Journal of Law and Policy

Volume 12
Issue 2
SCIENCE FOR JUDGES II:
The Practice of Epidemiology and Administrative
Agency Created Science

Article 2

2004

Overview of Research Design in Epidemiology

John Concato

OVERVIEW OF RESEARCH DESIGN IN EPIDEMIOLOGY

*John Concato, M.D., M.S., M.P.H.**

INTRODUCTION

The objectives of this paper are to first, provide background regarding a conceptual model of causality in epidemiology; second, describe common types of structure (architecture) used for research design in epidemiology, including descriptive, cohort, case-control, and cross-sectional studies; third, review frequently-encountered formats for reporting the results of such studies; and fourth, discuss the strengths and limitations of strategies used in epidemiology. The first and fourth objectives represent a big-picture assessment for interpreting epidemiologic studies; whereas the second and third objectives promote a nuts-and-bolts understanding of the studies themselves.

I. CONCEPTUAL MODEL OF CAUSALITY

A discussion of research design involves considerations of the concept of causality, i.e., what causes disease.[1] In this context,

* The author is an Associate Professor of Medicine in the Department of Internal Medicine at Yale University School of Medicine in New Haven, CT, and Director of the Clinical Epidemiology Research Center at VA Connecticut Healthcare System in West Haven, CT. The conclusions and opinions expressed in this article are those of the author and do not represent any official or unofficial position of Yale University or the U.S. Department of Veterans Affairs. Presented at the second Science for Judges Symposium, Brooklyn Law School, Brooklyn, NY, November 7, 2003.

[1] *See, e.g.*, Sir Austin Bradford-A.B. Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. R. SOC. MED. 295-300 (1965).

biological phenomena can be considered either deterministic or probabilistic. A deterministic situation exists when the exposure can be linked conclusively to an outcome (e.g., when a particular genetic rearrangement causes sickle cell anemia). In contrast, a probabilistic phenomenon occurs when exposure is said to be associated with an outcome (e.g., if hypertension (high blood pressure) is associated with stroke). In the first example, the genetic problem is always found with the disease, and vice versa. In the second example, hypertension increases the probability of stroke; but some patients with hypertension do not suffer a stroke, and some patients with stroke do not have antecedent hypertension.

A major role of epidemiological research design is to provide information for, or against, a probabilistic association. A conceptual (intellectual) model[2] has been developed for this purpose, and is often described with terms such as cause-effect research (Figure 1). The entity being evaluated as a possible cause of a disease, or other endpoint, is referred to as an exposure, but is not limited to environmental exposures, and can include a person's age, sex, personal habits (such as cigarette smoking), ingested medications, etc. The entity being assessed as a possible disease or other endpoint is referred to as an outcome, and can include the development of a disease, a quality of life measurement, death, etc.

FIGURE 1—CONCEPTUAL MODEL FOR CAUSE-EFFECT RESEARCH

Baseline state	→	Exposure	→	Outcome
Condition prior to exposure or outcome		*Factor(s) that may lead to outcome*		*Disease or entity of interest*

Example:

Healthy adults → Hypertension → Stroke

The model can be applied to the association of hypertension and stroke, addressing the question of whether hypertension can

[2] ALVAN R. FEINSTEIN, CLINICAL EPIDEMIOLOGY: THE ARCHITECTURE OF CLINICAL RESEARCH 50 (Saunders 1985).

"cause" stroke (in a probabilistic, not a deterministic, sense). In this example, adults, either with or without hypertension, do or do not develop subsequent stroke. The challenge involves determining whether stroke is more common among patients with hypertension; and if so, whether the corresponding evidence (in the form of data) is strong enough, and stable enough, to confirm that an association exists and is "real."

A useful aspect of cause and effect studies is the ability to summarize the association of interest with a simplified schematic (2 × 2 table) showing the relationship of exposure and outcome (Table 1A). Although most epidemiological investigations involve more complex designs, the focus of a study can be understood using this approach. This framework will be used throughout the subsequent text, with each of the types of research architecture, to demonstrate basic principles involved.

