OBJECTIVE:
To review advantages and disadvantages of commonly used study design strategies and examine issues in interpretation of study results including bias and confounding.

DATA SOURCES:
Clinical literature, established epidemiologic texts.

STUDY SELECTION:
Not applicable.

DATA EXTRACTION:
Not applicable.

DATA SYNTHESIS:
The increasing flow of research reports on various therapeutic agents and procedures places managed care pharmacists in the position of evaluating the clinical relevance of new information and its applicability to clinical decision-making in their organizations. A solid understanding of study design strategies such as cohort, case-control, and clinical trials, and the flaws inherent in each, is invaluable for accurate interpretation of any study results. Some design strategies are more appropriate than others to answer certain clinical questions. The randomized, controlled, blinded clinical trial is the most reliable design for assessing therapeutic effects of pharmacologic agents.

CONCLUSION:
Armed with a solid understanding of study design strategies and issues in interpretation, managed care pharmacists can proceed into the wealth of health care information, confident in their assessments of the validity and utility of research findings.

KEY WORDS:
Study design, Study methodology.

Table 1. Evidence for Cause and Effect in Observations

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporality</td>
<td>Cause precedes effect</td>
</tr>
<tr>
<td>Strength</td>
<td>Large relative risk</td>
</tr>
<tr>
<td>Dose-response</td>
<td>Larger exposures to cause associated with higher rates of disease</td>
</tr>
<tr>
<td>Reversibility</td>
<td>Reduction in exposure associated with lower rates of disease</td>
</tr>
<tr>
<td>Consistency</td>
<td>Repeatedly observed by different persons, in different places, circumstances, and times</td>
</tr>
<tr>
<td>Biologic plausibility</td>
<td>Makes sense, according to biologic knowledge of the time</td>
</tr>
<tr>
<td>Specificity</td>
<td>One cause leads to one effect</td>
</tr>
<tr>
<td>Analogy</td>
<td>Cause-and-effect relationship already established for a similar exposure or disease</td>
</tr>
</tbody>
</table>

\* Adapted from reference 2.

People today live in the age of information. With the arrival of each medical and pharmacy journal come myriad reports on diverse therapies for equally diverse medical conditions. As anyone in the health care field will attest, distilling meaningful findings from the reams of data that confront practitioners daily is our most challenging task. Nowhere does the successful completion of this task have more far-reaching implications than in managed care pharmacy. We must determine the relative importance of the “latest findings” to clinical decision-making in our respective organizations.

The first step in deciding whether an association made in a scientific study is because of cause-and-effect is to apply the criteria established in 1965 by Bradford-Hill (Table 1). Evaluation of all components of the study, including design, methods, data analysis, and interpretation of results, is integral to proper assessment of the findings. Clearly, no study is perfect; upon examination, any study will have some flaws. The pertinent question is whether the flaws are fatal to the findings. This article presents a summary of

Authors

DARLENE M. MEDNICK, R.Ph., is Vice President, Health and Utilization Management—Clinical Program Operations, Merck-Medco Managed Care, Inc., Montvale, New Jersey; and DAVID DAY, M.S., R.Ph., is Director Team Leader of Clinical Pharmacy, Pfizer Pharmaceuticals, New York, NY.

Author Correspondence: Darlene M. Mednick, Merck-Medco Managed Care, Inc., 100 Summit Avenue, Montvale, NJ 07645.

Copyright © 1997, Academy of Managed Care Pharmacy, Inc. All rights reserved.
design strategies and their advantages, disadvantages, and appropriateness for certain research subjects, as well as a review of issues in interpretation, including the roles of bias and confounding.

For the purposes of this discussion, unless otherwise noted, the term disease refers to a deviation from normal human functioning (e.g., congestive heart failure) or other outcomes of interest (e.g., myocardial infarction). The term exposure refers to a risk factor for developing a particular disease (e.g., high-fat diet) or to a measure taken to improve or prevent a disease or event (medication, exercise).

