13 Research Using Existing Data: Secondary Data Analysis, Ancillary Studies, and Systematic Reviews

Norman Hearst, Deborah Grady, Hal V. Barron, and Karla Kerlikowske

Many research questions can be answered quickly and efficiently using data that have already been collected. There are three general approaches to using existing data. **Secondary data analysis** is the use of existing data to investigate research questions other than the main ones for which the data were originally gathered. **Ancillary studies** add measurements of a small number of variables to a study, often in a subset of the participants, to answer a separate research question. **Systematic reviews** combine the results of multiple previous studies that have addressed a given research question to calculate a summary estimate of effect. Making creative use of existing data is an especially effective way for new investigators with limited time and resources to begin to answer important research questions.

■ ADVANTAGES AND DISADVANTAGES

The main advantages of using existing data are speed and economy. A research question that might otherwise require much time and money to investigate can sometimes be answered rapidly and inexpensively. For example, in the Multiple Risk Factor Intervention Trial (MRFIT), a large heart disease prevention trial, information about the smoking habits of the wives of the study subjects was recorded to examine whether this influenced the men's ability to quit smoking. After the study was over, one of the investigators realized that the data provided an opportunity to investigate the health effects of passive smoking (a more important research question!). A twofold excess in the incidence of heart disease was found in nonsmoking men married to smoking wives when compared with similar nonsmoking men married to nonsmoking wives (1).

Existing data sets also have limitations. The selection of which data to collect, the quality of data gathered, and how data were recorded are all predetermined. The investigator may have to settle for a variable that is not what he would prefer to have measured (e.g., history of hypertension, a dichotomous historical variable, in place of actual blood pressure). The quality of the data may be poor, with frequent missing or incorrect values. Important confounders and outcomes may not have been measured or recorded. All these factors contribute to the main disadvantage of using existing data: the investigator has little or no control over the data.

■ SECONDARY DATA ANALYSIS

Individual Data Sets

Secondary data sets are of two types: individual and aggregate. Individual data, in which separate information is available for each subject, may come from previous research studies, medical records, health care billing files, death certificates, and many other sources. In such a data set, associations between characteristics can be measured among individual members of the study sample, much as an investigator would do if gathering his own data.

One major category of individual data that may be useful to the clinical researcher is data collected in a **previous research study,** often at the investigator's institution. Many studies collect more data than the investigators can analyze and contain interesting findings that have gone unnoticed. Access to such data is controlled by the study's principal investigator; the new researcher should therefore seek out information about the work of senior investigators at his institution. One of the most important ways a good mentor can be helpful to a new investigator is by providing access to interesting data.

The second category includes large regional and national data sets that are publicly available and do not have a principal investigator. Computerized databases of this sort are a rapidly growing phenomenon and as varied as the reasons people have for collecting information. We will give two examples that deserve special mention, and readers can locate others in their own settings and areas of interest.

Tumor registries are government-supported agencies that collect complete statistics on cancer incidence, treatment, and outcome in defined geographic areas. These registries currently include about 15% of the U.S. population, and the area of coverage is expected to increase during the coming years. One of the purposes of these registries is to provide data to outside investigators. Combined data for all the registries are available from the Surveillance, Epidemiology, and End Results (SEER) Program (see the SEER Web site).

Death certificate registries can be used to follow the mortality of any cohort. The **National Death Index** includes all deaths in the United States since 1978. This can be used to ascertain the vital status of subjects of an earlier study or of those who are part of another data set that includes important predictor variables. An example is the follow-up of men with coronary disease who were treated with high-dose nicotinic acid (or placebo) to lower serum cholesterol in the Coronary Drug Project: Although there was no difference in death rates at the end of the 5 years of randomized treatment, a mortality follow-up 9 years later using the National Death Index revealed a significant difference (2). Whether an individual is alive or dead is public information, so follow-up was complete even for men who

had dropped out of the study. This was the first demonstration that cholesterol intervention can reduce total mortality.

The National Death Index can be used when any two of three basic individual identifiers (name, birth date, and Social Security number) are known. Ascertainment of the fact of death is 99% complete with this system, and additional information from the death certificates (notably cause of death) can then be obtained from state records. On the state and local level, many jurisdictions now have computerized vital statistics systems, in which individual data (such as information from birth or death certificates) are entered as they are received.

Aggregate Data Sets

When individual data are not available, aggregate data sets can sometimes be useful. The term **aggregate data** means that information is available only for groups of subjects (e.g., death rates from cervical cancer in each of the 50 states). With such data, associations can only be measured among these groups by comparing group information on a risk factor (such as tobacco sales) with the rate of an outcome. Studies using aggregate data are called **ecologic studies.**

The advantage of aggregate data is its availability. Its major drawback is the fact that associations are especially susceptible to confounding: Groups tend to differ from each other in many ways, not all of which are causally related. Furthermore, associations observed in the aggregate do not necessarily hold for the individual. For example, sales of cigarettes may be greater in states with high suicide rates, but the individuals who commit suicide may not be the ones doing most of the smoking. This situation is referred to as the **ecologic fallacy.** Aggregate data are most appropriately used to test the plausibility of a new hypothesis or to generate new hypotheses. Interesting results can then be pursued in another study that uses individual data.