TABLE 1A—SCHEMATIC OF CAUSE-EFFECT STUDY

	Outcome	No outcome
Exposure	a	b
No exposure	c	d

Exposure = factor that may "cause" outcome

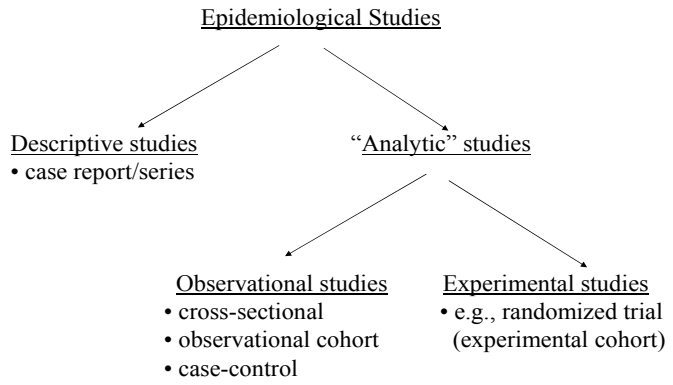
Outcome = entity (e.g., disease) of interest

II. SPECIFIC TYPES OF RESEARCH DESIGN

Unfortunately, no universally accepted terminology (classification scheme) exists for types of epidemiologic studies. Specific terms are often used to describe basic categories and formats of studies (Figure 2), but the terms may vary with different authors, leading to potential confusion.[3]

[3] For example, the terms prospective and retrospective are often used to describe studies, but these terms can be confusing because they are applied to

FIGURE 2—OVERVIEW OF RESEARCH ARCHITECTURE



A. Descriptive Studies

The first category of research design, descriptive studies, has a limited role in discussions of cause-effect research because such studies have limited implications regarding (probabilistic) causation. Descriptive studies are often presented as a case report, or a case series of patients, and are useful to inform clinical care, such as a recent report describing severe acute respiratory syndrome (SARS).[4] The strengths and limitations of descriptive studies in demonstrating causality is illustrated by a descriptive study reporting on 24 cases of valvular heart disease in patients taking fenfluramine-phentermine as a dietary suppressant.[5] This article can be considered a prominent, early publication linking dietary suppressants and valvular heart disease. As noted in the discussion section of that paper, however, "... definitive

different aspects of study architecture. In addition, I will use the association between dietary suppressants and heart disease hereafter in the text; I make no claims about the merit of any related court cases, and I do not imply to present a comprehensive nor complete assessment of the scientific evidence on this topic. Finally, I will use the term "patient" frequently, reflecting my training and experience as a physician; the term person may often be substituted.

[4] Nelson Lee et al., *A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong*, 348 NEW ENG. J. MED., 1986-94 (2003).

statements about a true association of valvular disease with fenfluramine-phentermine therapy cannot be made."[5] The investigation corresponds to the upper left hand "cell" (labeled a) in Table 1B, indicating that patients with valve defects who had taken fenfluramine-phentermine were the focus of investigation. Other descriptive studies may report on patients, for example, in the top row of cells, but not in all four cells. The importance of this feature as a limitation of descriptive studies is mentioned in the discussion of analytic studies.

TABLE 1B—EXAMPLE OF CAUSE-EFFECT STUDY

	Valve defects	No valve defects
Fen-phen	a	b
No fen-phen	c	d

Fenfluramine-phentermine (fen-phen) is the exposure;
valvular heart disease (valve defect) is the outcome.

B. Analytic Studies

A second, and more commonly encountered, type of study can be described as analytic (or comparative or cause-and-effect) studies. This category of study is so common that the term analytic is not often used; alternatively, terms such as etiologic (what causes disease), diagnostic (how is disease diagnosed), or prognostic (what happens to patients with disease), are used to indicate the main focus of the study, rather than the architecture itself.

The hallmark of analytic study is that two (or more) groups are compared to draw an inference regarding a cause-effect association. Thus, inclusion of a so-called control or comparison

[5] Heidi M. Connolly et al., *Valvular Heart Disease Associated with Fenfluramine-Phentermine*, 337 NEW ENG. J. MED., 581-88 (1997).

group is an indication of an analytic study. The presence of the control group provides a mechanism for establishing whether the outcome in the exposed group occurs at a higher (vs. lower or similar) rate, compared with the unexposed group.