ASSESSING DESIGN STRATEGIES

In evaluating a study, the key question to ask is whether the design employed in the study fits the subject. As shown in Table 2, some design strategies are more appropriate for certain subjects than others. Assuming all ethical and practical considerations are addressed, choice of study design depends on the characteristics of the disease and the exposure under investigation, available knowledge about both, and limitations such as time and resources.

Understanding basic study design types and the type of research best suited to each design type is imperative. Epidemiologic research studies fall into two categories: descriptive or analytic. Descriptive studies examine the distribution of a disease by searching for systematic variations in incidence and prevalence according to population subgroup, geographic location, or time period. Descriptive studies are used primarily to formulate a hypothesis about determinants of a disease.

Table 3. Overview of Epidemiologic Design Strategies

<table>
<thead>
<tr>
<th>Study Design Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive studies</td>
<td>Populations (correlational studies), Individuals, Case reports, Case series, Cross-sectional (prevalence) surveys</td>
</tr>
<tr>
<td>Analytic studies</td>
<td>Observational studies, Case-control studies, Cohort studies: retrospective and prospective, Intervention studies (clinical trials)</td>
</tr>
</tbody>
</table>

Correlational Studies

Correlational studies are used to compare disease frequencies (1) between different groups during the same time or (2) at different time periods in the same populations. Correlational studies may also be referred to as "ecologic" studies in the literature. For example, a study of this type might look at the correlation between red wine consumption and rates of heart disease in all men living in America versus all men living in France, or the correlation between red meat consumption and colon cancer rates in American versus Japanese populations. A correlational study is concerned with comparisons between populations as a whole. For this reason, a correlational study cannot be used to assess cause and effect. However, the results of a correlational study can form the basis for a hypothesis of a causal relationship.

Case Reports and Series

Compared with correlational studies, which are concerned with populations as a whole, case reports and case series are concerned with individuals. Case reports often appear as brief reports or letters to the editor in medical journals; they are detailed accounts of one particular patient. A case series is an expanded version of the case report. Clearly, case reports and series are not useful for determining a causal relationship because they are concerned with a single patient or, at most, several patients. The temporal relationship between an exposure and a disease in a few individuals may result from chance, or it may suggest a possible association between the exposure and the disease. The ultimate relevance of a case

Table 2. Matching the Strongest Research Designs to Clinical Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Design</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Prevalence</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Prevalence</td>
</tr>
<tr>
<td>Incidence</td>
<td>Cohort</td>
</tr>
<tr>
<td>Risk</td>
<td>Cohort, Case-control</td>
</tr>
<tr>
<td>Prognosis</td>
<td>Cohort</td>
</tr>
<tr>
<td>Treatment</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Prevention</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Cause</td>
<td>Cohort</td>
</tr>
</tbody>
</table>

* Adapted from reference 2.
report or case series is unknown, the experience reported may turn out to be limited, or it may be a red flag signaling the need for further investigation.

Cross-Sectional Studies
A prevalence, or cross-sectional, study attempts to determine the frequency of disease in a defined population at a certain point in time. A population under study could be the total U.S. population or a smaller subset of that population, such as nurses or teachers. In this type study, disease and exposure are assessed simultaneously. Using a randomly selected population sample, the investigator observes both the frequency of a particular disease and the possible determinants of the disease (e.g., age, gender, race, socioeconomic status, and lifestyle factors). Because a temporal relationship between possible determinants and disease status usually cannot be established in a cross-sectional study, cause and effect cannot be adequately assessed. For example, in a study assessing the prevalence of dementia in older British adults, 10% of those 75 years of age and older were found to be affected. Although this study could ascertain prevalence, it could not determine the cause of dementia. However, for any inborn condition (e.g., HLA type or blood type), cause and effect may be determined because the condition can be presumed to predate the outcome.