Getting Started

Secondary data analysis can begin in two ways. An investigator may start with a research question of interest to him and try to find a data set that can answer the question. This is the usual approach to clinical research (Chapter 2). The other approach, unique to secondary data analysis, is to begin with a data set and consider questions in the investigator's area of interest and expertise that it can answer. The challenge here is to discover meaningful findings among piles of information.

With either approach, the help of a senior colleague experienced in clinical research is invaluable. An experienced researcher has defined areas of interest in which he stays current and is aware of the important questions that need research. In secondary data analysis, this person can help not only in choosing a research question and designing a protocol, but also in identifying and gaining access to the appropriate database.

Finding Research Questions to Fit the Data

The process of finding research questions to fit the data, which can be particularly useful to a new investigator who has not yet settled on an area of interest, is summarized in Table 13.1. The investigator first identifies an available data set and familiarizes himself with the information that has been gathered. It may be useful to make a written list or flowchart of all the data collected, including the timing of variables that were measured more than once. The next step is to look

■ TABLE 13.1
Steps in Finding Research Questions to Fit an Existing Database

1. Choose a database.
2. Become thoroughly familiar with the database. Make a flow sheet of all variables and how they were measured.
3. Identify pairs or groups of variables whose association may be of interest.
4. Review the literature and consult experts to determine if these research questions would be novel and important.
5. Formulate specific hypotheses and settle on the statistical methods.
6. Analyze the data.

for pairs or groups of variables whose relation might be of interest. A brainstorming session involving others familiar with the data may help.

Especially in large cohort studies and clinical trials, relationships among many variables may never be assessed as part of the original study, simply because the investigators do not find time to analyze all combinations. For example, in the CARDIA cohort study of the antecedents of coronary risk factors, an investigator noticed that blood pressure was measured both sitting and standing. This led to the questions, "Was there any relation between a postural change in blood pressure at baseline and the subsequent incidence of hypertension?" and "Was there any difference between African Americans and whites in this regard?" The answer to both questions was yes. An orthostatic increase in systolic blood pressure predicted the 8-year incidence of diastolic hyptertension, especially among African Americans (3).

Large national data sets provide a similar opportunity. The National Health and Nutrition Examination Survey (NHANES) includes medical histories, nutritional questionnaires, and physical examinations performed on a probability sample of the U.S. population (4). The National Center for Health Statistics, which collects the NHANES data, encourages outside investigators to perform analyses, such as examining potential risk factors for diseases. Any scientist can arrange to purchase, at nominal cost, a computer file that includes any specified set of variables.

Finding Data Sets to Fit a Research Question

Many investigators prefer the approach of seeking a data set that can answer questions that emerge in a topic area of interest to them (Table 13.2). This approach

■ TABLE 13.2
Steps in Finding Databases to Fit a Specified Research Question

1. Choose a research question and review the literature thoroughly.
2. List combinations of predictor and outcome variables whose relationship might help answer the research question.
3. Identify databases that might include the variables of interest.
4. Become familiar with each of these databases and consult with individuals who know them well.
5. Choose the best database(s) and gain access to the data.
6. Formulate specific hypotheses and settle on the statistical methods.
7. Analyze the data.

is less constrained, leaving the investigator free to choose among an array of research questions and data sets. After choosing a research topic and becoming familiar with the literature in that area (including a thorough literature search and advice from a senior mentor), the next step is to investigate whether candidate research questions can be addressed with an existing database. It is important to do this carefully, spending days or weeks in an effort to find the best prospect, rather than rushing into something that will take months or years before ultimately being unsuccessful.

The best solution may be close at hand, a database (often attached to a candidate mentor) at the home institution. For example, a University of California, San Francisco (UCSF) fellow noticed in a review article about Lp(a), a little-known coronary disease risk factor, that one of the few interventions known to lower the level of this lipoprotein was estrogen. Knowing that HERS, a major clinical trial of hormone treatment to prevent coronary disease, was being managed at UCSF, the fellow approached the investigators with this interest. Since no one else had specifically planned to examine the relationship between hormone treatment, Lp(a), and coronary heart disease events, the fellow mastered the literature and designed an analysis plan (Appendix 13.1). After receiving permission from the HERS study leadership, he worked with coordinating center statisticians, epidemiologists, and programmers to carry out an analysis that he then wrote up and published (5).

Sometimes it is necessary to venture further afield. Working from a list of predictor and outcome variables whose relation might help to answer the research question, the next step is to locate databases that include these variables. Two great allies in this effort are the telephone and the Internet. Phone calls or e-mail messages to the authors of previous studies or to government officials might result in access to files containing useful data. It is essential to conquer any anxiety that the investigator may feel about contacting strangers to ask for help. Most people are surprisingly cooperative, either by providing data themselves or by suggesting other places to try.

It is sometimes possible to link two databases, one supplying information on the predictor variable and one on the outcome variable. This is relatively straightforward for ecologic studies, as long as both data sets use the same boundaries for the groups being compared (e.g., zip codes, states). Sources of individual data are more difficult to link unless both data sets include the same unique individual identifiers, such as Social Security number.

Once the data for answering the research question have been located, the next challenge is to obtain permission to use them. It is a good practice to use official letterhead on correspondence and to adopt any institutional titles that are appropriate (e.g., instead of "I'm an epidemiology student at . . .," say, "I'm calling from the Department of Epidemiology at . . ."). If someone suggested the contact, it is usually helpful to mention that person's name.

