

## Do Bariatric Surgeons and Co-Workers Have An Adequate Working Knowledge of Basic Statistics? Is It Really Important?

George Cowan, Jr., M.D., Professor Emeritus, Department of Surgery, University Tennessee Health Sciences Center, Memphis, Tennessee

Following my retirement from active bariatric surgery practice, I have tried to continue to contribute to our field by teaching courses in "Basic Statistics" and "Critical Review of the Literature (Applied Statistics)" at the American Society for Metabolic and Bariatric Surgery (ASMBS) and International Federation for the Surgery of Obesity (IFSO) meetings. In these courses, after reviewing a number of bariatric-focused papers, attendees were able to recognize correct and incorrect applications of several principles of basic statistics. These principles included: (1) use of non-ratio numbers with statistical tests or descriptors that require ratio numbers only; (2) avoiding bias of a research study in one direction or another by failing to completely and impartially randomize subjects into control and treated groups; and (3) precisely reporting the methods used in the research study, or providing reference(s) that describe the same, so that any other investigator can carry out a nearly identical study. The following is a brief discussion of the correct use of these three statistical principles:

### **(1) Use of non-ratio numbers with statistical tests or descriptors that require ratio numbers only.**

*Ratio numbers* are defined as numbers that follow one another in sequence along a numerical scale and that are equally spaced from one another. Ratio numbers also include an absolute zero value; that is, the series of values being measured diminishes to the point of reaching zero. For example, blood glucose values are ratio numbers. The 1 mg% difference between any two consecutive blood glucose numbers, such as, from 89 mg% to 90 mg%, is the same as the difference between, say, 149 mg% and 150 mg%. Also, although not compatible with life, it is possible for certain unique blood preparations, such as standards, to produce a result of zero glucose.



Jane N. Buchwald  
Viewpoint Editor  
About the Article

Reading or writing a research article for the scientific literature is markedly more complex today, at the end of first decade of the 21<sup>st</sup> century, than just 17 years ago, when the first bariatric surgery-focused journal was launched. A burgeoning of preclinical research, procedural innovation, advances in clinical care, and greater numbers of patients undergoing surgery and requiring follow-up have led to the need for more sophisticated analyses of treatment effect and risk. Certain bariatric study designs require a variety of analyses, and sometimes, rigorous statistics. Not only the author, but the reader of journal articles must knowledgeably assess the applicability and correct use of statistical approaches in the organization and analysis of data in order to judge an article's value to current practice.

### Survey Results

Do you feel that bariatric surgeons and co-workers have an adequate working knowledge of basic statistics? Is it really important?

20% Yes

40% No

40% Not Sure

*Non-ratio numbers* may be defined as a series of numbers, any number of which has a value that may not be equally different (or distant) from the number before or after it, and which may or may not include an absolute zero point (value). For example, nurses and others often use the numbers 0 through 10 when recording a patient's pain level following bariatric surgery. Like pain scores, questionnaires about products or continuing medical education courses may ask us to provide a set of choices, such as: "4=Strongly Disagree," "3=Disagree," "2=Agree," and "1=Strongly Agree." The results of these pain scores and questionnaires are non-ratio numbers.

Unlike serum glucose levels, the difference between numbers "4=Strongly Disagree" and "3=Disagree" is likely to be different from that between "2=Agree" and "1=Strongly Agree" and also "3=Disagree" and "2=Agree" for either of these scales. That is, each of these possible responses requires the individual to make a judgment about pain or agreement/disagreement that is subjective, a "guesstimate" of how she/he feels about an experience. These data do not consist of the purely objective number(s) we get from a blood glucose or blood chemistry analysis that have set standards for their scale of values. Therefore, the distance (value) between each of any two consecutive, "guesstimated," non-ratio numbers is not likely to be the same as the value between any two other non-ratio numbers. These non-ratio numbers are not properly analyzed by use of parametric statistics (defined below).

*Parametric statistics* includes the use of means and standard deviations, t-tests, linear regression analysis, and other similarly derived values and tests to analyze data composed of ratio numbers. This includes the above example of glucose concentrations in the blood, as well as temperature (Centigrade or Fahrenheit), body mass index (BMI), waist circumference (centimeters or inches), body weight (pounds or kilograms), and other like measures. Parametric statistics tests are sometimes improperly used on non-parametric types of data. *Non-parametric statistics* includes methods by which the above pain, questionnaire, and similar non-ratio numbers may be properly analyzed. Non-parametric statistics use medians and quartiles, quintiles, deciles, chi-square, and other similarly based values and tests to analyze non-ratio number data. It may also be used for ratio number data; however, especially for data composed of small groups, non-parametric statistics may be less able than parametric statistics to show differences between groups.