Although descriptive studies cannot be used to investigate a causal relationship between any given exposure and a disease, such studies should not be dismissed or considered less valuable than analytic studies. Prevalence studies, case reports, and case series raise questions about the possibility of an association between a disease and an exposure. They provide clues that may be used to generate a hypothesis of a causal relationship. Hypotheses are tested using analytic studies, including case-control studies, cohort studies, and clinical trials.

Case-Control Studies
In a case-control study, the disease has either developed or not developed in the subjects, and the investigator looks backwards in time for a possible cause. Case-control studies are frequently referred to as retrospective studies, but this is not quite correct, because cohort studies may also be retrospective (see section on cohort studies below).

In case-control studies, one group of subjects with a disease (cases) is compared with another group of subjects without the disease (controls) with respect to presence or absence of a particular exposure. Data on exposure status are usually collected through interviews, reviews of records, or biological assays.

▲ Advantages Case-control studies are both time and cost efficient. Case-control studies are useful for studying diseases with a long latency period because, at the time of the study, the disease and exposure have already occurred. The investigators do not have to wait for the disease to develop. A case-control design is also ideal for studying rare diseases because adequate numbers of cases can be obtained. In addition, case-control studies can evaluate a wide range of potential exposures and their interrelation.

▲ Disadvantages Case-control studies do have limitations because they are retrospective, therefore it is sometimes difficult to determine if an exposure preceded a disease. They are particularly susceptible to bias and confounding factors (discussed further below) because the groups being compared are usually constructed by the researcher and do not occur naturally. In addition, case-control studies usually cannot be used to determine incidence rates.

Cohort Studies
Cohort studies can be either retrospective (historical) or prospective (concurrent). Subjects are categorized by whether or not they have been exposed to a potential risk factor for a disease. In either retrospective or prospective designs, the cohort is assembled before the disease occurs and subjects are followed over a period of time to assess the development of the disease.

Both exposed and unexposed groups should be as similar as possible in all aspects except the exposure. Similarity between the groups helps control the influence of outside factors other than the exposure. Data for cohort studies are usually collected by periodic physical examination of study subjects, or review of records, disease registration forms, and death certificates.

▲ Advantages When the temporal sequence between an exposure and development of a disease is to be considered, a cohort study is useful because the investigators observe the development of the disease over time. Prospective cohort studies also provide the benefit of minimizing selection bias because the outcomes of interest have not yet occurred. In contrast to case-control studies, which are well-suited for studying rare diseases, cohort studies are useful for determining the effects of rare exposures because subjects can be selected on the basis of their exposure status. Additionally, a cohort study can examine multiple effects of a single exposure, whereas case-control studies can examine multiple exposures associated with a single effect. Finally, findings from cohort studies are more reliable and provide a stronger basis for cause and effect than case-control studies because cohort studies allow the investigator to observe the development of the disease over time after the exposure.

▲ Disadvantages Prospective cohort studies are time-consuming and expensive. Retrospective cohort studies are limited by the availability of adequate records. As mentioned above, cohort studies are usually not suited for evaluating rare diseases because it is difficult to assemble an adequate sample size of exposed and unexposed subjects. For example, study-
ing the development of lung cancer in nonsmokers exposed to passive smoke would be difficult with a cohort study. The exposure is so common, yet the disease is quite rare among nonsmokers. In addition, lung cancer takes several decades to develop, so an extremely large group of exposed individuals would have to be followed for a long period of time without the assurance that any would develop the disease. Finally, cohort studies are susceptible to different sources of bias including misclassification bias, and loss to follow-up, both of which can seriously undermine the validity of the results.

Clinical Trials

Clinical trials, also known as experimental studies or intervention studies, are actually a variation of a prospective cohort study. Clinical trials are unique because they are prospective, the investigator assigns the exposure status, and the subjects are followed to determine the effect of a treatment on their disease status. When subjects are assigned exposure status at random and investigators and subjects are blinded to the assignment, the study is known as a randomized, double-blind clinical trial; such trials are the most reliable source of evidence.