Along with opportunity to assess (probabilistic) causation, a problem called analytic bias can arise in the conduct of analytic studies. This problem occurs if (various types of) systematic error exists when assessing the association of interest. For example, recall bias is considered a problem if certain patients do not recall accurately their prior exposure. Susceptibility bias is encountered if compared groups of persons have unequal susceptibility for the outcome at baseline before considering the impact of a specific exposure. A general term for these various sources of bias is confounding; a problem that exists if an extraneous factor interferes with assessing the relationship between exposure and outcome, preventing an accurate determination of the true magnitude of association. A possible example of confounding involves the question of whether ingestion of dietary suppressants is associated with heart disease; specifically, whether the association is confounded by obesity. Obesity is related (obviously) to ingestion of these products. The problem of confounding would arise if obesity were also related independently to valve defects—then dietary suppressants could be a "marker" for obesity, rather than a causal factor in the development of valvular heart disease. A solution in this context is to use a procedure (e.g., adjusting, controlling, matching) that accounts for confounding variables. These techniques are usually "as good as" the logic and measurements used in identifying potential confounding factors; the mathematical procedures, although complex, are usually not a problem themselves.

1. Observational Studies

Analytic studies are typically described by another feature of their design; specifically, whether they are observational studies or experimental studies. Observational studies are investigations in which exposure is not assigned by a research investigator. Rather, the exposure occurs via "nature" or as a result of human

interventions in a non-research setting. For example, exposure to ambient air pollution would be assessed in an observational study, as would patients receiving certain medications in the context of their routine health care. Common examples of observational study architecture include cross-sectional, observational cohort, and case control studies.

a. Cross-Sectional Studies

The distinguishing feature of a cross-sectional study is that data on exposure status and outcome status are obtained at essentially the same point in time. In an example of a cross-sectional study, patients were identified from one of three previously conducted appetite suppressant studies, and new "control" subjects (not in the appetite studies) were selected based on matching on several factors (e.g., age, sex, physical features).[6] All of the participants were then assessed for valvular heart disease, and the corresponding four cells of Table 1B would represent a simultaneous exposure-outcome relationship. As noted in the corresponding discussion section of that paper, "the purpose of the study was to determine the prevalence and severity of valvular dysfunction in obese patients who had taken appetite suppressants and those who had not."[6] Although antecedent appetite suppressant use might result in subsequent valve defects, the study architecture could not exclude a converse scenario with the onset of heart disease preceding the ingestion. Thus, the authors were suitably cautious regarding any claim of causation.

b. Observational Cohort Studies

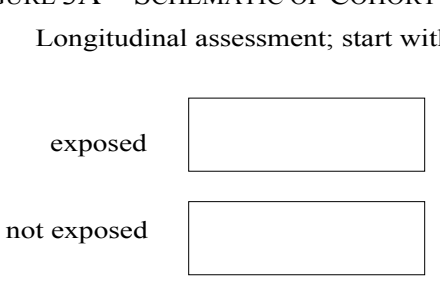
An observational cohort study differs from a cross-sectional study in that it involves a longitudinal assessment, usually based on assembling patients with regard to their status as exposed or not exposed (Figure 3A). Patients' exposure status is determined first,

---

[6] Mehmood A. Khan et al., *The Prevalence of Cardiac Valvular Insufficiency Assessed by Transthoracic Echocardiography in Obese Patients Treated with Appetite-Suppressant Drugs*, 339 NEW ENG. J. MED., 713-18 (1998).

their outcome status is determined subsequently, and then the association between exposure and outcome is assessed (Table 1B). A paragraph of an article describing the methods of an example of an observational cohort analysis stated that the investigators identified subjects who had been given fenfluramine-phentermine, and as a comparison group, they identified a group of subjects who had not received an appetite suppressant.[7] The article found that the risk of valvular disease in subjects who had received certain dietary suppressants was substantially higher over a five-year period, indicating the ability of the study to make a claim regarding causality.[8]

FIGURE 3A—SCHEMATIC OF COHORT ARCHITECTURE



c. Case-Control Studies

A case-control study also involves a longitudinal assessment, but starts with identifying patients based on their outcome status (Figure 3B). For example, in a case-control analysis, investigators compared patients with cardiac-valve abnormalities to patients without such abnormalities.[9] Antecedent ingestion of appetite suppressants was then determined to establish the association of interest (Table 1B). The finding that longer use of fenfluramine or

---

[7] H. Jick et al., *A Population-Based Study of Appetite-Suppressant Drugs and the Risk of Cardiac-Valve Regurgitation*, 339 NEW ENG. J. MED., 719-24 (1998). Of note, this article included both observational cohort and case-control analyses within one manuscript; representing an uncommon, but suitable, format.

