



โครงการพัฒนาศักยภาพบุคลากร

เรื่อง การประเมินความจำเป็นด้านสุขภาพ(Health Needs Assessment)

เพื่อรองรับการบริหารจัดการเขตสุขภาพ

นายแพทย์ธนรักษ์ ผลิพัฒน์

Evidence-Based Public Health

ROSS C. BROWNSON

ELIZABETH A. BAKER

TERRY L. LEET

KATHLEEN N. GILLESPIE

OXFORD

UNIVERSITY PRESS

2003

2

Assessing Scientific Evidence for Public Health Action

It is often necessary to make a decision on the basis of information sufficient for action but insufficient to satisfy the intellect.

—Immanuel Kant

In most areas of public health and clinical practice, decisions on when to intervene and which program or policy to implement are not simple and straightforward. These decisions are often based on three fundamental questions: (1) Should public health action be taken to address a particular public health issue (primarily based on Type I evidence)? (2) What action should be taken (primarily based on Type II evidence)? and (3) How can a particular program or policy most effectively be implemented and evaluated? This chapter primarily explores the first question. That is, it focuses on several key considerations in evaluating scientific evidence and determining when a scientific basis exists for some type of public health action. It deals largely with the interpretation of epidemiologic studies that seek to identify health risks associated with preventable morbidity and mortality. The second and third questions are explored in more detail in later chapters.

BACKGROUND

In this era when public and media interest in health issues is intense, the reasons for not taking action based on an individual research study, especially if it was carefully designed, successfully conducted, and properly analyzed and interpreted, need to be emphasized. Public health research is incremental, with a body of scientific evidence building up over years or decades. Therefore, while individual studies may contribute substantially to public health decision making, a single study rarely constitutes a strong basis for action. The example in Box

Box 2–1. Toxic Shock Syndrome in the United States

In the case of an infectious agent transmitted by a fomite (i.e., an inanimate object that may harbor a pathogen), the illness known as toxic shock syndrome was reported to the Centers for Disease Control by individual physicians and five state health departments beginning in October 1979.³⁵ Toxic shock syndrome began with high fever, vomiting, and profuse watery diarrhea and progressed to hypotensive shock. Among the first 55 cases, the case fatality ratio was 13%. The bacterium *Staphylococcus aureus* was found to be responsible for the syndrome. Through a nationwide case-control study of 52 cases and 52 matched controls, the mode of transmission was determined to be the use of high absorbency (fluid capacity) tampons in women.³⁶ The findings of epidemiologic studies led to public health recommendations to women regarding safe use of tampons, a voluntary removal of the Rely brand, and subsequent lowering of absorbency of all brands of tampons.³⁷ These actions in turn led to substantial reductions in the incidence of toxic shock syndrome since the early observations of the association between tampon use and toxic shock syndrome.

2–1 regarding toxic shock syndrome is unusual since rapid action was taken based on a small but convincing body of scientific evidence.

When considering the science, strong evidence from epidemiologic (and other) studies may suggest that control measures should be taken. Conversely, evidence may be equivocal, so that taking action would be premature. Often the strength of evidence is suggestive, but not conclusive; yet one has to make a decision about the desirability of taking action. Here, other questions come to mind:

- How serious are the consequences of taking some action or no action, and what other impact will the course of action have?
- Will the action reduce the frequency and/or severity of a serious disease?
- Are there few (if any) adverse effects of intervention?
- Is the action inexpensive and/or cost-effective?

If the answer to the final three questions is “yes,” then the decision to take action is straightforward. In practice, unfortunately, decisions are seldom so simple.

EXAMINING A BODY OF SCIENTIFIC EVIDENCE

As practitioners, researchers, and policy makers committed to improving population health, we have a natural tendency to scrutinize the scientific literature for new findings that would serve as the basis for prevention or intervention

programs.¹ In fact, the main motivation for conducting research should be to stimulate appropriate public health action. Adding to this inclination to intervene may be claims from investigators regarding the critical importance of their findings, media interpretation of the findings as the basis for immediate action, and community support for responding to the striking new research findings with new or modified programs. The importance of community action in motivating public health efforts was shown recently in the Long Island Breast Cancer Study. Community advocates in Long Island raised concerns about the high incidence of breast cancer and possible linkages with environmental chemicals and radiation. A series of studies are being conducted by the New York State Health Department, along with scientists from universities and the National Institutes of Health. In each Long Island-area county, breast cancer incidence increased over a ten-year period, while mortality from breast cancer decreased.² To date, a case-control study has shown that women living closer to chemical facilities have a higher risk of postmenopausal breast cancer, but linkages with specific environmental agents have not been demonstrated and several detailed studies are ongoing.

Finding Scientific Evidence

Chapter 6 describes systematic methods for seeking out credible, peer-reviewed scientific evidence. Modern information technologies have made searching the scientific literature quick and accessible. There are also hundreds of websites that summarize research and provide ready access to surveillance data. The ready access to information may also present a paradox, in that access to more information is assumed to be productive, yet bad science and bad advice are also found in the myriad of information on various topics. Often, various tools are helpful in examining an entire body of evidence, rather than reviewing the literature study-by-study. These summary approaches, described in Chapter 3, include systematic reviews of the literature, evidence-based guidelines, summaries of best practices, and economic evaluations.

The Roles of Peer Review and Publication Bias

In assessing evidence, it is important to understand the role of peer review. Peer review is the process of reviewing research proposals, manuscripts submitted for publication by journals, and abstracts submitted for presentation at scientific meetings. These materials are judged for scientific and technical merit by other scientists in the same field.³ Reviewers are commonly asked to comment on such issues as the scientific soundness of the methods used, originality, rele-

programs.¹ In fact, the main motivation for conducting research should be to stimulate appropriate public health action. Adding to this inclination to intervene may be claims from investigators regarding the critical importance of their findings, media interpretation of the findings as the basis for immediate action, and community support for responding to the striking new research findings with new or modified programs. The importance of community action in motivating public health efforts was shown recently in the Long Island Breast Cancer Study. Community advocates in Long Island raised concerns about the high incidence of breast cancer and possible linkages with environmental chemicals and radiation. A series of studies are being conducted by the New York State Health Department, along with scientists from universities and the National Institutes of Health. In each Long Island-area county, breast cancer incidence increased over a ten-year period, while mortality from breast cancer decreased.² To date, a case-control study has shown that women living closer to chemical facilities have a higher risk of postmenopausal breast cancer, but linkages with specific environmental agents have not been demonstrated and several detailed studies are ongoing.

Finding Scientific Evidence

Chapter 6 describes systematic methods for seeking out credible, peer-reviewed scientific evidence. Modern information technologies have made searching the scientific literature quick and accessible. There are also hundreds of websites that summarize research and provide ready access to surveillance data. The ready access to information may also present a paradox, in that access to more information is assumed to be productive, yet bad science and bad advice are also found in the myriad of information on various topics. Often, various tools are helpful in examining an entire body of evidence, rather than reviewing the literature study-by-study. These summary approaches, described in Chapter 3, include systematic reviews of the literature, evidence-based guidelines, summaries of best practices, and economic evaluations.

The Roles of Peer Review and Publication Bias

In assessing evidence, it is important to understand the role of peer review. Peer review is the process of reviewing research proposals, manuscripts submitted for publication by journals, and abstracts submitted for presentation at scientific meetings. These materials are judged for scientific and technical merit by other scientists in the same field.³ Reviewers are commonly asked to comment on such issues as the scientific soundness of the methods used, originality, rele-

vance, and appropriateness of a scientific article to the audience. Although peer review has numerous limitations, including a large time commitment, complexity, and expense, it remains the closest approximation to a gold standard when determining the merits of scientific endeavor.

Publication bias is the higher likelihood for journal editors to publish positive or “new” findings in contrast to negative studies or those that do not yield statistically significant results. Studies have shown that positive findings tend to get published more often and more quickly.⁴ There are numerous possible reasons for publication bias, including researchers’ tendency to submit positive rather than negative studies, peer reviewers who are more likely to recommend publication of positive studies, and journal editors who favor publication of positive studies.⁵ The net effect of publication bias may be an overrepresentation of false positive findings in the literature. It is also important to be aware of potential publication bias when reading or conducting meta-analyses that rely solely on the published literature and do not seek out unpublished studies. When a sufficient number of studies is available, funnel plots may be an effective method by which to determine whether publication bias is present in a particular body of evidence.^{5, 6} Figure 2–1 presents hypothetical data showing the effects of publication bias. In the plot on the right-hand side, smaller studies are represented in the literature only when they tend to show a positive effect. Thus, the left side of the inverted funnel is missing, and publication bias may be present.

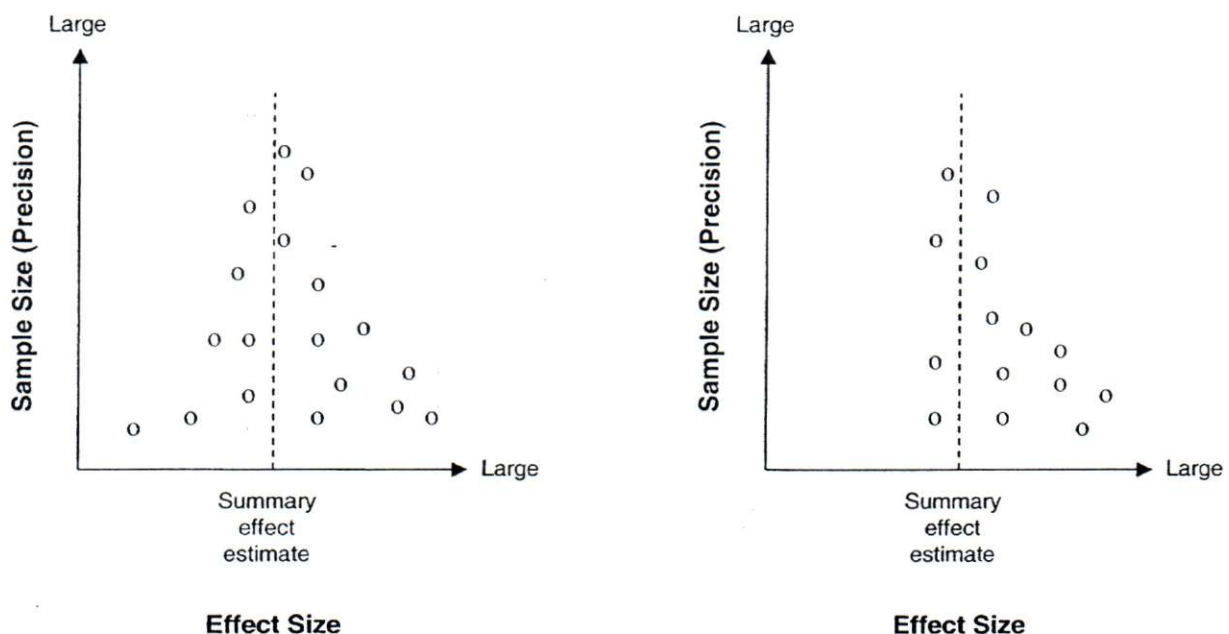


FIGURE 2–1. Hypothetical funnel plots illustrating the effect of publication bias.

STUDY DESIGN ISSUES: QUALITY AND GENERALIZABILITY

Public health information for decision making is founded upon science, and science is based on the collection, analysis, and interpretation of data.⁷ Data in public health are generally derived from two overlapping sources: research studies and public health surveillance systems. Here, we focus on information from research studies; an emphasis on public health surveillance is provided in Chapter 3. Research studies are primarily conducted in four broad areas: (1) to understand the etiologies of health conditions (Does fruit and vegetable intake influence the risk of coronary heart disease?); (2) to determine whether public health interventions are successful in meeting their stated objectives for risk reduction (Is a media campaign to increase fruit and vegetable intake effective?); (3) to explore the links between interventions and etiology (Does an intervention that increases consumption of fruits and vegetable result in a decline in coronary heart disease incidence?); and (4) to develop and test new research methods (What are the most valid and reliable methods by which to measure fruit and vegetable consumption?).

The quality of the evidence from a given study can be assessed based on the study design, execution of the study (internal validity), and generalizability (external validity). In public health research, a variety of study designs is used to assess health risks and to measure intervention effectiveness. Commonly, these are not “true” experiments in which study participants are randomized to an intervention or control condition. These generally quasi-experimental or observational designs are described in Chapter 5. A hierarchy of designs shows that a randomized trial tends to be the strongest type of study, yet such a study is often not feasible in community settings (Table 2–1).⁸ Interestingly, when summary results from the same topic were based on observational studies and on randomized controlled trials, the findings across study designs were remarkably similar.⁹

The quality of a study’s execution can be determined by many different standards. Individual studies are often judged on the basis of their internal validity. While it is beyond the scope of this chapter to discuss these issues in detail, an overview of key issues is provided, along with entry points into a larger body of literature. For a study or program evaluation to be internally valid, the study and comparison groups should be selected and compared in a way that the observed differences in dependent variables are attributed to the hypothesized effects under study (apart from sampling error).³ In other words, can the observed results be attributed to the risk factor being studied or intervention being implemented? These concepts are illustrated in Figure 2–2. In general, internal validity is threatened by all types of systematic error, and error rates are influenced by both study design and study execution. Systematic error occurs when

Table 2–1. Hierarchy of Study Designs

<i>Suitability</i>	<i>Examples</i>	<i>Attributes</i>
Greatest	Randomized group or individual trial; prospective cohort study; time series study with comparison group	Concurrent comparison groups and prospective measurement of exposure and outcome
Moderate	Case-control study; time series study without comparison group	All retrospective designs or multiple pre- or postmeasurements but no concurrent comparison group
Least	Cross-sectional study; case series; ecological study	Before-after studies with no comparison group or exposure and outcome measured in a single group at the same point in time

Source: Adapted from Briss et al., 2000.⁸

there is a tendency within a particular study to produce results that vary in a systematic way from the true values.¹⁰ Dozens of specific types of bias have been identified. Among the most important are

1. Selection bias: error due to systematic differences in characteristics between those who take part in the study and those who do not³
2. Information bias: a flaw in measuring exposure or outcomes that results in different quality (accuracy) of information between study groups³
3. Confounding bias: distortion of the estimated effect of an exposure on an outcome, caused by the presence of an extraneous factor associated with both the exposure and the outcome³

In *this* study:

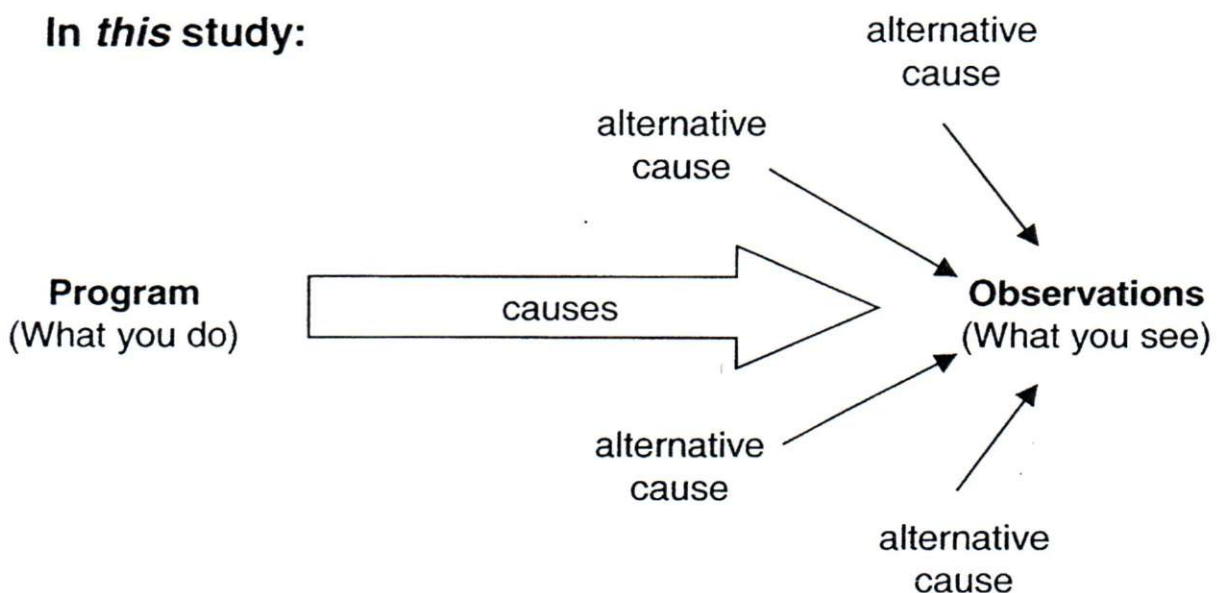


FIGURE 2–2. Illustration of internal validity in establishing a cause and effect relationship (Source: <<http://trochim.human.cornell.edu/kb/>>).

In ongoing work of the U.S. Public Health Service,^{11, 12} study execution is assessed according to six categories, each of which may threaten internal validity: (1) study population and intervention descriptions; (2) sampling; (3) exposure and outcome measurement; (4) data analysis; (5) interpretation of results (including follow-up, bias, and confounding); and (6) other related factors.⁸

External validity, synonymous with generalizability, relates to whether a study can produce unbiased inferences to other populations beyond the subjects in the study.³ Internal validity of a study is necessary for, but does not guarantee, external validity.¹⁰ For example, are the findings from a study of middle-aged white males applicable to ethnically diverse women? In practice, multicenter studies among diverse populations, using similar methods, are sometimes conducted to enhance the likelihood of attaining external validity.

ASSESSING CAUSALITY IN PUBLIC HEALTH RESEARCH

Any intervention program or public health action is based on the presumption that the associations found in epidemiologic studies are causal rather than arising through bias or for some other spurious reason.¹ A cause of a disease is an event, condition, characteristic, or combination of factors that plays an important role in the development of the disease or health condition.¹⁰ Unfortunately, in most instances in observational research, there is no opportunity to prove absolutely that an association is causal. Nonetheless, numerous frameworks have been developed that are useful in determining whether a cause-and-effect relationship exists between a particular risk factor and a given health outcome. This is one of the reasons for assembling experts to reach scientific consensus on various issues (discussed in Chapter 3).

Criteria for Assessing Causality

The earliest guidelines for assessing causality for infectious diseases were developed in the 1800s by Jacob Henle and Robert Koch. The Henle-Koch Postulates state that: (1) the agent must be shown to be present in every case of the disease by isolation in pure culture; (2) the agent must not be found in cases of other disease; (3) once isolated, the agent must be capable of reproducing the disease in experimental animals; and (4) the agent must be recovered from the experimental disease produced.^{3, 13} These postulates have proven less useful in evaluating causality for more contemporary health conditions since most non-infectious diseases have long periods of induction and multifactorial causation.¹⁴

Subsequently, the U.S. Surgeon General,¹⁵ Hill,¹⁶ Susser,¹⁷ and Rothman¹⁸ have all provided insights into causal criteria, particularly in regard to causation

In ongoing work of the U.S. Public Health Service,^{11, 12} study execution is assessed according to six categories, each of which may threaten internal validity: (1) study population and intervention descriptions; (2) sampling; (3) exposure and outcome measurement; (4) data analysis; (5) interpretation of results (including follow-up, bias, and confounding); and (6) other related factors.⁸

External validity, synonymous with generalizability, relates to whether a study can produce unbiased inferences to other populations beyond the subjects in the study.³ Internal validity of a study is necessary for, but does not guarantee, external validity.¹⁰ For example, are the findings from a study of middle-aged white males applicable to ethnically diverse women? In practice, multicenter studies among diverse populations, using similar methods, are sometimes conducted to enhance the likelihood of attaining external validity.

ASSESSING CAUSALITY IN PUBLIC HEALTH RESEARCH

Any intervention program or public health action is based on the presumption that the associations found in epidemiologic studies are causal rather than arising through bias or for some other spurious reason.¹ A cause of a disease is an event, condition, characteristic, or combination of factors that plays an important role in the development of the disease or health condition.¹⁰ Unfortunately, in most instances in observational research, there is no opportunity to prove absolutely that an association is causal. Nonetheless, numerous frameworks have been developed that are useful in determining whether a cause-and-effect relationship exists between a particular risk factor and a given health outcome. This is one of the reasons for assembling experts to reach scientific consensus on various issues (discussed in Chapter 3).

Criteria for Assessing Causality

The earliest guidelines for assessing causality for infectious diseases were developed in the 1800s by Jacob Henle and Robert Koch. The Henle-Koch Postulates state that: (1) the agent must be shown to be present in every case of the disease by isolation in pure culture; (2) the agent must not be found in cases of other disease; (3) once isolated, the agent must be capable of reproducing the disease in experimental animals; and (4) the agent must be recovered from the experimental disease produced.^{3, 13} These postulates have proven less useful in evaluating causality for more contemporary health conditions since most non-infectious diseases have long periods of induction and multifactorial causation.¹⁴

Subsequently, the U.S. Surgeon General,¹⁵ Hill,¹⁶ Susser,¹⁷ and Rothman¹⁸ have all provided insights into causal criteria, particularly in regard to causation

of chronic diseases such as heart disease, cancer, and arthritis. Although criteria have sometimes been cited as checklists for assessing causality, they were intended as factors to consider when examining an association: they have value, but only as general guidelines. Several criteria relate to particular cases of refuting biases or drawing on nonepidemiologic evidence. These criteria have been discussed in detail elsewhere.^{15-17, 19} In the end, belief in causality is based on an individual's judgment, and different individuals may in good faith reach different conclusions from the same available information. The six key issues below have been adapted from Hill¹⁶ and Weed.²⁰ Each is described by a definition and a rule of evidence. These are also illustrated in Table 2-2 by examining two risk-factor/disease relationships.

1. *Consistency*

Definition: The association is observed in studies in different settings and populations, using various methods.

Rule of evidence: The likelihood of a causal association increases as the proportion of studies with similar (positive) results increases.

2. *Strength*

Definition: This is defined by the size of the relative risk estimate. In some situations, meta-analytic techniques are used to provide an overall, summary risk estimate.

Rule of evidence: The likelihood of a causal association increases as the summary relative risk estimate increases. Larger effect estimates are generally less likely to be explained by unmeasured bias or confounding.

3. *Temporality*

Definition: This is perhaps the most important criterion for causality—some consider it an absolute condition. Temporality refers to the temporal relationship between the occurrence of the risk factor and the occurrence of the disease or health condition.

Rule of evidence: The exposure (risk factor) must precede the disease.

4. *Dose-response relationship*

Definition: The observed relationship between the dose of the exposure and the magnitude of the relative risk estimate.

Rule of evidence: An increasing level of exposure (in intensity and/or time) increases the risk when hypothesized to do so.

5. *Biological plausibility*

Definition: The available knowledge on the biological mechanism of action for the studied risk factor and disease outcome.