Clinical trials generally fall into one of the following three categories:

- **Therapeutic trials**: Studies in which a treatment, such as a medication or procedure, is administered to study subjects for the purpose of relieving symptoms, curing a disease, or prolonging survival.

- **Intervention trials**: Studies in which the investigator intervenes before a disease has developed in subjects at risk for developing the disease. Study subjects are in a preclinical phase of the disease at study time, and the intervention is meant to prevent or delay clinical disease or slow its progression.

- **Preventive or prophylactic trials**: Studies in which a preventive agent or procedure is administered to study subjects without a disease for the purpose of evaluating the efficacy of that agent or procedure.

Examples of each type of clinical trial are shown in Table 4.

The randomized, controlled, double-blind clinical trial is the gold standard for evaluating therapeutic effects of any given treatment. Although some degree of control for known factors is feasible in all study types, randomization in a clinical trial provides otherwise unobtainable control over factors unknown and unaccounted for in designing the study. Blinding protects the study against bias, because the actions or interpretations of investigators and participants are not influenced by knowledge of exposure status. The randomized controlled, double-blind clinical trial is the study design that most closely approximates the setting of a basic research science experiment in terms of control. Therefore, a randomized clinical trial is the ideal design for evaluating a causal relationship, especially that between a treatment and a disease.

<table>
<thead>
<tr>
<th>TRIAL TYPE</th>
<th>EXAMPLE OF TRIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapeutic</td>
<td>Zidovudine treatment for AIDS</td>
</tr>
<tr>
<td></td>
<td>Simple mastectomy for breast cancer</td>
</tr>
<tr>
<td>Intervention</td>
<td>Zidovudine treatment of HIV-positive subject without symptoms of AIDS</td>
</tr>
<tr>
<td></td>
<td>Mammography to detect asymptomatic breast cancer</td>
</tr>
<tr>
<td></td>
<td>Cholesterol-lowering drugs to decrease the risk of myocardial infarction</td>
</tr>
<tr>
<td>Preventive</td>
<td>Hepatitis B vaccination for prevention of hepatitis and hepatocellular carcinoma</td>
</tr>
<tr>
<td></td>
<td>Education in use of condoms to reduce risk of HIV transmission and infection</td>
</tr>
</tbody>
</table>

*Adapted from reference 4.*

Meta-analysis

A meta-analysis is not technically a study design; it is a statistical technique in which data from many trials are pooled and reanalyzed as if they had come from a single study. Meta-analysis provides a minimally biased method for estimating a treatment effect and its magnitude. Because the results of meta-analyses are often represented as findings from clinical trials, it is important to provide some discussion here of their relative merits and disadvantages. Meta-analysis can also be applied to observational studies.

**Advantages** A meta-analysis is technically most appropriate when, for whatever reason, adequate sample sizes are difficult to obtain, and studies of a particular exposure and disease show evidence of an association but fail to achieve statistical significance. In this case, pooling the data from a number of such studies and reanalyzing it may demonstrate the existence of an overall effect. Meta-analysis can also help clarify issues when results from a number of studies are inconsistent. A meta-analysis may prove useful when definitive randomized trials are not possible or the results are inconclusive. The major strength of a meta-analysis is the ability to discern small but potentially important clinical effects and add statistical significance to associations that show a trend.

Meta-analytic evidence has proved beneficial in cardiology. Use of beta blockers and thrombolytics in the treatment of acute myocardial infarction was first justified by results from meta-analyses of small randomized clinical trials and then later confirmed by larger randomized clinical trials.

