CHAPTER 13

Research Using Existing Data

Deborah G. Grady, Steven R. Cummings, and Stephen B. Hulley

Many research questions can be answered quickly and efficiently using data or specimens that have already been collected. There are three general approaches to using these existing resources. **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 one or more measurements 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 of a given research question, often including calculation of a summary estimate of effect that has greater precision than the individual study estimates. Making creative use of existing data and specimens is a fast and effective way for new investigators with limited resources to begin to answer important research questions, gain valuable experience in a research area, and sometimes have a publishable finding in a short time frame.

ADVANTAGES AND DISADVANTAGES

The main **advantages** of studies 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 database of the Study of Osteoporotic Fractures, a prospective cohort study originally designed to study risk factors for fracture, Yaffe and colleagues used repeated measurements that had been made of physical activity and of cognitive function to discover that women who walked more had a 36% lower risk of cognitive decline than women who walked less (1).

Studies using existing data or specimens also have **disadvantages**. The selection of the population to study, which data to collect, the quality of data gathered, and how variables were measured and recorded are all predetermined. The existing data may have been collected from a population that is not ideal (e.g., men only rather than men and women), the measurement approach may not be what the investigator would prefer (history of hypertension, a dichotomous historical variable, in place of actual blood pressure), and the quality of the data may be poor (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 what data have been collected, and how.

SECONDARY DATA ANALYSIS

Secondary data sets may come from medical records, health care billing files, death certificates, public databases, and many other sources, but other **research studies**, either conducted at the investigator's institution or elsewhere, are one of the richest sources of secondary data. Many studies collect more data than the investigators analyze and these data can be used to document interesting results that have gone unnoticed. Access to such data is generally controlled by the study's **principal investigator** or a **steering committee**; the new researcher should therefore seek out information about studies by other investigators that may have made measurements relevant to the research question. One of the most important ways a good mentor can be helpful

to a new investigator is by providing knowledge of and access to relevant data sets. Most **NIH-funded** studies are required to make their data publicly available.

Other fruitful sources of secondary data are large regional and **national data sets** that are publicly available and do not have a principal investigator. Computerized databases of this sort are as varied as the reasons people have for collecting information. We will give several examples that deserve special mention, and readers can locate others in their own 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 one quarter of the U.S. population, and the area of coverage is expected to increase during the coming years. One purpose 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. For example, investigators used the SEER registry of breast cancer diagnoses to find that the annual incidence of estrogen-receptor positive breast cancer declined 13% in postmenopausal women between 2001 and 2003; this trend paralleled the reduction in use of hormone therapy by postmenopausal women, suggesting that stopping hormone therapy reduced the risk of breast cancer (2).
- **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. A classic example is the follow-up of men with coronary disease who were randomly assigned to high-dose nicotinic acid or placebo to lower serum cholesterol in the Coronary Drug Project. No study had ever shown an effect of lipid treatment on mortality and there was no difference in death rates at the end of the 5 years of randomized treatment, but a mortality follow-up 9 years later using the National Death Index revealed a significant benefit (3). Whether an individual is alive or dead is public information, so follow-up was available even for men who had dropped out of the study.

  The National Death Index can be used when either the Social Security number or the name and birth date are known. Ascertainment of the fact of death is 99% complete with this system, and additional information from 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.
- **NHANES,** the National Health and Nutrition Examination Survey is a series of surveys that assess the health and nutritional status of both adults and children in the United States. The surveys employ population-based cluster random selection to identify a nationally representative sample, and include self-reported data (e.g., demographic, socioeconomic, dietary, and health-related behaviors), physical examinations, laboratory tests, and other measurements. NHANES data can provide population-based estimates of disease prevalence, risk factors, and other variables. For example, bone mineral density (BMD) of the hip was measured during two examinations: 1988–1994 and 2005–2006. The results provide normal values for women and men of various races in the United States that are used to define 'osteoporosis' as 2.5 standard deviations below the average BMD value for young adults in NHANES (4). Investigators also used the repeated measurements to discover that BMD has been improving and the prevalence of osteoporosis has been declining (5).

