DISTINGUISHING ASSOCIATION FROM CAUSATION:

A BACKGROUNDER FOR JOURNALISTS

Written for the American Council on Science and Health

By

Kathleen Meister, M.A.

Based on a technical paper by

William P. Kucewicz, John W. Morgan, Dr.PH.[1], and Diana M. Torres, BS[1]

October 29, 2007

1. Loma Linda University

TABLE OF CONTENTS

ACSH accepts unrestricted grants on the condition that it is solely responsible for the conduct of its research and the dissemination of its work to the public. The organization does not perform proprietary research, nor does it accept support from individual corporations for specific research projects. All contributions to ACSH—a publicly funded organization under Section 501(c)(3) of the Internal Revenue Code—are tax deductible.

Executive Summary

- Scientific studies that show an *association* between a factor and a health effect do not necessarily imply that the factor *causes* the health effect. Many such studies are preliminary reports that cannot justify any valid claim of causation without considerable additional research, experimentation, and replication.
- Randomized trials are studies in which human volunteers are randomly assigned to receive either the agent being studied or an inactive placebo, usually under double-blind conditions (where neither the participants nor the investigators know which substance each individual is receiving), and their health is then monitored for a period of time. This type of study can provide strong evidence for a causal effect, especially if its findings are replicated by other studies. Such trials, however, are often impossible for ethical, practical, or financial reasons. When they can be conducted, the use of low doses and brief durations of exposure may limit the applicability of their findings.
- The findings of animal experiments may not be directly applicable to the human situation because of genetic, anatomic, and physiologic differences between species and/or because of the use of unrealistically high doses.
- In vitro experiments are useful for defining and isolating biologic mechanisms but are not directly applicable to humans.

- Observational epidemiologic studies are studies in human populations in which researchers collect data on people's exposures to various agents and relate these data to the occurrence of diseases or other health effects among the study participants. The findings from studies of this type are directly applicable to humans, but the associations detected in such studies are not necessarily causal.
- Useful, time-tested criteria for determining whether an association is causal include:
-
    - *Temporality.* For an association to be causal, the cause must precede the effect.
    - *Strength.* Scientists can be more confident in the causality of strong associations than weak ones.
    - *Dose-response.* Responses that increase in frequency as exposure increases are more convincingly supportive of causality than those that do not show this pattern.
    - *Consistency.* Relationships that are repeatedly observed by different investigators, in different places, circumstances, and times, are more likely to be causal.
    - *Biological plausbility.* Associations that are consistent with the scientific understanding of the biology of the disease or health effect under investigation are more likely to be causal.

- New research results need to be interpreted in the context of related previous research. The quality of new studies should also be assessed. Those that include appropriate statistical analysis and that have been published in peer-reviewed journals carry greater weight than those that lack statistical analysis and/or have been announced in other ways.
- Claims of causation should never be made lightly. Premature or poorly justified claims of causation can mislead people into thinking that something they are exposed to is endangering their health, when this may not be true, or that a useless or even dangerous product may produce desirable health effects.

DISTINGUISHING ASSOCIATION FROM CAUSATION:

A BACKGROUNDER FOR JOURNALISTS

Introduction

Hardly a day goes by without a new headline about the supposed health risks or benefits of some food component, pharmaceutical product, environmental contaminant, dietary supplement, or other substance. But are these headlines justified? Often, the answer is no. Although the news reports are based on the results of scientific studies, in many instances the evidence is insufficient to justify the conclusion that the substance actually *caused* the health effect rather than merely being *associated* with it.

Journalists who report on health issues often face the problem of distinguishing association from causation. A study that shows an association between factor X and health effect Y in cultured cells, in experimental animals, or even in a human population group does not necessarily imply that X causes Y. Many such studies are preliminary reports that cannot justify any valid claim of causation without considerable additional research, experimentation, and replication. Reports on such studies can mislead readers or viewers into thinking that something they consume or something they are exposed to in their surroundings is endangering their health — or that some useless or even dangerous product may produce desirable health effects.

