UNM COLLEGE of PHARMACY

UNM RADIOPHARMACY 1973 2003 30 YEARS OF EXCELLENCE

.::VOLUME 17, LESSON 5::.

Evaluating the Medical Literature
Part 1: Basic Principles

Continuing Education for Nuclear Pharmacists
And
Nuclear Medicine Professionals

By

William G. Troutman, PharmD

The University of New Mexico Health Sciences Center, College of Pharmacy is accredited by the Accreditation Council for Pharmacy Education as a provider of continuing pharmacy education. Program No. 0039-0000-14-175-H04-P 2.5 Contact Hours or 0.25 CEUs. Initial release date: 8/15/2014

-- Intentionally left blank --

Evaluating the Medical Literature
Part1: Basic Principles

By
William G. Troutman, PharmD

Editorial Board
Stephen Dragotakes, RPh, BCNP, FAPhA
Michael Mosley, RPh, BCNP, FAPhA
Neil Petry, RPh, MS, BCNP, FAPhA
Janet Robertson, RPh, BCNP
Tim Quinton, PharmD, BCNP, FAPhA
Sally Schwarz, BCNP, FAPhA
Duann Vanderslice Thistlethwaite, RPh, BCNP, FAPhA
John Yuen, PharmD, BCNP

Advisory Board
Christine Brown, RPh, BCNP
Leana DiBenedetto, PharmD, BCNP
Dave Engstrom, PharmD, BCNP
Walter Holst, PharmD, BCNP
Scott Knishka, RPh, BCNP
Susan Lardner, BCNP
Vivian Loveless, PharmD, BCNP, FAPhA
Brigette Nelson, MS, PharmD, BCNP
Brantley Strickland, BCNP

Director, CENP
Kristina Wittstrom, PhD, RPh, BCNP, FAPhA
UNM College of Pharmacy

Administrator, CE & Web Publisher
Christina Muñoz, M.A.
UNM College of Pharmacy

While the advice and information in this publication are believed to be true and accurate at the time of press, the author(s), editors, or the publisher cannot accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty,
Copyright 2014
University of New Mexico Health Sciences Center
Pharmacy Continuing Education

Instructions:

Upon purchase of this Lesson, you will have gained access to this lesson and the corresponding assessment via the following link < https://pharmacyce.health.unm.edu >

To receive a Statement of Credit you must:

1. Review the lesson content
2. Complete the assessment, submit answers online with 70% correct (you will have 2 chances to pass)
3. Complete the lesson evaluation

Once all requirements are met, a Statement of Credit will be available in your workspace. At any time you may "View the Certificate" and use the print command of your web browser to print the completion certificate for your records.

NOTE: Please be aware that we **cannot** provide you with the correct answers to questions you received wrong. This would violate the rules and regulations for accreditation by ACPE. The system will identify those items marked as incorrect.

Disclosure:

The Author(s) does not hold a vested interest in or affiliation with any corporate organization offering financial support or grant monies for this continuing education activity, or any affiliation with an organization whose philosophy could potentially bias the presentation.

This lesson is a reprint of that initially released in 1998. The content is judged still relevant and useful to the basic understanding of relevant literature documenting scientific clinical trials involving diagnostic agents.

EVALUATING THE MEDICAL LITERATURE
PART 1: BASIC PRINCIPLES

STATEMENT OF LEARNING OBJECTIVES:

The goal of this correspondence continuing education lesson is to increase the reader's ability to evaluate a wide range of medical literature. The process of literature evaluation begins with the recognition of the type of study being evaluated and then proceeds to the examination of the individual components of the study.

Upon successful completion of this lesson, the reader should be able to:

1. Distinguish between descriptive, observational and experimental studies
2. Identify the objective of a study and its published report.
3. evaluate the following types of published reports:
    a. case reports/case series
    b. cross-sectional studies
    c. case control studies
    d. cohort studies
    e. experimental studies (clinical trials)
4. define unblended, single-blind and double-blind as they relate to study design
5. define placebo control, active treatment control and cross-over as they relate to study design
6. describe some of the problems encountered in evaluating the test methodologies used in studies
7. differentiate between statistical and clinical significance

COURSE OUTLINE

EVALUATION THE MEDICAL LITERATURE
PART1: BASIC PRINCIPLES

William G. Troutman, PharmD

INTRODUCTION

"Most people will agree that there is too much of the scientific literature, but nobody seems to have a convincing remedy." This statement reflects the dilemma facing health care practitioners today. We have no shortage of information available to us and we have multiple means by which to access that information. There is far more information available in our practice areas, even for subspecialist, than we can possibly read and comprehend. The challenge is to find continuing sources of applicable information and to evaluate the information available from those sources before applying it in patient care. The growth rate in the number of biomedical publications may be exceeding the growth rate of the production of high-quality manuscripts. This places additional importance on the health care practitioner's ability to evaluate literature. Literature evaluation is not difficult; it is a game which requires the player to exercise a mixture of applied common sense and healthy skepticism. This lesson will provide the reader with the basic skills needed for performing an evaluation of a published study. **A later lesson will focus specifically on the issues related to evaluating diagnostic studies**.

When appropriate, a hypothetical radiopharmaceutical product PhindALL® (^{99m}Tc-d-obfuscate) has been selected to provide the examples in this lesson. This new product is used for the imaging of the hepatobiliary system, with its greatest application in the diagnosis of gall bladder disease.

COMMON STUDY TYPES

There are several schemes for classifying studies. This lesson will use the scheme most frequently used by major texts which offer guidance on research design and literature evaluation.[2-4] Published reports of studies usually fit into one of three broad classes: descriptive reports, observational reports, or experimental reports. Each of these classes will be reviewed in depth later in this lesson, but merit some introduction here. Descriptive reports describe an observation. The authors generally played no intentional part in the circumstances preceding the event being described. The authors noticed something interesting and chose to write it up for publication. If a single patient is being described, the

report is called a case report. If several patients are described, the report is labeled a case series. Observation reports differ from descriptive reports in that there is a structure to the manner in which patients are selected and evaluated that allows assessment of causality. Observational studies include cross-sectional, case-control, and cohort study formats. They present a higher order of evidence for the association between a drug and an observed effect than descriptive studies. Experimental studies (i.e., clinical trials) include not only structured patient selection and evaluation, but also application of an intervention (such as drug administration) in a manner determined by the study designers. This greater control over the study circumstances makes the results of experimental studies generally accepted as the highest order of evidence for drug effects. For this reason, the evaluation of published experimental studies will be the major focus of this lesson. Some issues common to all study types (i.e., patient selection) will only be presented in the discussion of experimental studies in order to minimize repetition.

EVALUATION OF DESCRIPTIVE REPORTS

Case reports and series present unstructured observations of one or more patients. Often they present relationship between drugs and effects that were not expected. As such, they can bring potential drug problems or therapeutic breakthroughs to the attention of the reader. Much of what we know about drug therapy, especially adverse drug events started with an observation in a single patient. These reports may serve to stimulate researchers to design a more definitive study to investigate the observations described in the case report or case series. On the other hand these reports may only serve to muddle our understanding of a drug's effects by associating the drug with unrelated events. For example, if a case report describes a patient who suffered a stroke while receiving PhindAll®, it may not fully explore the patient's risk factors for stroke and its conclusion, that PhindAll® is responsible for the stroke may be incorrect. Descriptive reports cannot assess any level of a cause and effect relationship, but rather only suggest the possibility of a relationship between two or more events.[5] The key considerations for evaluating descriptive reports are presented in Table 1.