The investigator should be very specific about what information is sought and should confirm the request in writing. It is a good idea to keep the size of the request to a minimum and to offer to pay any cost of preparing the data. If the data set is controlled by another group of researchers, the investigator can suggest a collaborative relationship. In addition to providing an incentive to share the data, this can engage a coinvestigator who is familiar with the database. It is wise to clearly define such a relationship early on, including who will be first author of the planned publications. Serious arrangements of this sort will benefit from a face-to-face meeting.

Here is an example of the use of secondary data in a **natural experiment** that included randomization. In a natural experiment, people have received an intervention in the past for reasons unrelated to research, and a database is available with which to assess an outcome that might have been affected by this assignment. We studied the effect of the 1970 to 1972 draft lottery (involving 5.2 million 20-year-old men assigned randomly by date of birth) on delayed mortality (assessed by state death certificate registries) (6). The predictor variable in this study (date of birth) was a proxy for military service during the Vietnam era. Men who had been randomly assigned to be eligible for the draft had significantly greater mortality from suicide and motor vehicle accidents in the 10 years after they would have returned from the military. The study was done for less than $2,000 (not including the investigators' time), yet it was a more unbiased approach to examining the delayed effect of military service on specific causes of death than other studies of this topic with much larger budgets.

Use of Large Community-Based Data Sets

Secondary data can be especially useful for studies to evaluate patterns of utilization and clinical outcomes of medical treatment. This approach can complement the information available from randomized trials and examine questions that trials cannot answer. The types of existing data include (a) **administrative and clinical databases** such as those developed by Medicare, the Veterans Administration, Kaiser Permanente Medical Group, and the Duke Cardiovascular Disease Databank, and (b) **registries** such as the San Francisco mammography registry and the National Registry of Myocardial Infarction (NRMI). Information from any of these sources (many of which can be found on the Web) can be very useful for studying rare outcomes and for assessing real-world utilization and effectiveness of an intervention that has been shown to work in a clinical trial setting (Example 13.1).

Example 13.1. The National Registry of Myocardial Infarction (NRMI)

Large data sets are especially useful for studying rare events. For example, the NRMI was used to examine the risk factors for intracranial hemorrhage after recombinant tissue-type plasminogen activator (tPA) was given for acute myocardial infarction in patients receiving usual care (7). The registry included 71,073 patients who received tPA; among these, 673 had intracranial hemorrhage confirmed by computed tomography or magnetic resonance imaging. A multivariate analysis showed that a tPA dose exceeding 1.5 mg/kg was significantly associated with developing an intracranial hemorrhage. Given that the overall risk of developing an intracranial hemorrhage was less than 1%, a clinical trial collecting primary data to examine this outcome would have been prohibitively large and expensive.

Another valuable contribution from this type of secondary data analysis is a better understanding of the difference between efficacy and effectiveness. The randomized clinical trial is the gold standard for determining the **efficacy** of a therapy under highly controlled circumstances in selected clinical settings. In the "real world," however, patients and treatments are often different. The choice of drugs and dosage by the treating physician and the adherence to medications by the patient are much more variable. These factors often act to make the new

therapy less effective than demonstrated in trials. Assessing the **effectiveness** of treatments in actual practice can sometimes be accomplished through studies using secondary data. For example, primary angioplasty has been demonstrated to be superior to thrombolytic therapy in clinical trials of treating patients with acute myocardial infarction (8). But this may only be true when success rates for angioplasty are as good as those achieved in the clinical trial setting. Secondary analysis of a community data set has shown no benefit of primary angioplasty over thrombolytic therapy (9,10).

Secondary data analysis is often the best approach for studying the utilization of accepted therapies. Although clinical trials can demonstrate efficacy of a new therapy, this benefit can only occur if the therapy is adopted by practicing physicians. Understanding **utilization rates,** addressing regional variation and use in specific populations (such as the elderly, ethnic minorities, the economically disadvantaged, and women), can have major public health implications. For example, despite convincing data that angiotensin-converting enzyme inhibitors decrease mortality in patients with MI, a secondary analysis of community data has shown that many patients with clear indications for such therapy do not receive it (11,12).

As with other types of research, secondary data analyses are more convincing when an association found in one data set can be duplicated in another collected using a different approach or a different patient population. This reduces the chance that the association observed in the first database was a result of chance or bias. For example, Vaccarino et al. performed a retrospective cohort study using existing data from 15 Connecticut hospitals and found a higher mortality after MI in women compared with men in younger age groups but not in those who were older (13). Subsequently, a similar analysis in a larger and more diverse database found a similar interaction (14).

■ ANCILLARY STUDIES

Research with secondary data requires that the data needed to answer a research question are already available. In an **ancillary study,** the investigator adds a small number of measurements to an existing study to answer a different research question. For example, in the HERS trial of the effect of hormone therapy on risk for coronary events in 2,763 elderly women, an investigator added measurement of the frequency and severity of urinary incontinence. Adding a one-page questionnaire created a large trial of the effect of hormone therapy on urinary incontinence, with essentially no additional time or expense (15).

