CHAPTER NINE

STUDY DESIGN

In medical research, we attempt to study the multitude of constantly changing and interrelated biologic processes that comprise human physiology. In order to make meaningful conclusions from the abundance of physiologic data available, we need to carefully consider the design of our investigations. We must meticulously define our study hypotheses, patient population, and research methods for our conclusions to be valid and our study a useful addition to the medical literature.

Before data collection ever begins, therefore, much of the work involved in medical research is already completed. Failure to exercise such attention to detail and planning may result in faults in study design that are subsequently propagated throughout the study and impact on each step of the research process. No amount of statistical manipulation can correct for errors and biases introduced by a poorly designed study.

Initial planning
⇓
Research protocol design
⇓
Investigational review board (IRB) approval
⇓
Data collection
⇓
Data processing
⇓
Data analysis
⇓
Interpretation
⇓
Publication

Figure 9-1: The research process (adapted from reference 1)

"Study design" encompasses the deliberate planning of each level of the research process depicted in Figure 9-1. All levels of this process from initial planning to data analysis and interpretation are subject to the introduction of errors and statistical biases. We must anticipate and account for these potential sources of error at the outset if our study conclusions are to be valid. The goal in designing any research study, therefore, is to avoid systematic errors and biases as it is much easier to correct flaws in study design prior to beginning a study rather than during or after concluding a study.

Figure 9-2 illustrates the important elements to consider in the design of a research study. In the first eight chapters, we addressed the 4 elements on the right. This chapter will focus on the first four elements; namely, the various types of studies, methods of treatment allocation, sample size calculations, and creation of a research study protocol.

- Type of study
- Method of treatment allocation
- Sample size necessary
- Creation of study protocol
- Choice of outcome measures
- Nature of data required
- Validation of data collected
- Choice of statistical analysis

Figure 9-2: Elements of study design

TYPES OF STUDIES

Scientific studies can be divided into two basic types according to whether or not we administer an intervention or treatment. **Observational studies** involve the surveillance of one or more groups of patients to determine the effect of various patient characteristics on one or more outcome variables. These patients, however, receive no intervention or treatment. Observational studies demonstrate the "association", if any, between the patient characteristics and outcome variables of interest. An example of an observational study is the Framingham study. In this ongoing study, the inhabitants of Framingham, Massachusetts have been followed prospectively for the development of various diseases in order to identify any associated risk factors present.

Experimental studies involve the administration of an intervention or treatment to two or more groups of patients with attention being directed to identifying the impact the intervention has on a particular outcome. Experimental studies, through demonstrating a response to therapy, may prove "causation". A **clinical trial** is an experimental study that involves a clinically applicable therapy. Clinical trials generally involve the comparison of an experimental drug or therapy against a control (standard therapy or placebo). The Food and Drug Administration categorizes these into **Phase I**, **Phase II**, and **Phase III** trials. A **Phase I** trial is the initial introduction of a drug or treatment to humans. It identifies any associated toxicity and may begin to determine dosing levels and efficacy. Phase I trials are small, usually involving only 10 to 20 patients. A **Phase II** trial demonstrates the effectiveness and safety of a drug or treatment at specific dosages. These closely monitored trials involve 100 to 200 patients. After demonstrating a treatment's efficacy, an expanded **Phase III** trial is performed to identify the specific indications and uses of the drug or treatment as well as more precise information on related side effects. Figure 9-3 outlines the progression of a typical clinical trial.

Discovery of a drug or treatment
⇓
Formulation of the drug or treatment
⇓
Animal testing / Toxicology studies
⇓
Phase I trial
⇓
Phase II trial
⇓
Phase III trial
⇓
Clinical usage of the drug or treatment

Figure 9-3: Clinical Trials

We can also subdivide research studies by the time course of data collection (Figure 9-4). **Longitudinal studies** involve the analysis of data collected over an extended period of time and are thus able to identify the changing nature of a disease process. **Prevalence studies**, on the other hand, involve the analysis of data obtained at a single point in time and are useful in identifying the presence of disease in a particular patient population. **Retrospective studies** are longitudinal studies that look back in time to evaluate the factors affecting a particular variable of interest. **Prospective studies** are longitudinal studies that look