Descriptive reports are the best examples of the published literature's aversion to negative findings. While the case report claiming that PhindAll® caused a stroke may be published, a similar report presenting a case of a patient who did not have a stroke is unlikely to ever be published. While negative findings are important in clinical decision making the journals are reluctant to present them and many researchers are reluctant to even write them up.[6] This phenomenon is not unique to descriptive reports or to the biomedical literature as a whole. Daily newspapers and television news

programs routinely present accounts of terrible crimes occurring in the community while ignoring all of the people and neighborhoods that had ordinary, violence-free days. The descriptive literature, therefore, cannot be considered an accurate mirror of reality. In many ways, descriptive reports are better at raising questions than they are at answering them.

Table 1

CONSIDERATIONS IN THE EVALUATIONS OF DESCRIPTIVE REPORTS
1. The temporal relationship between the administration of the drug and the observed effect.
2. The likelihood that there may be other factors present which might cause the effect.
3. The results of withdrawing the drug (Did the effect disappear?).
4. The results of rechallenge, if any (Did the effect reappear when the drug was given again?).
5. The presence of other data supporting a relationship between drug and effect (Previous reports in the literature, etc.).
6. The biologic plausibility of the drug exposure and the effect.

A case report should not be confused with an *n-of-l* study which is a form of experimental study. In an n-of-l study, the investigators intentionally expose a patient to a drug with the purpose of observing one or more effects by comparing selected patient characteristics before and after the exposure. Unlike the typical case report which is conceived after the exposure has occurred, the n-of-l study is planned in advance of the exposure and appropriate monitoring takes place. N-of-l experiments can be quite convincing. The positive response of the first human to receive exogenous insulin established that the therapy could have a beneficial effect.[7] It remained to other researchers to determine how insulin therapy could be applied to the broad range of diabetics. N-of-l studies typically provide better evidence of a drug-effect relationship than case reports, but they share significant limitations with case reports on their extrapolation to other patients.

EVALUATION OF OBSERVATIONAL REPORTS

Cross-Sectional Studies

Cross-Sectional Studies are probably the most frequently encountered studies in the general population. While not labeled as such, consumer satisfaction and opinion surveys, political polls and even elections are forms of cross-sectional studies: they take a sample of a population and determine the presence of a characteristic at one point in time. In health care, they are useful for determining the prevalence of diseases, adverse reactions, or other medical events. Prevalence and incidence are frequently confused: see Table 2 for their definitions.[4, 8] Cross-sectional studies can be repeated to show trends over time. An example might be the annual determinations of the prevalence of a disease in a community. The results of the repeated cross-sectional studies might show that the prevalence is

rising, declining, or staying the same, but the results cannot provide an explanation for the observations.

Table 2

PREVALENCE VS. INCIDENCE		
Term	***Definition***	***Typical Expression***
Prevalence	Number of subjects with a disease at one point in time OR Number of subjects at risk at that time	percentage or fraction
Incidence	Number of new cases of a disease during a time interval OR Number of subjects at risk at the beginning of the interval	%/year OR Cases/1000/year

Cross-sectional studies have several advantages. They can be designed and implemented in a short period of time and their simplicity makes repetition easy. They are also inexpensive and the investigator doesn't have to be concerned about losing subjects for follow-up because there won't be any follow-up. Cross-sectional studies are well suited for the superficial study of rapidly changing conditions.

The temptations of the fast and easy cross-sectional study design must be tempered with an appreciation of its shortcomings. Cross-sectional studies only provide one look at the problem being investigated and may not, therefore, represent typical circumstances. It is possible that the same study conducted 15 minutes later might result in completely different findings. Cross-sectional studies are not particularly useful for investigating rare events, especially when they are conducted in the general population. Using targeted populations will help; investigators might choose to only interview or examine patients with characteristics thought to predispose them to the development of the condition under investigation. While patient selection is an issue with all study types, it creates special problems for cross-sectional studies. The subjects in a cross-sectional study might be stopped in the grocery store or handed a survey at their health clinic. Response rates are often quite low, and those who respond may not be representative of the population under study. Consumer surveyors often offer incentives for subjects to participate in their surveys. Is the person who completes a survey in exchange for a free product or small cash payment a good representative of the general population? Reports of well-conducted cross-sectional studies will provide detailed descriptions of the population under study and the survey methods.

Cross-sectional studies often serve as a basis for designing more definitive studies. Often, they provide the evidence needed to justify a study or the baseline characteristics of the population needed to calculate the appropriate number of subjects for a study.

Case-Control Studies

Case-control studies begin with an observation and look backward in time to try and determine the possible origins of the observation. Because they look back in time, case-control studies are sometimes called retrospective studies. Since any study, or even a case report, which looks backward in time could be considered a retrospective study, the terms retrospective and case-control are not synonymous. A case-control study begins with the identification of subjects who have the characteristic of interest (cases) and the identification of another group of subjects who do not have the characteristic (controls), but who otherwise closely resemble those who do. An investigation into the pasts of these subjects is then conducted to determine if there are events earlier in the lives of the case group which might explain the presence of the characteristic today, but which are absent from the earlier lives of the control group. For example, assume that case reports are beginning to appear in the literature describing the presence of a peculiar circular rash ("bull's eye" dermatitis) on the abdomens of patients who had their gall bladders surgically removed 5 years earlier. While the rashes don't itch or create any other serious problems for the patients, they do create some embarrassment during bathing suit season. A case-control study is designed to investigate the problem. A group of patients without gall bladders who have the rash are assembled (cases) and a similar group of ball bladder-free patients without the rash is identified. The investigators, preferably blinded, then read through all the medical records of the two groups of patients using a checklist they have developed to structure their investigation. They determine that most of the patients with rashes had received PhindALL® as part of the diagnostic procedure preceding their cholecystectomies while the patients without the rash had received other radiopharmaceutical drugs. If no other pattern is detected the investigators might then suggest that PhindAll® is associated with the delayed development of the unusual rashes.

The example study provides the opportunity to present the advantages and disadvantages of case-control studies. Case-control studies are useful for the study of effects that take time to develop. In this case, the rashes appeared five years after the cholecystectomies. It is unlikely that an investigator would establish a protocol which would require at least five years of observation before finding anything. A well-known example of this advantage of case-control studies was the association of unanticipated development of carcinomas of the cervix and vagina in young women who had been

exposed *in utero* to diethylstilbestrol (DES).[9] The case control design is also useful for studying rare events. By concentrating only on the subjects with the rare observation, the case-control study avoids the problem of having to follow hundreds or thousands of patients in the hope of finding a few cases of the rare condition. Investigators using the case-control design may choose to match the case and control groups for more characteristics than just the presence or absence of the rash or other primary characteristic. Patients could be matched by age and gender, for example, in order to make the case and control groups resemble each other even more closely. It is very difficult to identify an appropriate control group and the reader of a case-control study should pay special attention to the description of the subject selection process, determining if important characteristics have been included or excluded.