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. These types of existing data include **electronic** administrative and clinical **databases** such as those developed by Medicare, the Department of Veterans Affairs, Kaiser Permanente Medical Groups, the Duke Cardiovascular Disease Databank, and **registries** such as the San Francisco Mammography Registry and the National Registry of Myocardial Infarction. Information from these sources

(many of which can be found on the Web) can be very useful for studying **rare adverse events** and for assessing real-world utilization and effectiveness of an intervention that has been shown to work in a clinical trial setting. For example, the National Registry of Myocardial Infarction was used to examine risk factors for intracranial hemorrhage after treatment with recombinant tissue-type plasminogen activator (tPA) for acute myocardial infarction (MI). 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 when compared with lower doses (6). 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 in a select population under highly controlled circumstances in limited clinical settings. In the "real world," however, the patients who are treated, the choice of drugs and dosage by the treating physician, and adherence to medications by the patient are much more variable. These factors may make the application of therapy in the general population less effective than what is observed in trials. The **effectiveness** of treatments in actual practice can sometimes be studied using secondary data. For example, primary angioplasty has been demonstrated to be superior to thrombolytic therapy in clinical trials among patients with acute MI (7). But this may only be true when success rates for angioplasty are as good as those achieved in the clinical trial setting. Secondary analyses of community data sets have not found a benefit of primary angioplasty over thrombolytic therapy (8, 9). However, it is important to remember that observational studies of treatments have several limitations—most importantly potential confounding by differences in characteristics of those treated and those not treated. Bias and confounding are particularly difficult to assess using secondary databases that are not designed to study the effectiveness of treatments, and a randomized trial comparing treatments conducted in community settings is a better approach, when feasible.

Secondary data analysis is often the best approach for describing how therapies are used in clinical practice. 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 useful public health implications. For example, using publicly available data from a 5% random sample of Medicare beneficiaries, investigators demonstrated substantial regional variation in the prevalence of diagnosed glaucoma after adjustment for potential confounders, suggesting over- or underdiagnosis in certain regions of the country (10).

Two or more existing data sets may also be linked to answer a research question. Investigators who were interested in how military service affects health used the 1970 to 1972 draft lottery involving 5.2 million 20-year-old men who were assigned eligibility for military service randomly by date of birth (the first data set) and subsequent mortality based on **death certificate registries** (the second source of data). The predictor variable (date of birth) was a randomly assigned 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 ensuing 10 years (11). The study was done very inexpensively, yet it was a more unbiased approach to examining the effect of military service on specific causes of subsequent death than other studies of this topic with much larger budgets.

When individual data are not available, **aggregate data sets** can sometimes be useful. Aggregate data include information only for groups of persons (e.g., death rates from cervical cancer in each of the 50 states), not for individuals. With such data, associations can only be measured among these groups by comparing group information on a risk factor (such as tobacco sales by

region) with the rate of an outcome (lung cancer by region). Studies of associations based on aggregate data are called **ecologic studies**.

The advantage of aggregate data is its availability. Its major drawback is that associations are especially susceptible to confounding: Groups tend to differ from each other in many ways, not just with regard to the predictor variable of interest. As a result, 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 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

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 the research question can be addressed with an existing data set. The help of a **senior colleague** can be invaluable in finding an appropriate data set. An experienced researcher has defined areas of interest in which he stays current and is aware of important data sets and the investigators who control these data, both at his own institution and elsewhere. This person can help identify and gain access to the appropriate data. Often, the research question needs to be altered slightly (by modifying the definition of the predictor or outcome variables, for example) to fit the available data.

The best solution may be close at hand, a database at the **home institution**. For example, a University of California, San Francisco (UCSF) fellow who was interested in the role of lipoproteins in coronary disease noticed that one of the few interventions known to lower the level of lipoprotein(a) was estrogen. Knowing that the Heart and Estrogen/Progestin Replacement Study (HERS), a major clinical trial of hormone treatment to prevent coronary disease, was managed at UCSF, the fellow approached the investigators with his interest. Because no one else had specifically planned to examine the relationship between this lipoprotein, hormone treatment, and coronary heart disease events, the fellow designed an analysis and publication plan. After receiving permission from the HERS study leadership, he worked with coordinating center statisticians, epidemiologists, and programmers to carry out an analysis that he subsequently published in a leading journal (12).