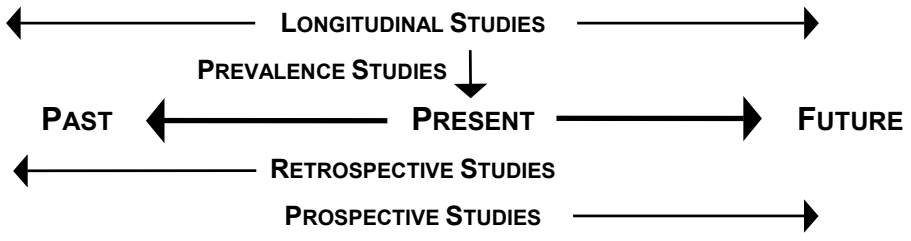

Figure 9-4: Study design and the time course of data collection

forward in time to analyze the impact of an intervention or disease process on the outcome variable(s) of interest. Retrospective studies are frequently easier to perform, but are more subject to potential errors and statistical biases. Prospective studies are usually more difficult to perform, more expensive, and more time consuming, especially if the disease of interest is rare. They provide more evidence for causality, however, than does a retrospective study.

OBSERVATIONAL STUDIES

Case-series, **case-control**, **cross-sectional**, and **cohort studies** are all examples of observational studies. **Case-series studies** are simply descriptive reports illustrating observations of interest on one or more, usually consecutive, patients. There are no control patients involved or initial hypotheses presented, although such studies frequently result in hypotheses that lead to further studies. Because they describe the disease course of the patients involved, they are longitudinal in design. An example of a case-series study is one that describes a hospital's last 50 patients with malignant thymoma, detailing their demographics, response to therapy, and survival.

Case-control studies are retrospective studies that identify a disease or outcome variable and look back in time to determine the risk factors that led to the disease. They compare a control population (patients without the disease) with a case population (patients with the disease) to identify factors that make the two groups different and may account for the occurrence of the disease. It is important to remember that observational studies, such as case-control studies, can only imply association and cannot prove causality. Case-control studies are also longitudinal studies as they review data from an extended period of time to arrive at their conclusions. An example of a case-control study is one that evaluates patients with and without colon cancer to determine whether differences in dietary or environmental factors may result in the development of malignancy.

Retrospective studies, such as the case-control design, are subject to a number of potential errors and statistical biases that are not present in prospective studies. One of the more practical problems in performing a study using historical control patients is that we must assume that the data were collected accurately and recorded correctly in the patient's chart. This is not always the case and, unlike in a prospective study, we do not have access to the patients to remeasure the necessary information. Another common problem is that the necessary data may either not be present in the records or was unavailable at that time. For example, historical controls from the 1960's will not have computed tomography reports to compare with those of patients in the 1980's as this technology did not exist until 1974.

The methods by which we diagnose and treat patients change with time. Thus, historical control patients may not have received the same level of care or treatment as the current study patients, and the two groups may, therefore, not be comparable. This difference in treatment due to time introduces the potential for statistical bias which may affect the validity of our final conclusions. Similar bias may be introduced when the study and control patients are treated at different institutions or by different groups of physicians with differing methods of management.

As physicians, we are much more efficient in our diagnostic skills than we were in the past due to the advent of new diagnostic tests and imaging techniques (i.e., computed tomography, magnetic resonance imaging, etc.). We are thus able to detect some disease processes sooner and at an earlier stage than could our predecessors. If we compare the 5 year survival rates for many cancers, therefore, we find that survival appears to have increased over the past few decades. In reality, however, survival has stayed constant; we are simply detecting disease earlier than in the past and patients appear to be living longer. This is known as "**zero-time shift**" or "**lead time bias**" (also known as the Will Rogers' phenomenon) referring to the shift towards earlier diagnosis of disease. "**Stage migration**" refers to the fact that we are diagnosing patients at earlier stages of disease and the incidence of disease by stage therefore appears to be changing. All of these factors can impact on the validity of conclusions made using historical controls.

Although at greater risk for the occurrence of errors and statistical biases, retrospective studies are still a useful study design as long as we carefully choose our historical controls. Retrospective studies are also frequently used as a less expensive, initial study to identify potential treatments that warrant further prospective investigation.