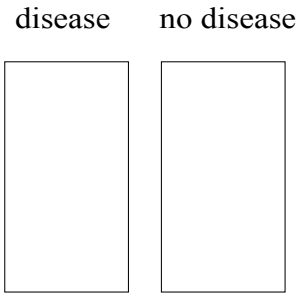
[8] *Id.*

[9] *Id.*

dexfenfluramine was associated with an increased risk of cardiac-valve disorders reflects the analytic nature of the study.

FIGURE 3B—SCHEMATIC OF CASE CONTROL ARCHITECTURE

Longitudinal assessment; start with outcome:



2. Experimental Studies

The hallmark of an experimental study is that exposure is assigned by a research investigator. In practical terms, a randomized controlled trial (or what could be called an experimental cohort) is the currently accepted experimental design in clinical research involving intact human beings. The importance of randomization is that it leads to a balance of risk factors in exposed and non-exposed groups, promoting an unbiased evaluation of exposure-outcome associations. An example of a randomized trial involving dietary supplements is an article that used data from an earlier randomized, controlled trial comparing the efficacy and safety of dexfenfluramine and placebo in treating obesity. [10] Taking advantage of previously collected data (before a possible link to heart diease was recognized), the authors reported a "small increase" in the prevalence of valvular disease in patients

[10] Neil J. Weissman et al., *An Assessment of Heart-Valve Abnormalities in Obese Patients Taking Dexfenfluramine, Sustained-Release Dexfenfluramine, or Placebo*, 339 NEW ENG. J. MED., 725-32 (1998). It is self-evident that designing a clinical trial to assess the association between dietary supplements and valvular heart disease would be unethical.

treated with fenfluramine.[11]

3. Other Study Designs

The research designs described in the previous sub-sections often go by different names, and many other types of research designs also exist. An example of another type of study is ecologic analysis, with data for geographic areas (e.g., countries) used in lieu of individual exposures. Other designs include meta analysis, in which mathematical pooling of available studies is done,[12] and decision analysis, in which a mathematical model is used to simulate results for hypothetical patients, based on data from patient-based studies and other sources.

III. FORMAT FOR REPORTING RESULTS

A discussion of research design would benefit from a description of the common formats used for reporting their results. In brief, the results can be thought of as being quantitative (clinical) and statistical (probabilistic). Quantitative results address the strength of an association, whereas statistical results address the stability of an association. As an example of hypothetical results (Table 2), an association between exposure and disease among 2,000 patients can be considered.

---

[11] *Id*.

[12] *See* John P.A. Ioannidis, M.D. & Joseph Lau, M.D., *Systematic Review of Medica Evidence*, 12 J.L. & POL'Y 509 (2004).

TABLE 2—EXAMPLE OF STATISTICAL RESULTS FOR HYPOTHETICAL STUDY

	Disease	No disease	
Exposed	60	940	1000
Not exposed	30	970	1000

Relative risk = (60/1000) ÷ (30/1000) = 6% ÷ 3% = 2.0;

95% confidence interval = 1.3 to 3.1; P = 0.001

A. Quantitative Results

A common format for reporting the quantitative strength of the association is the relative risk (Table 2), representing the rate of outcome among exposed patients (60/1000 = 6%), divided by the rate of outcome in the non-exposed group (30/1000 = 3%). The relative risk is calculated as 2.0 in this situation and would be stated as "exposed patients were twice as likely to have disease as non-exposed patients." A relative risk of 1.0 would indicate no association (i.e., identical outcome rates for the compared groups), and is therefore called the null value.