Rule of evidence: There is not a standard rule of thumb except that the more likely the agent is biologically capable of influencing the disease, then the more probable that a causal relationship exists.

6. *Experimental evidence*

Definition: The presence of findings from a prevention trial in which the factor of interest is removed from randomly assigned individuals.

Rule of evidence: A positive result (i.e., reduction in a health condition) after removal of the risk factor is strong evidence that the factor is causal.

In practice, evidence for causality is often established through the elimination of noncausal explanations for an observed association. For example, consider the evidence that alcohol use may increase the risk of breast cancer.¹ A series of further studies might confirm that this relationship is internally valid and not a result of confounding or other biases. By whittling away alternative explanations, the hypothesis that asserts alcohol use causes breast cancer becomes increasingly credible. It is the job of researchers to propose and test noncausal explanations, so that when the association has withstood a series of such challenges, the case for causality is strengthened.

Since most associations involve unknown confounders, a key issue becomes the extent to which causal conclusions or public health recommendations should be delayed until all or nearly all potential confounders are discovered and/or better measured.²¹ As noted earlier, those who argue that causality must be established with absolute certainty before interventions are attempted may fail to appreciate that their two alternatives—action and inaction—each have risks and benefits. When searching for causal relationships, researchers generally seek those that are modifiable and potentially amenable to some type of public health intervention. If researchers discovered that time of initiation of teen smoking was strongly related to the ethnicity of the teen and exposure to advertising, for example, the latter variable would be a likely target of intervention efforts.

RELATED ISSUES WHEN CONSIDERING PUBLIC HEALTH ACTION

In addition to understanding scientific causality and methods for accessing the published literature, several related issues are important to consider when weighing public health action.

Factors Influencing Decision Making in Public Health

There are many factors that influence decision making in public health (Table 2–3).²² Some of these factors are under the control of the public health practitioner, whereas others are nearly impossible to modify. A group of experts may

Table 2–2. Degree to Which Causal Criteria Are Met for Two Contemporary Public Health Issues

<i>Issue</i>	<i>Physical Activity and Coronary Heart Disease (CHD)</i>	<i>Extremely Low Frequency Electromagnetic Fields (EMFs) and Childhood Cancer^a</i>
Consistency	Approximately 50 studies since 1953; vast majority of studies show positive association	Based on a relatively small number of studies, the preponderance of the evidence favors a judgment of no association
Strength	Median relative risk of 1.9 for a sedentary lifestyle across studies, after controlling for other risk factors	Early studies showed relative risks in the range of 1.5 to 2.5. Most subsequent studies with larger sample sizes and more comprehensive exposure methods have not shown positive associations
Temporality	Satisfied, based on prospective cohort study design	Not satisfied; very difficult to assess because of ubiquitous exposure and the rarity of the disease
Dose–response relationship	Most studies show an inverse relationship between physical activity and risk of CHD	Since there is little biological guidance into what component(s) of EMF exposure may be problematic, exposure assessment is subject to a high degree of misclassification. True dose gradients are therefore very hard to classify reliably
Biological plausibility	Biological mechanisms are demonstrated, including: atherosclerosis, plasma/lipid changes, blood pressure, ischemia, and thrombosis	No direct cancer mechanism is yet known, as EMFs produce energy levels far too low to cause DNA damage or chemical reactions
Experimental evidence	Trials have not been conducted related to CHD but have been carried out for CHD intermediate factors such as blood pressure, lipoprotein profile, insulin sensitivity, and body fat	Numerous experimental studies of EMF exposure have been conducted to assess indirect mechanisms for carcinogenesis in animals and via <i>in vitro</i> cell models. The few positive findings to date have not been successfully reproduced in other laboratories

^aPredominantly childhood leukemia and brain cancer.

Table 2–3. Factors Influencing Decision Making among Public Health Administrators, Policy Makers, and the General Public

<i>Category</i>	<i>Influential Factor</i>
Information	<ul style="list-style-type: none"> • Sound scientific basis, including knowledge of causality • Source (e.g., professional organization, government, mass media, friends)
Clarity of contents	<ul style="list-style-type: none"> • Formatting and framing • Perceived validity • Perceived relevance • Strength of the message (i.e., vividness)
Perceived values, preferences, beliefs	<ul style="list-style-type: none"> • Role of the decision maker • Economic background • Previous education • Personal experience or involvement • Political affiliation or political climate • Willingness to adopt innovations • Willingness to accept uncertainty • Willingness to accept risk • Ethical aspect of the decision
Context	<ul style="list-style-type: none"> • Culture • Lobbying • Timing • Media attention • Administrative, financial, or political constraints

Source: adapted from Bero et al., 1998.²²

systematically assemble and present a persuasive body of scientific evidence such as recommendations for clinical or community-based interventions,^{11, 12, 16, 23} but even when they convene in a rational and evidence-based manner, the process is imperfect, participants may disagree, and events may become politically charged, as noted in Table 2–3 and Box 2–2.²⁴ In addition, one may have little control over the timing of some major public health event (e.g., prostate cancer diagnosis in an elected leader) that may have a large impact on the awareness and behaviors of the general public and policy makers.

Assessing Population Burden

As noted earlier, many factors enter into decisions about public health interventions, including certainty of causality, validity, relevance, economics, and political climate (Table 2–3). Measures of burden may also contribute substantially to science-based decision making. The burden of infectious diseases, such as measles, has been primarily assessed through incidence, measured in case numbers or rates. For chronic or noninfectious diseases like cancer, burden can be measured in terms of morbidity, mortality, and disability. The choice of measure should depend on the characteristics of the condition being examined. For ex-

Box 2–2. Establishing Breast Cancer Screening Guidelines

Breast cancer screening guidance for women ages 40 to 49 years has been the subject of considerable debate and controversy. Breast cancer is the most common cancer type among U.S. women, accounting for 182,800 new cases and 40,800 annual deaths.³⁸ It is suggested that appropriate use of screening mammography may lower death rates due to breast cancer up to 30%. Official expert guidance from the U.S. government was first issued in 1977 when the National Cancer Institute (NCI) recommended annual mammography screening for women ages 50 and older but discouraged screening for younger women.³⁹ In 1980, the American Cancer Society dissented from this guidance and recommended a baseline mammogram for women at age 35 years and annual or biannual mammograms for women in their 40s.⁴⁰ The NCI and other professional organizations differed on recommendations for women in their 40s throughout the late 1980s and 1990s. To resolve disagreement, the director of the National Institutes of Health called for a Consensus Development Conference in January 1997. Based on evidence from randomized, controlled trials, the consensus group concluded that the available data did not support a blanket mammography recommendation for women in their 40s. The panel issued a draft statement that largely left the decision regarding screening up to the woman (Table 2–4).⁴¹ This guidance led to widespread media attention and controversy. Within one week, the U.S. Senate passed a 98 to 0 vote resolution calling on the NCI to express unequivocal support for screening women in their 40s, and within 60 days, the NCI had issued a new recommendation.

ample, mortality rates are useful in reporting data on a fatal condition such as lung cancer. For a common, yet generally nonfatal condition such as arthritis, a measure of disability would be more useful (e.g., limitations in “activities of daily living”). When available, measures of the population burden of health conditions are extremely useful, e.g., quality-adjusted life years (QALYs) (see Chapter 3).

When assessing the scientific basis for a public health program or policy, quantitative considerations of preventable disease can help us make a rational choice. This can be thought of as “preventable burden.” When presented with an array of potential causal factors for disease, we need to evaluate how much might be gained by reducing or eliminating each of the hazards. For example, can we predict in numerical terms the benefits that one or more interventions might yield in the community?

Epidemiologic measures, such as relative risk estimates indicate how strongly exposure and disease are associated, but they do not indicate directly the benefits that could be gained through modifying the exposure. Of still greater potential value is the incorporation of information on how common the exposure is. Although some exposures exert a powerful influence on individuals (i.e., a large relative risk), they are so rare that their public health impact is minimal. Conversely, some exposures have a modest impact but are so widespread that their elimination could have great benefit. To answer the question, “What proportion

Table 2-4. Chronology and Selected Statements from the Development of Consensus Breast Cancer Screening Guidelines for Women Aged 40 to 49 Years, 1997

<i>Date</i>	<i>Source</i>	<i>Statement or Quote</i>
January 23, 1997	NIH Consensus Development Panel (called for and co-sponsored by the National Cancer Institute)	Every woman should decide for herself "based not only on objective analysis of scientific evidence and consideration of her individual medical history, but also on how she perceives and weighs each potential risk and benefits, the values she places on each and how she deals with uncertainty."
January 24, 1997	American Cancer Society	"The confusion surrounding the important question of whether women in their 40s should have regular mammograms had not been cleared up and perhaps was made even murkier by the recent announcement."
February 4, 1997	U.S. Senator Mikulski	"I could not believe when an NIH advisory panel decided that women in this age group might not need mammograms. This flies in the face of what we know."
February 4, 1997	U.S. Senator Snowe	"Women and their doctors look to the Nation's preeminent cancer research institute—the National Cancer Institute—for clear guidance and advice on this issue. . . . By rescinding its guideline, NCI produced widespread confusion and concern among women and physicians regarding the appropriate age at which to seek mammograms."
February 4, 1997	US Senate Resolution 47	"... we say enough is enough. We should take time out, go back to our science, go back to our research, go back to the National Institutes of Health and ask them to come up with the recommendation that we need."
March 27, 1997	National Cancer Institute	"The NCI advises women age 40–49 who are of average risk of breast cancer to have screening mammograms every year or two."

of disease in the total population is a result of the exposure?" the *population attributable risk* (PAR) is used. The PAR is calculated as follows:

$$\frac{P_e (\text{relative risk}_a - 1)}{1 + P_e (\text{relative risk}_a - 1)}$$

where P_e represents the proportion of the population that is exposed. Assuming that the relative risk_a of lung cancer due to cigarette smoking is 15 (i.e., smokers have 15 times the risk of lung cancer compared with nonsmokers) and that 30% of the population are smokers, the population attributable risk is 0.81 or 81%. This would suggest that 81% of the lung cancer burden in the population is

Table 2–5. Modifiable Risk Factors for Coronary Heart Disease, United States

<i>Magnitude</i>	<i>Risk Factor</i>	<i>Best Estimate (%) of Population Attributable Risk (Range)</i>
Strong (relative risk >4)	None	—
Moderate (relative risk 2–4)	High blood pressure ($\geq 140/90$ mm Hg)	25 (20–29)
	Cigarette smoking	22 (17–25)
	Elevated cholesterol (≥ 200 mg/dL)	43 (39–47)
	Diabetes (fasting glucose ≥ 140 mg/dL)	8 (1–15)
Weak (relative risk <2)	Obesity ^a	17 (7–32)
	Physical inactivity	35 (23–46)
	Environmental tobacco smoke exposure	18 (8–23)
Possible	Psychological factors	—
	Alcohol use ^b	—
	Elevated plasma homocysteine	—
	Infectious agents	—

Source: From Newschaffer et al.²⁵

^aBased on body mass index >27.8 kg/m² for men and >27.3 kg/m² for women.

^bModerate to heavy alcohol use may increase risk, whereas light use may reduce risk.

caused by cigarette smoking and could be eliminated if the exposure were eliminated. Table 2–5 describes a variety of risk factors for coronary heart disease.²⁵ This list demonstrates that the greatest population burden (PAR) would be affected by eliminating elevated cholesterol and physical inactivity, even though the relative risk values for these risk factors are in the moderate or weak range.

A related metric is the prevented fraction (PF). In an intervention in which “exposure” to a program or policy may protect against disease, the PF is the proportion of disease occurrence in a population averted due to a protective risk factor or public health intervention.²⁶ The PF is calculated as follows:

$$P_e (\text{relative risk}_b - 1)$$

where P_e represents the prevalence of exposure to the protective factor and relative risk_b is a protective effect estimate (i.e., exposure to the preventive measure protects against acquiring a specific health problem). This formula for the PF is the same one used to calculate vaccine efficacy and has also been used to estimate the benefits of disease screening programs.²⁷

Assessing Time Trends

There are numerous other factors that may be considered when weighing the need for public health action. One important factor to consider involves temporal

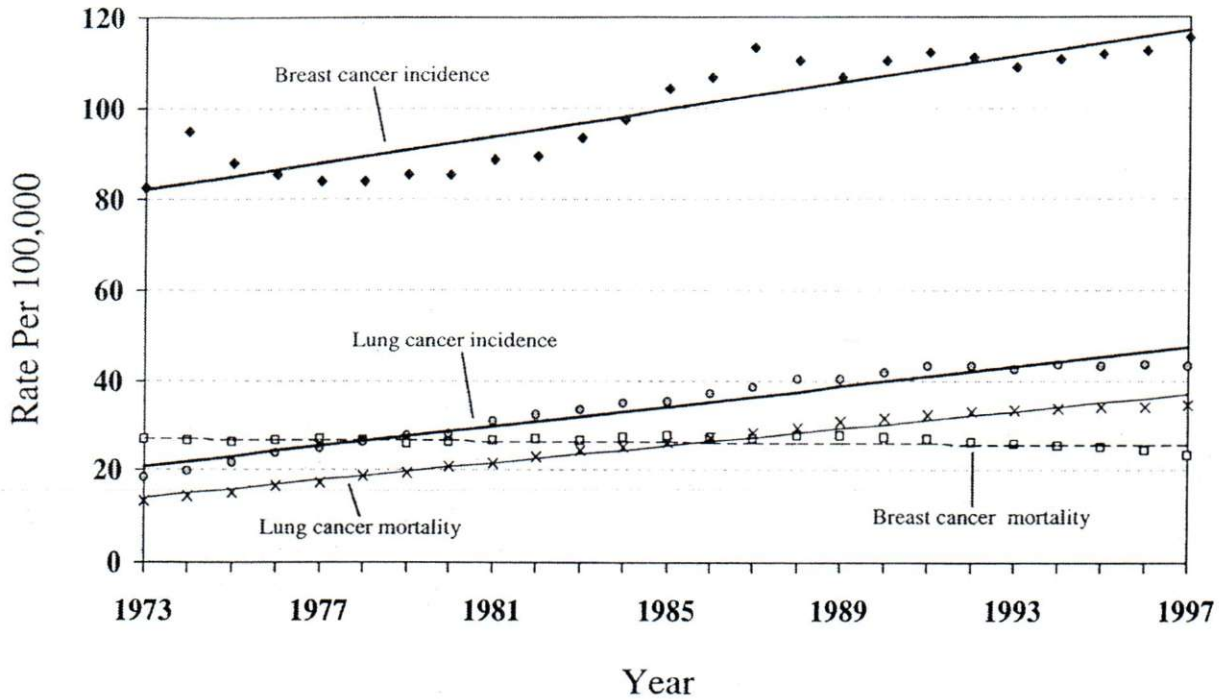


FIGURE 2–3. Trends in incidence and mortality for lung and breast cancer in women, United States, 1973–1997 (Source: Ries et al., 2000²⁸).

trends (see chapter 5). Public health surveillance systems can provide information on changes over time in a risk factor or disease of interest. Through use of these data, one may determine whether the condition of interest is increasing, decreasing, or remaining constant. One may also examine the incidence or prevalence of a condition in relation to other conditions of interest. For example, if a public health practitioner were working with a statewide coalition to control cancer, it would be useful to plot both the incidence and mortality rates for various cancer sites (Figure 2–3).²⁸ The researcher might reach different conclusions on the impact and changing magnitude of various cancers when examining incidence versus mortality rates across the state. When working at a local level, however, it would be important to note that sample sizes might be too small for many health conditions, making rates unstable and subject to considerable fluctuations over time. In addition, a formal time-series analysis requires numerous data points (approximately 50 for the most sophisticated statistical methods). A simple and often useful time-series analysis can often be conducted with ordinary least-squares regression techniques, which are amenable to fewer data points than formal time-series analyses.

Priority Setting via National Health Goals

Determining public health and health care priorities in a climate of limited resources is a demanding task. In some cases, priority setting from experts and

governmental bodies can help to focus areas for public health action. These efforts are particularly useful in evaluating Type I evidence (i.e., something must be done on a particular health topic). They are often less helpful for Type II evidence (i.e., this specific intervention should be conducted within a local area).

Public health leaders began to formulate concrete public health objectives as a basis for action during the post–World War II era. This was a clear shift from earlier efforts as emphasis was placed on quantifiable objectives and explicit time limits.²⁹ A few key examples illustrate the use of public data in setting and measuring progress toward health objectives. A paper by the Institute of Medicine³⁰ sparked a U.S. movement to set objectives for public health.²⁹ These initial actions by the Institute of Medicine led to the 1979 Surgeon General’s Report on Health Promotion and Disease Prevention, which set five national goals—one each for the principal life stages of infancy, childhood, adolescence and young adulthood, adulthood, and older adulthood.³¹

More recently, the U.S. Public Health Service established two overarching health goals for the year 2010: (1) increase quality and years of healthy life; and (2) eliminate health disparities.³² To achieve these two goals, a comprehensive set of objectives was established in twenty-eight focus areas. In support of these efforts, *Healthy People 2010* has also identified ten leading health indicators that will be used to set priorities and to measure the health of the United States over the next ten years (Table 2–6). The leading health indicators were selected on the basis of their ability to motivate action, the availability of data to measure progress, and their importance as public health issues.³² In its earlier version *Healthy People 2000*), progress toward meeting objectives was measured in annual reports. Establishment of national, quantifiable objectives has stimulated state and local efforts in program and organizational planning. For example, an estimated 70% of all U.S. local health agencies (from a total of about 3,000) have used *Healthy People 2000* objectives.³³

Table 2–6. Leading Health Indicators from *Healthy People 2010*

-
1. Physical Activity
 2. Overweight and Obesity
 3. Tobacco Use
 4. Substance Abuse
 5. Responsible Sexual Behavior
 6. Mental Health
 7. Injury and Violence
 8. Environmental Quality
 9. Immunization
 10. Access to Health Care
-

SUMMARY

The issues covered in this chapter highlight one of the continuing challenges for public health practitioners and policy makers—determining when scientific evidence is sufficient for public health action. In nearly all instances, scientific studies cannot demonstrate causality with absolute certainty.^{16, 34} The demarcation between action and inaction is seldom distinct and requires careful consideration of scientific evidence as well as assessment of values, preferences, costs, and benefits of various options. The difficulty in determining scientific certainty was eloquently summarized by A.B. Hill:¹⁶

All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

In many instances, waiting for absolute scientific certainty would mean delaying crucial public health action. For example, gaining a full understanding of the molecular biology of HIV/AIDS transmission prior to implementing population-based programs would have delayed important advances in prevention. Strong skills are crucial in understanding causality, interpreting the ever-expanding scientific literature, and seeking out summaries of evidence.

Other key points in this chapter:

- When considering public health action, it is helpful to consider the consequences of taking action or no action.
- Advances in epidemiology and public health research are generally incremental, suggesting the need for intervention as a body of literature accumulates.
- When evaluating literature, both the quality of component studies and their generalizability (external validity) should be considered.
- A set of standardized criteria can be useful in assessing the causality of an association.
- Many factors beyond science, such as resource constraints, sources of information, timing, and politics, influence decision-making in public health.

SUGGESTED READINGS AND WEBSITES

Readings

Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services—methods. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine* 2000;18(1 Suppl):35–43.

- Rothman KJ. Causes. *American Journal of Epidemiology* 1976;104:587–592.
- Savitz DA, Harris RP, Brownson RC. Methods in chronic disease epidemiology. In: Brownson RC, Remington PL, Davis JR, eds. *Chronic Disease Epidemiology and Control*. 2nd ed. ed. Washington, DC: American Public Health Association, 1998, pp. 27–54.
- Weed DL. On the use of causal criteria. *International Journal of Epidemiology* 1997; 26(6):1137–1141.

Selected Websites

Partners in Information Access for Public Health Professionals <<http://www.nlm.nih.gov/partners/>>. A collaborative project to provide public health professionals with timely, convenient access to information resources to help them improve the health of the American public.

The Research Methods Knowledge Base <<http://trochim.human.cornell.edu/kb/>>. The Research Methods Knowledge Base is a comprehensive web-based textbook that addresses all of the topics in a typical introductory undergraduate or graduate course in social research methods. It covers the entire research process including: formulating research questions; sampling (probability and nonprobability); measurement (surveys, scaling, qualitative, unobtrusive); research design (experimental and quasi-experimental); data analysis; and, writing the research paper. It uses an informal, conversational style to engage both the newcomer and the more experienced student of research.

UCSF School of Medicine: Virtual Library in Epidemiology <<http://chanane.ucsf.edu/epidem/>> Large listing of websites in epidemiology and related fields is provided. Among the categories are governmental agencies, quantitative epidemiology, data sources, publications, and university sites.

REFERENCES

1. Savitz DA, Harris RP, Brownson RC. Methods in chronic disease epidemiology. In: Brownson RC, Remington PL and Davis JR, eds. *Chronic Disease Epidemiology and Control*. 2nd ed. Washington, DC: American Public Health Association, 1998, pp. 27–54.
2. U.S. Dept. of Health and Human Services. *Interim Report of the Long Island Breast Cancer Study Project*. Bethesda, MD: National Institutes of Health, June 2000.
3. Last JM, ed. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press, 2001.
4. Friis RH, Sellers TA. *Epidemiology for Public Health Practice*. 2nd ed. Gaithersburg, MD: Aspen Publishers, Inc., 1999.
5. Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. Chicago: American Medical Association Press, 2002.
6. Petitti DB. *Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. 2nd ed. New York: Oxford University Press, 2000.