Cross-sectional studies are prevalence studies which collect data at a discrete point in time in order to answer particular questions about the status of the population at that instant. Disease prevalence studies,

surveys, questionnaires, and meta-analyses are examples of cross-sectional studies. Cross-sectional studies are also frequently used to show "association" and suggest future studies.

Cohort studies are longitudinal studies which follow a group of patients with a common characteristic and collect data prospectively to answer questions related to the outcome variables of interest. This is the strongest of the observational study designs in that it is the least subject to statistical biases and errors in data recall and collection due to its prospective nature. Studies that are intended to determine the natural history of a particular disease process, such as the Framingham study, are examples of cohort studies.

EXPERIMENTAL STUDIES

Experimental studies prospectively evaluate the efficacy of an intervention or treatment against another treatment. They can be either **uncontrolled** or **controlled**. **Uncontrolled studies** describe the effect of an intervention or treatment in a single group of patients. This type of study is uncontrolled because it makes no comparison with another treatment or no treatment at all. Thus, this design does not account for the effect of confounding variables on study outcome. It may, therefore, be difficult to interpret the results of such studies as we can never be sure that the treatment effect we see is due to the intervention and not due to the effect of a confounding variable. If the natural course of a disease process is predictable, a control group is less important. A study on the value of intravascular fluids in hypotensive shock, for example, would not necessarily need a control group. Without fluids, a hypotensive patient will almost certainly die (the natural course of the disease is predictable). An uncontrolled study design would therefore be appropriate in this circumstance.

If the natural course of disease is to improve spontaneously, however, a control group is important. We might, for example, perform an uncontrolled study to evaluate the effect of a drug on the common cold. When we analyze our results, however, we may interpret patient improvement as being due to our drug when, in reality, the patients were likely to improve regardless of our therapy. In this situation, we need to be able to compare the patients who received the drug with patients who did not to be able to make valid conclusions regarding the clinical benefit of the drug.

Uncontrolled studies are also subject to the "**Hawthorne effect**", which is the tendency of patients to change their behavior as a result of a study. In our hypothetical study on the common cold, for example, some patients, wishing to please their doctors, might report fewer symptoms and therefore appear improved. Because their attention was focused on their colds as a result of the study, they might also take care to get more sleep, eat properly, or drink more fluids. Any of these factors might result in improvement in their symptoms which would appear to be due to our drug, but, in fact, would not be. A controlled study would thus be necessary to account for these potential confounding variables.

"**Regression towards the mean**" is another potential error to consider in an uncontrolled study. This refers to the fact that there is a certain degree of error associated with any physiologic measurement we make. Suppose, for example, that we decide to study the effect of a drug on blood pressure and include only those patients with systolic blood pressures (SBP) over 150 mmHg. Some patients will truly have a SBP over 150 mmHg. Others, however, will have an erroneously high initial SBP due to measurement error and will subsequently have normal SBP measurements. In an uncontrolled study, we might interpret this apparent decrease in mean SBP (a "regression towards the mean") as being due to the treatment drug. In reality, however, it is due to the fact that our sample population does not accurately represent the population we wished to study (i.e., true hypertensive patients). A controlled study theoretically evenly distributes the effect of regression towards the mean between the study and control groups, and negates its confounding effect on study conclusions.

A final error that may occur with uncontrolled trials is the "**placebo effect**". Some patients who receive a drug may believe that the drug is making them better when it is not. In an uncontrolled trial, such as our common cold study where the outcome is largely subjective, it may be difficult to separate placebo effect from true clinical improvement.

This is not to say that uncontrolled studies have no role in medical research. In situations where the outcome without therapy is clearly predictable, an uncontrolled trial may be perfectly appropriate. Uncontrolled studies are a frequent first step in the preliminary evaluation of a therapy to determine safety

and technical information. The results of these preliminary studies may then suggest the need for larger controlled clinical trials.

