FELLOWS RESEARCH WORKSHOP:

DESIGNING AND CONDUCTING A CLINICAL RESEARCH STUDY

Alan R. Tait, Ph.D.
Professor
Department of Anesthesiology,
University of Michigan Health System
Ann Arbor MI

TEL (734) 763-8128, FAX (734) 764-9332
e-mail: atait@umich.edu

This lesson is designed to serve as a primer for how to critically review the medical literature and as an introduction to young investigators interested in designing and conducting a clinical research study.

STUDY DESIGNS

Study designs are categorized broadly into two types; descriptive and analytic.

Descriptive studies:

In the anesthesiology literature, descriptive studies usually take the form of case reports or case series. These are important in describing unusual cases, adverse events, or novel or untested approaches to anesthetic management, however, they have no basis for hypothesis testing or statistical analysis. Furthermore, since the findings in a single case report/series may result from chance or coincidence, care must be taken in extrapolating them to broader populations.

Analytical studies:

Unlike descriptive studies, analytical studies allow for hypothesis testing, statistical analysis and examination of cause and effect relationships. There are two main types of analytical studies; observational and experimental (interventional).

Observational studies are those in which there are no interventions or attempts by the investigator to alter the course or outcome of the study. In general, groups are selected depending on whether or not they have a particular characteristic, exposure, risk factor or disease (outcome). There are two types of observational studies typically seen in the anesthesiology literature; the retrospective (case-control or retrospective cohort) study and the prospective cohort study.

Retrospective studies:

These types of studies involve comparisons between groups of individuals who have the disease or outcome in question (cases) with those who do not (controls). Since randomization of subjects to cases and controls cannot be done "after the fact," controls

may be "matched" to the cases i.e., controls are selected that have similar attributes to the cases save for the outcome in question. Typically, controls are matched for attributes such as age, gender, and surgical procedure. It is important to note, however, that while increasing the number of "matches" improves the homogeneity of the cases and controls, it also reduces the size of the eligible subject pool (overmatching). Once the groups are identified, variables that may be associated with the outcome are sought retrospectively, typically from the medical records or other clinical databases.

Retrospective studies may also begin with groups that have or do not have a characteristic, risk factor, or exposure to determine if these variables result in a specific outcome. This retrospective cohort design follows the same procedures for matching and data collection as the case-control study. Of particular concern in retrospective studies is the issue of selection bias. This may occur since the criteria for group selection are made after-the-fact *(post hoc)*, and, therefore, may be subject to interpretation based on the quality of recorded data and/or the unintended bias of the investigator.

Prospective studies:

These types of studies involve comparisons between groups of individuals based on whether or not they have a characteristic, risk factor, or exposure. Individuals are followed up over time to determine the incidence of disease or outcome in the experimental versus the control group. Prospective studies offer an excellent avenue for hypothesis testing since cause precedes effect. However, they may not be suitable for studying rare outcomes and may suffer from loss to follow-up when subjects are observed over long periods of time.

Experimental Studies:

In these types of studies there is a deliberate action on the part of the investigator to intervene and alter the course of the study. The most common type of experimental study is the randomized clinical trial which is typically used to evaluate the effect of a new drug, procedure, or intervention on outcome. Randomization maximizes the comparability of groups and reduces confounding. An important concern regarding experimental studies is the potential for observer bias. As such, it is generally preferable, if possible, for both the subject (single-blind) and the investigator (double-blind) to be unaware as to the allocation of the subject to a particular study group. The ability of a researcher to remain objective in an unblinded study is open to speculation since investigators may have a vested interest in the outcome.

Meta-Analysis:

Meta-analyses are seen with increasing frequency in the anesthesia literature. These analyses represent a statistical overview utilizing data from previously published studies. This approach involving pooling of data from several sources increases the power of the study to detect true outcome effects and enhances the interpretation of the results. Although this method can be a powerful tool, the actual process can be time-consuming since the search for comparable studies must be exhaustive. Furthermore, considerable care must be taken to ensure that each study selected for inclusion in the meta-analysis is similar in terms of selection criteria, study design, and outcome measures.