Case-control studies are relatively fast and inexpensive. Since the data have already been generated, there is no need for additional expensive laboratory work-ups or drug regimens. They have the ability to look at multiple possible causes of the observed characteristic at once. They are also useful for collecting preliminary data in advance of a prospective study. On the other hand, case-control studies present some major challenges to the reader. They rely on medical records or other data created in the past. The people creating the data were not thinking about the needs of later researchers when they wrote their reports. In the PhindAll® - rash example, the failure to record that PhindAll® was used as a diagnostic agent could lead to the erroneous conclusion that a rash case had no prior exposure to the drug. Medical records from different institutions may record data in different ways and it may be difficult for investigators to find what they are looking for. The author of the case-control study should tell the reader specifically which risk factors were considered while the case reviews were being conducted. It is always possible that the cause of the problem under investigation is not even on the list of things the researchers are looking for. Conversely, some authors seem to look for everything in the hope of finding something that distinguishes the cases from the controls. The process is called "data mining" and frequently leads to erroneous conclusions. For example, the statement, "The people with the rashes drank more milk as children [$p < .05$]," leads to the conclusion that childhood milk consumption might be the cause of the rashes. Case-control studies which pursue a large enough number of subject characteristics will usually find statistically significant differences between the groups simply' due to the statistical inevitability of such differences. However, observational studies, including case-control studies, are always in danger of missing a key association between events because they didn't specifically look for it. [10] Case-control studies cannot establish temporal relationships or exclude other causes, but well conducted case-control studies can give strong

indications of the association of an earlier event and its later consequences. The combination of the strength of the association developed by a well conducted case-control study and the dangers of studying the problem prospectively may combine to make the case-study the definitive evaluation of a problem. No one is likely to intentionally expose unborn children to DES to confirm that it causes cancer later in life. Table 3 is a summary of points to consider in evaluating case-control studies. [11]

Table 3

CONSIDERATIONS IN THE EVALUATION OF CASE-CONTROL STUDIES[11]
1. Cases and controls should be selected from a common, defined population.
2. The criteria for selection of cases and controls should be predetermined. Similarly, exclusion criteria should be well defined.
3. The definition of what will constitute prior exposure to the suspected risk factors should be predetermined.
4. The cases and controls should be similar in important characteristics including demographics, drug-exposure recall, and prior medical surveillance.
5. Data collection should use a structured format and, if interviews are required, interviewers who are unaware of the assignment of the subject to either the case or control group.

Cohort Studies

While case-control studies are retrospective observational studies, cohort studies are prospective. Cohort studies begin with the premise that there may be a relationship between a suspected risk factor and later outcomes in patients with that risk factor. The stimulus for a cohort study may be case reports or a case-control study. Before starting a cohort study, the investigators should have complete assurance that the presence of a risk factor truly preceded the outcome. Then they enroll two groups of subjects, one group with the risk factor and an otherwise similar group without the risk factor. Cohort studies usually involve a large number of subjects and continue for months to years, sometimes even decades. Much of our understanding of the risk factors leading to premature death from cardiovascular events has come from long-term cohort studies involving thousands of subjects.

To continue the example presented in the case-control section, researchers could identify patients undergoing cholecystectomies after diagnosis using PhindAll® and an otherwise similar group of patients whose diagnoses were established with other imaging agents. All of these patients could then be followed into the future as the researchers take note of any "bull's eye" rashes or other side effects that may develop. It is important to stress here that the researchers did not influence the choice of the diagnostic radiopharmaceutical used in the patients. In fact, the study of the patients might not have begun until weeks, months or even years after the administration of the drugs. The most important consideration in the evaluation of cohort studies is the determination that the two subject groups are as

similar as possible except for the risk factor being studied. The author of the published study should provide narrative or, more likely, tabular data comparing the groups which provide reassurance to the reader of the similarity of the groups. Finding people who differ in only one characteristic is almost impossible, especially when that characteristic may contribute to the development of others. For example, if one were to study the association between regular exercise and myocardial infarction, other characteristics might be hard to match. Would we expect exercisers and non-exercisers to have the same body weight? Since lower body weight is often the result of regular exercise, one would expect the non-exercisers to be heavier than the exercisers.

The observation of subjects in a cohort study is not haphazard, but rather follows a protocol of observations established by the researchers. The ability to control many sources of bias related to the selection of subjects and the recording of measurements and the establishment of stronger association and temporal relationships between suspected risk factors and outcomes are advantages of cohort studies over the other study design presented thus far. The protocol may require physical examinations or laboratory tests at regular intervals. Subjects with and without the risk factor should be evaluated in the same way and at the same intervals. Testing should be standardized and the interpretation of clinical findings should follow uniform guidelines, especially if more than one or two researchers will be performing the evaluations. Unlike a case-control study which focuses on a single outcome, cohort studies can be used to delineate a variety of outcomes that may be associated with a single risk factor. Cohort studies are especially well-suited for studying the course of a condition. They are not, however, without their limitations. Unlike case-control studies, cohort studies may take a long time to complete and may be very expensive, especially when a large number of subjects are needed. The number and complexity of evaluations to be performed may be far in excess of what a subject might normally receive and the extra costs are usually borne by the sponsors of the study. Because of their long duration, cohort studies are vulnerable to patient attrition. Patients may leave the study because of family relocation, loss of interest in the protocol, or even death. Cohort study reports must explain what happened to the subjects who didn't finish the study.

As with case-control studies, cohort studies are not definitive proof of the connection between a risk factor and an outcome. Despite this limitation, cohort studies are valuable for exploring trends and the outcomes associated with a risk factor. Because not every clinical problem can be examined through an experiment, cohort studies provide strong evidence supporting theories of causality. Table 4 is a summary of points to consider in evaluating cohort studies. [11]

Table 4

CONSIDERATIONS IN THE EVALUATION OF COHORT STUDIES[11]
1. The cohort of patients being studied should be representative of the segment of the population to which the results will be applied.
2. The subjects with and without the risk factor being studied should be similar in important characteristics including demographics, drug exposure recall, and prior medical surveillance.
3. Drug exposure and drug compliance should be determined equally in both groups.
4. Physical examination, laboratory testing and other forms of data collection should use a structured format, equally applied in all patients.
5. Dropout rates and characteristics of dropouts should be similar for groups with and without the risk factor.

EVALUATION OF EXPERIMENTAL REPORTS

As has been mentioned previously, experimental studies are the highest order of evidence for establishing cause and effect relationships. Whenever practicable, experimental studies should be conducted to confirm or refute the results of observational studies. Experimental studies conducted in humans are called clinical trials. The following discussion of the evaluation of experimental studies will focus on clinical trial design and conduct. While clinical trials will be evaluated, the principles presented can be applied to other study designs. The discussion will be presented in approximately the same order in which the described elements appear in the typical published clinical trial. The introduction to the study is presented first, followed by the methods employed in the study, the presentation of the results, and the discussion and conclusions drawn from the results. The biomedical literature is not known for radical variations in style or language. Stylistically, its predictability can be used to the reader's advantage since you know what to expect from each of the sections of the article. It is imperative that you read and evaluate the entire article. Flaws in one part of a study may lead to errors in other parts. The perfect study has yet to be executed, and that fact must guide literature evaluation. We are most concerned about what we might term "fatal" errors, those which are of sufficient magnitude to render the study results practically useless. We cannot rely on journal editors or reviewers to weed out all fatally flawed studies.