Controlled studies compare the impact of an intervention on two or more groups of patients. They compare the treatment group which receives the intervention to a control group which does not or which receives a different intervention that is frequently the current standard of treatment. Properly designed controlled studies equally distribute the potential sources of error discussed for uncontrolled studies between both the treatment and control groups. The effect of these errors will then tend to "cancel" each other out in the statistical analysis. Controlled studies are therefore statistically stronger than uncontrolled trials and are frequently necessary to adequately prove causality.

In our example of fluid administration to hypotensive patients, we saw that when the outcome was obvious, an uncontrolled study was sufficient. Consider a study in which we wish to demonstrate whether crystalloid or colloid administration is superior in reversing hypotension in the same patients. It is not immediately obvious whether crystalloid or colloid is the better choice of resuscitation fluid. To perform this study, therefore, we would need to perform a controlled study where we can compare the effect of one intervention (crystalloid) with another (colloid) to determine which therapy is superior.

In the ideal controlled study, both the treatment and control groups are exactly the same except for the intervention to be received. In this way, we can attribute any differences that we detect between the groups to the intervention alone. We also minimize the effect of statistical bias and confounding variables on the validity of our conclusions. In reality, however, it is impossible to create two groups which are exactly alike, and the goal of designing a study is to minimize the group differences as much as possible.

TREATMENT ALLOCATION: RANDOMIZED VS NON-RANDOMIZED

The method by which we determine which patients will receive which treatment is known as **treatment allocation**. As discussed above, the goal in designing a controlled study is for both the treatment and control groups to be as similar as possible. The best way to accomplish this is to **randomly** assign patients to one group or another such that each patient has an equal chance of receiving either treatment. Randomization should equally spread any potential differences that might act as confounding variables between the groups, decreasing the likelihood that they will affect the data analysis and lead to erroneous conclusions. Randomization also tends to equally distribute the potential effect of statistical bias between the groups.

The method used to ensure accurate randomization is important. Assigning patients to one treatment or another by methods such as odd versus even days of entry, days of the week, birthdates, first letter of last name, etc., although seemingly "random", do not ensure accurate randomization and group equality. Patients may be more likely to be treated on certain days of the week than on others, for example, due to clinic schedules, etc. These methods also do not prevent selection bias on the part of the investigator who, intentionally or unintentionally, may assign a patient to the treatment he or she deems best. Even computer generated random numbers may not actually be "random". Care must therefore be taken to ensure that each patient has an equal chance of receiving the various treatments under study. The best way to randomize patients is through use of a random number table (such as those contained within most statistics textbooks). Even randomization does not necessarily ensure that the groups will be equivalent, however. Randomly occurring significant differences may still exist between the groups.

The ideal design for ensuring similar groups is a **double-blind, randomized trial** since it is least subject to statistical bias and is most likely to correctly identify causality. "Double-blind" studies are those in which neither the physician nor the patient knows to which group (treatment or control) the patient is assigned. This decreases the impact of the Hawthorne and placebo effects since the patient does not know which therapy, if any, he is receiving. Furthermore, it minimizes the introduction of selection bias on the part of the investigator since he or she does not need to decide which therapy a patient will receive. Even double-blind studies are not infallible, however, and several reports exist of investigators opening or transilluminating randomization envelopes in order to force the assignment of certain patients into one group or another. For this reason, investigators who have a strong preference for one treatment option over another should not participate in a randomized trial. The study should also not include patients for which one treatment is clearly medically preferable to the other.

A "single-blind" trial refers to a study in which only the patient does not know to which group he or she is assigned. This type of study is necessary when, for medical reasons, the investigator must know which therapy a patient is receiving. Some studies, due to the obvious nature of the disease or treatment, cannot be blinded at all.

It is not always possible to perform such a double-blind study nor is it always necessary. The more complicated the study design, the more costly the study, and the less likely it will be properly performed. A controlled study with validity in-between that of an unrandomized and a double-blind, randomized controlled study is a **cross-over study**. In this design, randomly chosen patients initially receive one of the interventions (either treatment or control). After a period of time, the intervention is stopped and patients are allowed a "washout period". Patients then receive the alternate treatment option for a similar period of time. Patient response to the various interventions is then compared with the patients serving as their own controls.