7. Nelson DE. Assessing the science, locating information, and translating data. In: Nelson DE, Brownson RC, Remington PL, Parvanta C, eds. *Communicating Public Health Information Effectively: A Guide for Public Health Practitioners*. Washington, DC: American Public Health Association, (in press).
8. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services—methods. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine* 2000;18(1 Suppl):35–43.
9. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine* 2000;342:1887–1892.
10. Beaglehole R, Bonita R, Kjellstrom T. *Basic Epidemiology*. Geneva, Switzerland: World Health Organization, 1993.
11. Pappaioanou M, Evans C. Developing a guide to community preventive services: A U.S. Public Health Service initiative. *Journal of Public Health Management and Practice*. 1998;4:48–54.
12. Truman BI, Smith-Akin CK, Hinman AR, et al. Developing the guide to community preventive services—overview and rationale. *American Journal of Preventive Medicine* 2000;18(1S):18–26.
13. Rivers TM. Viruses and Koch's postulates. *Journal of Bacteriology* 1937;33:1–12.
14. McKenna MT, Taylor WR, Marks JS, Koplan JP. Current issues and challenges in chronic disease control. In: Brownson RC, Remington PL and Davis JR, eds. *Chronic Disease Epidemiology and Control*. 2nd ed. Washington, DC: American Public Health Association; 1998 pp. 1–26.
15. U.S. Dept. of Health, Education, and Welfare. *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Vol. Publication (PHS) 1103. Washington, DC: Center for Disease Control, 1964.
16. Hill AB. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 1965;58:295–300.
17. Susser M. *Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology*. New York: Oxford University Press, 1973.
18. Rothman KJ. Causes. *American Journal of Epidemiology* 1976;104:587–592.
19. Kelsey JL. *Methods in Observational Epidemiology*. 2nd ed. New York: Oxford University Press, 1996.
20. Weed DL. Epidemiologic evidence and causal inference. *Hematology/Oncology Clinics of North America* 2000;14(4):797–807.
21. Weed DL. On the use of causal criteria. *International Journal of Epidemiology* 1997;26(6):1137–1141.
22. Bero LA, Jadad AR. How consumers and policy makers can use systematic reviews for decision making. In: Mulrow C, Cook D, eds. *Systematic Reviews. Synthesis of Best Evidence for Health Care Decisions*. Philadelphia: American College of Physicians, 1998, pp. 45–54.
23. U.S. Preventive Services Task Force. *Guide to Clinical Preventive Services*. 2nd ed. Baltimore: Williams & Wilkins, 1996.
24. Ernster VL. Mammography screening for women aged 40 through 49—A guidelines saga and a clarion call for informed decision making. *American Journal of Public Health* 1997;87(7):1103–1106.
25. Newschaffer CJ, Brownson CA, Dusenbury LJ. Cardiovascular disease. In: Brownson RC, Remington PL and Davis JR, eds. *Chronic Disease Epidemiology and Con-*

- trol.* 2nd ed. Washington, DC: American Public Health Association, 1998, pp. 297–334.
26. Gargiullo PM, Rothenberg RB, Wilson HG. Confidence intervals, hypothesis tests, and sample sizes for the prevented fraction in cross-sectional studies. *Statistics in Medicine* 1995;14(1):51–72.
 27. Straatman H, Verbeek AL, Peeters PH. Etiologic and prevented fraction in case-control studies of screening. *Journal of Clinical Epidemiology*. 1988;41(8):807–811.
 28. Ries LAG, Eisner MP, Kosary CL, et al., eds. *SEER Cancer Statistics Review, 1973–1997*. Bethesda, MD: National Cancer Institute, 2000.
 29. Breslow L. The future of public health: Prospects in the United States for the 1990s. *Annual Review of Public Health* 1990;11:1–28.
 30. Nightingale EO, Cureton M, Kamar V, Trudeau MB. *Perspectives on Health Promotion and Disease Prevention in the United States*. [staff paper]. Washington, DC: Institute of Medicine, National Academy of Sciences, 1978.
 31. U.S. Dept. of Health Education, and Welfare. *Healthy People. The Surgeon General's Report on Health Promotion and Disease Prevention*. Washington, DC: U.S. Dept. of Health, Education, and Welfare, 1979. Publication no. 79–55071.
 32. U.S. Dept. of Health and Human Services. *Healthy People 2010. Volume II. Conference Edition*. Washington, DC: U.S. Dept. of Health and Human Services, 2000.
 33. National Association of County and City Health Officials. *1992–1993 National Profile of Local Health Departments. National Surveillance Series*. Washington, DC: National Association of County and City Health Officials, 1995.
 34. Susser M. Judgement and causal inference: Criteria in epidemiologic studies. *American Journal of Epidemiology* 1977;105:1–15.
 35. Centers for Disease Control. Toxic-shock syndrome—United States. *Morbidity and Mortality Weekly Report* 1980;29:229–230.
 36. Shands KN, Schmid GP, Dan BB, et al. Toxic-shock syndrome in menstruating women: Association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *New England Journal of Medicine* 1980;303:1436–1442.
 37. Schuchat A, Broome CV. Toxic shock syndrome and tampons *Epidemiologic Reviews*. 1991;13:99–112.
 38. American Cancer Society. *Cancer Facts & Figures 2000*. Atlanta: American Cancer Society, 2000. 00–300M-No. 5008.00.
 39. Breslow L, et al. Final Report of NCI Ad Hoc Working Groups on Mammography in Screening for Breast Cancer. *Journal of the National Cancer Institute* 1977;59(2):469–541.
 40. American Cancer Society. Report on the Cancer-Related Health Checkup. CA: *A Cancer Journal of Clinicians* 1980;30:193–196.
 41. National Institutes of Health Consensus Development Panel. National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40–49, January 21–23, 1997. *Journal of the National Cancer Institute* 1997;89:1015–1026.

Prevention Effectiveness

*A Guide to Decision Analysis and
Economic Evaluation*

SECOND EDITION

Edited by

ANNE C. HADDIX, Ph.D.

STEVEN M. TEUTSCH, M.D., M.P.H.

PHAEDRA S. CORSO, Ph.D.

OXFORD
UNIVERSITY PRESS

2003

1

Introduction

STEVEN M. TEUTSCH
JEFFREY R. HARRIS

Prevention-effectiveness studies assess the impact of public health policies, programs, and practices on health outcomes¹ by determining their effectiveness, safety, and cost. The roots of prevention effectiveness lie in the assessment of medical technology and in research on health services and outcomes and build on operations research, epidemiology, statistics, economics, psychology, and other decision sciences. The results of prevention-effectiveness studies provide a basis for recommendations regarding public health programs, guidelines for prevention and control, and decision making about resource allocations; they are at the core of evidence-based public health. Sound decisions require timely, high-quality, comparable, and appropriately directed information. Prevention-effectiveness studies can provide this information.

For many years, particularly in the arena of medical care, it was sufficient to show that the benefits of a technology exceed its harms before using it. Now, in a world of limited resources for public health, officials must use resources as efficiently as possible and must demonstrate that a technology delivers value for the resources expended. By the same token, policies and programs should also be scrutinized for the value they deliver.

Prevention-effectiveness studies help meet these goals. They provide a systematic approach to organizing the available information about prevention strategies so that policy makers can have a scientific framework for making decisions. The concept pulls together information from epidemiology and public health surveillance, intervention studies, and economic analyses, using direct evidence when available and indirect evidence when necessary. It addresses basic questions, such as the following:

- What is the magnitude of the problem addressed by the prevention strategy (burden of illness and injury, descriptive epidemiology, and public health surveillance)?
- Can the intervention work (efficacy)?
- Does the intervention work (effectiveness)?

- What are the benefits and harms of the intervention (net benefits)?
- What does the intervention cost (cost analysis)?
- How do the benefits compare with the costs (cost-effectiveness, cost-benefit, and cost-utility analyses)?
- What additional benefit could be obtained with additional resources (marginal and incremental analyses)?

CONCEPTUAL MODEL FOR THE DEVELOPMENT OF PREVENTION STRATEGIES

Public health strategies evolve from basic science and applied research through community demonstrations into widespread use (Fig. 1.1). The information available at each stage and the methods for analyzing and synthesizing that information differ.

Information on biological risk factors is derived from basic research. Bench and epidemiologic research identify risk factors and the magnitude of their impact, as well as the biological and social underpinnings of disease and injury. Understanding these risk factors focuses attention on initial targets for potential intervention programs. Once major risk factors are identified (e.g., hypercholesterolemia for myocardial infarction, use of seat belts for automobile-crash injuries, or social isolation for mental health), potential interventions can be developed. Applied research, such as randomized controlled trials, can be conducted to provide information on the efficacy of these interventions. Research on efficacy shows the degree to which intervention strategies can work under idealized conditions with carefully selected populations and, in many cases, optimal resources.

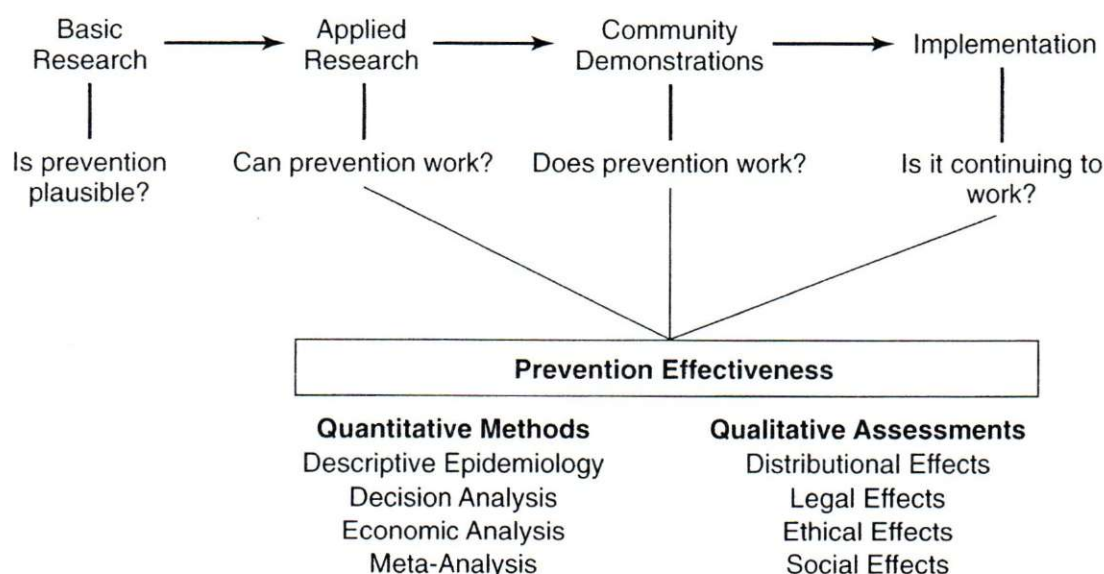


Figure 1.1 Development and implementation of prevention strategies and the role of prevention effectiveness. *Source:* Teutsch, SM. A framework for assessing the effectiveness of disease and injury prevention. *MMWR* 1992; 41 (No. RR-3).

After determining which strategies are efficacious, the next step is to determine how well these strategies actually work in community settings. Such community-based demonstrations are used to assess the real-world effectiveness of the prevention strategy. We define *effectiveness* as the impact that an intervention achieves in the real world under practical resource constraints in an entire unselected population or in specified subgroups of a population. It is axiomatic that effectiveness so defined will almost certainly be lower than efficacy because of resource constraints, individual adherence, and coverage of an intervention strategy, although there are some exceptions, such as herd immunity from immunizations.

Effectiveness research is outcome-oriented. Rather than focusing on the process of disease prevention and control (e.g., measuring how many people receive an intervention), prevention-effectiveness research seeks to directly link the intervention with the health outcome of interest (e.g., mortality, quality of life, or functional status). In this respect, the focus of prevention-effectiveness research differs from that of program evaluation. A prevention-effectiveness study would show, for example, how mortality is decreased by a particular intervention rather than how an intervention was administered.

As prevention strategies are implemented more widely, there is a growing need to maximize the intended impact with the resources available or to obtain a particular impact with as little expenditure of resources as possible. Thus, the efficiency of various approaches needs to be examined. There are always competing uses for resources, so the opportunity costs for each of our choices must be considered.

We enter the domain of prevention effectiveness as the results of applied research begin to demonstrate the efficacy of intervention strategies, including technologies, policies, and programs. The process continues throughout the development of the intervention into practical public health tools and their application in real settings (Fig. 1.1). Various methods are available for use at each stage of development and implementation.

TRADITIONAL APPROACHES

Attributable Risk and Prevented Fraction

In assessing an intervention strategy, one must know what it can realistically accomplish in terms of health outcomes. The first evidence usually comes from research on cause-and-effect relationships associated with health problems, in which the link between a risk factor and an outcome is identified, for example, the relationship between hypertension and coronary artery disease. The *relative risk* associated with the risk factor provides a traditional epidemiological measure of the potential impact. However, the overall public health impact is based not only on the relative risk but also on the frequency at which the condition occurs in the population. The impact is measured in terms of *attributable risk*, a measure of the amount of disease or injury that could be eliminated if the risk factor never occurred in a given population. It is the maximal limit of disease or injury that could be averted by avoiding a particular risk factor. In that sense, it is analogous to effi-

cacy (i.e., what could be achieved under ideal circumstances). In contrast, the *prevented fraction* is a measure of the amount of a health problem that has actually been avoided by a prevention strategy and reflects what can be achieved in a real-world setting (i.e., it is analogous to effectiveness).

Program Evaluation and Prevention Effectiveness

Program evaluation supports prevention-effectiveness research. *Program evaluation* assesses the structure, processes, and outcomes of intervention programs, with particular attention paid to the purposes and expectations of stakeholders.² Research on evaluation includes a complex array of experimental and quasi-experimental designs. These methods form the basis for determining the effectiveness and efficiency of prevention strategies by providing data on how programs are implemented and consumed.

ADDITIONAL METHODS FOR PREVENTION EFFECTIVENESS

Decision Models

Models are useful in conducting prevention-effectiveness studies, especially when evidence of effectiveness is indirect or uncertain. In some instances, a prevention strategy demonstrably improves a health outcome, and direct evidence of this is available. For example, mammography screening and follow-up for women over 50 years of age have been confirmed to reduce mortality from breast cancer. In many instances, however, such direct effects cannot be measured. When they cannot, we must rely on indirect evidence of the effectiveness of an intervention. For example, former smokers have lowered their risk of lung cancer, yet no studies have been conducted to show that a specific smoking-cessation strategy prevents lung cancer. A model would rely on indirect evidence that the smoking-cessation strategy decreases smoking and on the knowledge that cessation decreases the risk of developing lung cancer. Such indirect evidence can be used with confidence because each link in the chain of causality can be clearly documented.

Models can be very helpful in making assumptions explicit and in forcing examination of the logic, coherence, and evidence for each step in the process. The evidence for each step should be assessed using systematic methods and rules of evidence, such as systematic evidence reviews and, where needed, meta-analyses (see Chapters 3 and 10). Although assessments can also include literature reviews or the consensus of experts, such approaches may be subject to bias unless the rules of evidence are followed uniformly.

In addition to structuring effectiveness studies, models are useful in structuring other types of analysis. A basic decision analysis uses a decision-tree model to compare alternate strategies (see Chapter 7). Decision trees include information about the likelihood of each outcome (probabilities) and can incorporate preferences (utilities) or costs or both for different outcomes. Other approaches model more complex situations, such as infectious disease transmission. Markov models define sets of probabilities for transitions among health states for each

intervention to assess how outcomes are affected, Monte Carlo simulations provide a probabilistic approach, mathematical models incorporate complex mathematical relationships, and microsimulation attempts to replicate details of clinical processes.

In decision trees and other models, sensitivity analyses permit assessment of values for which there is uncertainty. Sensitivity analyses can identify critical steps that are likely to make a substantial difference in choosing one strategy over another.

Modeling may help to identify the important issues for which data are needed and thereby help to formulate a research agenda; it may also pinpoint issues for which more precise estimates will not affect a decision. Economic analyses are often based on such models. The use of models makes the decision process explicit and can help to clarify the criteria upon which decisions are based.

Economic Models

Resources are always constrained. Economics provides a range of tools to understand how resources are allocated as well as to help make choices about future allocation. The methods discussed in this volume are cost-benefit, cost-effectiveness, and cost-utility analyses. Each method allows comparison of different intervention strategies based on the resources they consume and the outputs they generate. Each requires a careful cost analysis (i.e., identification of costs associated with a prevention activity) and assessment of outcomes, both harms and benefits. The scope of an analysis usually determines the appropriate analytical method and range of consequences to consider. Cost-benefit analysis includes all costs, benefits, and harms and values them in dollars or another monetary unit. It includes costs of programs, costs to patients and others such as medical costs, direct out-of-pocket expenses, productivity and leisure losses, and intangible costs (e.g., grief, pain, suffering). It requires that the health outcomes be valued in monetary terms. Cost-benefit analysis is particularly suited to comparisons with interventions that include cross-sectoral considerations, for example, housing, education, or transportation interventions. Cost-effectiveness usually examines direct medical, non-medical, and productivity costs. It compares those costs with outcomes in standard health units, such as cost per case averted. It is most suitable when comparing interventions that have similar health outcomes. Cost-utility analysis compares direct medical and nonmedical costs with health outcomes converted to a standard health unit, often a quality-adjusted life year, which combines both morbidity and mortality. Because it provides a general health measure, cost-utility analysis is often used to compare health interventions which have different types of health outcome. Each of these types of analysis is covered in more detail in later chapters. Users should recognize that these analyses provide a great deal of information about interventions, how they can be targeted, modified, or made more efficient. Using them for this purpose can be particularly valuable in modifying program strategies. Our choices have opportunity costs: one choice inevitably means that we forego another. These techniques help us to understand the costs and consequences of our choices when allocating resources.

USES AND LIMITATIONS

Prevention-effectiveness studies are only tools in the decision-making process. They certainly do not make the decisions themselves. Decision processes should be based on solid technical information, such as that obtained from prevention-effectiveness studies; but quantitative information must be combined with an understanding of preferences and values of the stakeholders that are not intrinsically technical in nature. Judgments about those preferences, including acceptability, feasibility, and consistency with strategy, are central to effective decision-making processes. The information generated by these techniques can help to focus discussions among stakeholders and decision makers, allowing them to discern the trade-offs and consequences among alternative strategies, potentially redirecting energy toward a better understanding of the nature of the problem and identifying acceptable solutions.

The application of economic analysis in public health is still a relatively new and dynamic area. Although there is general agreement on many of the principles of economic analysis, controversy about their application in public health persists. Thus, issues such as choice of discount rates, valuation of life, and discounting of future benefits have not been fully resolved. Because public health economics is an evolving field, researchers are urged to consult economists or decision analysts when they begin to design a study. This collaboration can assure that acceptable methods and current and appropriate data sources are used.

Many of the basic decision techniques described in this book can be used at an individual level or a population level. A clinically relevant decision analysis can use individual patient preferences to help guide prevention and therapeutic choices. Similar types of analysis can inform policy choices as well.

SOCIAL, LEGAL, AND EQUITY ASPECTS

Although the decision-making aids described in this book can assist in the process of making policy decisions, they are still just aids. Other considerations must also be included in the decision-making process. Many prevention strategies have much broader effects than those directly related to the health outcomes at which they are aimed. This scope of effect is often not included explicitly in the decision-making process. The social impact of an intervention strategy is well illustrated by the many ramifications associated with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS). Intervention strategies implemented for AIDS have raised issues about civil rights among high-risk groups. Sex education and the distribution of condoms in public schools have raised a panoply of concerns reflecting conflicting social values involving students, parents, educators, and public health officials.

It is apparent that prevention strategies must be compatible with the law, but they do raise issues regarding regulations and have an influence on the legal system and precedents. Counseling and testing for HIV/AIDS, for instance, have raised important issues relating to the right to privacy (confidentiality). Many important

advances in public health have required changes in regulations (e.g., limitations on smoking in public places, the worksite, or commercial air travel).

Equity and distributional aspects must also be considered. In the context of limited resources, prevention programs can be focused in different ways. They can be directed intensively toward high-risk populations or, less intensively, toward an entire population. These alternate strategies may have very similar costs and benefits, yet the groups that benefit may differ substantially. Similarly, the group that benefits and the group that is harmed may be different, thus raising issues of equity. For example, fortification of flour with folic acid to prevent neural tube defects is accompanied by the risk of permanent neurological disease because diagnosis of vitamin B₁₂ deficiency may be delayed. The benefits accrue to infants and their families and the harms to a generally older population. Prevention-effectiveness studies can identify, but cannot resolve, these concerns.

TYPES OF PREVENTION STRATEGIES

Traditional biomedical studies on prevention focus on such clinical strategies to prevent disease and injury as surgical intervention and screening. Prevention-effectiveness studies, however, focus on prevention strategies that encompass the entire domain of public health practice. As a result, many elements of the outcomes and costs are measured differently in prevention-effectiveness studies compared to clinical prevention studies. In preventing lead poisoning, for example, costs related to a prevention strategy might include the nonmedical costs of removing lead paint from older buildings and replacing plumbing that contains lead, in addition to the medical costs of screening and treating persons with elevated blood-lead levels.

Prevention strategies often embody a variety of intervention approaches. In general, however, strategies can be classified as clinical, behavioral, environmental, or systemic. *Clinical* prevention strategies are those conveyed by a health-care provider to a patient, often within a clinical setting (e.g., vaccinations, screening and treatment for diabetic eye disease, and monitoring treatment for tuberculosis). *Behavioral* interventions require individual action, such as eating a healthful diet, exercising, stopping smoking, or wearing a bicycle helmet. They may employ a clinical or a population-based implementation strategy. *Environmental* strategies are those that society can impose and that may require little effort on the part of an individual. Examples of such strategies are laws that limit smoking in public places, dictate the removal of lead from gasoline, prescribe the addition of fluoride to public water supplies, and require the use of seat belts in motor vehicles. *Systemic* changes involve changing the fundamental community processes. For example, improved access to care may require basic changes in the health-care system and financing.

These four approaches should be distinguished from the traditional medical model of prevention based on three stages in disease and injury processes. *Primary prevention* targets risk factors to prevent occurrence of disease or injury. *Secondary prevention* targets subclinical disease through early identification and treatment.

Tertiary prevention is aimed at an established disease or injury to ameliorate progression and maximize function for the person affected.

In keeping with expectations for an introductory text, this book is not intended to be exhaustive, nor does it provide an in-depth theoretical basis for analytic methods. Similarly, this book does not provide an in-depth discussion of systematic literature reviews and synthesis. Such information is available elsewhere (see the bibliography at the end of each chapter). A guide to computer software is also included in Appendix C.

Although program evaluation supports prevention-effectiveness research, program-evaluation methods are not specifically addressed in this book, nor do we discuss in depth many social, legal, and distributional issues that are important parts of many policy decisions.

REFERENCES

1. Teutsch SM. A framework for assessing the effectiveness of disease and injury prevention. *MMWR Morb Mortal Wkly Rep* 1992; 41:i-iv, 1-12.
2. Milstein RL, Wetterhall SF. CDC. Framework for program evaluation in public health. *MMWR Morb Mortal Wkly Rep* 1999; 48.