Sometimes a research question can be addressed that has little to do with the original study. For example, another fellow from UCSF was interested in the value of repeated screening Pap tests in women over 65 years old. He realized that the mean age of participants in the HERS trial was 67 years, that participants were required to have a normal Pap test to enter the trial, and that participants then underwent screening Pap tests annually during follow-up. By following up on Pap test outcomes, he was able to document that 110 Pap tests were abnormal among 2,763 women screened over a 2-year period, and that only one woman was ultimately found to have abnormal follow-up histology. Therefore, all but one of the abnormal Pap tests were falsely positive (13). This study strongly influenced the next U.S. Preventive Services Task Force recommendation that Pap tests should not be performed in low-risk women over age 65 with previous normal tests.

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, an investigator can seek to locate databases that include these variables. Some studies have websites that provide free access to the study data without requiring permission. When the data are not available online, 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 about contacting strangers to ask for help. Most people are surprisingly cooperative, either by providing data themselves or by suggesting other places to try.

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 titles and your institutional domain name on correspondence or e-mail, and to copy your mentor as someone who will be recognized as an expert in the field. Young investigators should determine if their mentors are acquainted with the investigators who control the database, as an introduction may be more effective than a cold contact. It is generally most effective to work with an investigator, or a member of the study staff, who is interested in the research topic and involved in the study that has the data of interest. This investigator can facilitate access to the data, assure understanding of the study methods and how the variables were measured, and often becomes a valued colleague and collaborator. Data sets from multicenter studies and clinical trials generally have clear procedures for obtaining access to the data that include the requirement for a written proposal that must be approved by an analysis or publications committee.

The investigator should be very specific about what information is sought and confirm the request in writing. Many studies have guidelines for requesting data that specify what data are being requested, how the analyses will be done, and the timelines for completing the work. It is a good idea to keep the size of the request to a minimum and to offer to pay the cost of preparing the data. If the data set is controlled by a group of researchers, the investigator can suggest a collaborative relationship. In addition to providing an incentive to share the data, this can engage a co-investigator 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.

ANCILLARY STUDIES

Research using secondary data takes advantage of the fact that most of the data needed to answer a research question are already available. In an **ancillary study**, the investigator **adds** one or several **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 brief questionnaire at the next planned exam created a large trial of the effect of hormone therapy on urinary incontinence, with little additional time or expense (14).

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.

Ancillary studies have the problem that the measurements may be most informative when added before the study begins, and it may be difficult for an outsider to identify studies in the planning phase. Even when a variable was not measured at baseline, however, a single measurement during or at the end of a 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 (15).

A good opportunity for ancillary studies is provided by the banks of **stored sera**, **DNA**, **images**, and so on, that are found in most large clinical trials and cohort studies. The opportunity to propose new measurements using these stored 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 or case–cohort design (Chapter 8). In HERS, for example, a nested case–control study that carried out genetic analyses on stored specimens showed that the excess number of thromboembolic events in the hormone-treated group was not due to an interaction with factor V Leiden (16).

Getting Started

Opportunities for ancillary studies should be actively pursued, especially by new investigators who have 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, an investigator interested in the effect of weight loss on pain associated with osteoarthritis of the knee might start by identifying studies that include good measurement of painful osteoarthritis (by validated questionnaires) or databases with records of joint replacements that also have preceding measurements of weight. Additionally, the investigator may look for 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 (http://clinicaltrials.gov or http://report.nih.gov), 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 at a follow-up exam of subjects enrolled in these studies.