SAMPLE SIZE

Calculation of the number of patients or **sample size** necessary to perform a study should be one of the initial steps in study design. If such calculations (also known as **power analysis**) are not performed and insufficient data are collected, we may not have the statistical power to make accurate conclusions and we will have performed a useless study. At the same time, if we study more patients than are necessary to demonstrate a statistically significant difference, we will have performed a more expensive and time consuming study than was necessary.

In order to calculate the sample size necessary for a particular study, we must first answer four questions:

1. What is our desired significance level?
2. What is our desired level of statistical power?
3. What difference is clinically important?
4. What is the usual variability present?

As we saw in Chapter 1, the significance level, or alpha, is the probability of making a Type I error (i.e., rejecting the null hypothesis when there is no difference). Most researchers use a significance level of 0.05; that is, they wish to have 95% confidence that their conclusions are accurate and that they are not committing a Type I error.

Statistical power quantifies the ability of a study to detect a clinically significant difference when present. Power is defined as 1 - beta, where beta is the probability of making a Type II error (i.e., rejecting the alternate hypothesis when there is a difference). Most researchers would consider a power of 80-90% to be appropriate for most studies. It is important to remember that as we decrease the desired significance level or increase the desired power, we must increase our sample size.

The difference that is considered to be clinically important is determined by clinical experience. A treatment which results in an increase in cardiac index of 200 mL is probably not clinically important whereas an increase of 1000 mL is. The smaller the clinically important difference, the more difficult it will be to prove statistically, and the larger the sample size necessary. Similarly, the larger the difference, the easier it will be to prove, and the smaller the sample size necessary.

Quantifying the degree of usual variability present in the population of interest is perhaps the most difficult of the four questions to answer. It is best answered using clinical experience, if available, or results from previous studies in the literature.

The calculation of sample size is based on the equations used to calculate the critical z value which determines whether two populations are significantly different:

$$z = \frac{\overline{X} - \mu}{sd / \sqrt{n}}$$

where z = the critical value of the normal distribution, $\overline{X}$ = the mean of the sample, μ = the mean of the population, sd = the standard deviation of the sample, and n = the sample size

By rearranging and solving for n, we derive the following equation, which can be used to calculate the sample size necessary for any study in which there are two populations of patients designated population 1 and population 0.

$$n = \left[\frac{(z_\alpha - z_\beta)\,sd}{\mu_1 - \mu_0} \right]^2$$

where z_α = the critical value associated with the null hypothesis, z_β = the critical value associated with the alternate hypothesis, μ_1 = the mean of population 1, μ_0 = the mean of population 0, sd = the standard deviation of populations 1 and 0, and n = sample size

CALCULATING SAMPLE SIZES FOR A SINGLE MEAN

It is not uncommon to evaluate a new treatment in which we wish to determine whether a physiologic parameter is significantly different from a standard or normal value. For example, suppose we hypothesize that pulmonary embolism significantly increases the pulmonary vascular resistance index (PVRI). How many patients with pulmonary embolism would we need to observe to show that this is true?

In order to calculate the necessary sample size, we must first answer the four questions listed above. We will use a significance level of 0.05 since we wish to determine our results with 95% confidence that we are not committing a Type I error. We will use a power of 0.90 such that we only have a 10% chance (1-0.90) of making a Type II error. If we assume that a normal PVRI is 250 dynes·sec·cm^{-5}/m^2, we might decide that a significantly increased PVRI is one that is 50% above normal (or greater than 375 dynes·sec·cm^{-5}/m^2 in this example). With regards to the usual variability present, if we consult the medical literature we find that the usual standard deviation associated with PVRI is approximately 200 dynes·sec·cm^{-5}/m^2 for patients with pulmonary embolism.

From a z distribution table, we determine that the two-tailed critical z value associated with a significance level of 0.05 is 1.96. We must perform a two-tailed test since we do not know with 100% certainty that PVRI will increase as a result of pulmonary embolism (it might decrease). The lower one-tailed critical value of z for a beta of 0.1 (1-power) is -1.28. In sample size calculations, power is always considered as one-tailed. Using these values, and the mean and standard deviation information we determined above, we can now calculate the necessary sample size:

$$n = \left[\frac{(z_\alpha - z_\beta)\,sd}{\mu_1 - \mu_0} \right]^2 = \left[\frac{(1.96 - (-1.28))(200)}{(375 - 250)} \right]^2 = 26.9 \cong 27$$

Thus, to show that patients with pulmonary embolism have significantly elevated PVRI measurements, we will have to observe 27 patients with pulmonary embolism to have a 95% chance of detecting a significant increase in PVRI with 90% power.