Introduction

The introduction to a published clinical trial is the first part of the main body of the article. It is not the abstract or summary which may precede the main body. The introduction contains two very important types of information regarding the study. It presents background information on the problem being addressed through the study. The background information should not present the results of the study, but rather should provide enough information so that the reader can understand the factors which

motivated the investigators to conduct the study. The background presentation should cite key previous studies. The length of the background portion of an article's introduction may be dictated by the editorial style of the journal in which it is being published. Some editors prefer brief introductions with the bulk of the presentation of previous work appearing in the discussion section of the article. The second important feature of the background section is the statement of the study objective. The failure to present a clearly stated objective is a major obstacle to understanding the rest of the article and it will often contribute to the rejection of a manuscript by a journal. The ideal clinical trial objective clearly describes what is going to be measured, the subjects who will be evaluated, and the evaluation methodology. As with other ideal things in this world, the ideal objective is hard to find. Mention of the subjects may wait until the methodology section of the article and the manner of evaluation is routinely missing from the objective. Despite these shortcomings, it is important to locate and understand the objective before reading further in the article. Underlining the objective makes it easy to review when evaluating the other sections of the article. The reader should be able to determine if the objective of the study was conceived prior to data collection or if it evolved after the authors looked at their data. The latter is more likely to feature chance findings. An appropriate objective might be, "The objective of this study is to determine whether ^{99m}Tc-d-obfuscate (PhindAll®) is superior to ^{99m}Tc-mebrofenin for identifying the presence or absence of bile stasis in patients with suspected cholecystitis later confirmed by surgical pathology."

At least one major text[2] on literature evaluation advises the reader to bypass the introduction section of the article. Its premise is that the methods section of the article is where the important information will be found. On the other hand, without a clearly stated objective, how can we determine if the methodology is appropriate? You don't get onto an airliner without some idea of its destination.

Methodology

The methodology section of a study report is usually the most difficult to read and, therefore, the section most frequently skipped by readers. The publishers of the journals exacerbate this problem when they print the methodology section in a smaller font size than the rest of the article. The message to the reader is clear, "If you must read this section, go ahead, but we aren't going to waste large type on it because it isn't important." You must resist this message. The methodology section is the heart of the article. Mistakes made in the design and conduct of the study lead to erroneous results and conclusions. The following discussion presents the elements usually found in the methodology of an article describing a clinical trial. The topics are in the order in which they frequently appear in the

methodology section. Variation from this order of presentation should not be interpreted as evidence of a poor study design.

Site and Patients

The presentation of the methodology of a clinical trial often begins with a brief description of the site(s) of the trial. The site of the trial will influence the characteristics of the patients available for enrollment. A study conducted in a large, urban teaching hospital is likely to include patients who are not the same as those who might be recruited from rural private practices. Patient selection should be described in great detail. Specific and objective admission criteria must be developed before enrollment begins. The admission criteria define not only the patients who will be participating in the study, but also the patients to whom the results may apply. The admission criteria typically include the type and stage of disease that must be present as well as non-disease factors such as age, gender, ethnicity, socioeconomic status, and previous drug exposure. Investigators are now being encouraged to include a broader range of patients in their studies including more women and minorities as well as pediatric and geriatric patients when appropriate. Clinical trials often deal with patients with few health problems other than the disease under study. In this way, they differ from the patients encountered in practice who may have several related or unrelated health problems simultaneously. "Normal" volunteers are rarely normal if, for no other reason, because they are willing to submit themselves to the protocol. Reasons for patient participation range widely from an interest in helping advance medical knowledge to the promise of financial reward. Interference with work and lifestyle are the major reasons for patient non-participation.[12]

A common problem with conducting a clinical trial is determining how many patients to enroll. Too few patients, for example, may result in failure to distinguish the true difference between experimental and control groups, thus not achieving a statistically significant difference (a false negative, or type 2 error). One review of sample size problems determined that only 36% of the studies examined had sufficient sample sizes to detect a 50% difference between the groups being studied.[13] Because therapeutic differences are rarely of this magnitude, these studies were doomed to produce negative results before they started. The number of subjects needed for a study is determined by the level of statistical significance desired, the size of difference that the study should be able to detect, and the natural occurrence/variation of the characteristic being studied in the sample population. Increasingly, studies are reporting the results of their sample size calculation (power analysis) as part of their protocol. Of course, the actual number of patients enrolled will be influenced by other factors

including the difficulty in finding appropriate patients, their willingness to comply with the protocol, the cost per patient enrolled, and the length of the study. While "more is better" is a common principle in patient selection, the final number selected is usually a compromise. Sample size calculations are beyond the scope of this lesson, but are presented in an increasing number of books, journal article,[14] and computerized statistical programs.

Controls and Patient Allocation

Control groups provide a basis for comparison within a clinical trial. Three types of controls will be presented here: parallel (sometimes called concurrent) controls, sequential controls, and external controls. PhindAll® will again take the stage as our example study drug. In a study with parallel controls, the patients are assigned to receive either PhindAll® or some other agent during the same time period. It is possible that a patient who received PhindAll®would be in the same hospital room with a patient who received the other agent. Both patients are treated the same, differing only in the agent they received. It is unlikely that PhindAll® would be compared with a placebo since a placebo would not produce an image, but comparisons of therapeutic agents and placebos are common. The use of a placebo could establish whether a drug had therapeutic effects at all and whether the adverse effects associated with treatment were related to the drug or to the act of drug administration. Placebos are especially useful in the initial evaluation of new therapeutic agents. If a drug is supposed to lower blood pressure, a placebo-controlled clinical trial might establish that it has some hypotensive effects. The ideal placebo is inert; it should not be expected to do anything other than to establish the impact of the act of drug administration on the patient. Not surprisingly, placebos regularly have effects on medical conditions, especially those conditions with strong subjective elements such as pain. Placebos can also have side effects. This pattern of response is called the placebo effect and the use of a placebo prevents us from accidentally ascribing therapeutic effects to a drug when, in fact, patients respond similarly to an inert substance. For oral dosage forms, lactose is a commonly accepted placebo substance. Even lactose intolerant patients can usually take the small amount of lactose in a tablet or capsule without effect. Some studies have used "placebos" which were not inert, but which included small amounts of atropine or other drugs to mimic the side effects of the experimental drug. These formulations are no longer considered to be true placebos. The placebo should be identical in appearance to the experimental drug. If the experimental drug is a white tablet then the placebo should also be a white tablet of the same size, weight, and marking. Placebos should also be given on the same dosing schedule as the experimental drug. Of course, there are situations in which the use of a placebo would be scientifically, ethically, or even morally unacceptable. Cancer chemotherapy trials are

almost never placebo-controlled because of the consequences of withholding potentially life-prolonging treatment.

PhindAll® should be compared with an accepted imaging drug like ^{99m}Tc-mebrofenin. In this circumstance, ^{99m}Tc-mebrofenin would be called an active control. Comparison with active controls permits the establishment of the place of the experimental agent in the therapeutic or diagnostic scheme. In our example, if it turns out that PhindAll® is superior to ^{99m}Tc-mcbrofenin for identifying patients with pathologically confirmed biliary tract disease, we may consider using PhindAll® instead of ^{99m}Tc-mebrofenin. A final decision would, of course, take into account the relative safety, ease of use, and cost of the two drugs. In reading studies that use active controls it is important to determine if the control drug is a reasonable drug for the condition being studied, and that it is being used in a manner consistent with current clinical practices. A small dose of a control drug should not be compared with a full dose of the experimental agent. The ideal active control for a diagnostic agent would be the "gold standard" procedure which is accepted as the definitive evidence for the presence of the target condition.