In epidemiology, meta-analysis showed passive smoking to be a lung-cancer risk for nonsmokers. Many small case-control studies of passive smoking and lung cancer had sug-
suggested an association between exposure to someone else's cigarette smoke and the development of lung cancer, however, most of these studies failed to achieve statistical significance because sample sizes were too small. As was widely reported in the lay press, a meta-analysis using data pooled from a number of the studies showed a statistically significant association between passive smoking and development of lung cancer.  

**Disadvantages** Clearly, the results of an analysis of pooled data must be interpreted with caution, because of the wide variation in study subjects and conditions. The weaknesses of a meta-analysis include (but are not limited to) differences in the study designs among single studies from which the data were pooled, bias introduced in deciding which trials to include in the meta-analysis, and incorrect assumptions made in combining studies.

**ISSUES IN INTERPRETATION**

Interpretation of study results is as much an art as it is a science. Sometimes reported findings are no doubt valid, and managed care pharmacists need only ascertain the applicability of the findings to the needs of an organization. Frequently, however, research findings fall into a gray area, and interpretation of those findings will depend on basic scientific reasoning, logic, and, ultimately, one's own clinical judgment.

The clinical question a study proposes should first be determined, and then one can assess whether the design chosen is most suited to provide the answer. If the design fits the study subject, the next question is to judge the validity of the findings—i.e., are there alternate explanations for the study outcome? Chance, bias, and confounding all contribute to invalid results.

**Chance**

A review of probability functions is beyond the scope of this article. In general, it is most important to remember basic concepts, such as the importance of sample size (the larger the sample, the more likely it is that the experience of an unbiased sample is representative of the experience of the population from which it was drawn). Inferences based on smaller sample sizes are subject to greater variability; therefore, they are less reliable than inferences based on larger sample sizes. Always interpret studies using small sample sizes with caution. Although they may offer some clues worth pursuing, the findings from small studies generally should not be taken as conclusive. However, small sample sizes are sometimes highly significant because the measure of association was large and few other variables were present (e.g., diethylstilbestrol and clear-cell carcinoma).

Statistical analyses are designed to give some measure of the role played by chance in the findings of any given study. Statistical analyses make use of a probability statement, commonly referred to as the p value. The p value is an expression of the probability of obtaining by chance alone a result that is at least as extreme as that observed in a study, assuming that there is no actual relationship between the exposure and the disease. The p value that indicates that an association is statistically significant is, by convention, usually set at 0.05, although other levels may occasionally be chosen.

A p value of 0.05 means that there is, at most, a 5% probability of observing by chance alone an association as large or larger than that observed in the study. In other words, a smaller p value indicates that a relationship between an exposure and a disease is more likely responsible for the association found in a study and that chance is a less probable explanation.

In the medical literature, a wide variety of tests for statistical significance are employed. Other than biostatisticians, most people find it difficult to remember which tests are more suited to which types of data. It is a good idea to consult a recent textbook on biostatistics.

**Bias**

Bias is a systematic variation in the way in which two or more study groups are treated or evaluated, as opposed to chance, which is random variation. If bias is found to be present in a study, it will not be correctable since it will be part of the study design. The effect of bias on a study's results can be stated in terms of simple logic: if the premises are false, the conclusion will be invalid.

**Selection Bias** Selection bias occurs when there is any error in choosing a study population so that subjects in the comparison groups (e.g., cases versus controls) differ in factors, other than those under investigation, that affect the study outcome.

Case-control and retrospective cohort studies are susceptible to selection bias, because in these study designs both the disease and exposure have already occurred. In case-control studies, the risk is that cases and controls will be selected based on different criteria that are related to exposure status. A recent case-control study, for example, raises the possibility that calcium channel blockers given to hypertensive patients increase their risk for myocardial infarction. However, there is the likelihood that patients receiving the calcium channel blockers may have been systematically different from the patients given the other drugs. Likewise, in retrospective cohort studies, disease status may affect selection of subjects with and without a history of exposure. In either case, presence of selection bias in a study sheds considerable doubt on the findings.