Ancillary studies have many of the **advantages** of secondary data analysis with fewer constraints. They are both inexpensive and efficient, and the investigator can design a few key ancillary measurements specifically to answer the research question. Ancillary studies can be added to any type of study, including cross-sectional and case-control studies, but large prospective cohort studies and randomized trials are particularly well suited to such studies.

Ancillary studies in randomized trials have the problem that the measurements may be most informative when added before the trial begins, and it may be difficult for an outsider to identify trials in the planning phase. Even when a variable was not measured at baseline, however, a single measurement during or at the end of the trial can produce useful information. By adding cognitive function measures at the end of the HERS trial, the investigators were able to compare the cognitive function of elderly women treated with hormone therapy

for 4 years with the cognitive function of those treated with placebo (16). In the absence of a baseline measurement, *change* in cognitive function between the two groups (usually a more powerful outcome) could not be compared; however, the comparison at the end of the study preserved the value of randomization and produced an important null finding (16).

A good opportunity for ancillary studies is provided by the **banks of stored serum, DNA, images,** and so on, that are found in most large clinical trials and cohort studies. The opportunity to propose new measurements using these specimens can be an extremely cost-effective approach to answering a novel research question, especially if it is possible to make these measurements on a subset of specimens using a nested case-control design. In HERS, for example, it was possible to examine whether the excess number of thromboembolic events in the hormone-treated group was due to an interaction with Leiden factor V deficiency by genetic analyses of fewer than 100 cases and controls.

In most large prospective cohort studies, the investigators periodically add new measurements, providing an excellent opportunity for ancillary studies. For example, an ongoing prospective cohort study of risk factors for osteoporotic fractures among 9,700 elderly women (Study of Osteoporotic Fractures) included multiple measures of health and functional ability. By adding a measure of nursing home placement, an investigator was able to evaluate risk factors for institutionalization among elderly women.

Getting Started

Opportunities for ancillary studies should be actively pursued, especially by new investigators with limited time and resources. A good place to start is to identify studies with research questions that include either the predictor or the outcome variable of interest. For example, investigators interested in the effect of weight loss on pain associated with osteoarthritis might start by identifying trials of interventions (such as diet, exercise, behavior change, or drugs) for weight loss. Such studies can be identified by searching lists of studies funded by the federal government, by contacting pharmaceutical companies that manufacture drugs for weight loss, and by talking with experts in weight loss who are familiar with ongoing studies. To create an ancillary study, the investigator would simply add a measure of arthritis symptoms among subjects enrolled in these studies. Alternatively, he might identify studies that have joint pain as an outcome, and add change in weight as an ancillary measure.

After identifying a study that provides a good opportunity for ancillary measures, the next step is to obtain the cooperation of the study investigators. Most researchers are happy to add ancillary measures to an established study if they address an important question and do not substantially interfere with the conduct of the main study. For example, most researchers would be enthusiastic about adding a brief questionnaire or a measurement that participants find unintrusive and interesting. Investigators will be more reluctant to add measures that require a lot of the participant's time (cognitive function testing) or are dangerous or unpleasant (colonoscopy).

Generally, formal permission from the principal investigator or the appropriate study committee is required to add an ancillary study. Most large, multicenter studies have established procedures requiring a written application. The proposed ancillary study is generally reviewed by a committee that can approve, reject, or revise the ancillary study. Many ancillary measures require funding, and the ancillary study investigator must find a way to pay these costs. Of course, the

cost of an ancillary study is much less than the cost of conducting the same study independently. Some large studies may have their own mechanisms for funding ancillary studies, especially if the research question is important and considered relevant by the funding agency.

The **disadvantages** of ancillary studies are few. In some cases there may be practical problems in obtaining official permission to perform the study, training those who will make the measurements, and obtaining separate informed consent from participants. Because the ancillary study investigator may not have designed or conducted the main study, it may also be difficult to obtain access to the full database for analysis. These issues, including a clear understanding of authorship of scientific papers that result from the ancillary study and the rules governing their preparation and submission, need to be clarified before starting the study.

■ SYSTEMATIC REVIEWS

Systematic reviews identify completed studies that address a research question, and evaluate the results of these studies to arrive at conclusions about a body of research. In contrast to other approaches to reviewing the literature, systematic reviews use a well-defined and uniform approach to identify all relevant studies, display the results of eligible studies, and, when appropriate, calculate a summary estimate of the overall results. The statistical aspects of a systematic review (calculating summary effect estimates and variance, statistical tests of heterogeneity, and statistical estimates of publication bias) are called **meta-analysis.**

A systematic review can be a good opportunity for a new investigator. Although it takes a surprising amount of time and effort, a systematic review generally does not require substantial financial or other resources. Completing a good systematic review requires that the investigator become intimately familiar with the literature regarding the research question. For new investigators, this detailed knowledge of published studies is invaluable. Publication of a good systematic review can also establish a new investigator as an "expert" on the research question. Moreover, the findings, with power enhanced by the larger sample size available in the combined studies and peculiarities of individual study findings revealed by comparison with the others, often represent an important scientific contribution. Systematic review findings can be particularly useful for developing practice guidelines.