A critical step in the conduct of a controlled clinical trial is the allocation of patients to treatment groups. Unfortunately, many articles dispose of this step in a sentence or two ("The patients were randomly allocated to receive either drug A or drug B"). The allocation of patients should be done in a manner that gives each patient an equal, unbiased chance of being selected for either the experimental or control group. This may be accomplished in a number of ways whose description is beyond the scope of this lesson, but more familiar means of randomization would include coin tosses, drawing names or numbers from a hat, using a table of random numbers, etc. Some would argue that previously determined and equally distributed characteristics such as odd or even Social Security numbers are unacceptable as methods of assignment because the patient never had an opportunity to be in the other group. Table 5 presents some tempting, but unacceptable, methods of randomizing patients in a clinical trial.

Table 5

UNACCEPTABLE METHODS OF PATIENT RANDOMIZATION
1. Every other patients seen in the clinic (too easily manipulated)
2. Odd and even hospital admission dates (group assignment of late evening admissions can be controlled by speeding or delaying admission)
3. Clinic A vs. clinic B (unlikely that the clinic patients are comparable)
4. Assignment by physician's impression of which treatment group would be best for the patient.
5. Assignment based on severity of disease (sicker patients get more intense therapy)

There are two additional features of randomization that merit discussion here: block randomization and stratification. In block randomization, a block size is determined (i.e., 12 patients) and when sufficient patients have been enrolled to fill the block, they are then allocated to one of the treatments within the block (i.e., 6 to the experimental group and 6 to the control group). The advantage of block randomization is even if the study falls short of its enrollment goal, it will at least have an equal number of patients in each treatment group. For this reason, block randomization is increasingly encountered in reports of clinical trials. Sometimes, the patients are stratified before randomization. Stratification involves separating patients into subgroups based on age, gender, or other characteristics before randomization. Then the members of each subgroup are randomly allocated to the study groups. This is intended to prevent a maldistribution between groups of an important co-factor for the disease under study, leading it to unduly influence the results. For example, the results of our study of the diagnostic prowess of PhindAII® and ^{99m}Tc-mebrofenin might be affected if all of the alcoholic patients were assigned to the ^{99m}Tc-mebrofenin group. Separating the alcoholic and non-alcoholic patients prior to randomization would evenly distribute this potentially important patient characteristic. While the previous example used stratification to promote homogeneity within the study sample, stratification may also be used to create heterogeneity. Stratification of the patients in a study by age could be used to determine if a new therapy is more effective in some age groups than in others, perhaps identifying efficacy that may be missed when all ages are considered together. The identification of too many subgroups can produce meaningless results as each subgroup fails to have enough patients in it to allow for the detection of any differences. A pretrial power analysis can be used to determine how many patients will be needed for each subgroup in the same manner in which it can tell researchers how many patients must be enrolled in an unstratified study.

When sequential controls are used, each patient serves as his own control. In the simplest form, a group of patients are studied to establish their baseline characteristics. They then receive a form of therapy and the same assessments are repeated, determining if the therapy had an effect on the disease.

An antibiotic might be studied in this way with the eradication of the infecting organism being taken as evidence that the antibiotic worked. This type of sequential control is suited for conditions in which the therapy may irreversibly change the underlying condition, like curing an infection. Of course, this same problem could also be studied in a parallel control design with a proportion of the patients receiving a placebo or another antibiotic. Another type of sequential control is better suited to circumstances in which the therapy affects the symptoms of a disease rather than curing it. This is the cross-over control design in which each patient receives each therapy in sequence. Some patients may receive the experimental therapy for a period of time, followed by the control for a comparable time, while others may receive the control therapy first, followed by the experimental. Cross-over studies are commonly conducted in chronic, unremitting diseases for which treatment suppresses or controls the symptoms without significantly affecting the underlying condition. Examples would include treatment of diabetes mellitus or Parkinson's disease. In these examples, removal of therapy results in the prompt reemergence of the disease. Cross-over studies are attractive to researchers because fewer total patients are needed and the experimental and control groups are well matched (since they are the same patients), leading some readers to relax their literature evaluation standards when reading them. Cross-over studies rarely produce results that could not be obtained by a well-designed, but more expensive parallel-control study. External controls represent data derived from sources other than the present study. The control data might be from concurrent experience at other facilities unconnected with the present study, or it might be from data developed in the past. The latter data would constitute an historical control group. External controls are less desirable than concurrent, parallel controls. The problems of external controls are most obvious when the data are old. If a study uses an external control, the author must be able to convince the reader that the control patients are truly comparable to the current experimental patients. A major problem with using historical controls is phase migration, the result of improvements in diagnostic technology. If the outcome of a study is expressed as survival time since diagnosis, the ability of new diagnostic techniques to identify patients earlier in the course of their disease might make a new treatment with no improved benefit appear to be superior. The patient groups would not be similar; the older data would be from patients with more advanced disease at the time of their diagnosis.

Blinding

Blinding in clinical trials describes the process of actively withholding information from participants in the study in order to minimize the influence of their pre-existing biases or expectations on the results. There are three basic types of blinding to be found in studies. In the unblinded trial, no blinding exists;

everybody knows what's going on. Single-blind describes studies in which either the patient or the evaluator has been kept in the dark regarding the conduct of the study. Usually, it is the patient who is unaware of the assigned therapy in single-blind trials. In double-blind trials, neither the patient nor the evaluator knows who is in the experimental or control groups. Studies should be blinded when the measurement of response has a subjective component. In the case of our study of the ability of PhindAll® and ^{99m}Tc-mebrofenin to identify patients with cholecystitis, it would be very important to prevent those interpreting the images and surgical specimens from knowing which drug was used in which patients because such interpretations have a large subjective component. Different studies require different levels of blinding. Studies of the pharmacokinetics of a drug rarely need blinding, but the evaluation of headache relief almost always needs blinding. Legitimate exceptions to blinding would include studies involving surgical procedures or physical manipulations where blinding would be impossible. Blinding can be broken unintentionally by minor differences in the appearance of the experimental and control drugs, or the presence of side effects characteristic of one of the drugs. Not all clinical trials are double-blinded. The author must specifically declare the type of blinding used. Unfortunately, the reader is rarely told what specific steps were taken to assure the blinding.

Drug Considerations

An article describing a clinical trial should present the details of the drug regimen used for the experimental and control groups. The dosage schedule should be clearly presented, with both groups following the same schedule. The dosages of the drugs should be those currently in accepted use or those likely to be used in the case of drugs in development. In most studies, the dosages used will be within the range approved by the FDA but exceptions occur. The objective of the study may require the use of a nonstandard dosage. Authors should clearly justify the use of the dosage regimen they have selected. The route of administration should be defined. Studies which use routes of administration that will not be available in clinical practice are obviously limited in their applicability. The timing of the assessment of the patients relative to the administration of a drug merits the attention of the reader. For short-acting drugs, was the patient evaluated during peak or trough blood levels? For slow acting drugs, was enough time allowed to pass before patient assessment to allow the full drug effect to develop? For diagnostic agents were the patients assessed at the optimal time for imaging? The duration of drug effect is also a concern in cross-over studies where a drug-free period between treatments is commonly used to allow the effect of one drug to dissipate before the administration of another. The duration of this interval (a wash-out period) should be sufficient to allow nearly complete dissipation of effects. For many drugs, this interval may correspond to the time

needed for the disappearance of the drug from the serum (i.e., 4 or 5 half-lives), while others may have effects which outlast detectable serum concentrations.