Although we chose the commonly used values for significance level (0.05) and power (90%) in the above example, we might have chosen to use other values. Remember from Chapter One that in order to decrease the probability of making both a Type I error (i.e., by decreasing the significance level) and a Type II error (i.e., by increasing the power), the sample size has to increase. Consider the alterations in sample size that occur as we vary the significance level and power utilized in the above sample size calculation:

Changes in Sample Size with Alterations in Significance Level

Significance level	Power	Clinically Significant Difference	Standard deviation	Sample size
0.10	0.90	50%	200	**22**
0.05	0.90	50%	200	**27**
0.01	0.90	50%	200	**39**
0.005	0.90	50%	200	**43**

Changes in Sample Size with Alterations in Power

Significance level	Power	Clinically Significant Difference	Standard deviation	Sample size
0.05	**0.95**	50%	200	**34**
0.05	**0.90**	50%	200	**27**
0.05	**0.80**	50%	200	**21**
0.05	**0.70**	50%	200	**16**

The necessary sample size for any study will also vary depending on what we define as being a clinically significant difference as well as on the usual variability in the population. The smaller the difference we wish to detect, the larger the sample size necessary to prove that the difference has not occurred by chance. Similarly, the less variable the data, the easier it will be to detect a difference compared to data that are more variable. Consider the changes in sample size that occur as we manipulate the clinically significant difference and standard deviation for the pulmonary embolism example:

Changes in Sample Size with Alterations in the Clinically Significant Difference

Significance level	Power	Clinically Significant Difference	Standard deviation	Sample size
0.05	0.90	**10%**	200	**672**
0.05	0.90	**25%**	200	**108**
0.05	0.90	**50%**	200	**27**
0.05	0.90	**100%**	200	**7**

Changes in Sample Size with Alterations in Standard Deviation

Significance level	Power	Clinically Significant Difference	Standard deviation	Sample size
0.05	0.90	50%	**100**	**7**
0.05	0.90	50%	**200**	**27**
0.05	0.90	50%	**300**	**61**

CALCULATING SAMPLE SIZES FOR TWO MEANS

We can apply a similar equation to calculate the sample size necessary to determine whether two population means are significantly different. To do so, we must first determine our desired significance level, our power, the mean for each population, and an estimate of the variability in each population. In practice, we assume that the "variability" or standard deviation is the same in each population. The equation we use takes the following form:

$$n = 2\left[\frac{\left(z_\alpha - z_\beta\right)}{\mu_1 - \mu_0}\right]^2$$

where n = the sample size necessary in each population

Consider the example from Chapter Six in which we compared the effect of positive end-expiratory pressure (PEEP) on arterial oxygen tension (PaO_2). Using a single-sample, paired t-test, we found that there was a highly significant difference in PaO_2 after the application of PEEP ($p<0.0001$). Suppose we wanted to know how many patients we needed to just show a significant difference using a significance level of 0.05 and a power of 0.90. In the example from Chapter Six, the pre-PEEP mean PaO_2 was 59.1 torr and the post-PEEP mean was 74.5 torr. Since the standard deviations of the two populations were similar (10.7 and 9.9 torr), we will simply use the average of the two as our estimated standard deviation. We now have the necessary information to calculate the sample size required.

$$n = 2\left[\frac{\left(z_\alpha - z_\beta\right)sd}{\mu_1 - \mu_0}\right]^2 = 2\left[\frac{(1.96 + 1.28)(10.3)}{(75.4 - 59.1)}\right]^2 = 8.4 \cong 9$$

where z_α = the two-tailed critical z value for a significance level of 0.05 and z_β = the critical one-tailed z value for a power of 0.90

Although the actual sample size is 8.4 patients, we always round up to the next integer. Thus, in order to be able to detect a significant difference between the means of 75.4 and 59.1 torr (standard deviation of 10.3 torr) with a significance level of 0.05 and power of 0.90, we would need 9 patients.