HYPOTHESIS TESTING AND STATISTICS

The scientific method requires that the investigator state the problem, develop a hypothesis, test the hypothesis, and draw conclusions. The primary purpose of the hypothesis is to establish the basis for tests of statistical significance before the study begins. Hypotheses are generally stated as the null hypothesis or, more typically, the research (alternative) hypothesis. The null hypothesis asserts that there are no anticipated differences between study groups or associations between the predictor and outcome variables. The null hypothesis is the basis on which many statistical tests are based (e.g., the *t* test). The alternative hypothesis anticipates that there will be a difference between study groups.

There are two important errors associated with hypothesis testing i.e., the Type I (α) error and the type II (β) error. The type I error is the probability of rejecting the null hypothesis when it is, in fact, true (false positive). This often results when an observed difference between study groups occurs by chance. In general, the type I error is set at 5% i.e., the investigator is willing to accept a 5% chance that any observed difference between study groups is actually false. The type II error is the probability of failing to reject the null hypothesis when it is, in fact, false (false negative). This typically occurs when the study sample is too small (under-powered) to detect a difference between study groups when a difference really exists. By convention the type II error is set at 10 or 20% i.e., the investigator is willing to accept a 10-20% chance that an observed negative result (no difference between groups) is false. The power of a study is also an important concept. Power is the ability of a study to detect a difference between study groups when one truly exists. Power = 1- β error and, therefore, by convention is set at 80-90%.

The P value is common to all statistical analyses and represents the mathematical probability that a study finding is due to chance. By convention a P value of less than 0.05 is considered statistically significant i.e., there is a 1 in 20 probability that a study finding is due to chance. In conducting research, it is also important to understand different types of data. Basically, data are classified as either continuous or categorical (discrete). Continuous data are those which take on any value within a defined range e.g., height, weight. Categorical data are either nominal i.e., named categories with no implied order (e.g., eye color, gender, race) or ordinal i.e., ordered categories (e.g., ASA status, tumor staging). Data can either follow a normal distribution (parametric data) or not (non-parametric data). Regardless of whether the data are continuous or categorical, parametric or non-parametric, there are specific statistical tests that can be used to analyze each type.

CRITICALLY REVIEWING THE LITERATURE

Although the anesthesiology literature represents but a small fraction of the total medical literature, the number of articles available to the average anesthesiologist interested in staying "current" is often overwhelming. The ability to select articles of interest and critically review them is therefore important to the busy care provider. This is particularly important for articles that may influence clinical practice.

In order to fully understand and interpret a research article it is important to break it down into its component parts such that each part can be individually examined and critiqued. Riegelman and Hirsch present a very functional framework in which to do this. This framework consists of a number of elements that must be evaluated individually in order to make valued judgments regarding the validity and importance of the entire

article. Deficiencies in any of these elements may invalidate the findings of the study. This framework can be applied to different study designs. Although readers may wish to develop their own framework, the following adapted from Riegelman and Hirsch represents a logical approach, as follows:

Study design: The study type, e.g., observational, experimental, prospective, etc.
Assignment*:* A description of the study sample and how it was derived.
Assessment*:* The outcomes of the study and how they were measured
Analysis: The statistics, sample size
Interpretation*:* What the results mean for the study sample
Extrapolation*:* What the results mean for other populations.

Having identified an article of interest, the reader should use the framework as a paradigm for identifying the salient features for review.

The Abstract

The abstract serves to provide a comprehensive synopsis of the article and to attract the reader to study the article in more depth. The abstract, in effect, acts as the "carrot" to entice the reader to read on. The abstract is not only important when reviewing the literature but is a critical element when writing one's own research paper.

The Introduction

The introduction should provide the background and rationale for the study and an explanation of why it is important. It is here that the first element of the framework is usually introduced i.e., the ***study design*****.** It is important that the study design be appropriate to the question being asked. For example, a retrospective study, despite its inherent limitations, may be much more appropriate when the outcome measure is rare than a prospective study. The introduction should contain a clear description of the aims of the study but more importantly, an hypothesis. Unfortunately, a statement of hypothesis is sadly lacking in much of the anesthesiology and medical literature. This is disappointing since a clearly defined hypothesis allows the reader to know exactly what the investigator hopes to accomplish. Examination of the results then allows the reader to definitively determine if the hypothesis was accepted or rejected. One should be beware of studies that suggest a "fishing trip" with no clearly defined aims since they frequently produce clinically ambiguous or questionable results.