Patient Assessment

One of the most critical steps in the design of a clinical research study is the selection of patient assessment methods. Just as it is important to choose the right patients for a study to represent the target population, the scans, biopsies, physiological measurements, laboratory tests, and interviews used to assess the patients must be selected to reflect the important features of the condition under study and its progression or regression. The test methods should be described in sufficient detail to theoretically permit the reader to reproduce the experiment. This description presents a challenge to the reader since the degree of detail provided makes for difficult reading. Unless the test methods themselves are the subject of the study, they should be methods which are well established. Experimental therapies should not be evaluated with experimental methods since the reader may find it difficult to determine whether the results of the study came from the therapy or some weakness in the assessment methodology. In the case of our study of PhindAII® vs. ^{99m}Tc-mebrofenin, actual inspection of the gall bladder tissue has been selected as the assessment technique. This is a good example of an assessment technique which provides direct evidence for the presence or absence of the condition under study. For most studies, the assessment techniques provide less direct evidence. Table 6 presents several short questions that the reader of a published study can use to evaluate patient assessment processes.

Table 6

QUESTIONS FOR EVALUATING PATIENT ASSESSMENT
1. Does the assessment technique selected measure changes in the variable stated in the objective? Is the study actually measuring what it said it would measure?
2. Does the test reflect all or only part of the pathologic process? In evaluating liver disease, for example, focusing only on bile flow without considering hepatocellular damage may not be appropriate.
3. Can the test detect incremental changes in the variable? How much better (or worse) must the patient be before the test detects the change?
4. Can the test results vary depending on who is performing the test? This is a major concern in studies which depend on interviews or other potentially subjective assessment methods. If 10 people evaluate the same patient will all of them agree on the results? Studies using multiple evaluators should explain how they addressed the issue of inter-rater variability.
5. Can the same circumstances presented to the same evaluator produce different results? Evaluators, like the rest of us, have good days and bad days. The more objective the assessment technique is, the less likely it is that the mood of the evaluator will affect it.
6. Is the range of the test being used suitable for the experiment? Most evaluation techniques have a range over which they are most reliable. Is your car's speedometer accurate below 10 mph?

7. Are there factors present which may interfere with the test? These considerations can include drugs or concomitant diseases.
8. Have the patients been evaluated often enough or over a long enough period of time to detect changes? This requires a balance between our desire for continuous monitoring and the cost, invasiveness or danger of the assessment.

Patient assessment in clinical trials cannot be considered without recalling other aspects of the study. For example, if patients find one of the regimens easier or more pleasant to comply with, the results of the assessment may reflect compliance more than therapeutic effect.[15] Recall bias may play havoc with assessment techniques which include interviews. Patients who have had a very positive or very negative reaction to their treatment are more likely to recall specific information than those with less dramatic responses. Sometimes, doing anything results in a positive outcome. The Hawthorne effect was first noted at a Western Electric plant in Chicago in the 1930s. The management of the plant made several changes in the working environment to improve productivity. The changes, even minor ones, all seemed to change worker behavior. Paying attention to people changes them, often in a positive way, potentially creating improvement in placebo-treated patients' conditions.[16]

Patient assessment includes not only the evaluation of the condition under study, but also other aspects of the drug use as well. Perhaps the most important is evaluation of any adverse medical events encountered during or, occasionally, even after the study. Adverse medical events are any undesirable outcomes associated with the protocol. Some, like side effects, are readily associated with the drug, while others, like increases in the prevalence of other medical conditions, may be more difficult to associate. A superior drug that produces intolerable side effects is unlikely to be adopted for use. It is important that the methodology section of a published study describes what efforts were made to gather information about adverse medical events. In general, an active process such as asking specific questions of the patients is preferred over a passive one in which the investigators wait for the patient to complain. For example, asking patients specifically about the presence or absence of PhindAII® - associated “bull's eye" dermatitis should discover more cases of the adverse effect than simply asking patients if they have any side effects.

An important component of the study protocol is the description of the statistical methods that will be used in the evaluation of the results. It is common for authors to consult with a statistician to determine the most appropriate statistical tests to apply.

Results

The results section of a published study presents the data which were gathered during the conduct of the study. The manner of presentation depends on the nature of the data. Some studies show all of the individual findings while others show only summary data. A combination of individual findings and summary data often provides the reader with the most useful data. The summary data provide an overview of the results describing the outcomes in the study groups as units while the individual data will help the reader gain an appreciation for the variability in the data. The most important results are those which directly address the objective of the study. In the case of our PhindAII® study, the results should tell us about the ability of our study drugs to identify "the presence or absence of bile in patients with suspected cholecystitis later confirmed by surgical pathology." The results section should clearly present the results of the pathological examination of the tissue. [17]

One of the first things that the reader should encounter in the results section of a published study is the description of the patients from whom the data are derived. In the methodology section, the criteria for enrollment were described and the results section should describe the characteristics of the patients actually enrolled. Sometimes the two sections are in disagreement. The presentation of patients is commonly done as a table listing patient groups (or individual patients) along one axis and their characteristics along the other. The table also reflects the success or failure of the randomization process in creating two or more equivalent groups for study. The baseline characteristics of the study groups should be the same; otherwise one group may start the study with an advantage. The authors should perform statistical analysis of the important characteristics of the patients to assure the reader that no significant differences exist between the groups.[18] The authors should also account for all of the patients enrolled in the trial. If 100 patients were enrolled, but only 50 are represented in the results, the reader needs to know what happened to the other 50 (dropouts). If they all died from complications of the treatment they received, that might influence our willingness to use that treatment despite the wonderful results seen in the survivors.

Table 7

CONSIDERATIONS IN THE EVALUATION OF RESULTS
1. Is there consistency in the numbers? All patients should be represented in the data. Check row and column totals.
2. Are the observations counted correctly and are they clearly identified as observations rather than patients? Some studies make repeated observations and report the number of positive observations rather than the number of patients in whom positive observations have been made. Do not confuse the two.
3. If graphs arc used are they clearly labeled and appropriately scaled? To emphasize a small difference between groups, some authors will begin the axis showing the data at some point other than 0. This technique makes little difference appear large.
4. Are the results of compliance assessment presented?
5. Are the adverse medical events encountered in the study presented?

The presentation of the study results may be in text, tables or graphs. Regardless of method of presentation the reader should be able to easily comprehend the results. Extensive guidelines are available to assist authors with the selection of the best way to present their data.[19] Some considerations in the review of the presentation of results are given in Table 7.

Table 8

BASIC STATISTICAL TERMS	
Term	***Definition***
Mean	The mean is the arithmetic average of all of the observations. Advantages: everybody knows what it means and most people can calculate it. Disadvantage: it is heavily influenced by outliers (findings which are especially distant from the center of the group). Like weight on a lever, the further out you are, the more influence you have.
Median	The middle observation. Half of the observations are smaller and half are larger. Advantages: resistance to the influence of outliers and, generally, it may be a better measure of central tendency than the mean. Disadvantage: it can obscure outliers when they may be very important.
Percentile	An indication of the percentage of a data distribution which is equal to or which lies below a particular finding.
Range	The difference between the largest and smallest observations. A large range indicates a wide spread of findings while a small range suggests tightly grouped data.
Standard Deviation (SD)	The most commonly used measure of dispersion of biomedical data. The standard deviation is a measure of the spread of data about their mean. The SD deals with individual data points, not the likely mean of the group. A small SD implies a tight grouping of data while a large SD implies widely scattered findings. It is commonly presented as a ± value following the mean. (Note: ± one SD includes about 67% of the findings while ± two SD includes about 95%.)
Standard Error of the	Like standard deviation except that it reflects the likely location of the group

Mean (SEM)	mean rather than an individual finding. It is derived from the SD and the number of findings in the data array. A big SEM reduces confidence in the reproducibility of the data. Some researchers prefer to report the SEM for the wrong reason (because it is a smaller number and they think it makes their data look better).