Calculation of Sample Size for Proportions

In some studies, we are interested in the percent change that occurs as a result of a particular intervention or treatment rather than the absolute change in the means. We might, for example, wish to study whether preoperative antibiotics decrease the incidence of wound infections in trauma patients with penetrating bowel injuries. If we know that 30% of such patients will develop a wound infection without preoperative antibiotics, we might wish to determine how many patients who receive preoperative antibiotics would be required to demonstrate a 50% reduction in the wound infection rate.

To calculate this sample size, we use a modification of the sample size equation we have used previously, adapted for use with proportions. The equation for comparing a single proportion (π_1) against a known value (π_0) is as follows:

$$n = \left[\frac{z_\alpha \sqrt{\pi_0(1-\pi_0)} - z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2$$

Assuming a significance level of 0.05 and power of 0.90, our critical z_α and z_β values will be 1.96 and -1.28 respectively. Our known proportion, π_0 (the true wound infection rate), is 0.30 and our study proportion, π_1, is 50% of 0.30 or 0.15. Using this information we calculate our necessary sample size as:

$$n = \left[\frac{1.96\sqrt{(0.30)(1-0.30)} + 1.28\sqrt{(0.15)(1-0.15)}}{0.30 - 0.15} \right]^2 = \left(\frac{1.36}{0.15} \right)^2 = 83$$

In order to be able to detect a 50% reduction in wound infections due to the use of preoperative antibiotics, therefore, we would need to study 83 patients before we would have sufficient data to be able to detect a significant difference with 95% confidence and 90% power. If we perform the study with 80% power, we would only need to study 65 patients, but we would have to accept the higher risk of making a Type II error.

CREATION OF A RESEARCH PROTOCOL

Once we have chosen an appropriate study design, determined our method of treatment allocation, and calculated the sample size necessary to identify a significant result, it is time to perform the study "on paper". Creation of a research protocol allows us to carefully consider and define each step of the study, anticipating potential problems and sources of error before the actual data collection begins. It is also a necessary first step in applying for research funding and investigational review board (IRB) approval. A typical research protocol considers the elements listed in Figure 9-5, each of which will be addressed in detail.

- **Purpose**
- **Patient population**
- **Study hypotheses**
- **Methods (including treatment)**
- **Inclusion and exclusion criteria**
- **Outcome variables**
- **Data to be collected**
- **Statistical methods to be applied**
- **Anticipated economic cost**

Figure 9-5: Elements of a research protocol

- Purpose: Any research protocol should begin by considering the question “Is this study necessary and, if so, why?” This is especially true in this day and age where research funding is becoming more scarce, competition for funding more fierce, and IRB approval more difficult to obtain. The rationale for performing the study should be clearly defined and previous studies in the medical literature which support the need for this study identified. The scientific background for the study should be outlined such that a grant reviewer or IRB committee, upon reading the protocol, will have sufficient knowledge to understand the study and its hypotheses.

- Patient Population: The patient population of interest should be described in detail to provide a clear image of to which population the study conclusions will be applicable. This is also an essential step in ensuring that the method of treatment allocation will result in a representative patient sample being chosen.

- Study Hypotheses: The primary research hypothesis and all secondary hypotheses should be clearly stated. Some studies will only have a single hypothesis while others will have several. It is generally a good idea to limit the number of hypotheses in a study to a primary hypothesis and no more than 4 or 5 secondary hypotheses. Multiple hypotheses can result in an unwieldy study that is complicated and therefore difficult to perform accurately. Simplicity is best.

- Methods: This is the most important part of the research protocol. It is here that the details of the study are defined and potential problems identified. All necessary equipment and supplies should be listed as well as the manner in which data will be measured and collected. The treatment being administered, if any, should also be clearly described including the method by which treatment will be allocated.