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 will consider adding brief ancillary measures to an established study if they address an important question and do not substantially interfere with the conduct of the main study. Investigators will be reluctant to add measures that require a lot of the participant's time (e.g., cognitive function testing) or are invasive and unpleasant (colonoscopy) or costly (positron emission tomography scanning).

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 often 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 marginal cost of an ancillary study is far less than the cost of conducting the same study independently. Ancillary studies are also very well suited for some types of NIH funding that provide only modest support for measurements and analyses but substantial support for career development (Chapter 19). Some large studies 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. If the study will be collecting data from participants, new measures can be added, but variables already being measured generally cannot be changed. In some cases there may be practical problems in obtaining permission from the investigators or sponsor to perform the ancillary study, in training those who will make the measurements, or in obtaining separate informed consent from participants. 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 a set of completed studies that address a particular 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, a systematic review uses a well-defined approach to identify all relevant studies, display the characteristics and 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 great 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 on 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 from 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 13.1. A good source of information on methods for conducting high-quality systematic reviews can be found in the *Cochrane Handbook for Systematic Reviews* (http://handbook.cochrane.org). Just as for 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

A good systematic review has a well-formulated, clear research question that meets the **FINER** criteria (Chapter 2). Feasibility depends largely on the 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 with acute coronary syndrome, does treatment with aspirin plus intravenous heparin reduce the risk of myocardial infarction and death during the hospitalization more than treatment with aspirin alone?"

This research question led to a meta-analysis that found that adding aspirin to heparin improved outcomes, which was published in a top medical journal (17) and had an important impact on practice patterns.

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 may not list non-English-language references. Depending on the research question, electronic databases such as AIDSLINE, CANCERLIT, and EMBASE should 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. The search strategy should be clearly described so that other investigators can replicate the search.

TABLE 13.1 ELEMENTS OF A GOOD SYSTEMATIC REVIEW

1. Clear research question
2. Comprehensive and unbiased identification of completed studies
3. Clear 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 weighted 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

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*** (Table 13.2). 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 studies are identified, these 30 studies should be fully referenced and a reason 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 (Table 13.3). The data abstraction forms should include any data that will subsequently appear in the text, tables, or figures describing the studies included in the systematic review, or in tables or figures presenting the outcomes. When the two abstractors disagree, a third abstractor can 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

TABLE 13.2 CRITERIA FOR INCLUDING OR EXCLUDING STUDIES FROM META-ANALYSES

CRITERIA	EXAMPLE–OMEGA-3 FATTY ACIDS AND CARDIOVASCULAR EVENTS*
1. Period during which the studies were published	Studies published before August 2012
2. Study design	Randomized, controlled trials implemented in primary or secondary cardiovascular disease prevention settings
3. Study population	Studies of adults randomized to omega-3 fatty acids or control
4. Intervention or risk factor	Omega-3 fatty acid administration, either by diet or supplements, any dose, administered for at least one year
5. Acceptable control groups	A non-omega-3 fatty acid diet or supplement
6. Other study design requirements (e.g., blinding for trials or control for specific potential confounders for observational studies)	None
7. Acceptable outcomes	All-cause mortality, cardiac death, sudden death, myocardial infarction, and stroke
8. Maximal acceptable loss to follow-up	Not stated
9. Minimal acceptable length of follow-up	Not stated

*This example of how these criteria are used is drawn from a published meta-analysis showing no effect of omega-3 fatty acids in preventing cardiovascular disease events. (24)

TABLE 13.3 ELEMENTS TO INCLUDE ON DATA ABSTRACTION FORMS FOR META-ANALYSES

1. Eligibility criteria (how does the study meet pre-established eligibility criteria?)
2. Design features (study design, control group, blinding, control for confounding, etc.)
3. Characteristics and number of participants in each study group (demographics, illness severity, etc.)
4. Intervention (for trials) or risk factors (for observational studies).
   - For interventions—dose, duration of treatment, etc.
   - For observational studies—type and level of risk factor, etc.
5. Main outcome, secondary outcomes, and outcomes in pre-established subgroups
6. Elements to allow assessment of quality of included studies (randomization, blinding, adherence, loss to follow-up, control for confounding, etc.)