The results for this same hypothetical study can be reported, however, in many different ways. For example, the relative risk (just calculated as 2.0) is often replaced by other so-called point estimates, such as an odds ratio, hazard ratio, rate ratio, or rate per 100 person-years of follow-up (e.g., 10 people for 10 years, or 50 people for two years), depending on the particular research design. Even if one focuses on a relative risk, however, it should be appreciated that the comparison of 6% vs. 3% could be described as an absolute risk difference, calculated as (6% - 3% =) a 3% increase; or a proportionate difference, calculated as ([6% - 3%] ÷

3% =) a 100% increase.[13] In addition, other formats for reporting results have been devised, such as the number needed to treat (or in this situation, the number needed to harm).[14]

Importantly, no single scientific threshold exists for quantitative significance, such as determining whether a 6% vs. 3% outcome is big enough. In a legal context, however, a relative risk of greater than 2 has been cited as representing a situation where "were the exposures more likely than not the cause"[15] of the outcome. The logic behind this statement is shown in Table 2, using the same hypothetical study. The explanation is that among the 60 patients who were exposed and who experienced the outcome, 30 would have the baseline rate (or expected) outcome, and 30 more could be attributed to the exposure. Thus, if more than 60 patients are in the upper left hand cell, it is more likely than not that any given patient's outcome is due to exposure. From an epidemiologic perspective, however, this threshold value of 2.0 is quite arbitrary. The results of any study are subject to statistical variation as well as bias, such that a relative risk of 2.1 vs. 1.9 would not be considered substantially different by most epidemiologists.

B. Statistical Results

The same hypothetical study would typically have additional results reported in Probability (P) values or confidence intervals. In contrast to quantitative results, threshold values do exist for these statistical results. Specifically, a P value less than or equal to 0.05, and a 95% confidence interval that excludes 1.0 (or another null value) are used to determine that results are "statistically

---

[13] Depending on one's perspective—especially in a legal setting—one party might be inclined to report "only" a 3% increase, whereas the other party may claim a "whopping" 100% increase, for the exact same findings.

[14] Ignoring the mathematical calculations involved (1/[0.06 – 0.03] = 1/0.03 = 33.3); the interpretation is that when approximately 33 patients are exposed (or not) in each arm of the study, one extra outcome would be observed in the exposed group

[15] FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 384 (2d ed. 2000).

significant."

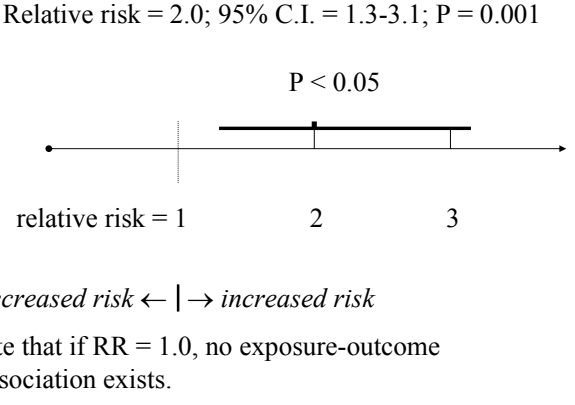
1. P Values

A P value can be thought of as the probability of an observed result occurring by chance alone. The P value could range from 0 to 1, where "lower is better," and P less than or equal to 0.05 is a threshold value as mentioned previously. If the P value is equal to 0.05, we are acknowledging that there is a 1 in 20 probability of the result occurring by chance alone, i.e., being a false-positive result. Using an example from the mathematical concept of probability, the P value for obtaining 5 consecutive heads when flipping a fair coin would be $(0.5)^5 = 0.031$, or approximately 3 in 100. For 10 consecutive heads, the P value would be $(0.5)^{10} = 0.00098$, or less than 1 in 1000. One might "worry" about a coin having two heads if heads appeared five times in a row, and even more so for ten times in a row, but either event could happen potentially. The corresponding interpretation of a study with P = 0.031 would be that the observed result linking exposure to outcome was unlikely due to chance (less than 0.05); and the interpretation for a study with P = 0.00098 would be that the observed result was very unlikely to represent a chance occurrence (much less than 0.05). The calculated P value for the data in Table 2 is 0.001, indicating a statistically significant (stable) result.