- Inclusion and exclusion criteria: The criteria which will be used to identify patients eligible for the study (“inclusion criteria”) should be identified as well as those criteria which will eliminate a patient from being in the study (“exclusion criteria”). It should be kept in mind that the more specific the inclusion criteria, the smaller the population to which the results will be applicable. For example, if we include “all adult males” in our study, the conclusions that we make will be applicable to a large patient population. If, however, we limit the study to “all males between 18 and 25 years who have hypertension," the population to which our conclusions will be relevant is a much smaller one. We must therefore take care to ensure that our inclusion criteria include the population to which we wish to apply our results.

  Our choice of exclusion criteria is just as important, if not more so. Exclusion criteria represent those criteria by which we will reject patients from entering the study who otherwise meet the entrance criteria. It is essential that the exclusion criteria be defined prior to and followed carefully during the study in order to prevent the introduction of selection and assembly bias.

- Outcome variables: Those variables which will be used to define response to the treatment (or lack thereof) should be identified. These variables might include survival, presence of organ failure, successful extubation, decreased length of stay, or any of a multitude of other outcome measures.

- Data to be collected: All of the independent variables to be evaluated during the study should be listed. It is important to consider, before the study begins, which types of data will be necessary to answer the study hypotheses. It is frequently difficult, if not impossible, to obtain or recreate data after data collection is completed. Care should therefore be taken to anticipate and collect all of the data that may be required during data analysis. It is always better to have too much data than too little. By the same token, data collection should be limited to those variables which are pertinent to the study hypotheses. The collection of extraneous data is time consuming and costly, and tends to decrease the validity of the data collection overall.

- Statistical Analysis: Considering the types of data variables which are being studied, it is generally a good idea to list the statistical methods that will be used during data analysis. By considering the statistical analysis before the study, the need for inclusion of further data variables may become apparent.

- Anticipated Economic Cost: All costs that will be incurred by performance of the study should be listed including labor and equipment costs, pharmacy costs, laboratory fees, radiologic procedure charges, etc. Care should also be taken to account for unforeseen costs that may arise once the study begins.

There are several benefits of taking the time to prepare a research protocol prior to beginning a study. First, as previously mentioned, it allows us to perform the study "on paper" in such a manner that we can anticipate and correct for problems, potential errors, and sources of statistical bias before the actual study begins. Second, it is a necessary requirement for applying for research funding or IRB approval. Third, having drafted the research protocol, the preparation of abstracts and manuscripts for publication is simplified. The background literature research has already been performed and the Introduction and Methods sections have essentially been written. Thus, as we will see in the next chapter, the abstract and manuscript for a study have largely been composed before the study is ever performed.

SUGGESTED READING

1. Altman DG. Statistics and ethics in medical ressearch:misuse of statistics is unethical. Brit Med J 1980; 281:1182-1184.
2. Altman DG. Statistics and ethics in medical ressearch: Study design. Brit Med J 1980; 281:1267-1269.
3. Altman DG. Statistics and ethics in medical ressearch: How large a sample?. Brit Med J 1980; 281:1336-1338.
4. Altman DG. Statistics and ethics in medical research: Collecting and screening data. Brit Med J 1980; 281:1399-1401.
5. Dawson-Saunders B, Trapp RG. Basic and clinical biostatistics (2nd Ed). Norwalk, CT; Appleton and Lange, 1994:6-19, 199-120, 156-157.
6. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. NEJM 1987;317:426-432.
7. O'Brien PC, Shampo MA. Statistics for clincians: 11. Survivorship studies. Mayo Clin Proc 1981;56:709-711.
8. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233-240.
9. Arkin CF, Wachtel MS. How many patients are necessary to assess test performance? JAMA 1990; 263:275-278.
10. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266:93-98.
11. Wassertheil-Smoller S. Biostatistics and epidemiology: a primer for health professionals. New York; Springer-Verlag, 1990: 6, 7, 86-103.
12. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology - the essentials. Baltimore; Williams and Wilkins, 1982: 131-142.
13. Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon: stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. NEJM 1985; 312:1604-1608.
14. Ellenberg JH. Clinical trials. Neurologic Clinics 1990, 8:15-30.
15. van der Linden W. Pitfalls in randomized surgical trials. Surgery 1980; 87:258-262.
16. Lawrence W. Some problems with clinical trials. Arch Surg 1991; 126:370-378.