BIBLIOGRAPHY

Prevention Effectiveness

- Bierman H Jr, Bonini CP, Hausman WH. *Quantitative Analysis for Business Decisions*, 8th ed. Homewood, IL: Irwin, 1991.
- Bunker JP, Barnes BA, Mosteller F, eds. *Costs, Risks, and Benefits of Surgery*. New York: Oxford University Press, 1977.
- Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*, 2d ed. Oxford: Oxford Medical Publications, 1997.
- Eddy DM, Hasselblad V, Shachter R. *Meta-analysis by the Confidence Profile Method*. Boston: Academic Press, 1992.
- Halperin W, Baker EL, Monson RR. *Public Health Surveillance*. Boston: Little, Brown, 1992.
- Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. Boston: Academic Press, 1985.
- National Research Council. *Valuing Health Risks, Costs, and Benefits for Environmental Decision Making: Report of a Conference*. Washington DC: National Academy Press, 1990.
- National Research Council Institute of Medicine. *Toward a National Health Care Survey: A Data System for the 21st Century*. Washington DC: National Academy Press, 1992.
- Petitti DB. *Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*, 2d ed. New York: Oxford University Press, 2000.
- Russell LB. *Is Prevention Better than Cure?* Washington DC: Brookings Institution, 1986.
- Russell LB. *Evaluating Preventive Care: Report on a Workshop*. Washington DC: Brookings Institution, 1987.
- Scriven M. *Evaluation Thesaurus*, 4th ed. Newbury Park, CA: Sage, 1991.

- Sox H, Blatt M, Higgins M, Marton K. *Medical Decision Making*. Stoneham, MA: Butterworth-Heinemann, 1988.
- Stokey E, Zeckhauser R. *A Primer for Policy Analysis*. New York: Norton, 1978.
- Teutsch SM. A framework for assessing the effectiveness of disease and injury prevention. *MMWR Morb Mortal Wkly Rep* 1992; 41:i-iv, 1-12.
- Teutsch SM, Churchill RE. *Principles and Practice of Public Health Surveillance*, 2d ed. New York: Oxford University Press, 2000.
- Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. Philadelphia: Saunders, 1980.
- Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press, 1986.

Economic Analysis

- Borus MEJ, Buntz CG, Tash WR. *Evaluating the Impact of Health Programs: Primer*. Cambridge: Cambridge University Press, 1982.
- Cohen DR, Henderson JB. *Health, Prevention and Economics*. Oxford: Oxford University Press, 1982.
- Drummond MF. Survey of cost-effectiveness and cost-benefit analyses in industrialized countries. *World Health Stat Q* 1985; 38:383-401.
- Drummond MF, Davies L. Economic analyses along side clinical trials. *Int J Technol Assess Health Care* 1991; 7:561-73.
- Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*, 2d ed. Oxford: Oxford Medical Publications, 1997.
- Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. *JAMA* 1989; 262:2879-86.
- Eisenberg JM, Koffer H, Finkler SA. Economic analysis of a new drug: potential savings in hospital operating costs from the use of a once-daily regimen of a parenteral cephalosporin. *Rev Infect Dis* 1984; 6:S909-23.
- Feldstein P. *Health Care Economics*, 3rd ed. New York: John Wiley and Sons, 1988.
- Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- Gramlich EM. *A Guide to Benefit-Cost Analysis*, 2d ed. Englewood Cliffs, NJ: Prentice-Hall, 1990.
- Harrington W, Krupnick AJ, Spofford WO. *Economics and Episodic Disease: The Benefits of Preventing a Giardiasis Outbreak*. Washington DC: Resources for the Future, 1991.
- Hartunian NS. The incidence and economic costs of cancer, motor vehicle injuries, coronary heart disease and stroke: a comparative analysis. *Am J Public Health* 1980; 70:1249-60.
- Luce BR, Elixhauser A. *Standards for Socioeconomic Evaluation of Health Care Products and Services*. New York: Springer-Verlag, 1990.
- Manning WG, et al. The taxes of sin: do smokers and drinkers pay their way? *JAMA* 1989; 261:1604-9.
- Mills A. Survey and examples of economic evaluation of health programmes in developing countries. *World Health Stat Q* 1985; 39:402-31.
- Mohr LB. *Impact Analysis for Program Evaluation*. Pacific Grove, CA: Dorsey, 1988.
- Petitti DB. *Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*, 2d ed. New York: Oxford University Press, 2000.
- Rice DP, Hodgson TA, Kopgtein AN. The economic costs of illness: a replication and update. *Health Care Financ Rev* 1985; 7:61-80.
- Richardson AW, Gafni A. Treatment of capital costs in evaluating health-care programs. *Cost Manage* 1983; 26-30.

- Robinson JC. Philosophical origins of the economic valuation of life. *Milbank Q* 1986; 64:133-55.
- Scitovsky A, Rice D. Estimates of the direct and indirect costs of acquired immunodeficiency syndrome in the U.S. 1985, 1986, 1987. *Public Health Rep* 1987; 102:5-17.
- Scitovsky AA. Estimating the direct costs of illness. *Milbank Memorial Fund Q* 1982; 60:463-91.
- Shepard DS, Thompson MS. First principles of cost-effectiveness analysis. *Public Health Rep* 1979; 94:535-43.
- Warner KE, Luce BR. *Cost-Benefit and Cost-Effectiveness Analysis in Health Care*. Ann Arbor, MI: Health Administration Press, 1982.
- Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; 236:716-21.



The “Million Hearts” Initiative - Preventing Heart Attacks and Strokes



Perspective

The “Million Hearts” Initiative — Preventing Heart Attacks and Strokes

Thomas R. Frieden, M.D., M.P.H., and Donald M. Berwick, M.D., M.P.P.

Each year, more than 2 million Americans have a heart attack or stroke, and more than 800,000 of them die; cardiovascular disease is the leading cause of death in the United States and the largest

cause of lower life expectancy among blacks. Related medical costs and productivity losses approach \$450 billion annually, and inflation-adjusted direct medical costs are projected to triple over the next two decades if present trends continue.¹

To reduce this burden, the Department of Health and Human Services (DHHS), other federal, state, and local government agencies, and a broad range of private-sector partners are today launching a “Million Hearts” initiative to prevent 1 million heart attacks and strokes over the next 5 years by implementing proven, effective, inexpensive interventions (see table).

Cardiovascular prevention works in two realms: the clinic and the community. Clinical and community interventions each contributed about equally to the 50% reduction in U.S. mortality due to heart attacks between 1980 and 2000.² If used consistently, proven interventions could prevent more than half of heart attacks and strokes. It’s time to take the next big step.

In the clinical realm, Million Hearts will improve management of the “ABCS” — aspirin for high-risk patients, blood-pressure control, cholesterol management, and smoking cessation. As for community-based prevention, the initiative will encourage efforts to

reduce smoking, improve nutrition, and reduce blood pressure. It will implement the cardiovascular-disease-prevention priorities of the National Quality and National Prevention Strategies and help in meeting targets set by Healthy People 2020.

Improving management of the ABCS can prevent more deaths than other clinical preventive services.³ Patients reduce their risk of heart attack or stroke by taking aspirin as appropriate. Treating high blood pressure and high cholesterol substantially and quickly reduces mortality among high-risk patients. Even brief smoking-cessation advice from clinicians doubles the likelihood of a successful quit attempt, and the use of medications increases quit rates further.

Currently, less than half of people with ischemic heart disease take daily aspirin or another

The Million Hearts Initiative: Principles and Examples of Interventions.*

Million Hearts Principles and Approaches	Effect on Quality of Care or Prevention	Sample Activities
Improving clinical management of aspirin use, blood-pressure control, cholesterol management, and smoking cessation (ABCS)		
Focus	Communication, clinical measurement, and reporting by physicians, health care facilities, and health care systems will emphasize improving ABCS care	<ul style="list-style-type: none"> Incorporation of simple, consistent ABCS indicators into the Physician Quality Reporting System, Medicare Part D and Medicare Advantage plan ratings, EHR meaningful use criteria, community clinic measures, and guidelines from private-sector organizations
Health information technology (HIT)	HIT enables providers and facilities to improve cardiovascular care and target intervention to patients in need of intensified care through registries and EHR functions used at the point of care	<ul style="list-style-type: none"> Meaningful use criteria include clinical quality measures for hypertension and cholesterol control. Meaningful use of EHRs can include routine assessment of cardiac risk; use of patient recall, reminders, decision support, order sets; and monitoring of medication adherence HIT Regional Extension Centers and Beacon Communities, which reach nearly 100,000 primary care providers, will support providers and institutions in using EHRs to improve ABCS management
Clinical innovations	Innovations such as team-based care, patient-centered medical homes, and interventions to promote adherence will be supported, evaluated, and disseminated rapidly to increase use of effective ABCS care practices	<ul style="list-style-type: none"> In 2011, a pharmacist-led campaign will provide materials and facilitate patient counseling about hypertension control The CDC and the Agency for Healthcare Research and Quality will identify and disseminate strategies that improve ABCS delivery
Expanding community initiatives to reduce smoking, improve nutrition, and reduce blood pressure		
Policies and programs designed to reduce tobacco use and exposure to second-hand smoke	Lowering exposure reduces heart attack risk among both smokers and nonsmokers	<ul style="list-style-type: none"> Beginning September 2012, the FDA will require large, prominent health warnings on all cigarette packaging and advertisements in the United States Grants to communities will address tobacco-use prevention and cessation Mass-media campaigns will aim to reduce smoking initiation and promote cessation
Policies for reducing sodium content of food	Facilitating healthier choices by consumers reduces risk of hypertension and cardiovascular disease	<ul style="list-style-type: none"> Menu-labeling requirements in chain restaurants will help people make more informed choices about what they eat The CDC will increase public and professional education regarding sodium The CDC's NHANES will begin collecting information on sodium consumption
Policies aimed at eliminating artificial trans fats from diet	Further reducing intake of artificial trans fat, which increases LDL cholesterol levels, decreases HDL cholesterol levels, and increases risk of heart attacks	<ul style="list-style-type: none"> The CDC and the FDA will work with industry to expand voluntary food reformulation initiatives The CDC will monitor trans fat levels through NHANES

* EHR denotes electronic health record, HDL high-density lipoprotein, LDL low-density lipoprotein, and NHANES National Health and Nutrition Examination Survey.

antiplatelet agent; less than half with hypertension have it adequately controlled; only a third with hyperlipidemia have adequate treatment; and less than a quarter of smokers who try to quit get counseling or medications. As a result, more than 100 million people — half of American adults — smoke or have uncontrolled high blood pressure

or cholesterol; many have more than one of these cardiovascular risk factors. Increasing utilization of these simple interventions could save more than 100,000 lives a year.³ Measuring and monitoring can encourage providers to improve preventive care.⁴

Improving care is particularly critical in light of increases in the prevalence of obesity and dia-

betes. Obesity and physical activity are currently being addressed by complementary efforts designed to improve understanding, implement pilot or community-based programs, and evaluate outcomes. The First Lady's "Let's Move" campaign is a comprehensive initiative with the goal of ending childhood obesity — a precursor to cardiovascular disease — with-

in a generation by fostering environments that support increased physical activity and improved nutrition for children and families. And public and private partners are working to expand the Diabetes Prevention Program, which promotes weight loss, improved nutrition, and increased physical activity among people at highest risk.

The Affordable Care Act (ACA) provides a strong foundation for Million Hearts by increasing coverage and facilitating improved care. It waives patient cost sharing for preventive services, including blood-pressure and cholesterol screening and smoking-cessation counseling and treatment, for enrollees in new private insurance plans. The new annual wellness visit for Medicare beneficiaries will help physicians focus on reducing cardiovascular risk and target interventions appropriately. Eliminating Medicare's "doughnut hole" in prescription-drug coverage will increase access to blood-pressure, cholesterol-lowering, and smoking-cessation medications. Covering 32 million currently uninsured Americans will reduce financial barriers to preventive care, and expanding community health centers will increase access to care and reduce health disparities. In addition, electronic health records (EHRs) will support improved clinical decision making.

Additional means of increasing control of the ABCS include reducing or eliminating copayments for medications, once-a-day dosing, team-based care approaches, stepwise care management, and new forms of payment and delivery for higher-quality, higher-value, and coordinated care, such as those envisioned for accountable care organizations.

Expanding use of prevention-oriented EHRs will enable providers and health systems to track and improve management of the ABCS. Incorporating core ABCS-related quality measures and decision-support tools into the 2013–2014 criteria for "meaningful use" of information technology and providing technical assistance through quality-improvement organizations in all states, the 62 Health Information Technology Regional Extension Centers (which reach nearly 100,000 primary care doctors), and Beacon Communities will reach more than 100 million patients within the next few years.

Million Hearts will work to standardize core ABCS indicators across medical practices, insurers, institutional providers, and systems in public and nonpublic settings. Standardization will facilitate public reporting and identification and diffusion of best practices and will reduce providers' burden by streamlining quality measurement and improvement. The initiative will be linked to quality-recognition programs (e.g., the Physician Quality Reporting System and star ratings for Medicare Part D and Medicare Advantage plans) and may eventually support approaches in which providers are paid more for better preventive care.

Community-based prevention works by facilitating healthy choices. Important community-based prevention initiatives include those funded by the American Recovery and Reinvestment Act's Communities Putting Prevention to Work program and programs supported by the ACA's Prevention and Public Health Fund, including Community Transformation Grants, initiatives for tobacco control and chronic-

disease prevention and control, many National Prevention Strategy initiatives, and state and local actions addressing tobacco use, nutrition, and the linkage between clinical and community-based prevention.

Reductions in smoking, sodium consumption, and trans fat consumption can substantially and rapidly improve cardiovascular health. Warning people about the harms of tobacco use through mass media and other measures, as well as package labeling as enabled by the Family Smoking Prevention and Tobacco Control Act, and creating smoke-free public places and workplaces, as detailed in the National Prevention Strategy and facilitated through ACA-funded community grants, should further reduce smoking rates by discouraging smoking initiation and encouraging cessation.

Reducing sodium intake, another key National Prevention Strategy intervention, reduces risks of hypertension and cardiovascular disease. Because most dietary sodium comes from processed and restaurant foods, it's difficult for Americans to limit their sodium consumption. Procurement guidelines from the DHHS and the General Services Administration and proposed school-food standards from the Department of Agriculture include a focus on sodium reduction. Menu-labeling requirements in chain restaurants will help people make more informed choices. The Centers for Disease Control and Prevention (CDC) is increasing public and professional education regarding sodium, and the CDC's National Health and Nutrition Examination Survey (NHANES) will begin collecting information on sodium consumption.

Consumption of artificial trans fat increases the risk of cardiovascular disease by raising low-density lipoprotein (LDL) cholesterol levels and lowering high-density lipoprotein (HDL) cholesterol levels. Replacing artificial trans fat with heart-healthy oils is feasible and does not increase the cost or change the flavor or texture of foods. Since the Food and Drug Administration began requiring listing of trans fat content on food labels, the industry has voluntarily reformulated foods, and according to CDC data, Americans' trans fat consumption has decreased by at least half. Elimination of such consumption could prevent 50,000 deaths per year.⁵

Million Hearts will leverage, focus, and align existing investments and generally not require

new public spending. Voluntary initiatives will simplify, harmonize, and automate clinicians' reporting requirements, decrease administrative burden, improve the quality of prevention and care, and inform the public more fully. Improvements in control of the ABCs, nutrition, and smoking are projected to prevent more than a million heart attacks and strokes over the initiative's first 5 years. By focusing our initial efforts where they will save the most lives, we aim to make progress toward a health system that will serve Americans' needs in the 21st century.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

Dr. Frieden is the director of the Centers for Disease Control and Prevention, Atlanta; Dr. Berwick is the administrator of the Cen-

ters for Medicare and Medicaid Services, Baltimore.

This article (10.1056/NEJMp1110421) was published on September 13, 2011, at NEJM.org.

1. Heidenreich PA, Trogdon JD, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation* 2011;123:933-44.
2. Ford ES, Ajani UA, Croft JB, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N Engl J Med* 2007;356:2388-98.
3. Farley TA, Dalal MA, Mostashari F, Frieden TR. Deaths preventable in the U.S. by improvements in use of clinical preventive services. *Am J Prev Med* 2010;38:600-9.
4. Mehta RH, Peterson ED, Califf RM. Performance measures have a major effect on cardiovascular outcomes: a review. *Am J Med* 2007;120:398-402.
5. Danaei G, Ding EL, Mozaffarian D, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med* 2009;6(4):e1000058.

Copyright © 2011 Massachusetts Medical Society.

หลักการประเมินคุณภาพของงานวิจัยที่ลงตีพิมพ์
Critical Appraisal of Published Research



หลักการประเมินคุณภาพของงานวิจัยที่ลงตีพิมพ์

Critical Appraisal of Published Research

พญ.พรพันธุ์ บุญยรัตพันธุ์

ภาควิชาระบาดวิทยา

คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล

นักวิชาการจะต้องติดตามงานวิจัยต่างๆในสาขาที่ตนสนใจอยู่ตลอดเวลา เพื่อพัฒนาความรู้และความชำนาญ มีวารสารทางวิชาการและเรื่องวิจัยมากมายที่จะต้องติดตาม ในสาขาระบาดวิทยามีวารสารทั้งในประเทศและต่างประเทศเป็นจำนวนมากที่ลงตีพิมพ์การศึกษาวิจัยในสาขานี้ เช่น

- วารสารวิทยาการระบาด ของภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหิดล
- American Journal of Epidemiology
- Epidemiologic Reviews
- International Journal of Epidemiology
- Journal of Epidemiology and Community Health
- Journal of Clinical Epidemiology
- Journal of Chronic Diseases
- Annals of Epidemiology
- Bulletin of the World Health Organization
- New England Journal of Medicine
- World Health Statistics Quarterly
- American Journal of Public Health

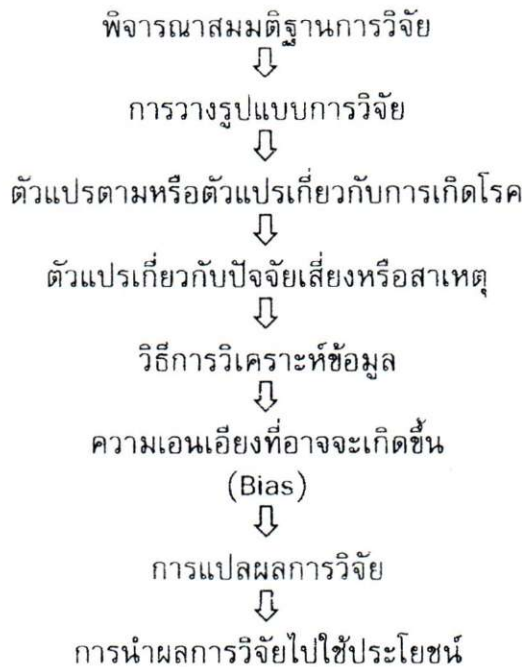
ฯลฯ

การพัฒนางานด้านการป้องกัน ควบคุมโรคหรือวิธีการรักษาพยาบาลที่มีประสิทธิภาพจะได้มาจากความรู้เกี่ยวกับงานวิจัยเหล่านี้ ดังนั้นนักวิชาการผู้อ่านผลงานการวิจัยต่างๆ จำเป็นจะต้องใช้ความพินิจพิเคราะห์โดยพื้นฐานของวิชาการเพื่อสามารถสรุปได้ว่าจะนำผลการวิจัยของวารสารเหล่านี้ไปใช้ประยุกต์ในการศึกษาของตนหรือศึกษาให้ลึกซึ้งต่อไปได้มากน้อยเพียงใดหรือไม่

โดยทั่วไปผลงานวิจัยทางระบาดวิทยาอาจแบ่งได้กว้างๆ เป็น

- การศึกษาธรรมชาติของการเกิดโรค
- การศึกษาสาเหตุของโรคหรือปรากฏการณ์ที่เป็นปัญหาสาธารณสุขต่างๆ
- การประเมินผลการรักษาพยาบาล หรือการวินิจฉัยโรคต่างๆ
- การหาประสิทธิภาพและประสิทธิผลของวิธีการหรือโครงการในการควบคุมและป้องกันโรค

ในการประเมินคุณภาพของรายงานการศึกษาวิจัยที่ลงตีพิมพ์ในวารสารต่างๆ มีขั้นตอนและหลักการ โดยทั่วไปในการวิเคราะห์ดังต่อไปนี้



1. การพิจารณาสมมติฐานการวิจัย (Research Hypothesis)

ควรพิจารณาในหัวข้อต่อไปนี้คือ

- มีการระบุที่ชัดเจนเกี่ยวกับสมมติฐานหรือวัตถุประสงค์การวิจัยหรือไม่
- การวิจัยนั้นได้ตอบคำถามที่มีความสำคัญทางด้านการแพทย์และการสาธารณสุขหรือไม่ ในด้านของการควบคุม ป้องกัน หรือพัฒนาวิธีการรักษาพยาบาล/การวินิจฉัยโรคที่มีประสิทธิภาพ

2. การพิจารณารูปแบบการวิจัย (Study design)

- การวางรูปแบบการวิจัยนั้นเหมาะสมกับสมมติฐานหรือวัตถุประสงค์ของการวิจัยหรือไม่
- รูปแบบการวิจัยที่ใช้ในการศึกษาเรื่องดังกล่าวดีกว่าที่เคยศึกษามาในอดีตหรือไม่
- เป็นการศึกษาแบบ Experiment หรือ Observational study

3. ตัวแปรเกี่ยวกับการเกิดโรค (Outcome variable)

- โรคที่ศึกษา (หรืออุบัติการณ์ต่างๆ เช่น ปัญหาสาธารณสุขต่างๆ) มีความสำคัญเหมาะสมแก่การศึกษาวิจัยหรือไม่
- เกณฑ์ที่ใช้ตัดสินการเป็นโรค
- ความถูกต้องแม่นยำในการตัดสินว่าเป็นโรคหรือไม่เป็นโรค

4. การพิจารณาปัจจัยเสี่ยง (Predictor variable(s) risk factors) ควรพิจารณาในข้อต่อไปนี้คือ

- มีปัจจัยเสี่ยงที่เป็นปัจจัยที่ศึกษาในการวิจัยครั้งนี้
- เกณฑ์ในการตัดสินว่ามีปัจจัยหรือไม่มีปัจจัย มีอะไรบ้าง
- เกณฑ์เหล่านั้นถูกต้องแม่นยำหรือไม่เพียงใด
- มีการวัดปริมาณของปัจจัยเสี่ยงหรือไม่ เช่น สูบบุหรี่วันละกี่มวน เป็นเวลานานเท่าใด เพื่อศึกษา Dose-response relationship
- ในการวัดบุคคลหนึ่งว่ามีปัจจัยเสี่ยงหรือไม่ ได้มีการใช้เครื่องวัดที่แม่นยำ เช่น Biologic markers หรือไม่

5. การพิจารณาวิธีการในการวิเคราะห์ข้อมูล (Methods of Data Analysis)

- วิธีการทางสถิติที่ใช้เหมาะสมกับชนิดของตัวแปรหรือไม่ เช่น ตัวแปร ordinal, nominal หรือ interval scale
- ได้มีการกล่าวถึง Type I หรือ Type II error อย่างถูกต้องเหมาะสมหรือไม่