The Methods

Here we learn about the sample and the population from which it was derived **(*Assignment*)**. Firstly, it is important to ascertain how the sample was obtained e.g., random selection, consecutive patients, or a convenience sample of patients seen during a specific time period. The inclusion and exclusion criteria should also be clearly described.

Control groups, if used, should be derived where possible from the same population as the experimental group. In general, the control group should be as similar as possible to the experimental group except for the variable(s) under study. If the study is a clinical trial, it is important to determine if there was appropriate randomization and blinding to reduce the potential for selection and observer biases, respectively.

Randomization affords each subject an equal probability of being selected to a particular study group. For case-control studies it may be important to match (since randomization cannot be performed retrospectively) the control with the experimental group (cases).

The methods section should also clearly describe the outcome(s) and the method(s) by which they will be measured **(*Assessment*)**. Measurements may involve objective measures e.g., blood pressure, blood loss, etc, subjective measures e.g., pain, anxiety etc, or they may rely on information reported by subjects in a questionnaire or from information documented in the patient's medical record. Measurements should be complete such that there is a minimal amount of missing data. Studies with large amounts of missing data may provide biased estimates of outcome. All measurements must be accurate, reliable and valid. This caveat applies equally to items in a questionnaire as to measurements made using a device or item of equipment.

Statistics **(*Analysis*)** are often overwhelming to the average reader, therefore, in order to be able to critically review an article, it may be necessary to consult with a statistician (not as scary as it seems). However, there are a few statistical concepts that are common to most research articles and, as such, are important to understand in order to interpret a research article. Firstly, it is important to know if the investigator has determined a sample size that is large enough to detect a difference between groups should one exist. *A priori* sample size determinations protect against the possibility of creating a type II error. In general, sample size determinations are based on the anticipated difference in outcome between study groups that the investigator believes is clinically important to detect. The methods section should also contain a detailed description of the statistics used. As described earlier, the statistics should be appropriate to the type of data to be analyzed.

The Results

The results section should be clear and well organized. It is important to determine if statistically significant differences were observed between groups (i.e., $P<0.05$). If there are no differences between groups, it is important to know if the sample size was large enough to have detected a difference had one existed. If the study has a stated hypothesis, the results should determine if the hypothesis was accepted or rejected. In interpreting the results one should also consider the potential for bias and whether or not the investigator attempted to limit its effect.

The Discussion

The discussion summarizes and interprets the findings of the study **(*Interpretation*)**. It is important here to be sure that the conclusions are consistent with the results. Investigators in their enthusiasm sometimes have a tendency to overplay the generalizability and importance of their results. In interpreting the results, it is also important to remember that statistical significance may not be synonymous with clinical significance. Small differences between groups may be statistically significant (particularly when the sample size is large) but may be of limited clinical significance.

The final consideration is to determine the generalizability of the findings and whether or not they can be extrapolated to other populations **(*Extrapolation*)**. This may be important in extrapolating adult data to the pediatric setting or animal data to the clinical arena. In general, it is easier to extrapolate data to populations that are similar to

the population that was studied. In some studies it may be possible to calculate the attributable risk between the factor of interest and outcome. This allows one to calculate the percentage of the outcome that is attributable to the factor and draw conclusions regarding the impact of an intervention e.g., removing the factor, on outcome in other populations.

REFERENCES:

1. Riegelman RK. Studying a study and testing a test: How to read the medical evidence. 5th edition, Lippincott Williams & Wilkins, Philadelphia, 2005.
2. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 2nd edition, B.C. Decker Inc., Hamilton Ontario, 2000.
3. Ryan-Wegner, NM. Guidelines for critique of a research report. Heart and Lung 21:394-401, 1992.
4. Rosenberg AL, Wei JT. Clinical study designs in the urologic literature: A review for the practicing urologist. Urology 55:468-476, 2000.