence or absence of disease, again, A and B versus C and D. An odds ratio, different from relative risk, however, is derived by multiplying  $A \times D$ . The presence of disease with exposure to a risk factor times the absence of disease without exposure to the risk factor over  $B \times C$ , the absence of disease with exposure to the risk vs. the presence of disease without exposure to the risk. The odds ratio again is  $A \times D$  divided by  $B \times C$ .

How do we use odds ratio and relative risk in interpreting tests of association? It is actually quite easy. When the odds ratio or relative risk is less than one, there is a negative association or protective factor for that risk for the disease. When the result is one, there is no relationship or association either for odds ratio or for relative risk. But when the ratio is greater than 1, there's a positive association for either odds ratio or relative risk.

This has been an introduction to some of the concepts underlying the use of bio statistics in clinical trials. In the next segment we will look at some examples of how these and other tests are used in analyzing data from clinical investigations.