		ความจริง	
		มีความแตกต่าง	ไม่มีความแตกต่าง
ผลการวิจัย	มีความแตกต่าง	A = ถูกต้อง	B = Type I error (α error) 0.05
	ไม่มีความแตกต่าง	C = Type II error (β error)	D = ถูกต้อง

ตัวอย่าง

ต้องการศึกษาอิทธิพลของปัจจัยเสี่ยงต่อการเกิดโรค

Ho: ไม่มีความแตกต่างในการเกิดโรกระหว่างกลุ่มที่มีปัจจัยเสี่ยงและไม่มีปัจจัยเสี่ยง

- ขนาดตัวอย่างที่ใช้เพียงพอที่จะตอบคำถามหรือวัตถุประสงค์การวิจัยหรือไม่ ถ้าต้องการให้ type I และ type II error น้อยลง ขนาดของตัวอย่างก็ต้องใหญ่ขึ้น
- ในการใช้ Statistical test ต่างๆ นั้นมี assumption ในการใช้ที่ถูกต้องหรือไม่เช่นในการใช้ χ^2 test

ตัวแปรมีการกระจายเป็น Normal distribution หรือไม่

ชนิดของนัยสำคัญ (Significance) ในการศึกษาวิจัยทางคลินิก

ชนิด	ความหมาย	การวัด
Statistical (ทางสถิติ)	ความสัมพันธ์นั้นไม่มี chance error เข้ามาเกี่ยวข้อง	Statistical test ต่างๆ
Clinical (ทางคลินิก)	มีความสำคัญเพียงพอที่จะเปลี่ยนแปลงการรักษาโรค	ขนาดของการเปลี่ยนแปลงทางคลินิกหรือการเกิดโรคอันเนื่องมาจาก intervention ที่ให้ไป
Biologic (ทางชีวภาพ)	ผลการวิจัยช่วยให้ความกระจ่างเกี่ยวกับกลไกการทำงานทางด้านชีวภาพ	เปรียบเทียบผลจากการศึกษาในหลอดทดลอง ในสัตว์ทดลองและในคนจากห้องปฏิบัติการ

6. พิจารณาเกี่ยวกับความเอนเอียงที่อาจจะเกิดขึ้น (Possible sources of bias or systematic errors)

- การเลือก subject ในการศึกษา มี bias หรือไม่
- การวัดหรือการจัดกลุ่มเกี่ยวกับกลุ่มที่มีปัจจัยเสี่ยง/ไม่มีปัจจัยเสี่ยง เป็นโรคหรือไม่เป็นโรค มีการผิดพลาดที่จะทำให้ผลการวิจัยเกิด bias หรือไม่
- ผลของการวิจัยเกิดจาก Confounding variable หรือไม่
- Bias ต่างๆ ที่เกิดขึ้นมีอิทธิพลทำให้ผลการวิจัยเอนเอียงไปทางใด (พิจารณาว่า น่าจะทำให้มากหรือน้อยเกินไป)

7. การพิจารณาการแปลผลการวิจัย (Interpretation of results) ควรพิจารณาด้านต่าง ๆ ต่อไปนี้

- ผลของการวิจัยเช่นความสัมพันธ์ระหว่างปัจจัยเสี่ยงและโรคมัขนาดเท่าใดหรือมากน้อยเพียงใด
- มีปรากฏการณ์ที่แสดงถึง dose-response relationship หรือไม่
- ผลของการวิจัยที่ได้สอดคล้องกับการวิจัยในสัตว์ทดลองหรือในห้องปฏิบัติการหรือไม่
- สามารถจะอธิบายผลได้ตามหลักการทางวิทยาศาสตร์ชีวภาพหรือไม่ (Biological plausible)
- ถ้าผลการวิจัยระบุไม่มีความสัมพันธ์ statistical power มีมากน้อยเพียงใด มากพอที่จะ detect ความแตกต่างหรือไม่

8. การนำผลการวิจัยไปประยุกต์ใช้

- ผลการวิจัยคล้ายคลึงกับการศึกษาวิจัยอื่นๆ ที่มีวัตถุประสงค์เดียวกันหรือไม่
- สามารถนำผลการวิจัยนี้ไปขยายผลในประชากรกลุ่มอื่นๆ ได้หรือไม่
- ผลการศึกษานี้สามารถนำไปประยุกต์ใช้ให้เกิดการเปลี่ยนแปลงในทางวิธีการทางสาธารณสุขหรือการแพทย์ได้หรือไม่

ในการประยุกต์ใช้ผลการวิจัย จะต้องมีการพิจารณาจากการวิจัยหลายๆ เรื่องด้วยกันที่มีวัตถุประสงค์เดียวกัน ซึ่งบางครั้งอาจจะได้ผลที่แตกต่างกันไปบ้าง ผู้วิจัยบางครั้งจะรวมผลการวิจัยจากหลายเรื่องในทำนองเดียวกันเข้าด้วยกัน และนำมาวิเคราะห์ในเชิงปริมาณใหม่ เรียก Meta-analysis



References

1. Greenberg RS. et al *Medical Epidemiology* 1993 Ch. 11
2. Schechter MT, Lebranc F.E. *Critical Appraisal of Published Research in : Principles and Practice of Research* edited by Troide H et al

พื้นฐานระบาดวิทยา

Basics of Epidemiology

บรรณาธิการ นายแพทย์คำนวณ อึ้งชูศักดิ์
 นายแพทย์ปฐม สวรรค์ปัญญาเลิศ
 นายแพทย์วิทยา สวัสดิ์วุฒิมงคล
 แพทย์หญิงจุฬารัตน์ จิระพงษ์

จัดพิมพ์โดย สมาคมนักระบาดวิทยาภาคสนาม
 88/20 อาคาร 4 ชั้น 6 ตึกสำนักงานปลัดกระทรวงสาธารณสุข
 ช.สถาบันบำราศนราดูร ถ.ติวานนท์ ต.ตลาดขวัญ อ.เมือง นนทบุรี 11000
 โทรศัพท์ 0-2590-1734-5 โทรสาร 0-2591-8581

จัดจำหน่ายโดย สมาคมระบาดวิทยา (ประเทศไทย)

พิมพ์ครั้งที่ 1 พฤษภาคม 2557 จำนวน 2,500 เล่ม

ราคา 350 บาท

ISBN 978-616-91574-1-0

ข้อมูลทางบรรณานุกรมของสำนักหอสมุดแห่งชาติ
National Library of Thailand Cataloging in Publication Data

พื้นฐานระบาดวิทยา (Basics of Epidemiology).--นนทบุรี : สมาคมนักระบาดวิทยาภาคสนาม, 2557.
318 หน้า.

1. ระบาดวิทยา. I. ชื่อเรื่อง.

614.4

ISBN 978-616-91574-1-0

รวบรวม นางสาวลัษณพร เจียรระโนรุงโรจน์
 นางสาวละอองดาว บุญยืน

ออกแบบ นางสาวชญาบุญ บุญศิริวรรณ

พิมพ์ที่ โรงพิมพ์แคนนา กราฟฟิค เขตภาษีเจริญ กรุงเทพฯ โทร. 0-2865-8454-5

“เรารู้ว่าหนังสือไม่ใช่วิธีการที่จะให้คนอื่นมาคิดแทนเรา ในทางตรงกันข้ามมันคือเครื่องมือที่กระตุ้นให้เราคิดได้ไกลมากยิ่งขึ้น” (อัมเบอร์โต อีโค นักเขียนนวนิยาย)

ความอยากรู้อยากเห็นเป็นคุณลักษณะที่ดีอย่างหนึ่งของผู้ทำงานระบาดวิทยา การอ่านช่วยเปิดโลกของเราไปสู่สิ่งที่คนอื่นได้เรียนรู้และปฏิบัติมา เพื่อนำมาประยุกต์ใช้กับชีวิต การอ่านผลงานทางวิชาการที่ดีพิมพ์ด้วยวัตถุประสงค์ใดๆ ก็ตาม เช่น ท่านต้องการทบทวนองค์ความรู้ที่ปฏิบัติมา เพื่อใช้ในการวิจัยทางสาธารณสุขของท่าน หรือการทบทวนผลการสอบสวนโรคที่ผ่านมาเพื่อเป็นแนวทางต่อการสอบสวนโรคในพื้นที่ของท่าน เป็นต้น เพื่อให้เกิดประโยชน์มากกว่าการเพิ่มพูนองค์ความรู้ ท่านควรกระตุ้นความคิดของท่านด้วยการตั้งคำถามต่อผลงานนั้นๆ ด้วย

โครงสร้างของเอกสารตีพิมพ์ทางวิทยาศาสตร์มีโครงสร้างคล้ายกัน คือ บทคัดย่อ ความเป็นมาและวัตถุประสงค์ วิธีการดำเนินการ ผลการศึกษา วิเคราะห์และวิจารณ์ผล สรุปผลและข้อเสนอแนะ เพื่อให้เกิดประโยชน์ในการอ่านผลงานทางวิชาการ ท่านสามารถนำหลักการคิดทางระบาดวิทยาเข้ามาประยุกต์ใช้ได้

เมื่ออ่านบทความทางวิชาการ ควรเริ่มพิจารณาตั้งแต่ชื่อเรื่อง ผู้แต่ง และบทคัดย่อ เพื่อทราบเรื่อง ผู้วิจัย และที่มาโดยย่อของผู้วิจัย คำถามการวิจัย วัตถุประสงค์ วิธีการศึกษา ผลการศึกษา และสรุปผลโดยย่อ ว่าตรงกับ ความสนใจของท่านหรือไม่ หากตรงกับความสนใจของท่านจึงพิจารณาอ่านเนื้อหาภายในต่อไป

เมื่อได้บทความที่ต้องการ ผู้อ่านควรพิจารณาประเด็นดังต่อไปนี้เพิ่มเติม เพื่อให้การอ่านเกิดประโยชน์ ดังนี้

1. ความเป็นมา และวัตถุประสงค์

ควรพิจารณาว่าคำถามของการศึกษานั้นคืออะไร โดยทบทวนเหตุผล ความเป็นมา และวัตถุประสงค์ของการศึกษา ผู้อ่านควรจะทราบที่มาและวัตถุประสงค์ของการศึกษานั้นๆ เพื่อให้เข้าใจว่าที่มาของการศึกษานี้เป็นอย่างไร และปัญหาที่ผู้เขียนนำมาทำการศึกษามีความสำคัญอย่างไร และสำคัญอย่างยิ่งที่จะต้องทราบ วัตถุประสงค์ของการศึกษา เนื่องจากวัตถุประสงค์เป็นตัวกำหนดทิศทางและวิธีการของการศึกษา

2. วิธีการศึกษา

เป็นส่วนที่มีความสำคัญมาก และผู้อ่านควรให้ความสนใจในการอ่านเพราะวิธีการศึกษาเป็นตัวบ่งชี้ว่า ผลการศึกษาที่ท่านจะอ่านต่อไปน่าเชื่อถือหรือไม่เพียงไร และสามารถใช้ตอบคำถามของการศึกษาได้หรือไม่ คำถามที่ควรพิจารณาเมื่ออ่านวิธีการศึกษา เช่น

- วิธีการศึกษาเป็นการศึกษาแบบใด ซึ่งผู้อ่านควรเข้าใจข้อจำกัดและข้อดีของการศึกษาทางระบาดวิทยา แต่ละชนิดด้วย เช่น การศึกษาทางระบาดวิทยาเชิงพรรณนา (descriptive epidemiological study) การศึกษาแบบภาคตัดขวาง (cross-sectional study) การศึกษาแบบเคส-คอนโทรล (case-control study) หรือการศึกษาแบบโคฮอร์ต (cohort study) เป็นต้น

- กลุ่มประชากรที่ใช้ในการศึกษาเป็นใคร โดยดูจากนิยามในการคัดเลือกประชากรในการศึกษาทั้งคัดเข้า และคัดออกจากการศึกษา

- วิธีการเลือกประชากรในการศึกษา และประชากรที่เลือกนั้นสามารถเป็นตัวแทนของประชากรที่สนใจได้หรือไม่

- ขนาดของตัวอย่างเหมาะสมหรือไม่

- มีความเป็นไปได้ที่จะเกิดความลำเอียงในการศึกษาหรือไม่ เช่น วิธีการเลือกตัวอย่าง วิธีการเก็บ

ข้อมูลมีความเหมาะสมหรือไม่ เช่น กรณีการศึกษาโคฮอร์ต มีผู้เข้าร่วมในการศึกษาจริงๆ ประมาณเท่าไร เมื่อเปรียบเทียบกับประชากรเป้าหมาย หรือมีการติดตามจนทราบผลของการเกิดหรือไม่เกิดโรคได้เท่าไร หรือหากเป็นการศึกษาแบบเคส-คอนโทรลมีวิธีการเลือกผู้ป่วยและกลุ่มควบคุมอย่างไร เป็นต้น

- วิธีการจำแนกผู้ป่วยหรือผู้ไม่ป่วยเหมาะสมหรือไม่ และสามารถตอบคำถามงานวิจัยได้หรือไม่ เช่น นิยามผู้ป่วยใช้เฉพาะกลุ่มอาการ หรือมีการตรวจทางห้องปฏิบัติการร่วมด้วย และวิธีการตรวจทางห้องปฏิบัติการมีความน่าเชื่อถือ ความไว และความจำเพาะมากน้อยเพียงใด เป็นต้น

- วิธีการจำแนกปัจจัยที่สนใจในการศึกษา เช่น ตัวแปรที่สนใจในการศึกษาได้แก่อะไรบ้าง วิธีการนิยามปัจจัยที่สนใจในการศึกษา มีการเก็บข้อมูลตัวแปรอื่นๆ ที่อาจเป็นตัวแปรกวนหรือไม่ และมีความเป็นไปได้ที่จะมีความลำเอียงจากการเก็บข้อมูลปัจจัยที่สนใจในการศึกษา เช่น ความลำเอียงจากการจดจำในอดีต (recall bias) หรือไม่ และผู้เขียนมีวิธีการอย่างไรในการลดความลำเอียงเหล่านั้น

- วิธีการเก็บข้อมูลเป็นอย่างไร เหมาะสมหรือไม่ เช่น วิธีการได้มาซึ่งข้อมูลมาจากการสัมภาษณ์ หรือให้ผู้เข้าร่วมการศึกษาตอบแบบสอบถามเอง ลักษณะของแบบสอบถามที่ใช้และชนิดของคำถาม เป็นต้น

- วิธีการวิเคราะห์ข้อมูลเพื่อตอบคำถามงานวิจัยเหมาะสมกับวิธีการศึกษาหรือไม่ มีการควบคุมตัวแปรกวนหรือไม่ อย่างไร และวิธีการเหมาะสมหรือไม่

3. ผลการศึกษา ควรพิจารณาประเด็นต่างๆ เช่น

- มีการนำเสนอผลการศึกษาสอดคล้องและครอบคลุมกลุ่มประชากรที่ทำการศึกษา ซึ่งอธิบายไว้ในวิธีการศึกษาหรือไม่

- มีการนำเสนอค่าทางสถิติที่ถูกต้องเหมาะสมกับชนิดของตัวแปรนั้นๆ หรือไม่

- ผลการศึกษาที่ได้ เช่น มีจำนวนข้อมูลที่ขาดหาย (missing data) มากหรือไม่ กรณีทำการศึกษาระบาดวิทยาเชิงวิเคราะห์ (analytical study) ค่าความสัมพันธ์ เช่น อัตราส่วนความเสี่ยงที่ได้ หรือผลที่ได้จากการศึกษาเป็นอย่างไร มีการคำนึงถึงความสัมพันธ์ระหว่างปริมาณกับการตอบสนอง (dose response relationship) หรือไม่ ผลที่ได้สามารถอธิบายได้ตามพื้นฐานความรู้ทางวิทยาศาสตร์ที่มีอยู่หรือไม่ ค่าความสัมพันธ์ที่ได้มีนัยสำคัญทางสถิติหรือไม่ หากไม่อาจเกิดจากสาเหตุอะไรได้บ้าง เช่น จำนวนผู้ที่เข้าร่วมในการศึกษาน้อย ทำให้ statistical power ต่ำ และกรณีที่ไม่นับนัยสำคัญทางสถิติควรพิจารณาว่าค่าความสัมพันธ์ยังมีความสำคัญทางสาธารณสุขหรือไม่

4. การวิเคราะห์วิจารณ์

เป็นการนำผลการศึกษาที่ได้มาวิจารณ์เพื่ออธิบายความเป็นไปได้ที่พบจากการศึกษานี้ ข้อที่ควรพิจารณา เช่น มีการวิเคราะห์วิจารณ์สอดคล้องกับผลการศึกษาที่ได้นำเสนอมาหรือไม่ เหตุผลที่ใช้ในการอธิบายสมเหตุสมผลหรือไม่ ผู้เขียนมีการทบทวนการศึกษาเรื่องใกล้เคียงกัน หรือลักษณะเดียวกันก่อนหน้านี้บ้างหรือไม่ และมีการอ้างอิงจากแหล่งที่เหมาะสมหรือไม่ เป็นต้น

5. การสรุปผลและข้อสรุป

การสรุปผลการศึกษา คือ การนำผลการศึกษาที่สำคัญมาสรุปอีกครั้งหนึ่ง เพื่อให้ผู้อ่านได้จับประเด็นผลที่ได้จากการศึกษานี้ และให้ข้อเสนอแนะ ประเด็นที่ควรพิจารณา เช่น ประเด็นสำคัญที่สรุปคืออะไร สิ่งที่สรุปได้มาจากผลการศึกษาที่กล่าวไปก่อนหน้านี้หรือไม่ มีเหตุผลสนับสนุนในการสรุปนี้หรือไม่ โดยอิงผลที่พบ

จากการศึกษานี้ รวมทั้งข้อเสนอแนะในการศึกษาสอดคล้องกับสิ่งที่พบจากการศึกษานั้นๆ หรือไม่ และผลจากการศึกษาสามารถนำไปใช้กับกลุ่มประชากรอื่นๆ ได้หรือไม่ และสามารถนำสิ่งที่ได้จากการศึกษานี้ไปใช้ประโยชน์ต่อไปได้หรือไม่ อย่างไร

พิจารณาคำถามของการศึกษา และวัตถุประสงค์ของการศึกษา

พิจารณาชนิดของการศึกษา เช่น cohort study, case-control เป็นต้น

พิจารณาผลลัพธ์ของการศึกษา (outcome) เช่น ป่วย/ไม่ป่วย เป็นต้น
รวมถึงนิยามและวิธีการวัดผลลัพธ์ของการศึกษา

พิจารณาปัจจัยต่างๆ ที่ผู้ศึกษาสนใจ (exposure/predictor)
รวมถึงนิยามและวิธีการวัดปัจจัยนั้นๆ

พิจารณาวิธีการวิเคราะห์ข้อมูล

พิจารณาว่ามีความเป็นไปได้หรือไม่ที่จะเกิดความลำเอียงในการศึกษา (bias)

พิจารณาผลการศึกษา ครอบคลุมเนื้อหาในวัตถุประสงค์ที่ระบุไว้
วิธีการนำเสนอผลการศึกษา และแปลผลการศึกษา

พิจารณาว่าบทความนั้นๆ สามารถนำมาประยุกต์ใช้ประโยชน์ได้อย่างไร

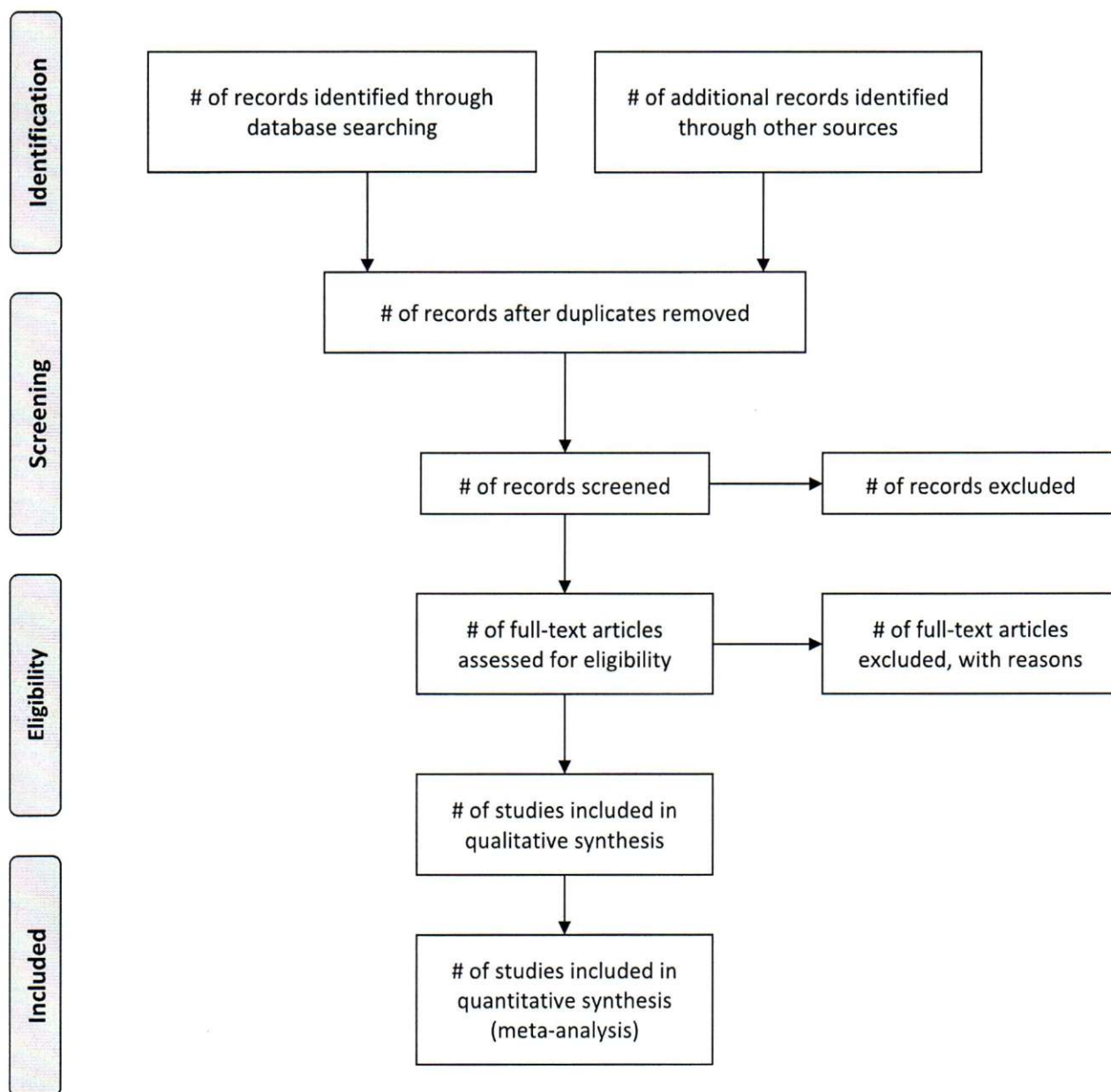
รูปที่ 4 สิ่งที่ควรพิจารณาในการอ่านบทความทางวิชาการทางระบาดวิทยา
ที่มา: ดัดแปลงจาก Medical Epidemiology โดย Greenberg RS



PRISMA 2009



PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

Different forms of economic evaluation

Study type	Measurement of benefits	Question posed
Cost minimisation analysis	Benefits found to be equivalent	Which is the most efficient way achieving a given goal (or objective)?
Cost effectiveness analysis	Natural units (eg life years gained)	or
Cost-utility analysis	Healthy years (eg quality adjusted life years, healthy years equivalents)	What is the most efficient way of spending a given budget?
Cost-benefit analysis	Monetary terms	Should a given goal (or objective) be pursued to a greater or lesser extent?