The elements of a good systematic review are listed in Table 13.3. Just as for

■ TABLE 13.3

Elements of a Good Systematic Review

1. Clear research question
2. Comprehensive and unbiased identification of completed studies
3. Definition of inclusion and exclusion criteria
4. Uniform and unbiased abstraction of the characteristics and findings of each study
5. Clear and uniform presentation of data from individual studies
6. Calculation of a summary estimate of effect and confidence interval based on the findings of all eligible studies when appropriate
7. Assessment of the heterogeneity of the findings of the individual studies
8. Assessment of potential publication bias
9. Subgroup and sensitivity analyses

other studies, the methods for completing each of these steps should be described in a written protocol before the systematic review begins.

The Research Question

As with any research, a good systematic review has a well-formulated, clear research question that meets the usual FINER criteria (feasible, interesting, novel, ethical, and relevant; see Chapter 2). Feasibility depends largely on the prior existence of a set of studies of the question. The research question should describe the disease or condition of interest, the population and setting, the intervention and comparison treatment (for trials), and the outcomes of interest. For example, "*Among persons admitted to an intensive care unit unstable angina, does treatment with aspirin plus intravenous heparin reduce the risk of myocardial infarction and death during the hospitalization more than treatment with aspirin alone* (17)?"

Identifying Completed Studies

Systematic reviews are based on a comprehensive and unbiased search for completed studies. The search should follow a well-defined strategy established before the results of the individual studies are known. The process of identifying studies for potential inclusion in the review and the sources for finding such articles should be explicitly documented before the study. Searches should not be limited to MEDLINE, which includes only about half of all published English-language clinical research studies and often does not list non-English-language references. Depending on the research question, other electronic databases such as AIDSLINE, CANCERLIT, and EMBASE can be included, as well as manual review of the bibliography of relevant published studies, previous reviews, evaluation of the Cochran Collaboration database, and consultation with experts.

Criteria for Including and Excluding Studies

The protocol for a systematic review should provide a good rationale for including and excluding studies, and these criteria should be established a priori. Criteria for including or excluding studies from meta-analyses typically designate the period during which studies were published, the population that is acceptable for study, the disease or condition of interest, the intervention to be studied, whether blinding is required (for trials), acceptable control groups, required outcomes, maximal acceptable loss to follow-up, and minimal acceptable length of follow-up. Once these criteria are established, each potentially eligible study should be reviewed for eligibility independently by two or more investigators, with disagreements resolved by another reviewer or by consensus. When determining eligibility, it may be best to blind reviewers to the date, journal, authors, and results of trials.

Published systematic reviews should list studies that were considered for inclusion and the specific reason for excluding a study. For example, if 30 potentially eligible trials are identified, these 30 trials should be fully referenced and reasons should be given for each exclusion.

Collecting Data from Eligible Studies

Data should be abstracted from each study in a uniform and unbiased fashion. Generally, this is done independently by two or more abstractors using predesigned forms that include variables that define eligibility criteria, design features,

the population included in the study, the number of individuals in each group, the intervention (for trials), the main outcome, secondary outcomes, and outcomes in subgroups. The data abstraction forms should include any data that will subsequently appear in tables describing the studies included in the systematic review or in tables or figures presenting the outcomes. When the two abstractors disagree, a third abstractor may settle the difference, or a consensus process may be used. The process for abstracting data from studies for the systematic review should be clearly described in the manuscript.

The published reports of some studies that might be eligible for inclusion in a systematic review may not include important information, such as design features, risk estimates, and standard deviations. Often it is difficult to tell if design features such as blinding were not implemented or were just not described in the publication. The reviewer can sometimes calculate relative risks and confidence intervals from crude data presented from randomized trials, but it is generally unacceptable to calculate risk estimates and confidence intervals based on crude data from observational studies unless there is sufficient information to adjust for potential confounders. Every effort should be made to contact the authors to retrieve important information that is not included in the published description of a study. If this information cannot be calculated or obtained, the study findings are generally excluded.

Presenting the Findings Clearly

Systematic reviews generally include three types of information. First, important characteristics of each study included in the summary review are presented in tables. These often include the study sample size, number of outcomes, length of follow-up, characteristics of the population studied, and methods used in the study. Second, the review displays the results of the individual studies (risk estimates and confidence intervals) in a figure. Finally, in the absence of significant heterogeneity (see below), the review presents summary estimates and confidence intervals based on the findings of all the included studies as well as sensitivity and subgroup analyses.

The summary effect estimates represent the main outcomes of the systematic review but should be presented in the context of all the information abstracted from the individual studies. The characteristics and findings of individual studies included in the systematic review should be displayed clearly in tables and figures so that the reader can form opinions that do not depend solely on the statistical summary estimates.

Meta-Analysis: Statistics for Systematic Reviews

Summary Effect Estimate and Confidence Interval. Once all completed studies have been identified, those that meet the inclusion and exclusion criteria have been chosen, and data have been abstracted from each study, a summary estimate (summary relative risk, summary odds ratio, etc.) and confidence interval may be calculated. The summary effect is essentially an average effect weighted by the size of each study. Methods for calculating the summary effect and confidence interval are discussed in Appendix 13.1.