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 because there is not 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 necessary 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 systematic review are presented in tables. These often include characteristics of the study population, sample size, number or rate of outcomes, length of follow-up, and methods used in the study. Second, the review displays the **analytic findings** of the individual studies (relative risk, odds ratio, risk difference, and confidence intervals or *P* values) in a table or figure. Finally, in the absence of significant heterogeneity (see below), the meta-analysis 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 a main outcome of the meta-analysis, 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, summary risk difference, etc.) and **confidence interval** are generally calculated. The summary effect is essentially an average effect weighted by the inverse of the variance of the outcome of each study. Methods for calculating the summary effect and confidence interval are discussed in Appendix 13. Those not interested in the details of calculating mean weighted estimates from multiple studies should at least be aware that different approaches can give different results. For example, recent meta-analyses of the effectiveness of condoms for preventing heterosexual transmission of HIV have given summary estimates ranging from 80% to 94% decrease in transmission rates, although they are based on the results of almost identical sets of studies (18, 19).
- **Heterogeneity**. Combining the results of several studies is not appropriate if the studies differ in clinically important ways, such as the population, intervention, outcome, control

condition, 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 whether methods and findings are similar enough to combine into summary estimates? 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, it is 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), but ultimately, the **assessment of heterogeneity** requires judgment. Every reported systematic review should include some discussion of heterogeneity and its effect on the summary estimates.

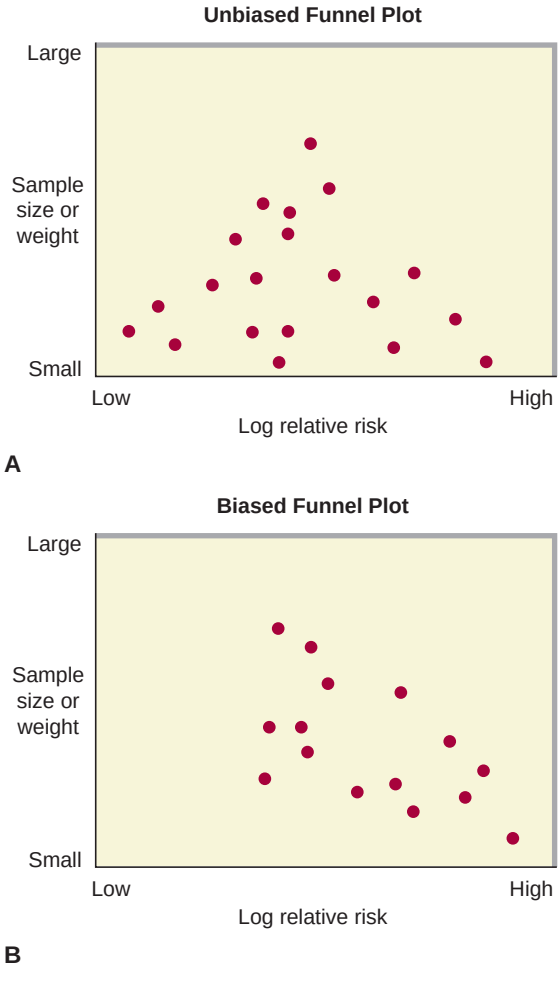
Assessment of Publication Bias

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 studies 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 (which, lacking the rigor of peer review, may be inferior). 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 (or the variance of the outcome) and the findings. The degree of this association is often measured using **Kendall's Tau**, a coefficient of correlation. A strong or statistically significant correlation between study outcome and sample size suggests publication bias. In the absence of publication bias, a plot of study sample size 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 Figure 13.1A suggests that there is little publication bias because small studies with both negative and positive findings were published. The plot in Figure 13.1B, on the other hand, suggests publication bias because the distribution appears truncated in the corner that should contain small, negative studies.

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.



■ **FIGURE 13.1** **A:** Funnel plot that does not suggest publication bias because 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 because few of the smaller studies report low relative risks.

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 (20).

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.