Referees' checklist (also to be used, implicitly, by authors)

Study design

- (1) The research question is stated
- (2) The economic importance of the research question is stated
- (3) The viewpoint(s) of the analysis are clearly stated and justified
- (4) The rationale for choosing the alternative programmes or interventions compared is stated
- (5) The alternatives being compared are clearly described
- (6) The form of economic evaluation used is stated
- (7) The choice of form of economic evaluation is justified in relation to the questions addressed

Data collection

- (8) The source(s) of effectiveness estimates used are stated
- (9) Details of the design and results of effectiveness study are given (if based on a single study)
- (10) Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)
- (11) The primary outcome measure(s) for the economic evaluation are clearly stated
- (12) Methods to value health states and other benefits are stated
- (13) Details of the subjects from whom valuations were obtained are given
- (14) Productivity changes (if included) are reported separately
- (15) The relevance of productivity changes to the study question is discussed
- (16) Quantities of resources are reported separately from their unit costs
- (17) Methods for the estimation of quantities and unit costs are described
- (18) Currency and price data are recorded
- (19) Details of currency of price adjustments for inflation or currency conversion are given
- (20) Details of any model used are given
- (21) The choice of model used and the key parameters on which it is based are justified

Analysis and interpretation of results

- (22) Time horizon of costs and benefits is stated
- (23) The discount rate(s) is stated
- (24) The choice of rate(s) is justified
- (25) An explanation is given if costs or benefits are not discounted
- (26) Details of statistical tests and confidence intervals are given for stochastic data
- (27) The approach to sensitivity analysis is given
- (28) The choice of variables for sensitivity analysis is justified
- (29) The ranges over which the variables are varied are stated
- (30) Relevant alternatives are compared
- (31) Incremental analysis is reported
- (32) Major outcomes are presented in a disaggregated as well as aggregated form
- (33) The answer to the study question is given
- (34) Conclusions follow from the data reported
- (35) Conclusions are accompanied by the appropriate caveats

MAKING CHOICES IN HEALTH:

**WHO GUIDE TO
COST-EFFECTIVENESS ANALYSIS**

EDITED BY

T. TAN-TORRES EDEJER, R. BALTUSSEN, T. ADAM,
R. HUTUBESSY, A. ACHARYA, D.B. EVANS AND
C.J.L. MURRAY



World Health Organization
Geneva

8 REPORTING CEA RESULTS

Reports on CEA results must provide sufficient information to enable independent analysts to critically evaluate the estimates of the costs and effectiveness of the interventions studied. In addition, they should be able to interpret the findings of the CEA and assess the possibility of generalizing them to their own decision-making context.³⁶ Since it may not always be possible to document this information in a journal article, additional information should be provided in background reports or on the World Wide Web. To enhance transparency and ensure accountability, all reports and all data inputs, including assumptions, used in deriving the estimates, should be placed in the public domain.

A CEA report usually contains, or indicates sources for, a detailed description of the inputs and methods used to estimate costs, effectiveness and cost-effectiveness ratios of the interventions studied. The ten-point checklist introduced by Drummond et al. (17) or a similar format may be used as a guide to analysts seeking to improve the quality of their study reports.

The following section outlines the key information to be reported with respect to the elements of CEA. A short description of WHO's approach to reporting GCEA results can be found in Annex A.

8.1 COST INFORMATION

Reports should contain or discuss:

- information on unit prices and quantities for the main factor inputs used to estimate programme costs (e.g. personnel, vehicles, office space etc.);
- how patient costs were estimated—for example the cost per visit or bed-day, the costs of laboratory tests—and what assumptions were used, including questions of intervention coverage levels, capacity utilization, depreciation rates used to obtain capital costs, etc.;
- whether the costs used in the study have face validity, in terms of other costs reported in the literature, for example, and whether they were

obtained from a sample of costs that are likely to be representative rather than based on a single observation;

- results of sensitivity and uncertainty analysis; and
- space permitting (e.g. for web-based presentation of results), a detailed listing of quantities and prices of factor inputs used in the analysis.

8.2 EFFECTIVENESS INFORMATION

Reports should contain or discuss:

- whether a systematic search for evidence on baseline epidemiology and effectiveness was undertaken, the criteria used for selecting sources, the assumptions made, etc.;
- quantitative documentation of the sources and assumptions used for: (1) the main input variables in the analysis such as prevalence, incidence or remission rates and relative risk ratios, all of which should be reported for both the null and the intervention scenarios; (2) how the effectiveness of each intervention was modelled (e.g. through a decrease in incidence, in duration, in remission, or in mortality rates); and (3) other factors related to modelling health effects such as intervention coverage rates, patient adherence to medicines and follow-up visits and quality of services provided;
- the healthy life years lived by the population under both the null and interventions scenarios, and the difference between the two scenarios—representing the health gain of the intervention; and
- results of sensitivity and uncertainty analysis.

8.3 COST-EFFECTIVENESS RATIOS

Reports should contain or discuss:

- both a numerical and a graphical documentation of cost-effectiveness ratios;
- cost-effectiveness ratios compared to the null for all interventions studied, and incremental cost-effectiveness ratios for those interventions on the expansion path;
- expansion paths clearly identified either in tabular or graphical form, for each set of inter-dependent interventions;
- results of uncertainty analysis including use of stochastic league tables where appropriate.

SUMMARY OF RECOMMENDATIONS

1. Reports on CEA results must provide sufficient information in the public domain to enable independent analysts and policy-makers to critically evaluate the validity of the estimates of the costs and effectiveness of the interventions studied.

*Economics notes***Using cost effectiveness information**

Andrew Briggs, Alastair Gray

This is the seventh in a series of occasional notes on economics

Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford OX3 7LF

Andrew Briggs
joint MRC/Anglia and Oxford Region training fellow
Alastair Gray
director

Correspondence to Dr Briggs

These notes are edited by James Raftery (J.PRAFTERY@bham.ac.uk)

BMJ 2000;320:246

How should the results of economic evaluations be interpreted and used by decision makers in health care? In cost benefit analyses the decision rule is in principle straightforward: if benefits exceed costs then the programme should be implemented; if not, it should be rejected. However, the use of cost benefit analysis is limited by the need to place monetary valuations on health outcomes, and cost utility analyses are more widely used, with results presented in terms of the cost per QALY (quality adjusted life year).

Unfortunately, no clear decision rule exists for cost utility analyses. Some analysts have suggested setting a threshold value for the cost per QALY that represents the willingness of society to pay for additional QALYs. But others argue such thresholds could lead to uncontrolled expenditure growth if new procedures deliver QALYs at less than the threshold.²

Incremental cost per QALY figures are often grouped in league tables, which imply that interventions at the top (with lower cost per QALY figures) should take priority over those further down (see table). Many commentators have cautioned against the unthinking use of league tables because of non-comparability of methods, inappropriate comparators, and non-generalisability of results.⁴ Even if these problems were solved, however, league tables would still need additional information to be useful to decision makers. In the original from which the table is constructed, Williams was considering whether the programme for coronary artery bypass grafting in the United Kingdom should be expanded.³ Each figure in the table represents the incremental cost effectiveness⁵ of bypass grafting compared with medical management: benefits declined as the programme was expanded to include patients with less severe disease. The incremental cost per QALY for bypass grafting for severe angina with left main vessel disease was 10 times less than for mild angina with double vessel disease.

This is an example of a changing marginal incremental cost per QALY. The importance of the margin is paramount in economic thinking. In the table marginal changes in the incremental cost effectiveness ratio take place at the "clinical margin"—that is, as the same intervention is expanded to cover individuals with less severe clinical disease. Age, sex, or risk factors could be seen as clinical margins when expanding programmes. For example, in a recent study of statin treatment for reducing cholesterol concentrations, the average incremental cost effectiveness for patients with pre-existing heart disease and a cholesterol concentration of >5.4 mmol/l was £32 000 per life year gained.⁶ But this average hides differences in patient subgroups of £6000 to £361 000 per life year.

Ideally, a league table should include marginal incremental cost effectiveness data by having separate entries for different subgroups. The clinical margin has major implications for league tables: in the statin example the authors estimated 48 differing cost effectiveness figures for differing subgroups.

Besides the clinical margin, an intensity margin may also be identified. Interventions may be offered at different levels of intensity to the same patient groups—for example, annual or biannual breast screening, or low dose versus high dose antiviral therapy. Here the incremental cost effectiveness ratio must be calculated along this intensity margin: for example, in a breast cancer screening evaluation the analyst should be interested in comparing screening every three years compared with no screening, a two year compared with a three year screen, and a one year compared with a two year screen. To compare an annual screening programme with no programme will be misleading, as many of the benefits of an annual screen could potentially be achieved by a two year screen—that is, at a lower intensity point at the margin.

The implications for the "league table" approach are that data are required on patient subgroups at the clinical margin of the same intervention, and between the same patients at the intensity margin of an intervention. This requires more information about each margin. As many evaluations already provide subgroup analyses, a first step is to make better use of available information. The real choices are not about blanket exclusions but about assessing incremental effectiveness and costs at the margin.

An example of an incremental cost per QALY league table

Intervention	Cost per QALY
Pacemaker for atrioventricular heart block	£700
Hip replacement	£750
Valve replacement for aortic stenosis	£900
CABG (severe angina; left main disease)	£1 040
CABG (severe angina; triple vessel disease)	£1 270
CABG (moderate angina; left main disease)	£1 330
CABG (severe angina; left main disease)	£2 280
CABG (moderate angina; triple vessel disease)	£2 400
CABG (mild angina; left main disease)	£2 520
Kidney transplantation (cadaver)	£3 000
CABG (moderate angina; double vessel disease)	£4 000
Heart transplantation	£5 000
CABG (mild angina; triple vessel disease)	£6 300
Haemodialysis at home	£11 000
CABG (mild angina; double vessel disease)	£12 600
Haemodialysis in hospital	£14 000

CABG=coronary artery bypass grafting. Adapted from Williams³

1 Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J* 1992;146:473-81.

2 Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. *Can Med Assoc J* 1993;148:913-7.

3 Williams A. The economics of coronary artery bypass grafting. *BMJ* 1985;291:326-9.

4 Gerard K, Mooney G. QALY league tables: handle with care. *Health Economics* 1993;2:59-64.

5 Palmer S, Raftery J. Opportunity cost. *BMJ* 1999;318:1551-2.

6 Pharos PDP, Hollingworth W. Cost effectiveness of lowering cholesterol concentration with in patients with and without pre-existing coronary heart disease: life table method applied to health authority population. *BMJ* 1996;312:1443-8.



EXERCISE

Vitamin Supplements and the Incidence of Colds in High School Basketball Players

A Preliminary Report

FRANK E. BARNES, JR., M. D.

SMITHFIELD

The incidence of colds among players on high school basketball teams can be a very real problem from the point of view of winning games and league championships. This is important to the teen-ager and to the school, and is part of the larger problem of keeping this age group in good health and able to do its best both in athletic and academic activities.

It has been recognized for some time that during periods of rapid growth nutritional requirements are greatly increased. The nutritional needs of infants and young children during their development have received much attention, but the adolescent does not always get his share of attention. Many adolescents eat too little, and many more eat poor combinations of food¹. A number of surveys of the diets of teen-agers indicate that they do not receive enough vitamins. One survey² indicated that supplemental vi-

amins and minerals given to school children resulted in significant gains in both academic attainment and growth. The ratio of growth increased from the tenth year to the degree that "in the girl's thirteenth year, and the boy's sixteenth year, the ratio of gain is double that seen in the early years of school life."³ In discussing adolescents, Fischer⁴ states that "subclinical vitamin deficiency is quite common in this age group and in some instances there may even be typical avitaminoses."

A recent report⁵ on the basis for concern about the diet of teen-agers states that "there is little question that this age group as a whole practices limited judgement in the choice of total food eaten." Two of the shortcomings cited are suboptimal supplies of vitamin A and insufficient ascorbic acid. The stresses which occur during adolescence are also discussed—accelerated growth and

Table 1
Composition of Theragran-M Tablets

	<i>Contents of One Tablet</i>	
Vitamin A	25,000	U.S.P. units
Vitamin D	1,000	U.S.P. units
Vitamin K	2	mg.
Thiamine mononitrate	10	mg.
Riboflavin	10	mg.
Pyridoxine hydrochloride	5.0	mg.
Vitamin B12 activity concentrate	5	mcg.
Folic acid	0.1	mg.
Niacinamide	100	mg.
d-Calcium pantothenate	20	mg.
Ascorbic acid	200	mg.
Vitamin E	5	International units
Calcium	105	mg.
Iodine	0.15	mg.
Iron	15	mg.
Potassium	5	mg.
Copper	1	mg.
Manganese	1	mg.
Magnesium	6	mg.
Zinc	1.5	mg.

inefficient use of certain important substances as a result of frequent, important emotional problems.

Although boys and girls playing basketball are usually in good physical condition as compared with the majority of high school students, they have the added stress of greater and more frequent exposure to changes in temperature, since they play in hot gymnasiums, take showers, and go out into cold air on repeated occasions.

Materials and Methods

In North Carolina it is well known that high school students eat a poorly balanced diet. Recent studies made by student nurses from the University of North Carolina Nursing School revealed that the majority of the students have an inadequate diet. A very small percentage drink a glass of milk a day, but many seem to survive on hot dogs and cola drinks. As a means of supplementing these diets and of observing the effects of multivitamins in fortifying the students against winter ills, a program was worked out with the staff and students of the Smithfield, North Carolina, High School. The school officials were enthusiastic, and cooperated in keeping daily charts on the progress of the study.

High levels of ascorbic acid were considered to be necessary to ward off the usual

run of colds, a fact that was taken into consideration in the choice of a supplemental vitamin preparation. Theragran-M* was chosen after the many vitamin preparations available were evaluated for the present purpose. The composition of this preparation is given in table 1.

The boys' and girls' basketball teams, 26 players in all, were examined for upper respiratory infections, and histories of recent colds or sore throats were taken before the study was begun. When first seen, 7 boys and 4 girls had colds or evidence of recent upper respiratory infection. None was severely ill.

The coaches of the teams were put in charge of dispensing the vitamin tablets, one of which was given to each player every day for approximately seven weeks. If a cold or sore throat was reported either at practice or before a game, this was recorded. Each player was given a supply of tablets to take over the week-end while he was at home. An attempt was made to keep the players on the vitamin preparation for the full seven days of every week. Sixteen other boys and girls of the same age and background as the players served as a control group and reported daily to the coaches.

* Supplied by The Squibb Institute for Medical Research, New Brunswick, New Jersey.

Table 2
The Effect of Vitamin Supplements on the Incidence of Colds
in High School Students
(Total Study Period, 848 Days)

Control Group (8 boys, 8 girls)				Group Receiving Vitamins					
January 13 to March 6				Boys (10)		Girls (13)			
Sex	No. Days with Colds	No. Days Absent	Boys	Days of Medication	Days with Colds	Girls	Days of Medication	Days with Colds	
JB*	F	10	..	EB	38	PA	43	9**
SC	F	7	2	RB	47	2	AB	43	1**
BC	F	4	..	CW	47	1	MAB	43	5**
BCr	F	9	3	PE	47	2***	LB	43	0
MD	F	4	..	MG	47	3***	BLC	43	0
SNH	F	SBMc	46	1	PE	43	0
KJ*	F	19	..	JO	42	2	PH	41	0
JO	F	7	..				NL	43	3
				TH	47	1			
TE	M	6	..	DL	46	3	MC	43	0
BG	M	11	..	SS	46	1	AC	43	0
JOg	M	8	1				LH	43	1
TO	M	9	2				SH	42	0
CO*	M	7	..				BB†	22	0†
JS	M						
BW	M						
CY*	M	9	..						
16		110	8	10	454	16	13	535	18

* At least two colds

** Very light colds

*** One day each, mild colds

† Added to group during third week of study
 "Flu" for five days

Records of absences from classes for all students in both groups were also kept. The parents and, if necessary, the family physician were consulted before the medication was initiated. Only two parents asked that their children not be given the vitamins, since these children were on special diets and were receiving other vitamin preparations at that time.

Results

Experimental groups

The results are summarized in table 2. Almost from the start of the experiment there were complaints among the boys who had been receiving their tablets just before practice, of abdominal cramps and some diarrhea (6 boys complained of this condition during the first week). None of the girls who received the tablets just after practice had these symptoms. When the

coaches changed the procedure and gave both groups their vitamin preparation after the practice sessions or games, the symptoms subsided. One boy, however, stopped taking the preparation during the third week of the study at the request of his parents, who felt that it might be causing his vague abdominal complaints. Another boy stopped taking the vitamins during the fourth week of the study because he thought the tablets were causing indigestion. The rest of the players experienced no other symptoms during the remainder of the study.

The girls' basketball team was made up of 13 players, one of whom refused to take the preparation being tested since she was on a diet. The group as a whole, consisting of 12 players, had 19 days of colds and five days of "flu." One girl reported a cold for nine days, another a cold for five days. Two girls,

therefore, were responsible for 14 of the total of 19 days when colds were reported. One of these two girls on four days, had only suggestive symptoms of a cold. It is noteworthy that only two days were missed by the entire group of 12 girls during their seven weeks of school. A new girl joined the team in the third week of practice and promptly had a case of influenza lasting five days. During the time of the study, 535 tablets were dispensed to the girls, and no ill effects were observed in this group.

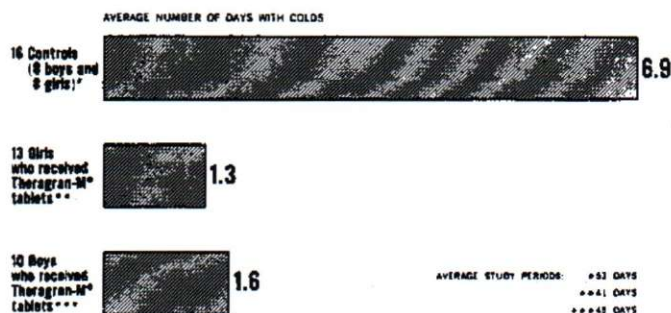


Fig. 1

The boys' basketball team was also made up of 13 players at the beginning of the study, but 3 boys stopped taking the vitamins before it was completed. A total of 454 tablets were dispensed. The majority of colds in the boys' basketball team were noted during the first week of practice—four days of colds involving 3 boys—and only two days of school were missed out of the entire eight weeks.

An unusual circumstance was that 7 boys reported with colds on the same day in the fourth week, but none was noted to have any cold the rest of the week. They had played a hard, rough game the night before, a circumstance that was repeated during the sixth week, when 6 players reported for practice with colds on a Monday, but no colds appeared during the rest of the week. It was questioned whether this was just the effect of strenuous exercise or undue activity over the week-end, since the colds completely disappeared in a 24-hour period. A total of 13 days with colds were reported by this group in two days, and none thereafter. An aggregate of 16 days with colds was reported by members of the boys' basketball team.

Control Group

The control group, made up of 8 boys and 8 girls, differed markedly from the two groups receiving vitamins as to the total number of days on which colds were reported, and the days missed from school. This control group reported 110 days with colds out of a total number of 848 days. Five of the group had colds twice during the period of the study, and only 3 out of the entire group had no colds or other sickness at any time during the study. Eleven had at least one week of colds. Three had 10 or more days of colds or evidence of influenza. The group as a whole missed eight days of school.

Comment

The coaches and the school officials noted a definite improvement in the students given the vitamin supplementation, both in their attitudes and playing ability, from that observed in previous years when no supplements had been given. The coaches were enthusiastic about the greatly reduced number of absences from practice and games. They are interested in continuing the vitamin supplements in future years, not only with the basketball teams but also with the football teams, since the bitter, cola weather of October and November is usually productive of colds and disrupts the athletic program.

Summary and Conclusions

1. Records were kept of the incidence of colds for approximately eight weeks in a total of 39 high school students. Of this group, 13 girls and 10 boys received Thera-M (a multivitamin supplement), and regularly 10 boys and girls of comparable age and background served as controls. The two groups who received the vitamins were members of the basketball teams.

2. The control group reported with colds on an aggregate of 110 days during the study period. The 10 boys who received 454 Thera-M tablets reported colds on 16 days, and the 13 girls who received 535 tablets reported colds on 18 days. This was an average incidence of 6.9 days of colds for the controls, but for those who received vi-

tamins it was 1.6 days for the boys and 1.3 days for the girls.

3. Although the number of students in the two groups was small, it is believed that the vitamins definitely improved the health of the students. It is suggested that all high school students need multivitamin supplements, particularly during strenuous athletics in the fall and winter months.

References

1. Smith, C. A.: Current Trends in Feeding of Children, J.A.M.A. 161:729 (June) 1956.
2. Byrd, O. E.: Leland Stanford University, Palo Alto, California, report of October 19, 1958.
3. McLester, J. S., and Darby, W. J.: Nutrition in Health and Disease, ed. 6, Philadelphia, W. B. Saunders Co., 1952.
4. Fischer, C. C.: Nutritional Problems in Adolescence, Am. Pract. & Digest. Treat. 3:968-972 (Dec.) 1952.
5. Everson, G. J.: Bases for Concern about Teen-agers' Diets, J. Am. Dietet. Ass. 36:17-21 (Jan.) 1960.

Correspondence

VITAMIN SUPPLEMENT FOR HIGH SCHOOL STUDENTS

To the Editor:

I write with reference to the article by Dr. Frank E. Barnes, Jr., "Vitamin Supplements and the Incidence of Colds in High School Basketball Players: A Preliminary Report." Dr. Barnes is to be commended for his desire to investigate the possibility that vitamin supplements might benefit the health of high school students. He ends, ". . . it is believed that the vitamins definitely improved the health of the students. It is suggested that all high school students need multivitamin supplements, particularly during strenuous athletics in the fall and winter months."