Now, to address the second central statistical principle:

**(2) Avoiding bias of a research study in one direction or another by failing to completely and impartially randomize subjects into control and treated groups.** Stated differently, randomization in a study involves randomly assigning the usual heterogeneous group of bariatric patients to different groups completely impartially in order to avoid work biased or prejudiced in one direction or another. Why is this important? The main objective of randomization is to obtain two, or more, groups for a study that are as reasonably identical with one another as possible. One group can then be designated the control, and the other, the treatment group, to determine whether a difference between the two groups exists.

A well-controlled, properly randomized study ought to provide a “clean” statistical result at the usual  $p < 0.05$  (probability) level if a difference exists between the two groups. Thus, if one were to repeat the same study a total of twenty times, at least nineteen of these studies would show the same result; however, one of these studies, purely by random chance, would have a probability of showing no difference between the groups. If the groups’ randomization was not properly done, the result would likely be “unclean,” and, therefore add to the always-present random chance, such that two, three, or even more of these studies would show no difference. An improper randomization can bias a study, with a greater likelihood of ruining its value as a useful addition to the literature than if the groups were correctly randomized.

To illustrate, here are some possible, though questionable, applications of randomization. You and your colleagues decide to perform a gastric bypass with the gastroenterostomy fashioned by:

(1) a hand-sewn gastroenterostomy (Group A) or a circular-cutter stapler-made gastroenterostomy (Group B) “randomized,” in strict sequence, to every other patient as they come through the operating room door and then study the incidence of complications;

- (2) the same as #(1) except that Group A and Group B are randomized, in strict sequence, by having alternate patients who are to be scheduled for future surgery by your office to have one of these anastomotic techniques;
- (3) the same as #(1) except that Group A and Group B were randomized by having Group A taken sequentially from all patients scheduled for surgery by your office who show up during morning clinics, and Group B will be taken sequentially from all patients scheduled for surgery by your office who show up during afternoon clinics of the same day of the week.

These vignettes may appear to provide sufficient randomization. However, they do not. Why? *Group (1) Answer:* You may, for example, happen to reschedule some of the smaller, more favorable patients to the hand-sewn or stapler group. You designed and know the study scheme involving the surgery patients, and you can change their order of surgery for any given day. The nurses may schedule the larger, "riskier" patient as the first of two patients you operate upon. Therefore, the groups could become biased, with larger patients in one and smaller patients in the other. *Group (2) Answer:* Patients may be scheduled for later surgery, such that almost all of Group A patients are in the first month and mostly Group B patients in the next month. During this time, many variables may develop (e.g., equipment changes, different scrub nurses, perhaps you experience some back pain during one of these months that lengthens the time taken for surgery, etc). *Group (3) Answer:* Patients get scheduled at different times of day for many reasons, many of them not random. Heavier patients, diabetics, and others may tend to attend your clinics with a preference for mornings or afternoons. A clear bias would likely emerge in your groups as a result.

Randomization requires avoiding the possibility of anyone having the ability to alter or modify the order or outcome of properly randomized subjects. A variety of methods exist to support proper randomization (e.g., blinding).

Finally, we address a third critical statistical principle:

**(3) Precisely reporting the methods used in the research study, or providing reference(s) that describe the same, so that any other investigator could carry out a nearly identical study.**

It is not unusual to review a paper that does not include the actual method of how patients were randomized; that is, such papers claim to have performed randomization of compared groups, but do not tell the reader how this was accomplished in careful detail. This unfortunate circumstance, or any other omission of description of essential study methodology, renders the value of the study impossible to determine. Randomization requires a fair amount of work and planning; this is what imbues the findings of a properly designed trial with higher scientific and clinical value than many other study designs.

If you are not involved in research, but are a reader of key medical journals, this concern over statistics principles seems of little importance. Yet, we are all, as readers, *involved* in research every time we read a bariatric surgery article that contains research data about our field. Without having some means of interpreting the data, we are left to read these papers' conclusions on a "take it or leave it" basis. Since it is not infrequent that papers contain design and statistical flaws, some fundamental, we owe it to patients to be able to interpret the data in the literature for ourselves and to be able to draw informed conclusions to support our clinical practice.

For those who perform research, but prefer to "leave the statistics to the statistician," this may place an inappropriate burden upon the statistician's head. He/she is likely as unfamiliar with bariatric surgery as you presumably may be with statistics. A gap of misunderstanding, unknown to either yourself or the statistician, can seriously flaw the experimental design or the analysis of data for the study. It is far preferable to be familiar with basic statistics to understand the different tests the statistician recommends and to mutually come to an understanding concerning the statistics to be employed. It will definitely be helpful to answer questions that may be asked about your data. If you are on the Editorial Board, or are a reviewer for *Obesity Surgery* or *SOARD*, it is essential to review papers with an adequate understanding of how the data were analyzed. Obtaining a working knowledge of Basic Statistics skills is not only critically important for those in positions of editorial responsibility, but for those who read and write papers.