2. Confidence Intervals

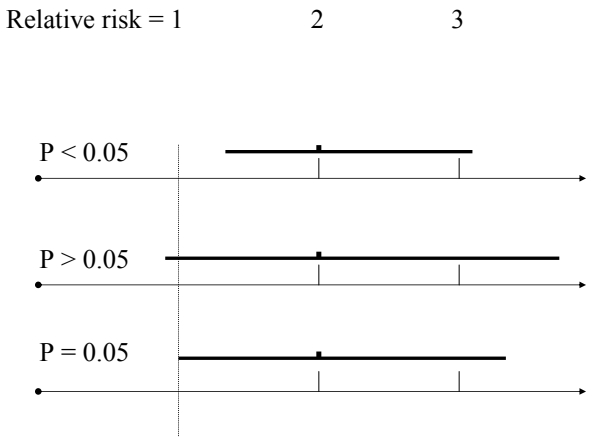
The results from Table 2 are illustrated again in Figure 4, showing the relative risk of 2.0 and a 95% confidence interval from 1.3 - 3.1; indicating the strength of association may be as low as 1.3 or as high as 3.1, but the best estimate is 2.0. In this manner, confidence intervals are shown to express stability of results in terms of "units" of relative risk (or other point estimates), based on the same mathematical information as P values. For example, the observation that the entire 95% confidence interval, including the lower bound at 1.3, does not overlap the null value of a relative risk of 1.0, indicates a statistically significant exposure-outcome association ($P < 0.05$).

FIGURE 4—P VALUE AND 95% CONFIDENCE INTERVAL (C.I.) FOR TABLE 2



The general relationship of P values and confidence intervals is shown in Figure 5; e.g., if a P value is greater than 0.05 (not statistically significant), the 95% confidence interval includes the null value of one.

FIGURE 5—RELATIONSHIP OF P VALUES AND 95% CONFIDENCE INTERVALS



3. Other Considerations

A more in-depth understanding of statistical results can be aided by evaluating "results" from baseball. When evaluating a comparison of batting averages, one can consider whether a .333 average is better than a .250 average (Table 3). The answer would almost certainly be "yes" regarding the quantitative difference, but most people would agree that results would be unstable (i.e., not likely to persist) if based, respectively, on 1 hit in 3 at-bats vs. 1 hit in 4 at-bats, perhaps based on one game at the end of the previous season. That instability is represented by a calculated P value of 0.81. Importantly, the P value decreases for the same comparison of batting averages as the denominator (number of at bats) increases. Thus, by the time these players have 300 and 400 at bats, respectively, the P value of 0.02 indicates that that difference is unlikely due to chance. If the batters end up having several thousand at-bats over the course of their career, the P value shrinks to a very small number (less than 0.0000001) shown in Table 3; indicating that the averages are unlikely to be the same, even if the .333 batter went in to a prolonged slump, and the .250 batter had a sustained hitting streak. The take-home message is that the larger the sample size, the smaller the P value, for constant proportions being compared.

TABLE 3—EXAMPLE OF "LARGE" QUANTITATIVE DIFFERENCE

Player A = .333	Player B = .250	P Value
1/3	1/4	0.81
10/30	10/40	0.45
100/300	100/400	0.02
1000/3000	1000/4000	<0.0000001

Note: The larger the sample size, the smaller the P value, for constant proportions.

A second example involving batting averages illustrates the same issue, with a different perspective (Table 4). In this scenario, one player has a .288 batting average, where as the second player has a .282 batting average. The P value of 0.77 indicates that the two numbers are not statistically significantly different based on 1,000 at bats; and the difference might not have been considered quantitatively important to start with. Yet, if the number of at-bats is increased sufficiently, up to (an unrealistic) value of 100,000, a P value of 0.003 can be obtained. The take-home message here is that a quantitatively unimportant difference can be statistically significant (i.e., less than 0.05), with a large enough sample size. These examples illustrate that a P value and or confidence intervals can help in evaluating the stability of results, but they do not themselves address directly the validity or trustworthiness of the data.

TABLE 4—EXAMPLE OF "SMALL" QUALITATIVE DIFFERENCE

Player A = .288	Player B = .282	P Value
288/1000	282/1000	0.77
2,880/10,000	2,820/10,000	0.35
28,800/100,000	28,200/100,000	0.003

Note: A quantitatively unimportant difference can be made statistically significant, with a large enough sample size.