Unfortunately, these conclusions are not justified in view of the design of the investigation. Undoubtedly the vitamin manufacturers are likely to quote Dr. Barnes' conclusions in their advertising. Again, we should be concerned about the possibility of unnecessary expenditure of money on vitamins by lay persons who are not qualified to interpret the meaning of such a report, and who may be unduly impressed by the conclusions. For these reasons the study needs to be redesigned and repeated. A survey of the article leads to the follow-

ing essential comments. The 16 "other boys and girls of the same age and background as the players" who served as a "control group" unfortunately have one very great difference which was ignored in the study, namely, that they are not basketball players. It is possible that non-players would experience a different number of colds even if they had also been receiving a vitamin supplement. In fact they received nothing, not even a placebo. In any case, the common cold leads a fairly typical course and it is doubtful if the "colds" of one day's duration are the same kind of illness as those of longer duration. Certainly, the presence or absence of a cold is a very subjective factor which must not be left uncontrolled. Other factors include the official enthusiasm described by Dr. Barnes and possibly the day-to-day successes or failures of the teams. It is interesting to notice Dr. Barnes's statement that on one day 7 boys reported with colds following a "hard, rough game the night before, a circumstance that was repeated during the sixth week, when 6 players reported for practice with colds on a Monday, but no colds appeared during the rest of the week." Unfortunately, he does not tell us whether the teams were having a winning or losing streak at that time, but this might have important effects upon the incidence of illness.

All the variables can be fairly adequately taken care of by setting up a carefully controlled study using the double blind technique. It is for this reason that I am writing this letter, not in a spirit of criticism but in the hope that Dr. Barnes will feel encouraged to continue with his work and to repeat it under more scientifically rigorous conditions. A research study such as he is conducting is worth going to some trouble over as the results can prove to be most enlightening. With a minimal degree of assistance, it should be possible for Dr. Barnes to divide his basketball players randomly into two groups who would then receive coded vitamin supplements or identical-looking tablets of no therapeutic value, while neither the players nor those involved in evaluating the study are aware of which people are getting which prepara-

tion. A study of this sort would be open to accurate statistical evaluation and a most important publication would result.

In North Carolina there are three medical schools each with many workers (psychologists, biostatisticians, and others) who are qualified to help design a suitable controlled study. I hope that Dr. Barnes will avail himself of such assistance and I am looking forward to seeing a further article by him.

John A. Ewing, M.D.

Department of Psychiatry

University of North Carolina

*(Dr. Barnes' reply to
Dr. Ewing follows.—Ed.)*

Smithfield

Dear Dr. Ewing:

Your comments for a double-blind control study as well as a study done on a bigger scale are certainly worth while criticisms. This sort of thing I have in mind at the present time. I hope to go ahead next year with a study of the Selma and the Smithfield high schools, using one as a control. The study that I undertook at the Smithfield High School was done more out of curiosity than anything else, as I was quite curious about the actual benefits of vitamins on the general public or on growing students. I have been deluged with vitamin literature by the different drug companies, and when Squibb offered me sufficient vitamins to carry out this study I thought I would just follow it through and see if there was any advantage in this type of medication on a growing boy or girl. This study has already been criticized by the A.M.A.'s Council on Nutrition. They felt that a growing boy or girl does not need vitamins. They might be correct. But after the study that was done by the student nurses from the University of North Carolina on our high school students, I was quite convinced that these kids were not on an adequate diet that gave them all the necessary vitamins. We found that a great percentage of these children did not eat a balanced diet, and it is amazing that they can even exist on what they put into their stomachs. I am also convinced that the two sessions of supposedly sore throats or colds that happened on two different oc-

casions were nothing more than evidences of fatigue, as the symptoms disappeared so quickly. I suppose that an ideal study at the present time would be to give the children in the gym classes the vitamins and keep the high school basketball athletes on just the regular diets they receive at home. If I can get the cooperation of the school officials and the parents, I certainly intend to follow up this sort of study as I think it is worth while and will either condemn or show the need for vitamin supplements to our growing high school students.

I am not sure whether vitamins are doing these kids any good or not, but they seem to get through school very well, without colds, and our girls just won the district championship. I would welcome any further investigations by people at schools such as the University of North Carolina or at other high schools. I hope that somebody else will pick this up and continue the study. I know that several of the universities have their athletes on vitamins. Ohio University had their basketball team on vitamins when they went to the National Championship last year. There has also been a study done at the Naval Academy where it was found that huge doses of vitamin C have cut down the incidences of bruises on the football players. All these things will have to be investigated further before we come up with a definite conclusion and I hope that the vitamin distributors will not take advantage of this article to try to impress the public that everybody needs vitamins. Certainly the doctors have been deluged with this kind of thing and I am sure they will not swallow the whole article as a final conclusion.

Frank E. Barnes, Jr., M.D.

A Meta-analysis of Coffee, Myocardial Infarction, and Coronary Death

Sander Greenland

This paper presents a meta-analysis of 22 studies of coffee use and myocardial infarction or coronary death. In the eight case-control studies, a fairly homogeneous increased risk was found among coffee users (geometric mean rate ratio of 1.42 for 5 cups per day vs none, with 95% confidence limits of 1.30, 1.55, homogeneity *P*-value of 0.89). The 14 cohort studies tended to exhibit lower but very heterogeneous rate ratios, with a trend toward larger rate ratios in studies with longer follow-up periods and later publication dates (geometric mean rate ratio of 0.92 for the five cohort studies published up to

1981, 1.27 for the nine cohort studies published in 1986 or later; overall homogeneity *P*-value of 0.0008). The evidence thus remains ambiguous regarding both the existence and size of a coffee effect, and although a rate ratio of over 1.5 for 5 cups per day appears unlikely, stronger effects for 10-cup-per-day drinkers cannot be ruled out. (*Epidemiology* 1993;4:366-374)

Keywords: coffee, meta-analysis, myocardial infarction.

In 1987 I published a statistical summary of nine studies of coffee and myocardial infarction as an illustration of meta-analytic methods.¹ Since that publication, several new, large, and well-analyzed studies have appeared on the topic of coffee and myocardial infarction or coronary death,²⁻⁷ as well as updates of older studies.⁸⁻¹¹ Most of these studies have found a positive association between coffee and coronary disease. Also, further studies linking coffee use to alteration of blood pressure and serum lipid patterns have appeared, including randomized trials, although the results have been inconsistent (see, for example, Refs 12-18).

The present report updates the 1987 meta-analysis¹ by presenting a statistical summary of 21 studies of coffee use and myocardial infarction or sudden coronary death published from 1968 to 1992, plus an unpublished case-control study conducted by Ulrik Gerdes in Denmark.

Materials and Methods

Table 1 provides details of the studies analyzed here. I searched the literature using MEDLINE and direct examination of reference lists of recovered articles. The search cutoff date was July 31, 1992. Unlike the

1987 meta-analysis,¹ studies of coffee and coronary mortality that did not separate myocardial infarction deaths from other coronary deaths were used here to allow inclusion of a much larger number of cohort studies. The original studies by Rosenberg *et al*,¹⁹ Yano *et al*,²⁰ the Western Electric Study group,²¹ and Snowdon *et al*²² are here replaced by later analyses of the same subjects.⁸⁻¹¹ The updates of the Framingham study²³ and the Wilhelmsen *et al* cohort study²⁴ presented insufficient data for inclusion here: the Framingham update gave results only for all vascular disease combined (including angina, congestive heart failure, intermittent claudication, and stroke) and gave insufficient data to compute standard errors, whereas the Wilhelmsen update²⁴ provided neither estimates of relative risk nor data from which estimates could be computed. The point estimates from the Framingham update were, however, virtually the same as those in the original report.

For studies that only presented detailed results separately for caffeinated and decaffeinated coffee,^{2,3,9} the results for caffeinated coffee were used; otherwise, total coffee was used. For the cohort studies,^{6-8,10,25-31} baseline coffee consumption (consumption at start of follow-up) was used. In the case-control studies used here,^{2,3,9,27,32-36} coffee questions referred to preinfarction consumption. Two of the case-control studies^{27,35} used only population controls (Wilhelmsen *et al*²⁷ reported both cohort and case-control studies), one (Gerdes) used a mix, and the remainder used only hospital controls. The record-based study of Klatsky

From the Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90024-1772.

Submitted September 21, 1992; final version accepted January 20, 1993.

© 1993 Epidemiology Resources Inc.

TABLE 1. Summary of Re-analyses for Studies of Coffee and Myocardial Infarction or Coronary Death (1968–1992) (Coefficients Represent Increment in Log Rate or Logit Risk per Cup of Coffee per Day)

Study	Outcome*	No. of Cases†	% ≥ n Cups‡,§	Coefficient§	Standard Error§	Weight	Rate Ratio for 5 Cups/Day	Years Follow-up (Cohort)
Cohort (in order of publication)								
Klatsky <i>et al.</i> ^{25¶}	M	464	22 ≥ 7	−44	29	1,171	0.80	6
Dawber <i>et al.</i> ²⁶	M	322	15 ≥ 5	−39	40	625	0.82	12
Wilhelmsen <i>et al.</i> ²⁷	M	60	50 ≥ 5	109	153	43	1.72	12
Heyden <i>et al.</i> ²⁸	D	36	13 ≥ 5	−44	76	5	0.80	4.5
Murray <i>et al.</i> ²⁹	D	721	31 ≥ 5	−4	19	2,921	0.98	11.5
Jacobsen <i>et al.</i> ³⁰	D	941	37 ≥ 5	−3	19	2,887	0.99	11.5
La Croix <i>et al.</i> ³¹	C	37	13 ≥ 5	86	58	294	1.54	25
Yano <i>et al.</i> ⁸	C	730	25 ≥ 5	49	36	766	1.28	15
LeGrady <i>et al.</i> ¹⁰	D	232	53 ≥ 4	82	27	1,351	1.51	19
Klatsky <i>et al.</i> ⁴	M	724	17 ≥ 4	68	19	2,746	1.40	5
Grobbbee <i>et al.</i> ⁵	DM	221	20 ≥ 4	4	30	1,076	1.02	2
Tverdal <i>et al.</i> ⁶	D	184	57 ≥ 5	86	35	825	1.54	6.4
Rosengren ⁷	DM	399	44 ≥ 5	31	26	1,504	1.17	7.1
Lindsted <i>et al.</i> ¹¹	D	NG	10 ≥ 3	86	21	2,166	1.53	26
Case-control (in order of publication)								
Boston Program ³²	M	276	9 ≥ 6	66	36	772	1.39	
Jick <i>et al.</i> ^{33,34}	M	440	11 ≥ 6	79	24	1,736	1.48	
Hennekens <i>et al.</i> ³⁵	D	649	NG	30	43	538	1.16	
Wilhelmsen <i>et al.</i> ²⁷	M	230	50 ≥ 5	62	23	1,890	1.36	
Rosenberg <i>et al.</i> ⁹	M	491	22 ≥ 5	70	22	2,141	1.42	
Robenberg <i>et al.</i> ²	M	1,541	28 ≥ 5	74	15	4,526	1.45	
La Vecchia <i>et al.</i> ^{3,36}	M	262	23 ≥ 4	120	46	465	1.82	
Gerdes (unpublished)	M	57	61 ≥ 5	31	52	367	1.17	

* C = coronary disease incidence; D = coronary death; M = myocardial infarction (La Croix *et al.*⁶⁷ included angina); DM = coronary death or myocardial infarction.

† NG = quantity not given in study report.

‡ Percentage of subjects (cohort) or controls (case-control) drinking given number of cups per day; for example, 22 ≥ 5 indicates 22% of subjects reported drinking ≥ 5 cups per day.

§ Times 1,000.

¶ Case-control within cohort (nested case-control) study.

|| "Corrected" coefficient: uncorrected coefficient plus external adjustment.

*et al.*²⁵ was treated as a cohort study because it was methodologically identical to a historical cohort study, except for sampling from existing records to obtain matched control data. In particular, the entire source population for the study was available and used as a sampling frame, and all coffee and covariate data were obtained from pre-outcome records rather than post-outcome interviews.

I excluded the earliest case-control studies^{37–39} because no coefficient and standard error estimate could be calculated from the reports or because the coffee questions referred to consumption at time of interview. The study by Howell and Wilson⁴⁰ was also excluded for these reasons, and because it was a prevalence study of heart disease. Two hospital-based case-control studies from Greece^{41,42} were excluded because their case series were mixtures of infarct and coronary arteriogram patients; both of these studies exhibited larger positive associations than any study included here. One

historical cohort study of caffeine consumption and mortality among hypertensives⁴³ was excluded because a coffee regression coefficient could not be reliably estimated from the published data (coffee use was not directly reported). Two ecologic studies of coffee and coronary mortality^{44,45} were excluded on the same grounds. Finally, the study of Hrubec⁴⁶ was excluded because it examined angina prevalence only.

UNPUBLISHED MATERIAL

The published reports of Klatsky *et al.*⁴ LeGrady *et al.*¹⁰ and Dawber *et al.*²⁶ were supplemented by data kindly supplied by M. A. Armstrong, A. Dyer, and P. Sorlie. The cases in the unpublished Gerdes study were 57 men ages 35–55 admitted to an intensive care unit who survived at least 1 day and were diagnosed with acute myocardial infarction according to World Health Organization criteria. Exclusion criteria included psychosis, alcoholism, and foreign birth. Con-

trols were age-matched men free of heart disease selected from admissions to the orthopedic unit during the same period (29 controls) or from the population register for the hospital's service area (34 controls).

STATISTICAL METHODS

The objective of this meta-analysis was to examine between-study variation in the log relative risk regression coefficient b for the effect of coffee on myocardial infarction and/or coronary death risk. For cohort studies, b is the coefficient for the effect of each cup of coffee per day in a Cox proportional hazards, logistic regression, or Poisson exponential regression model; for case-control studies, b is a logistic regression coefficient.⁴⁷ In either case, e^{5b} is the approximate relative risk of myocardial infarction or sudden coronary death in five-cup-per-day drinkers compared with nondrinkers. For studies that did not directly supply such coefficient estimates,^{2,3,5,8,9,11,25-30,32-36} I computed estimates from published data using the methods given in articles by Greenland¹ and Greenland and Longnecker.⁴⁸ A description of the bias adjustment methods used below is given in the 1987 review.¹

For studies reporting category-specific relative risks rather than overall coefficients (slopes), it was necessary to assign each category a value in cups per day. For categories of 2 cups width or less, I assigned category midpoints. For wider categories, I assigned category means computed from published distributions of coffee consumption. For Yano *et al.*,⁸ the distribution published in their earlier report²⁰ was used to compute the category means for that report. For other U.S. studies, category means were computed using published survey distributions of coffee consumption⁴⁹; for Scandinavian studies, the distribution given by Wilhelmsen *et al.*²⁷ was used.

The basic summary meta-analysis statistics used here are inverse variance weighted averages and heterogeneity tests,¹ with random effects used when indicated. Also, coefficient estimates were modeled using a "meta-regression" approach.¹ Models were limited to fixed effects linear regression models of the form

$$E(b) = \alpha + \beta_1 x_1 + \dots + \beta_n x_n,$$

where $E(b)$ represents the parameter (coefficient) being estimated by a study with design features x_1, \dots, x_n . These models were fit by weighted least squares using inverse variance weights. (Random effects were not needed in the final model because of the small residual error.)

For interpretability, I transformed regression coefficient estimates to relative risks for 5 cups per day vs

none. The latter value was chosen because it was the most common cutpoint across studies, because it is a fairly common but relatively high consumption level in U.S. studies, and because model-based estimates for this level of consumption do not involve extrapolation beyond the data of any study. Readers who desire estimates for another consumption level, say x cups per day, may obtain them simply by computing e^{bx} . The exponential dose-response assumption implicit in the summary estimates was checked using scatterplot smoothers⁵⁰ fit to scatterplots of the coffee category-specific log rate ratio estimates, separately for case-control and cohort studies that supplied such estimates. Since these revealed no departures from monotonicity and only small downward departures from log-linearity, for brevity they are not presented.

Quality scoring was not used because, in my opinion, such scores inappropriately and subjectively combine heterogeneous predictors of study results. Instead, separate indicators of specific design features were examined as possible predictors of the estimated coefficients.

Statistical Results

Table 1 presents a summary of results from the studies in the present analysis. The estimated rate ratio for 5 cups per day is the antilog of $5 \times$ the corrected coefficient estimate, and the study weight is the inverse of the squared standard error. For example, for the Dawber *et al.* study,²⁶ $0.82 = e^{5(-0.039)}$ and $625 = 1/0.040^2$. For the most part, the weights reflect the number of cases in the study; exceptions occur when only a crude indicator of coffee consumption was analyzed by the study (such as "use/no use" or " ≤ 3 cups/ >3 cups"), which results in imprecise estimates of the regression slope b .

External adjustments were made to five of the study-specific coefficient estimates: Dawber *et al.*²⁶: -30; Heyden *et al.*²⁸: -69; Boston program³²: -31; Jick *et al.*^{33,34}: -31; Wilhelmsen *et al.*,²⁷ case-control only: -69. -30 is adjustment for lack of smoking control based on comparison of age- and age-smoking-adjusted results in Framingham mortality data.²⁶ -69 is adjustment for lack of covariate control based on comparison of crude and multivariate-adjusted results in the Wilhelmsen *et al.* cohort data.²⁷ -31 is the sum of adjustments of -27 for limited smoking control and -4 for selection bias; see Ref 1 for details of this correction. Because all of these adjustments are negative, their elimination would have increased the summary estimates, albeit slightly.

Because different studies used different category cutpoints in presenting coffee distributions, the column showing the percentage of subjects or controls

using substantial amounts of coffee (" $\geq n$ cups") had to employ a varying cutpoint n . Although this creates some difficulty in comparing studies, it is clear that coffee use in the Scandinavian studies (7, 27, 31, and Gerdes) is far above that in other studies. Also, boiled coffee consumption is far more prevalent among Scandinavian populations and has been more closely linked to elevated serum total cholesterol than filtered coffee.^{12,13} Despite the latter relation, the Scandinavian studies do not exhibit systematically higher estimates than the other studies and, in fact, are quite heterogeneous among themselves. Unfortunately, none of the studies presented results for different coffee preparation methods, and so preparation method could not be directly analyzed.

From Table 1, it appears that the case-control studies have tended to yield higher estimates than the cohort studies, and that the cohort study estimates have tended to increase over time. The later cohort studies also appear to show less consistency among themselves. The remaining statistical analyses confirm these impressions.

Table 2 presents a basic statistical summary of the studies in Table 1, stratified by design (case-control vs cohort), and the cohort studies further stratified by publication year ("earlier" = published 1981 or earlier,²³⁻²⁷ and "later" = published 1986 or later^{4-8,10,11,25-31}; 1984 is the mean publication year for the cohort studies). The ratio of the case-control and all-cohort geometric mean relative risks is 1.20, with 95% limits of 1.01, 1.43. Because of the severe heterogeneity among the cohort studies, random effects methods were used to calculate these statistics and the summary statistics for cohort studies.

The difference between the case-control and later

cohort studies does not appear large: the point estimate and 95% confidence limits for the ratio of relative risks from the case-control and later cohort studies are 1.12 and 0.99, 1.26 ($P = 0.08$). In contrast, the point estimate and 95% confidence limits for the ratio of relative risks from the later cohort studies and earlier cohort studies are 1.39 and 1.18, 1.64 ($P = 0.0001$), which is larger than the ratio of the case-control and later cohort results. A global test for residual heterogeneity within the three groups of studies (case-control, later cohort, and early cohort) yields an 19 degree-of-freedom chi-squared statistic of 22.7, $P = 0.25$. These results parallel the informal impressions obtained from Table 1.

The separation of cohort results into earlier and later publications was suggested by the data, and thus one could argue that the statistics based on this grouping overstate the significance of the differences. Nonetheless, even with adjustment for this problem (such as penalizing the test statistic by adding degrees of freedom) one still obtains an extremely small P -value for the difference between the earlier and later cohort estimates. The same results would be obtained using any cutpoint between 1981 and 1986.

Whereas the case-control results are quite consistent given their statistical variability, the cohort results are quite heterogeneous. To explore further the sources of heterogeneity among the cohort studies, I analyzed the cohort studies in Table 1 using a meta-regression that included a location indicator (1 if Scandinavia,^{6,7,27,30} 0 otherwise), follow-up time, and year of publication minus 1973, hereafter termed "year."

Table 3 presents the results for this meta-regression. Regression adjustment does not alter the basic differences apparent in Table 1: the Scandinavian studies have a 13% lower geometric mean rate ratio than the

TABLE 2. Summary Statistics for Meta-analysis (All Computations Are Based on the Corrected Coefficients)

	Case-Control	Cohort		
		All	Later (≥ 1986)	Earlier (≤ 1981)
Average coefficient (\bar{b})*	70	32.5	48.5	-17.4
SE of (\bar{b})†	9	13.3	8.6	14.5
RR for 5 cups/day‡	1.42	1.18	1.27	0.92
95% confidence limits for RR	1.30, 1.55	1.03, 1.34	1.17, 1.39	0.80, 1.06
Z statistic (\bar{b}/SE)	7.8	2.46	5.65	-1.20
Homogeneity chi-squared	2.9	35.1	17.4	2.34
Degrees of freedom	7	13	8	4
P-value	0.89	0.0008	0.03	0.67

* Times 1,000. \bar{b} is weighted average coefficient using inverse variance weights (with random effect for cohort studies).

† Standard error (SE) is square root of inverse sum of weights.

‡ Estimated weighted geometric mean rate ratio (RR) $e^{\bar{b}}$.

TABLE 3. Results of Multiple Linear Regression of Cohort Study Coefficients on Location (1 if Scandinavia, 0 Otherwise), Years Follow-up, and Year of Publication (Minus 1973)*

	Location	Years Follow-up	Publication Year
Meta-coefficient ($\hat{\beta}$)†	-26.8	1.08	6.41
Standard error ($\hat{\sigma}$)†	17.0	1.04	1.39
Estimated ratio of rate ratios‡	0.87	1.10	1.67
95% confidence limits	0.74, 1.03	0.92, 1.33	1.34, 2.08
Z statistic ($\hat{\beta}/\hat{\sigma}$)	-1.58	1.04	4.60
P-value for Z (2-sided)	0.12	0.30	<0.0001

* Residual chi-squared = 9.7 on 10 degrees of freedom, $P = 0.47$.

† Times 1,000.

‡ Estimated ratio of rate ratios for the effect of 5 cups of coffee per day. Compares Scandinavian vs non-Scandinavian studies for location, 20 years vs 2 years for follow-up, and 1990 vs 1974 for publication year; computed as $e^{\hat{\beta}}$ for location, $e^{\hat{\beta} \times 5}$ for follow-up, and $e^{\hat{\beta} \times 16}$ for publication year.

other studies, and the rate ratios for cohort studies rose an average of $e^{5(0.00641)} - 1 = 3\%$ per year over the review period. Surprisingly, cohort studies with longer follow-up times tended to have higher estimates.