Meta-analyses can increase the **power** to answer a research question, but have the disadvantage that they do not include individual-level data to allow adjustment for potential confounding or to perform individual subgroup analyses. In some situations, it may be possible to obtain the individual-level data from the relevant individual studies and perform **pooled analyses**. In these cases, the pooled data from individual studies can be used to adjust for confounding or

assess subgroup effects just as would be done in a large single study. For example, the Early Breast Cancer Trialists Collaborative Group pooled individual-level data from 123 randomized trials to evaluate the efficacy of different chemotherapy regimens for early breast cancer (21). However, it is generally difficult to obtain individual-level data from relevant studies, and uncommon that these studies have measured variables in ways that are similar enough to be combined into one data set.

Garbage In, Garbage Out

The biggest drawback to a systematic review is that it can produce a reliable-appearing summary estimate based on the results of individual studies that are of poor quality. There are several approaches used to assess the quality of different study designs in meta-analyses, but 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. A special instance of this problem is encountered in systematic reviews of observational data. If the results of these studies are not adjusted for potential confounding variables, the results of the meta-analysis will also be unadjusted and potentially confounded.

■ SUMMARY

This chapter describes three approaches to making creative use of existing data and specimens, a fast and effective way for new investigators with limited resources to acquire valuable experience and an early publication.

Secondary Data Analysis

1. This approach to using existing data sets has the **advantage** of greatly reducing the time and cost of doing research and the **disadvantage** of providing little or no control over the study population, design, or measurements.
2. Sources of data for secondary analysis include **existing research projects**, **electronic medical records**, **administrative databases** and public databases such as **tumor registries**, **death certificate registries**, and national surveys such as **NHANES**.
3. Large community-based data sets are useful for studying **effectiveness** (the real-world effects of an intervention in various communities); for assessing **utilization rates** and **regional variation**, and for discovering **rare adverse events**.
4. Studies of associations based on **aggregate data** are called **ecological studies**; these can provide useful information but are subject to special biases termed **ecological fallacies**.

Ancillary Study

1. An ancillary study is a secondary data analysis in which the investigator makes one or more **new measurements** to answer a new research question with relatively **little cost and effort**.
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 new research question.
3. **Stored serum**, **DNA**, **images**, and so on, provide the opportunity for nested case–control designs.
4. Most large studies have written **policies** that allow investigators (including outside scientists) to propose and carry out secondary data analyses and ancillary studies.

Systematic Review

1. A good systematic review, like any other study, requires a **written protocol** before the study begins that includes the **research question**, methods for **identifying all eligible studies**, methods for **abstracting data** from the studies, and **statistical methods**.
2. The statistical aspects of combining studies on a topic, termed **meta-analysis**, include the **summary effect estimate and confidence interval**, tests for evaluating **heterogeneity** and potential **publication bias**, and **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. A major challenge is assessing **quality** of the studies in a systematic review, which can strongly influence the findings of the review.

APPENDIX 13
Statistical Methods for Meta-Analysis

SUMMARY EFFECTS AND CONFIDENCE INTERVALS

The primary goal of meta-analysis is to calculate a **summary effect estimate** and confidence interval. An intuitive way to do this is to multiply each study outcome, such as the relative risk (an effect estimate), by the sample size (a weight that reflects the precision 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 precision 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 log 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). Therefore, 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 (22, 23). Most use something that approximates the inverse of the sum of the weights of the individual studies ($1/\Sigma\ \text{weight}_i$). 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-EFFECTS MODELS

There are multiple statistical approaches available for calculating a summary estimate (22, 23). The choice of statistical method is usually dependent on the type of outcome (relative risk, odds ratio, risk difference, etc.). In addition to the statistical method, the investigator must also choose to use either a fixed-effects or random-effects model. The **fixed-effects** model simply calculates the variance of a weighted summary estimate based on the inverse of the sum of the weights of each individual study. The **random-effects** 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-effects model, but the variance of the summary effect is greater in the random-effects model to the degree that the results of the individual studies differ, and the confidence interval around the summary effect is correspondingly larger, so that summary results are less likely to be statistically significant. Many journals require authors to use a random-effects model because it is considered "conservative" (i.e., less likely to find a statistically significant effect if one does not exist). Meta-analyses should state clearly whether they used a fixed- or random-effects model.