IV. STRENGTHS AND LIMITATIONS OF STRATEGIES IN EPIDEMIOLOGY

An overview of the strategies used in epidemiology can be discussed in terms of the four stages of analytic study assessing probabilistic causation: framing a research question; developing a

study design;[16] collecting data; and completing a statistical analysis.[17]

A. Framing a Research Question

Issues regarding framing the research question include whether the topic is relevant and whether the question being posed is cogent. In medical research, the assessment of importance is often determined via granting agencies (e.g., the National Institutes of Health) that provide funding to investigators, or the peer review process that determines which manuscripts are published in various medical journals. In legal settings, the adjudication of a controversy in the courtroom would indicate de facto a substantial level of importance. Questions and controversies may still arise, however, regarding, for example, how to interpret studies that examined similar, but not identical, dietary suppressant drugs, or studies that did not perform standard diagnostic tests to find valvular disease among all patients, etc.

B. Developing a Study Design

Randomized controlled trials are less vulnerable to confounding when compared to observational studies because randomization balances potential confounding factors with regard to exposure. In this context, the conventional wisdom for the past several decades has been that observational studies are always inferior to randomized trials, due largely to the problem of confounding. Work done by our group and others,[18] however, has shown that contrary to prevailing beliefs, results from well-

---

[16] *See supra* Part II.

[17] *See supra* Part III.

[18] Kjell Benson & Arthur J. Hartz*, A Comparison of Observational Studies and Randomized, Controlled Trials*, 342 NEW ENG. J. MED., 1878-86 (2000); John Concato et al*., Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs*, 342 NEW ENG. J. MED., 1887-92 (2000); Martin McKee, et al., *Methods in Health Services Research: Interpreting the Evidence: Choosing Between Randomised and Non-Randomised Studies*, 319 BRIT. MED. J. 312-5 (1999).

designed observational studies (with a cohort or case-control design) did not systematically overestimate the magnitude of the associations between exposure and outcome compared with randomized, controlled trials. The results suggested that suitable precautions against bias were taken in higher quality observational studies, making them comparable to clinical trials in terms of producing valid results.

This work indicates that epidemiologists—and judges—should be flexible in interpreting evidence from various types of study design. Unfortunately, it is not a simple matter to determine whether a particular study is valid or not. Various lists (e.g., using "evidence-based medicine")[19] that attempt to provide a hierarchical assessment of research design should be viewed with caution.

C. Collecting Data

Data collection is a frequently overlooked activity, often with assumptions that the activity has been done in a trustworthy manner. Yet, it is important to note that high quality data are crucial to ensure the validity of a study, and such data can involve considerable effort to obtain. For example, large databases often collect information for one purpose (e.g., health insurance documentation) and are used for another (medical research), with incomplete measurements for the factors of interest.[20] In contrast, a review of medical records, or direct examination of patients, might be more difficult and expensive, but more appropriate. In legal settings, the collection of data can become an intense focus of scrutiny, with the motives of investigators called into question.

D. Statistical Analysis

In terms of assuring that a study is done well, statistical analyses, per se, are—perhaps surprisingly—not a critical issue, in

---

[19] Evidence-Based Medicine Working Group, *Evidence-Based Medicine: a New Approach to Teaching the Practice of Medicine*, 268 J. AM. MED. ASS'N 2420-25 (1992).

[20] *See, e.g.*, John Concato et al., *Problems of Comorbidity in Mortality after Prostatectomy*, 267 J. AM. MED. ASS'N 1077-82 (1992).

most situations. The most difficult aspect of statistical analyses can be understanding what particular statistical test (procedure) was used, and why.[21] Regardless of the statistical test, however, a point estimate, and a P value or confidence interval, is usually generated as a final product of the analysis. Although a particular statistical approach might obscure the understanding of a study's results, a competent biostatistician or other professional with suitable training should be able to describe the process in "plain English."

CONCLUSION

Epidemiology is a rigorous and important scientific discipline, but "truth" is difficult to establish. Specifically, studies differ regarding characteristics of patients, assessments of exposure and outcome, possible sources of methodololgical bias, as well as the potential influence of personal or political views. Contradictory results from multiple epidemiologic studies should therefore be resolved by scientific process of reconciling disagreement (e.g., evaluating the quality of research methods). This process, challenging under any circumstances, is made more complex when legal issues are involved.

[21] *See* John Concato et al., *The Risk of Determining Risk with Multivariable Models*, 118 ANNALS INTERNAL MED. 201-10 (1993) (discussing popular techniques).