Although the model fits very well, the P -values for fit and for the year coefficient cannot be regarded as valid because the data led to the definition and inclusion of year. Nevertheless, the results do suggest that the heterogeneity seen in Table 1 cannot be "explained" (in the statistical sense) without year in the model: the model fit was very poor ($P = 0.001$) when year was deleted. These findings were little changed when no correction factors were used in the analysis or when alternate category codes were used; for example, the year coefficient was 5.99 per 1,000 when no corrections were used. Further regressions (not shown) indicate that other variables (type of outcome, proportion subjects male, mean age of cohort) cannot explain the heterogeneity among the cohort studies and that the results are not attributable to any one study. There appeared to be no important association of any variable with the case-control results; variables examined were proportion male, mean age of subjects, source of controls (population, hospital), and year of study.

Only three studies^{2,5,9} presented separate results for decaffeinated coffee. Also, because of the lower frequency of decaffeinated coffee use, the estimates of decaffeinated coffee effects were very imprecise. Therefore, a meta-analysis was not attempted for decaffeinated coffee. The results of these studies of decaffeinated coffee are, however, within the ranges of the later studies in Table 1. Interestingly, Grobbee *et al*⁵ found a stronger association for decaffeinated coffee (estimated rate ratio for ≥ 4 cups of decaffeinated coffee per day of 1.5, compared to 1.1 among caffeinated coffee users), whereas Rosenberg *et al*^{2,9} found a

weaker association for decaffeinated coffee (estimated rate ratios for ≥ 5 cups decaffeinated coffee per day of 1.2 among women and 1.8 among men, compared to 1.7 and 2.1 among caffeinated coffee users). La Vecchia *et al*³ also presented an analysis for decaffeinated coffee but failed to exclude caffeinated coffee users from the reference (no-use) level; thus, their result (relative risk of 0.9 for ≥ 1 cup vs none) is probably biased downward due to the positive association in their study of caffeinated coffee and myocardial infarction.

Finally, four of the studies^{2-4,10} presented detailed results for smokers and nonsmokers separately. The differences within these studies were generally small and were not consistent across the studies, suggesting no important modification of coffee effect by smoking.

Discussion

The above analysis does not indicate reasons for the conflicts among the results of the case-control studies, the later cohort studies, and the early cohort studies (a conflict that persists even if no preliminary corrections are made to the study coefficients). It has been suggested that the case-control studies have exhibited higher estimates because of poorer control of confounding, selection bias, or recall bias.^{5,28} Most of the available evidence does not support these suggestions. For example, some case-control studies used only myocardial infarction survivors as cases, leading to suggestions that the apparent association is due to coffee use improving myocardial infarction survival. This explanation conflicts with the fact that survivor-based studies do not yield higher associations than the positive cohort studies and that coffee does not appear protective for coronary mortality.²⁶

The evidence also conflicts with the hypothesis that confounding is solely responsible for the higher case-

control estimates: control of confounding tended to be poorer in earlier than later cohort studies (for example, no diet variables were controlled in the earlier studies), yet the later studies have tended to give higher estimates, and several of these^{6,10,11,31} are higher than most of the case-control estimates.

Biased selection of controls has also been suggested as a source of the positive results in the early case-control studies, and external data have been presented to support this view.^{51,52} Nevertheless, alteration of control composition or selection within studies has not changed the case-control results in any important fashion,^{2,34} and Table 1 reveals no evidence of unusually low coffee use among the controls in the case-control studies relative to the cohorts. With regard to possible recall bias among case-control results, the Hennekens *et al* study³⁵ was based on recall by wives of coronary death cases, rather than recall by myocardial infarction survivors (as in all of the other case-control studies). Interestingly, this study gave one of the lowest case-control estimates. Nevertheless, one may doubt whether recall by the wife of a coronary death victim would be less biased than recall of an infarction survivor. Furthermore, unlike the cohort studies, later case-control studies do not tend to show higher estimates than the earliest case-control studies, despite media identification of coffee as a potential hazard in the wake of the earliest studies (and thus greater potential for recall bias in the ensuing studies).

One-half of the cohort studies have yielded relative risks for 5 cups per day of 1.3-1.5, entirely within the range of estimates obtained from case-control studies. Nevertheless, the general tendency of cohort studies to produce lower estimates than case-control studies is compatible with the hypothesis that coffee has an acute effect (perhaps in addition to a chronic effect due, for example, to effects on serum cholesterol). Resolution of this issue could be provided by cohort studies that analyze the association between measurement lag (time between measurement and outcome) and the relative risk for coffee effect: if the acute effect hypothesis is correct, the lag-relative risk association within each cohort should be negative (note that this is a prediction about the statistical interaction of time of measurement and magnitude of measurement within cohorts).

At least four cohort studies have published data bearing on the acute effect hypothesis. In the study by La Croix *et al*,³¹ coffee consumption and other possible risk factors (including smoking) were ascertained at 5-year intervals. Because of the small number of cases (only 37 in the multivariate analysis), results were imprecise. Nevertheless, results were consistent with

the acute effect hypothesis: the adjusted relative risk for the effect of ≥ 5 cups per day *vs* none increased from 1.8 (95% confidence limits 0.8, 4.0) at baseline to 2.5 (95% confidence limits 1.1, 5.8) when using the measurement nearest the outcome. In proportionate terms, this change is similar to the discrepancy between the early cohort and case-control results. In the study by Klatsky *et al*,⁴ relative risks were examined separately for myocardial infarction within 3 years of baseline measurement and for 3 or more years after baseline. Again, the results were consistent with the acute effect hypothesis: the adjusted relative risk for ≥ 4 cups per day *vs* none increased from 1.2 (95% limits 0.9, 1.7) for events 3 or more years after baseline, to 1.7 (95% limits 1.1, 2.4) for events within 3 years of baseline. Finally, in the study by Tverdal *et al*,⁶ county-specific relative risks were inversely correlated with average time from interview to end of follow-up, also consistent with the acute effect hypothesis.

In contrast to these studies, Grobbee *et al*⁵ found no association of total coffee use with any cardiovascular endpoint, despite the fact that coffee measurements were taken much nearer the outcome events than in other cohort studies. In a case study of coffee consumption patterns in myocardial infarction patients during the 26 hours before infarction, Mittleman *et al* (Mittleman M, Maclure M, Sherwood J, Goldberg R, Tofler G, Muller JE, unpublished) found no association of consumption patterns with onset of infarction (although there were few heavy coffee drinkers in this study). Finally, the follow-up time coefficient in the present meta-analysis did not show the negative sign one would expect if the acute effect hypothesis were true. Thus, the evidence regarding the hypothesis is inconsistent.

Particularly puzzling are the differences among the cohort studies. These differences can be "explained" (in the statistical sense) by the publication year variable, but the latter variable has no obvious biologic or epidemiologic meaning. One possibility is that the degree of confounding in the cohort study estimates has changed over time. For example, secular trends in cigarette consumption could alter the degree of residual confounding by smoking, especially if the trends differed by coffee use. Another possibility is that the heterogeneity is primarily due to variations in the degree of nondifferential misclassification of coffee use. This hypothesis fits with the fact that the coffee questions used in some early studies were relatively crude (only dichotomies in Klatsky *et al*²⁵ and Heyden *et al*²⁸). Nevertheless, even if one considers only the later cohort studies, one still sees considerable conflict, and

naive combination of results into a single summary estimate is improper.

Under certain assumptions, severe publication bias could explain the observed heterogeneity among the cohort studies. If results showing weak positive associations are much more likely to go unpublished than null or strongly positive results (perhaps because their results appear ambiguous), the pattern seen in Table 1 could result if all of the underlying effects were weakly positive. For example, suppose there were 12 additional cohort studies with coefficients of 30, 40, and 50 per 1,000 (four studies at each value) and standard errors of 25 per 1,000. When averaged with the rest of the cohort studies, the cohort studies would yield a homogeneity *P*-value of 0.06 and a geometric mean rate ratio of 1.2 with 95% limits of 1.1, 1.3. Such computations show that, in order for publication bias to explain the heterogeneity, nearly half of all cohort studies of coffee and heart disease must be unpublished, and all of these unpublished studies must fall in a narrow, weakly positive range.

One possibility is that heterogeneity among the cohort results is due to differences in the outcome variables used in the studies (for example, myocardial infarction, coronary death). Unfortunately, as may be apparent from Table 1 and confirmed by further regressions, the differences in outcome variables as summarized in Table 1 do not account for the heterogeneity.

Setting aside the conflicts among the studies, one may note that all except two relatively imprecise estimates of the 5-cup relative risk^{3,27} fall around 1.5 or less. Using the arguments given by Cornfield *et al.*,⁵³ it is not difficult to show that small relative risks could arise from combined residual confounding by such factors as exercise, diet, and smoking (in the best studies, only gross summaries of smoking history, such as cigarettes per day, were controlled). On the other hand, there are many plausible routes for small coffee effects on heart disease. Possible coffee effects on serum lipids and blood pressure provide mechanisms for chronic effects on coronary disease.¹²⁻¹⁸ Although the lipid effects appear most strongly associated with consumption of boiled coffee, which is uncommon in the U.S., the coffee effects on calcium balance⁵⁴ would provide another pathway for chronic effects if calcium is preventive for coronary disease, as has been hypothesized. Possible coffee effects on cardiac arrhythmias⁵⁵ may suggest mechanisms for acute coffee effects on coronary death.

Assuming a coffee effect on the order of a 5% rate increase per cup per day (the mean effect in later

cohort studies) would imply a considerable number of myocardial infarctions attributable to coffee use, given the frequencies of heavy coffee use and of myocardial infarction. Under an exponential dose-response assumption, it would also imply a 65% rate increase among persons drinking ≥ 10 cups per day. Further cohort studies of coffee and cardiovascular disease would thus be worthwhile, provided that they obtain longitudinal measurements of coffee (separately for caffeinated and decaffeinated), serum lipids, and blood pressure and analyze them in a manner that accounts for the possible intermediate nature of cholesterol and blood pressure.⁵⁶

The present meta-analysis will inevitably be contrasted with that of Myers and Basinski,⁵⁷ who reported point estimates of 1.01 and 1.09 for 4-6 and 6 cups per day (95% limits of 0.93, 1.11 and 0.97, 1.22). In my opinion, the latter analysis was improperly conducted and interpreted. Perhaps the most severe problem of the Myers and Basinski analysis is that it combined the cohort studies into a single, summary estimate, without even a random effect to account for the heterogeneity. Such a procedure underestimates the standard error and produces invalidly narrow confidence limits. Also, Myers and Basinski summarized individual study results via "vote counting," which is especially misleading when small associations are present (for then most studies will be "negative" by the statistical significance criterion used by Myers and Basinski).^{1,58} Both problems encourage the conclusion preferred by Myers and Basinski, that coffee and heart disease have no association. Their conclusion appears to be based on the fact that the confidence intervals from their analysis include 1, but it ignores the fact that their results are also compatible with small positive relative risks.

A technical source of disparity is the fact that the current analysis is based on combining coefficients taken from the individual studies, whereas Myers and Basinski⁵⁷ entered tabulated data reported by each study into a single logistic regression. The two approaches will not be equivalent unless confounding is absent from all studies and the true dose-response relations are homogeneously logistic. Neither assumption can be exactly true, and the first is very doubtful. I would argue that the adjusted coefficient reported by a study represents a less confounded source of information than the summary tabular data, since only the former is adjusted for age, smoking, and other covariates entered into the regression.

Perhaps the most serious source of conflict between the present analysis and that of Myers and Basinski is the choice of studies to include. I chose to use the

earlier²⁶ rather than later²³ Framingham data, because the former allowed detailed age adjustment and used a more specific outcome. Since both studies yielded near-null estimates, this choice probably made little difference. Myers and Basinski excluded the very positive findings of the Adventist study^{11,22} because the latter could not be incorporated into the logistic regression used by Myers and Basinski, whereas the present analysis excluded the null findings of Martin *et al.*,⁴³ because no coffee-myocardial infarction coefficient could be reliably estimated from the published report. Inclusion of both cohorts in the same analysis would only further increase the already strong appearance of heterogeneity among the cohort studies.

Finally, Myers and Basinski summarily excluded all case-control data, without evaluating whether biases were actually present in such data. I do not think the case-control studies should be ignored unless the biases hypothesized for these studies have been shown to be responsible for their results.

Acknowledgments

The author would like to thank Mary Anne Armstrong, Alan Dyer, Gary Friedman, Ulrik Gerdes, Matthew Longnecker, and Paul Sorlie for providing unpublished information used in this analysis, and Alan Dyer, Matthew Longnecker, Malcolm MacLure, Charles Poole, and Lynn Rosenberg for their helpful comments.

References

- Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987;9:1-30.
- Rosenberg L, Palmer JR, Kelly JP, Kaufman DW, Shapiro S. Coffee drinking and nonfatal myocardial infarction in men under 55 years of age. *Am J Epidemiol* 1988;128:570-578.
- La Vecchia C, Gentile A, Negri E, Parazzini F, Franceschi S. Coffee consumption and myocardial infarction in women. *Am J Epidemiol* 1989;130:481-485.
- Klatsky AL, Friedman GD, Armstrong MA. Coffee use prior to myocardial infarction restudied: heavier intake may increase risk. *Am J Epidemiol* 1990;132:479-488.
- Grobbee DE, Rimm EB, Giovannucci E, Colditz G, Stampfer M, Willett W. Coffee, caffeine, and cardiovascular disease in men. *N Engl J Med* 1990;323:1026-1032.
- Tverdal A, Stensvold I, Solvoll K, Foss OP, Lund-Larsen P, Bjartveit K. Coffee consumption and death from coronary heart disease in middle-aged Norwegian men and women. *Br Med J* 1990;300:566-569.
- Rosengren A, Wilhelmsen L. Coffee, coronary heart disease and mortality in middle-aged Swedish men: findings from the Primary Prevention Study. *J Intern Med* 1991;230:67-71.
- Yano K, Reed DM, MacLean CJ. Letter to Editor. *N Engl J Med* 1987;316:946.
- Rosenberg L, Werler MM, Kaufman DW, Shapiro S. Coffee drinking and myocardial infarction in young women: an update. *Am J Epidemiol* 1987;126:147-149.
- LeGrady D, Dyer AR, Shekelle RB, Stamler J, Liu K, Paul O, Lepper M, Shryock AM. Coffee consumption and mortality in the Chicago Western Electric Company Study. *Am J Epidemiol* 1987;126:803-812.
- Lindsted KD, Kuzma JW, Anderson JL. Coffee consumption and cause-specific mortality: association with age at death and compression of mortality. *J Clin Epidemiol* 1992;45:733-742.
- Stensvold I, Iverdal A, Foss OP. The effect of coffee on blood lipids and blood pressure: results from a Norwegian cross-sectional study, men and women, 40-42 years. *J Clin Epidemiol* 1989;42:877-884.
- Pietinen P, Aro A, Tuomilhto J, Uusitalo V, Korhonen H. Consumption of boiled coffee is correlated with serum cholesterol in Finland. *Int J Epidemiol* 1990;19:586-590.
- Salvaggio A, Periti M, Miano L, Anaglia G, Marzorati D. Coffee and cholesterol, an Italian study. *Am J Epidemiol* 1991;134:149-156.
- Bak AA, Grobbee DE. The effect on serum cholesterol levels of coffee brewed by filtering or boiling. *N Engl J Med* 1989;67:97-103.
- Superko HR, Burtz W, Albers JJ, Wood PD. Lipoprotein and apolipoprotein changes during a controlled trial of caffeinated and decaffeinated coffee drinking in men (Abstract). *Circulation (suppl II)* 1989;80:II-86.
- Bak AA, Grobbee DE. A randomized study on coffee and blood pressure. *J Hum Hypertens* 1990;4:259-264.
- Fried RE, Levine DM, Kwiterovich PO, Diamond EL, Wilder LB, Moy TF, Pearson TA. The effect of filtered-coffee consumption on plasma lipid levels: results of a randomized clinical trial. *JAMA* 1992;267:811-815.
- Rosenberg L, Slone D, Shapiro S, Kaufman DW, Stolley PD, Miettinen OS. Coffee drinking and myocardial infarction in young women. *Am J Epidemiol* 1980;111:675-681.
- Yano K, Rhoads GG, Kagan A. Coffee, alcohol and risk of coronary heart disease among Japanese men living in Hawaii. *N Engl J Med* 1977;297:405-409.
- Paul O. Stimulants and coronaries. *Postgrad Med* 1968;44:196-199.
- Snowdon DA, Phillips RL, Fraser GE. Meat consumption and fatal ischemic heart disease. *Prev Med* 1984;13:490-500.
- Wilson PWF, Garrison RJ, Kannel WB, McGee DL, Castelli WP. Is coffee consumption a contributor to cardiovascular disease? *Arch Intern Med* 1989;149:1169-1172.
- Welin L, Svardsudd K, Tibblin G, Wilhelmsen L. Coffee, traditional risk factors, coronary heart disease, and mortality. In: McMahon B, Sugimura T, eds. *Coffee and Health*. Banbury Report 17. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1984;219-229.
- Klatsky AL, Friedman GD, Siegel AB. Coffee drinking prior to acute myocardial infarction. *J Am Med Assoc* 1973;226:540-543.
- Dawber TR, Kannel WB, Gordon T. Coffee and cardiovascular disease. *N Engl J Med* 1974;291:871-874.
- Wilhelmsen L, Tibblin G, Elmfeldt D, Wedel H, Werkö L. Coffee consumption and coronary heart disease in middle-aged Swedish men. *Acta Med Scand* 1977;201:547-552.
- Heyden S, Tyroler HA, Heiss G, Hames CG, Bartel A. Coffee consumption and total mortality. *Arch Intern Med* 1978;138:1472-1475.
- Murray SS, Bjelke E, Gibson RW, Schuman LM. Coffee consumption and mortality from ischemic heart disease and other causes: results from the Lutheran Brotherhood Study. *Am J Epidemiol* 1981;113:661-667.
- Jacobsen BK, Bjelke E, Kvale G, Heuch I. Coffee drinking, mortality, and cancer incidence: results from a Norwegian prospective study. *J Natl Cancer Inst* 1986;76:823-831.
- La Croix AZ, Mead LA, Liang KY, Thomas CB, Pearson TA. Coffee consumption and the incidence of coronary heart disease. *N Engl J Med* 1986;315:977-982.
- Boston Collaborative Drug Surveillance Program. Coffee drinking and acute myocardial infarction. *Lancet* 1972;2:1278-1281.

33. Jick H, Miettinen OS, Neff RK, Shapiro S, Heinonen OP, Slone D. Coffee and myocardial infarction. *N Engl J Med* 1973;289:63-67.
34. Jick H. Re: coffee and myocardial infarction (Letter). *Am J Epidemiol* 1981;113:103-104.
35. Hennekens CH, Drolette ME, Jesse MJ, Davies JE, Hutchison GB. Coffee drinking and death due to coronary heart disease. *N Engl J Med* 1976;294:633-636.
36. Gramenzi A, Gentile A, Fasoli M, Negri E, Parazzini F, La Vecchia C. Association between certain foods and risk of acute myocardial infarction. *Br Med J* 1990;300:771-773.
37. Brown A. Coronary thrombosis: an environmental study. *Br Med J* 1962;567-573.
38. Yudkin J, Roddy J. Levels of dietary sucrose in patients with occlusive atherosclerotic disease. *Lancet* 1964;2:6-8.
39. Little JA, Shanoff HM, Csima A, Redmond SE, Yano R. Diet and serum lipids in male survivors of myocardial infarction. *Lancet* 1965;1:933-935.
40. Howell RW, Wilson DG. Dietary sugar and ischaemic heart disease. *Br Med J* 1969;3:145-148.
41. Kalandidi A, Tzonou A, Toupadaki N, Lan S-J, Koutis C, Drogari P, Notara V, Hsieh C-C, Toutouzias P, Trichopoulos D. A case-control study of coronary heart disease in Athens, Greece. *Int J Epidemiol* 1993;21:174-180.
42. Trichopoulos D, Tzonou A, Christopoulos C, Havatzoglou S, Trichopoulou A. Siesta and risk of coronary heart disease. *Stress Med* 1988;4:143-148.
43. Martin JB, Annegers JF, Curb JD, Heyden S, Howson C, Lee ES, Lee M. Mortality patterns among hypertensives by reported level of caffeine consumption. *Prev Med* 1988;17:310-320.
44. Hemminki E, Pesonen T. Regional coffee consumption and mortality from ischemic heart disease in Finland. *Acta Med Scand* 1977;201:127-130.
45. Brummer P. Coronary mortality and living standard. II. Coffee, tea, cocoa, alcohol and tobacco. *Acta Med Scand* 1969;186:61-63.
46. Hrubec Z. Coffee drinking and ischemic heart disease (Letter). *Lancet* 1973;1:548.
47. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. New York: Chapman and Hall, 1989.
48. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* 1992;136:1301-1309.
49. Bonham GS. *Use Habits among Adults of Cigarettes, Coffee, Aspirin, and Sleeping Pills*. Vital Health Statistics No. 131. Washington DC: U.S. Government Printing Office, 1976.
50. Hastie T, Tibshirani R. *Generalized Additive Models*. New York: Chapman and Hall, 1990.
51. Rosenberg L, Slone D, Shapiro S, Kaufman DW, Miettinen OS. Case-control studies on the acute effects of coffee upon the risk of myocardial infarction. *Am J Epidemiol* 1981;113:646-652.
52. Silverman DT, Hoover RN, Swanson M, Hartge P. The prevalence of coffee drinking among hospitalized and population-based control groups. *JAMA* 1983;249:1877-1880.
53. Cornfield J, Haenszel W, Hammond EC, Lillienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: evidence and a discussion of some questions (Appendix A). *J Natl Cancer Inst* 1959;22:173-203.
54. Kiel DP, Felson DT, Hannan MT, Anderson JJ, Wilson PWF. Caffeine and the risk of hip fracture: the Framingham study. *Am J Epidemiol* 1990;132:675-684.
55. Prineas RJ, Jacobs DR, Crow RS, Blackburn H. Coffee, tea, and VLB. *J Chron Dis* 1980;33:67-72.
56. Robins JM. The control of confounding by intermediate variables. *Stat Med* 1989;8:679-701.
57. Myers MG, Basinski A. Coffee and coronary heart disease. *Arch Intern Med* 1992;152:1767-1772.
58. Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press, 1984.



สมาคมข้าราชการวิทยาสานาม

โทร.0 2590 1790

โทรสาร 0 2590 1784