Simply using a random-effect model does not obviate the problem of heterogeneity. If the studies identified by a systematic review are clearly heterogeneous, a summary estimate should not be calculated.

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-squared test is commonly used (22). If the data do support the null hypothesis (P value ≥0.10), the investigator accepts that the studies are homogeneous. If the data do not support the hypothesis (P value <0.10), he rejects the null hypothesis and assumes that the study findings are heterogeneous. In other words, there are meaningful differences in the populations studied, the nature of the predictor or outcome variables, or the study results.

All meta-analyses should report tests 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 of 0.10 rather than 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. Yaffe K, Barnes D, Nevitt M, et al. A prospective study of physical activity and cognitive decline in elderly women: women who walk. *Arch Intern Med* 2001;161:1703–1708.
2. Kerlikowske K, Miglioretti D, Buist D, et al. Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J Natl Cancer Inst.* 2007;99:1335–1339.
3. Canner PL. Mortality in CDP patients during a nine-year post-treatment period. *J Am Coll Cardiol* 1986;8:1243–1255.
4. Looker AC, Johnston CC Jr., Wahner HW, et al. Prevalence of low femoral bone density in older U.S. women from NHANES III. *J Bone Miner Res* 1995;10:796–802.
5. Looker AC, Melton LJ, Harris TB, et al. Prevalence and trends in low femur bone density among older US adults: NHANES 2005–2006 compared with NHANES III. *J Bone Miner Res* 2010;25:64–71.
6. 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.
7. 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;278:2093–2098; published erratum appears in *JAMA* 1998;279:876.
8. 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;335:1253–1260.
9. 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;31:1240–1245.
10. Cassard SD, Quigley HA, Gower EW, et al. Regional variations and trends in the prevalence of diagnosed glaucoma in the Medicare population. *Ophthalmology* 2012;119:1342–1351.
11. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality: a randomized natural experiment. *N Engl J Med* 1986;314:620–624.
12. 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;283:1845–1852.
13. Sawaya GF, Grady D, Kerlikowske K, et al. The positive predictive value of cervical smears in previously screened postmenopausal women: the Heart and Estrogen/Progestin Replacement Study (HERS). *Ann Intern Med* 2000;133:942–950.
14. Grady D, Brown J, Vittinghoff E, et al. Postmenopausal hormones and incontinence: the Heart and Estrogen/Progestin Replacement Study. *Obstet Gynecol* 2001;97:116–120.
15. Grady D, Yaffe K, Kristof M, et al. Effect of postmenopausal hormone therapy on cognitive function: the Heart and Estrogen/Progestin Replacement Study. *Am J Med* 2002;113:543–548.
16. Herrington DM, Vittinghoff E, Howard TD, et al. Factor V Leiden, hormone replacement therapy, and risk of venous thromboembolic events in women with coronary disease. *Arterioscler Thromb Vasc Biol* 2002;22:1012–1017.
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;276:811–815.

18. Pinkerton SD, Abramson PR. Effectiveness of condoms in preventing HIV transmission. *Soc Sci Med* 1997;44:1303–1312.
19. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. *Cochrane Database Syst Rev* 2002;(1):CD003255.
20. Grady D, Gebretsadik T, Kerlikowske K, et al. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstet Gynecol* 1995;85:304–313.
21. Peto R, Davies C, Godwin J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet* 2012;379:432–441.
22. Petitti DB. *Meta-analysis, decision analysis and cost effectiveness analysis: methods for quantitative synthesis in medicine,* 2nd ed. New York: Oxford University Press, 2000.
23. Cooper H, Hedges LV. *The handbook of research synthesis*. New York: Russell Sage Foundation, 1994.
24. Rizos EC, Ntzani EE, Bika E, et al. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events. *JAMA* 2012;308:1024–1033.