Unless the results of the study are overwhelmingly obvious, the authors will present the results of the statistical analysis of their data. Some basic statistical terms are presented in Table 8. Unless the reader uses sophisticated statistical techniques daily, this part of the article can be very challenging. While a reader may not be able to reproduce the calculations that were performed, there is no reason to be intimidated by the results of those calculations. Many commonly used statistical tests generate a number called a *p* value. For example, the results of our study of PhindAll® and ^{99m}Tc-mebrofenin might show that "...^{99m}Tc-d-obfuscate was more successful in identifying patients with bile flow stasis than ^{99m}Tc-mcbrofenin ($p<0.05$)." This means that mathematically PhindAII® was superior to the conventional agent in this one aspect of the study and the probability that the observed results were due only to chance is less that 5%. That is, there is less than a 5% chance of the result of this calculation producing a false positive error. False positives, the perception of difference when no true difference exists, are called type 1 or α error. False negative errors (type 2 or β errors) occur when the results report no difference when, in fact, a true difference exists. Type 2 errors are often the result of enrolling too few patients. By convention, authors should not make a claim of statistical significance for their results unless the calculated *p* value is .05 or less. There is still some debate about the size of type 2 error that should be allowed with some favoring 10% and others 20%. The statistics used to evaluate the results of a study assume that the methodology was designed and conducted perfectly. All too often one hears "The study was badly conducted, but the results were statistically significant." Good statistics should never be allowed to save a bad study. The results may have been statistically significant because the study was poorly conducted, not in spite of it. The most important thing for the reader to realize is that the statistical analysis of the data is a mathematical process. It does not help us to determine the size of any differences that may exist and, most importantly, it does not represent the clinical importance of the findings. Patient care decisions are made using clinical, rather than statistical judgment. A useful tool for evaluation of the results of a study is to ask the simple question "So what?" If the results are clearly of clinical importance, then the answer to the question will come quickly. It is important for the reader to establish his own impression of the clinical importance of the findings before moving to the discussion and conclusion section of the article.

Discussion and Conclusion

In many articles, the discussion and conclusion section is the longest section. This is the part of the article where the authors present their version of the answer to the "So what?" question we posed at the end of the results section. The conclusions reached by the authors and the recommendations they make should be limited to ones which are supported by the data presented in the study. The creation of post-study subgroups of patients and presenting the findings from those subgroups is a dangerous and misleading practice of some authors. It is as if to say, "Well, the regular protocol failed to find what we wanted, but if we rearrange the data in just the right way we can achieve statistical significance for one or more subgroups."[20]

In this section, the authors will also compare their results to those encountered in related studies and offer an explanation if their results are different. The more familiar the reader is with the literature in the area, the easier it is to navigate through this section. The authors should present an unbiased interpretation of the related literature, but do not always do so. The biased selection and presentation of related literature occurs too frequently for comfort and contributes to the perpetuation of poor quality information. A related issue is the incorrect description of data taken from the articles listed in the study's reference list (quotation errors). Another common problem is the author who cites data from one article which merely repeated data from a third article (source errors). One study found 35% quotation errors and 41% source errors when reviewing the citations of articles from three established jounals.[21] If a piece of data from the discussion section is critical to the application of the results of the study, the information should be checked carefully against its original source.

Some authors treat their study like it was one of their children, incapable of doing wrong. Good authors will clearly present the limitations of their studies. This not only helps the reader by pointing out things that the reader may not have considered, it also improves the author's image as a thoughtful scientist. Rather than being viewed as a weakness of a study, a realistic presentation of limitations strengthens the report of the study and enhances its usefulness for the reader. Another characteristic of overly proud authors is the use of biased language. Phrases such as "remarkable improvement" or "profound increase in diagnostic power" or even "amazing recovery" serve to warn the reader that he is dealing with such an author. Similarly, attempting to salvage disappointing results with phrases such as "Although not statistically significant, the results show a clear trend toward the superiority of PhindAll®" should set off alarms. Some authors will even use an article describing a study as a platform for offering recommendations on issues not associated with the study.

CONCLUSION

Drug literature evaluation is not particularly complicated or mysterious. As this lesson has shown, it is the application of common sense coupled with a healthy degree of skepticism. Awareness of the possible pitfalls associated with different study types and a general understanding of the subject being discussed in a published study will serve the reader well. Recall that the perfect study has yet to be conducted and that small problems are inevitable. The reader is never obliged to accept the authors' interpretation of the data as presented in a published study without reservation. Studies must be interpreted in the individual context in which the data are to be used. The exercise of professional judgment becomes critically important if the data available to us are to be used for the improvement of the care of our patients.

REFERENCES

1. Maddox J. Why the pressure to publish [editorial]? *Nature* 1988;333:493.
2. Gehlbach SH. Interpreting the medical literature. 3rd edition. New York: McGraw-Hill; 1993.

3. Riegelman RK, Hirsch RP. Studying a study and testing a test: how to read the health science literature. 3rd edition. Boston : Little, Brown: 1996.

4. Hulley SB, Cummings SR, editors. Designing clinical research. Baltimore: Williams & Wilkins; 1988.

5. Jones JK. Determining causation from case reports. In: Strom BL, editor. Pharmacoepidemiology. 2nd ed. New York; John Wiley & Sons: 1994. p. 365-78.

6. Spilker B. Guide to clinical interpretation of data. New York: Raven; 1986.

7. Banting FG, Best CH. Collip JB. Campbell WR, Fletcher AA. Pancreatic extracts in the treatment of diabetes mellitus. *Can Med Assoc J* 1922;12:141-6.

8. Dawson-Sanders B, Trapp RG. Basic and clinical biostatistics. 2nd ed. Appleton & Lange: Norwalk (CT); 1994.

9. Herbst AL, Ufelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971;284:878-81.

10. Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med* 1991;115:901-5.

11. Hartzema AG. Guide to interpreting and evaluating the pharmacoepidemiologic literature. Ann Pharmacother 1992; *26:96-8.*

12. Cunny KA, Miller HW. Participation in clinical drug studies: motivations and barriers. *Clin Therap* 1994;16:273-82.

13. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-4.

14. Florey CD. Sample size for beginners. *BMJ* 1993;306:1181-4.

15. Freedman LS. The effect of partial noncompliance on the power of a study. *Control Clin Trials* 1990;11:157-68.

16. Roethlisberger FJ, Dickson WJ, Wright HA. Management and the worker: an account of a research program conducted by Western Electric Company, Hawthorne Works, Chicago. Cambridge (MA); Harvard University Press: 1939.

17. Jaeschke R, Guyatt G. Sackett DL. User's guide to the medical literature III. How to use an article about a diagnostic test A. Are the results of the study valid? *JAMA* 1994;217:389-91.