Heterogeneity. Combining the results of several studies is not appropriate if the studies differ in clinically important ways, such as the intervention, outcome, controls, blinding, and so on. It is also inappropriate to combine the findings if the results of the individual studies differ widely. Even if the methods used in

the studies appear to be similar, the fact that the results vary markedly suggests that something important was different in the individual studies. This variability in the findings of the individual studies is called **heterogeneity** (and the study findings are said to be **heterogeneous**); if there is little variability, the study findings are said to be **homogeneous.**

How can the investigator decide if the studies used the same methods and had similar findings? First, he can review the individual studies to determine if there are substantial differences in study design, study populations, intervention, or outcome. Then he can examine the results of the individual studies. If some trials report a substantial beneficial effect of an intervention and others report considerable harm, heterogeneity is clearly present.

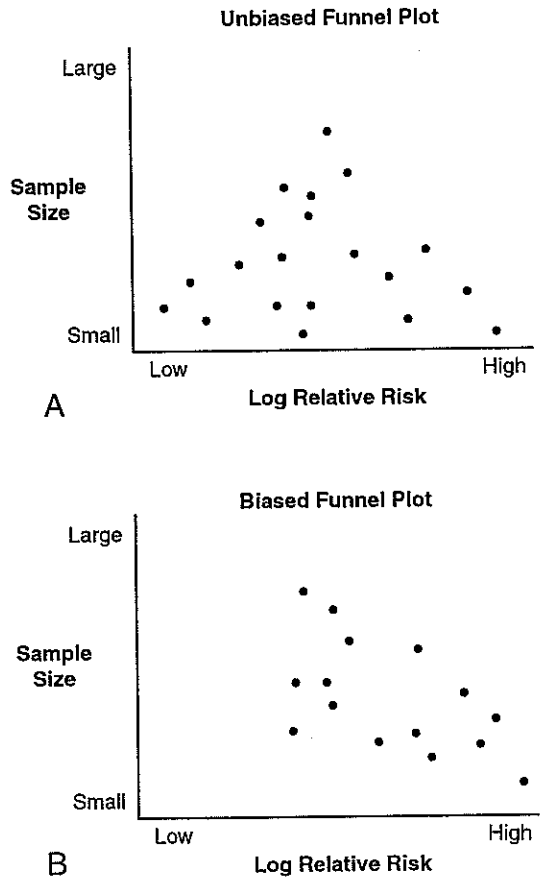
Sometimes, however, it is more difficult to decide if heterogeneity is present. For example, if one trial reports a 50% risk reduction for a specific intervention but another reports only a 30% risk reduction, is heterogeneity present? Statistical approaches (tests of homogeneity) have been developed to help answer this question (Appendix 13.1).

Assessment of Publication Bias

Publication bias occurs when published studies are not representative of all studies that have been done, usually because positive results tend to be submitted and published more often than negative results. There are two main ways to deal with publication bias. **Unpublished studies can be identified** and the results included in the summary estimate. Unpublished results may be identified by querying investigators and reviewing abstracts, meeting presentations, and doctoral theses. The results of unpublished studies can be included with those of the published trials in the overall summary estimate, or sensitivity analyses can determine if adding these unpublished results substantially changes the summary estimate determined from published results. However, including unpublished results in a systematic review is problematic for several reasons. It is often difficult to identify unpublished studies and even more difficult to abstract the required data. Frequently, inadequate information is available to determine if the study meets inclusion criteria for the systematic review or to evaluate the quality of the methods. For these reasons, unpublished data are not often included in meta-analyses.

Alternatively, the extent of potential **publication bias can be estimated** and this information used to temper the conclusions of the systematic review. Publication bias exists when unpublished studies have different findings from published studies. Unpublished studies are more likely to be small (large studies usually get published, regardless of the findings) and to have found no association between the risk factor or intervention and the outcome (markedly positive studies usually get published, even if small). If there is no publication bias, there should be no association between a study's size and findings. A strong correlation between study outcome and sample size suggests publication bias. In the absence of publication bias, a plot of study sample size (or study weight) versus outcome (e.g., log relative risk) should have a bell or funnel shape with the apex near the summary effect estimate.

The funnel plot in Fig. 13.1A suggests that there is little publication bias because small studies with both negative and positive findings were published. The plot in Fig. 13.1B, on the other hand, suggests publication bias because the distribution appears truncated in the corner that should contain small, negative studies.



■ **FIGURE 13.1**

A: Funnel plot suggestive of minimal publication bias since there are studies with a range of large and small sample sizes, and low relative risks are reported by some smaller studies. **B:** Funnel plot suggestive of publication bias since the smaller studies primarily report high relative risks.

When substantial publication bias is likely, summary estimates should not be calculated or should be interpreted cautiously. Every reported systematic review should include some discussion of potential publication bias and its effect on the summary estimates.

Subgroup and Sensitivity Analyses

Subgroup analyses may be possible using data from all or some subset of the studies included in the systematic review. For example, in a systematic review of the effect of postmenopausal estrogen therapy on endometrial cancer risk, some of the studies presented the results by duration of estrogen use. Subgroup analyses of the results of studies that provided such information demonstrated that longer duration of use was associated with higher risk for cancer (18).

Sensitivity analyses indicate how "sensitive" the findings of the meta-analysis are to certain decisions about the design of the systematic review or inclusion of certain studies. For example, if the authors decided to include studies with a slightly different design or methods in the systematic review, the findings are

strengthened if the summary results are similar whether or not the questionable studies are included. Systematic reviews should generally include sensitivity analyses if any of the design decisions appear questionable or arbitrary.