**Misclassification Bias** When an error is made in categorizing study subjects by disease status or exposure history, misclassification bias has occurred. However, the effect of misclassification bias depends on whether the error in categorization occurred at random—i.e., if proportions of misclassified
Table 5. Guides for Evaluation of Clinical Literature

**QUESTION: WERE THE RESULTS OF THE STUDY VALID?**

**Primary concerns**
- Was the assignment of patients to treatments randomized?
- Were all patients who entered the trial properly accounted for and attributed at its conclusion?
- Was follow-up complete?
- Were patients analyzed in the groups to which they were randomized?

**Secondary concerns**
- Were patients, health workers, and study personnel “blind” to treatment?
- Were the groups similar at the start of the trial?
- Aside from the experimental intervention, were the groups treated equally?

**QUESTION: WHAT WERE THE RESULTS OF THE STUDY?**
- How large was the treatment effect?
- How precise was the estimate of the treatment effect?

**QUESTION: WILL THE RESULTS HELP IN CARING FOR PATIENTS?**
- Can the results be applied to patient care?
- Were all clinically important outcomes considered?
- Are the likely treatment benefits worth potential harms and costs?

Adapted from reference 16

subjects in each group are equal. Because nondifferential misclassification reduces differences between exposed and nonexposed groups, the reported association between the exposure and disease will be weaker than the true association. Differential misclassification—when proportions of misclassified subjects in each group are unequal—may result in the reported association being either weaker or stronger than the true association. Examples of this type of bias are discussed below.

▲ Observation Bias Observation, or information, bias occurs when there are systematic differences in the manner in which exposure or disease-outcome data are gathered. Such differences are only important to consider if they exert unequal influence on comparative groups in a study. Observation bias can take several forms: recall bias, interviewer bias, and loss to follow-up bias.

In recall bias, study subjects differ in the completeness and accuracy of their memory of being exposed to a particular risk factor. Diseased subjects may be more likely to remember an exposure than nondiseased subjects because the disease gives them greater motivation to search their memories for a possible exposure. Interviewer bias occurs when the interviewer questions study groups in systematically different ways. Loss to follow-up bias is introduced when study subjects lost to follow-up differ significantly in terms of exposure history and development of the disease.

▲ Confounding Confounding is a major concern in any epidemiologic study. A confounder is both a factor associated with the exposure and an independent risk factor for the disease. A confounder may be the actual cause of the disease, while the exposure is not a true risk factor. The confounder cannot be a step in the causal chain between the disease and the exposure under investigation.

Confounding may exist in any study involving human subjects. For example, basic characteristics such as age, sex, and smoking status should always be considered potential confounders because they are associated with many diseases and exposures. Confounders can either inflate or suppress the real measure of association between an exposure and disease. In a study to determine whether regular exercise lowers the risk of heart disease, the group that exercised had lower rates of disease. However, the association between exercise and heart disease may have resulted from less smoking (a risk factor for coronary disease) in the exercise group. Evaluating study findings for the role of confounding is often a difficult and complicated process. The effect of confounding is assessed during data analysis.

**CONCLUSION**

Knowledge of both study design and sources of error is essential to determining the validity of research findings. Table 5 presents a quick reference guide to assessing validity of findings from studies reporting on therapy or prevention. The questions posed in Table 5 correlate closely with the material reviewed in this article.

Clearly, when evaluating a treatment or procedure for potential benefit, the most reliable source of information comes from randomized, controlled, blinded clinical trials. This type of study provides the most control over all sources of error and the most accurate assessment of cause and effect.

Managed care pharmacists are on the front lines of the conflict over various treatment options. We must sift through the reams of data confronting us every day and glean the information most vital to the health of our customers and the optimal functioning of our organizations. Armed with a solid understanding of study design strategies and issues in interpretation, we can proceed into the wealth of health care information facing us, confident in our assessments of the validity and utility of research findings.
References


16. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. 11. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993; 270: 2598-601.