18. Elenbaas JK, Cuddy PG, Elenbaas RM. Evaluating the medical literature, part III: results and discussion. *Ann Emerg Med* 1983;12:679-86.

19. Spilker B, Schoenfelder J. Presentation of clinical data. New York; Raven Press: 1990.

20. Abrahamson NS, Kelsey SF, Safar P, Sutton-Tyrell K. Simpson's paradox and clinical trials: what you find is not necessarily what you prove. *Ann Emerg Med* 1992;21.1480-2.

21. Goldberg R, Newton E, Cameron J, Jacobson R, Chan L, Bukata WR et al. Reference accuracy in the emergency medicine literature. *Ann Emerg Med* 1993;22:1450-4.

ASSESSMENT QUESTIONS

1. A recently published study compared the effects of two drugs on patients' blood pressure. The results section of the study states that drug A was found to decrease the patients' diastolic blood pressure by 10mm Hg while drug B decreased it by 3 mm Hg. These findings are reported as being statistically significant (p<.05>. The p value reported means that:

    a. The likelihood of the difference between the groups being due to chance is less than 5%.
    b. The probability of repeating the findings in a similar group of patients is less than 5%
    c. The results represent a true difference between the groups of less than 5%
    d. There is less than a 5% chance that this is a false negative result

2. In a single-blind study, which people are typically not aware of who is in the experimental group and who is in the control group? That is, which people are usually "blinded"?

    a. The people analyzing the data
    b. The researchers
    c. The researchers and the subjects
    d. The subjects

3. Which one of the following would most appropriately be investigated using a crossover study design in which subjects are exposed to two or more Tc-99m radiopharmaceuticals?

    a. Relatively stable conditions such as arterial stenosis
    b. Unstable diseases such as vasospasm
    c. Rapidly terminal diseases such as massive myocardial infarction
    d. None of the above

4. A researcher conducted a study comparing a new headache medication to acetaminophen. Four patients were randomly assigned to receive doses of either the new medication or acetaminophen for their next five headaches. At the end of the study period, the researcher's results show no statistically significant difference between the pain relief provided by the two drugs. This study is at risk for:

    a. A type 1 error in the results
    b. A type 2 error in the results
    c. Both a & b
    d. None of the above

5. A research study of a new drug has collected a diverse group of patients. Their ages are given below. Select the response which is **closest** to the median age of the group.

    79, 37, 59, 50, 20, 41, 15, 96, 32

    a. 35
    b. 45
    c. 55
    d. 65

6. Using the same patient ages as in the questions above, select the response which is **closest** to the mean age of the group.

   a. 35
   b. 45
   c. 55
   d. 65

7. When an article is described as having a high rate of quotation errors, this means that:

   a. Authors who take information from this article often do so incorrectly
   b. Many of the entries in its reference list are incorrect or incomplete
   c. Much of the data taken from the articles in its reference list are incorrectly presented in the article
   d. There are discrepancies between statements in different parts of the article

8. Active treatment controls can be reasonable used in studies:

   a. In which the efficacy of the control drug has not been established.
   b. Which compare one therapy with another
   c. Which have failed to detect a difference with a placebo
   d. Without independent samples

9. Some studies use historical controls. All of the following are characteristics of an appropriate historical control **except:**

   a. No new, more sensitive diagnostic methods have entered into general use since the historical data were collected
   b. The author's recollection of past experiences forms a solid base for comparison
   c. The data on the historical control patients are reliable and complete
   d. The morbidity and mortality of the disease are predictable

10. All of the following are components of a well-constructed study objective **except**:

    a. Description of how measurements will be made
    b. A description of the study results
    c. A description of the study subjects
    d. A description of what will be measured

11. An n-of-1 study differs from a case report in that the N-of 1 study:

    a. Describes intentional exposure to drug
    b. Is published in a different journal
    c. Presents only one patient
    d. Provides poorer evidence for a drug effect

12. When compared to an experimental treatment, a suitable placebo has all of the following characteristics **except**:

    a. An identical appearance
    b. An identical dosing schedule
    c. An identical monitoring plan
    d. An identical side effect profile

13. In an attempt to study the long-term side effects of ^{131}I compared with another anti-thyroid drug, investigators enrolled two groups of patients, one which had just received ^{131}I and another which had just been treated with the other anti-thyroid drug. These two groups will be followed for 5 years. This is an example of which one of the following study types?

    a. Case-control
    b. Clinical trial
    c. Cohort
    d. Cross-sectional

14. What if the investigators located patients who had received ^{131}I or the other anti-thyroid drug 5 years earlier and asked them a series of questions about their health?

    a. Case-control
    b. Clinical trial
    c. Cohort
    d. Cross-sectional

15. What if the investigators assigned patients to receive either ^{131}I or the other anti-thyroid drug and then followed them for 5 years?

    a. Case-control
    b. Clinical trial
    c. Cohort
    d. Cross-sectional

16. What if the investigators identified a group of patients with one of the suspected ^{131}I-induced side effects and a similar group without the side effect and then checked their medical records for the possible exposure to anti-thyroid drugs?

    a. Case-control
    b. Clinical trial
    c. Cohort
    d. Cross-sectional

17. Patients selected for participation in clinical trials may differ from those who are likely to receive the drug in actual practice in several ways. All of the following are possible differences **except*:***

    a. Study patients may be more healthy
    b. Study patients may be younger
    c. Study patients may have fewer total medical conditions
    d. Study patients may not have the target disease

18. An active treatment control compares the experimental therapy to:

    a. A conventional treatment
    b. An inert substance
    c. Another experiment therapy
    d. The patient's natural response

19. A researcher has assembled 100 patients at two clinics for the purpose of evaluation a new radiopharmaceutical drug in comparison with an established drug. All of the following are acceptable methods of assigning patients to the treatment groups **except**:

    a. Clinic physicians picking who they want to receive each drug
    b. Drawing patient names out of a hat
    c. Randomizing patients within each clinic
    d. Using a table of random numbers to identify treatment assignments

20. A new drug is being evaluated for its ability to reverse body mass loss associated with AIDS. Which of the following would be the **best** to measure the effect of the drug?

    a. Asking patients if they feel bulkier
    b. Bathroom scale
    c. Photographs of the patients
    d. Physicians' clinical impressions

21. A well designed nationwide study reports that drug A was better than drug B for curing the common cold (21% vs. 19%) and that the advantage was statistically significant. Your response to this data should be to:

    a. Abandon use of drug B
    b. Examine other aspects of the study results
    c. Recalculate the study's results
    d. Switch patients from drug B to drug A

22. A letter to the editor in nuclear medicine journal describes a patient who experienced hair loss while being treated with a radiopharmaceutical drug. Your questions might include all of the following **except**:

    a. Were there any other potential causes of hair loss?
    b. Has this been reported before?
    c. Was the patient losing hair before the treatment?
    d. Why was the patient concerned about hair loss?

23. "At present, 13% of patients treated with the new drug experiences hair loss" expresses:

    a. incidence
    b. likelihood
    c. prevalence
    d. risk

24. All of the following represent external control groups **except**:

    a. Data from half the current study patients
    b. Data from medical records of similar patients from 1993-98
    c. Data from patients studies at another clinic
    d. Data from the medical literature

25. In a double-blind study, all of the following should be prevented from knowing the treatment assignments **except:**

    a. The evaluators
    b. The patients
    c. The readers
    d. Answers (b) and (c) are both correct