Garbage In, Garbage Out

The biggest drawback to a systematic review is that it can produce a very reliable-appearing summary estimate based on the results of individual studies that are of poor quality. The process of assessing quality is complex and problematic. We favor relying on relatively strict criteria for good study design when setting the inclusion criteria. If the individual studies that are summarized in a systematic review are of poor quality, no amount of careful analysis can prevent the summary estimate from being unreliable.

■ SUMMARY

Secondary Data Analysis

1. Secondary data analysis has the **advantage** of greatly reducing the time and cost of doing research and the **disadvantage** of providing the investigator with little or no control over the data.
2. One good source of data for secondary analysis is an **existing project** at the investigator's institution; another is the large number of **databases** now available from many sources. Although **individual data are preferable,** aggregate data can also sometimes be useful for ecologic analysis.
3. Investigators may begin either by looking for research questions to fit an existing database or by looking for a database that can answer a particular research question.
4. Large community-based data sets are useful for studying the **effectiveness** and **utilization** of an intervention in the community, and for discovering **rare adverse events.**

Ancillary Studies

1. A clever ancillary study can answer a new research question with **little cost and effort.** As with secondary data analyses, the investigator **cannot control the design,** including the population and many of the variables measured, but he is able to **specify a few key additional measurements.**
2. Good opportunities for ancillary studies may be found in **cohort studies** or **clinical trials** that include either the predictor or outcome variable for the research question of interest. **Stored banks of serum, DNA, images,** and so on, provide the opportunity for cost-effective nested case-control designs.
3. Most large studies have written procedures that allow investigators (including outside scientists) to propose and carry out ancillary studies.

Systematic Reviews

1. A good systematic review, like any other study, requires a complete **written protocol** before the study begins. The protocol should include the **research question,** methods for **identifying all eligible studies,** methods for **abstracting data** from the studies, and **statistical methods.**
2. The statistical aspects of a systematic review, termed **meta-analysis,** include the **summary effect estimate and confidence interval,** tests for evaluating **heterogeneity** and **potential publication bias,** and planned **subgroup** and **sensitivity analyses.**

3. The **characteristics** and **findings** of individual studies should be displayed clearly in tables and figures so that the reader can form opinions that do not depend solely on the statistical summary estimates.
4. The biggest drawback to a systematic review is that the results can be no more reliable than the **quality** of the underlying studies upon which it is based.

EXERCISES

1. The research question is, "Do Latinos in the United States have higher rates of gallbladder disease than whites, African Americans, or Asian Americans?" What existing databases might enable you to determine race-, age- and sex-specific rates of gallbladder disease at low cost in time and money?

2. A research fellow became interested in the controversial question of whether high serum triglyceride levels increase risk for coronary heart disease. Because of the expense and difficulty of conducting a study to generate primary data, he wanted an existing database that contained the variables he needed to answer his research question but was generated for other reasons (so that it was unlikely that someone else had already studied the research question with that data set). He found an investigator at his institution who had access to the database of a large multicenter randomized controlled trial of the effects of interventions on several risk factors for coronary heart disease (the Multiple Risk Factor Intervention Trial).

Fortunately, the MRFIT investigators agreed to allow him access to the data. The necessary predictors and outcome measures were present in the database. The fellow and his advisor then requested the analyses that they wanted from the MRFIT Coordinating Center.

a. What are the advantages of this approach to study this question?

b. What are the disadvantages of this approach?

References

1. Svendsen KH, Kuller LH, Martin MJ, et al. Effects of passive smoking in the Multiple Risk Factor Intervention Trial (MRFIT). *Am J Epidemiol* 1987;126:783–95.
2. Canner PL. Mortality in CDP patients during a nine-year post-treatment period. *J Am Coll Cardiol* 1986;8:1245–55.
3. Thomas R, Hulley S. Personal communication.
4. See National Health and Nutrition Examination Survey (NHANES) website.
5. Shlipak M, Simon J, Vittinghoff E, et al. Estrogen and progestin, lipoprotein (a), and the risk of recurrent coronary heart disease events after menopause. *JAMA* 2000;283:1845.
6. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality: a randomized natural experiment. *N Engl J Med* 1986;314:620–4.
7. Gurwitz JH, Gore JM, Goldberg RJ, et al. Risk for intracranial hemorrhage after tissue plasminogen activator treatment for acute myocardial infarction. Participants in the National Registry of Myocardial Infarction 2. *Ann Intern Med* 1998;129:597–604.
8. Weaver WD, Simes RJ, Betriu A, et al. Comparison of primary coronary angioplasty and intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review. *JAMA* 1997;278:2093–8 [published erratum appears in *JAMA* 1998;279:1876].
9. Every NR, Parsons LS, Hlatky M, et al. A comparison of thrombolytic therapy with

primary coronary angioplasty for acute myocardial infarction. Myocardial Infarction Triage and Intervention Investigators. *N Engl J Med* 1996;335:1253–60.
10. Tiefenbrunn AJ, Chandra NC, French WJ, et al. Clinical experience with primary percutaneous transluminal coronary angioplasty compared with alteplase (recombinant tissue-type plasminogen activator) in patients with acute myocardial infarction: a report from the Second National Registry of Myocardial Infarction (NRMI- 2). *J Am Coll Cardiol* 1998;31:1240–5.
11. Barron HV, Bowlby LJ, Breen T, et al. Use of reperfusion therapy for acute myocardial infarction in the United States: data from the National Registry of Myocardial Infarction 2. *Circulation* 1998;97:1150–6.
12. Barron HV, Michaels AD, Maynard C, et al. Use of angiotensin-converting enzyme inhibitors at discharge in patients with acute myocardial infarction in the United States: data from the National Registry of Myocardial Infarction 2. *J Am Coll Cardiol* 1998;32:360–7.
13. Vaccarino V, Horwitz RI, Meehan TP, et al. Sex differences in mortality after myocardial infarction: evidence for a sex-age interaction. *Arch Intern Med* 1998;158:2054–62.
14. Vaccarino V, Parsons L, Every NR, et al. Sex-based differences in early mortality after myocardial infarction. National Registry of Myocardial Infarction 2 Participants. *N Engl J Med* 1999;341:217–25.
15. Grady D, Brown J, Vittinghoff E, et al. Postmenopausal hormone therapy for urinary incontinence, in press.
16. Grady D, Yaffe K, Kristof M, et al. Effect of postmenopausal hormone therapy on cognitive function: the Heart and Estrogen/Progestin Replacement Study, in press.
17. Oler A, Whooley M, Oler J, et al. Heparin plus aspirin reduces the risk of myocardial infarction or death in patients with unstable angina. *JAMA* 1996;276:811–15.
18. Grady D, Gebretsadik T, Kerlikowske K, et al. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstet Gynecol* 1995;85:304–13.

■ APPENDIX 13.1. STATISTICAL METHODS FOR META-ANALYSIS

Summary Effects and Confidence Intervals

The primary goal of meta-analysis is to calculate a summary effect size and confidence interval. An intuitive way to do this is to multiply each trial relative risk (an effect estimate) by the sample size (a weight that reflects the accuracy of the relative risk), add these products, and divide by the sum of the weights. In actual practice, the inverse of the variance of the effect estimate from each individual study ($1/\text{variance}_i$) is used as the weight for each study. The inverse of the variance is a better estimate of the accuracy of the effect estimate than the sample size because it takes into account the number of outcomes and their distribution. The weighted mean effect estimate is calculated by multiplying each study weight ($1/\text{variance}_i$) by the log of the relative risk (or any other risk estimate, such as the odds ratio, risk difference, etc.), adding these products, and dividing by the sum of the weights. Small studies generally result in a large variance (and a wide confidence interval around the risk estimate) and large studies result in a small variance (and a narrow confidence interval around the risk estimate). Thus, in a meta-analysis, large studies get a lot of weight (1/small variance) and small studies get little weight (1/big variance).

To determine if the summary effect estimate is statistically significant, the variability of the estimate of the summary effect is calculated. There are various formulas for calculating the variance of summary risk estimates (1,2). Most use something that approximates the inverse of the sum of the weights of the individual studies (1/sum weights). The variance of the summary estimate is used to calculate the 95% confidence interval around the summary estimate ($\pm 1.96 \times \text{variance}^{1/2}$).

Random- versus Fixed-Effect Models

There are multiple statistical approaches available for calculating a summary estimate (2). The choice of statistical approach is usually dependent on the type of outcome (relative risk, risk reduction, difference score, etc.) and does not usually make much difference. However, one statistical issue that can affect the outcome of a meta-analysis of clinical trials is whether a fixed-effect or random-effect model is chosen. The fixed-effect model simply calculates the variance of a summary estimate based on the inverse of the sum of the weights of each individual study. The random-effect model adds variance to the summary effect in proportion to the variability of the results of the individual studies. Summary effect estimates are generally similar using either the fixed- or random-effect model, but the variance of the summary effect is greater in the random-effect model to the degree that there is heterogeneity among studies, and the confidence interval around the summary effect is correspondingly larger and more likely not to be statistically significant. Many journals now require authors to use a random-effect model because it is considered "conservative" (i.e., investigators are less likely to find a statistically significant result than when using the fixed-effect model). Meta-analyses should state clearly whether they used a fixed- or random-effect model.

Statistical Tests of Homogeneity

Tests of homogeneity assume that the findings of the individual trials are the same (the null hypothesis) and use a statistical test (test of homogeneity) to

determine if the data (the individual study findings) refute this hypothesis. A chi square test is commonly used (1). If the data do support the hypothesis ($P > 0.10$), the investigator accepts that the studies are homogeneous. If the data do not support the hypothesis ($P < 0.10$), he rejects the null hypothesis and assumes that the study findings are heterogeneous (apples and oranges). In other words, there are meaningful differences in the populations studied or in the nature of the predictor or outcome variables.

All meta-analyses should report a test of homogeneity with a *P*-value. These tests are not very powerful and it is hard to reject the null hypothesis and prove heterogeneity when the sample size—the number of individual studies—is small. For this reason, a *P*-value somewhat higher than the typical value of 0.05 is typically used as a cutoff. If substantial heterogeneity is present, it is inappropriate to combine the results of trials into a single summary estimate.

References

1. Petitti D. *Meta-analysis, decision analysis and cost effectiveness analysis.* New York: Oxford University Press, 1994.
2. Cooper H, Hedges LV. *The handbook of research synthesis.* New York: Russell Sage Foundation, 1994.