The purpose of this report is to provide insight into how to use the methods of science itself to help distinguish association from causation. The report will address such questions as

- What kinds of studies provide evidence for causation?
- What criteria can be used to assess whether an association is causal?
- What kinds of factors other than a causal relationship might be responsible for an association between an exposure and a health effect?

To address these complex issues, it is necessary to understand the different types of studies that may link a factor with a health effect and the methods, such as statistical analysis and peer review, that scientists use to judge the validity of a new research finding.

Randomized Trials

Randomized trials are designed to isolate the effect of a single factor —X — and to eliminate or control extraneous factors that might affect the results. In a trial of this type, volunteers are randomly assigned to receive either X or an identical-appearing inactive placebo, and neither the study participants nor the investigators who examine the subjects for health effects know which substance each participant is receiving. At the end of the study, the data are analyzed to determine whether the frequency of effect Y in those receiving substance X differed from that in those receiving the placebo.

A randomized trial can provide strong evidence for a causal effect, especially if its findings are replicated by other studies. But unfortunately, in the real world, except for relatively short-term trials conducted as part of the evaluation of the safety and efficacy

of new drugs, this type of study is usually not possible for ethical, practical, or financial reasons. Among these practical limitations, potentially harmful substances cannot be tested in randomized trials, unless the harm is minimal and the duration of exposure is brief, because deliberately exposing people to serious harm is unethical. For the same reason, substances for which there is compelling evidence of benefit cannot be tested in randomized trials because subjects assigned to the placebo group are denied the protective benefit.

Many exposures (e.g., exercise, dietary changes) cannot be evaluated in a double-blind trial because they cannot be concealed from the study participants and because 100% compliance by study subjects cannot be expected. Additionally, the levels of exposure to potentially harmful agents are necessarily minimized in trials that hold safety among study subjects above other considerations. This practical restriction limits randomized trials to assessment of exposures that immediately precede development of diseases like cancer, heart disease, and stroke, which develop over long periods of time. For example, in three randomized trials that evaluated the possible preventive effect of beta-carotene against lung cancer, participants received beta-carotene supplements for an average of 4, 6, and 12 years, respectively, but lung cancer develops over a period of 20 years or more. Findings from randomized trials assessing recent and low-dose exposures will frequently differ from those attributable to higher concentrations of the same agent and to exposures that occurred decades earlier.

Although randomized trials can only be used in restricted settings, they provide information that is useful when making statements about cause-and-effect. For example, randomized trials established the benefit of aspirin in preventing heart attacks — a

finding that has prompted widespread use of aspirin for this purpose. Randomized trials altered thinking about the once-promising hypothesis that beta-carotene could reduce lung cancer risk. This led to recognition by researchers and clinicians that the role of beta-carotene in cancer prevention is more complex than previously believed, i.e., supplementation with beta-carotene did not support the hypothesis generated from the observational studies.

Other Types of Studies

Since randomized trials are usually unavailable and their findings are imperfect, scientists must explore causation on the basis of other study types, including animal experiments, *in vitro* (test tube or cell culture) experiments, and observational epidemiologic studies in human populations. It is important to note that *no single study of any type can justify a claim that factor X causes health effect Y.* Instead, any new finding must be considered in conjunction with the entire body of scientific evidence on the topic to determine whether causality is likely.

Animal Experiments

Scientists can perform experiments on animals that would be impossible in humans. They can deliberately expose them to dangerous substances, often using doses much higher than those to which humans would ever be exposed. They can sacrifice the animals and examine tissues from their bodies in minute detail. If they choose a short-lived species, they can expose the animals to a substance for a lifetime — or even for the lifetimes of several successive generations. From such studies, scientists can generate large amounts

of precisely measured and potentially useful data, although the relevancy of these findings to humans is often questionable.

People are not big rats. Species differ in a variety of genetic, anatomic, and physiologic properties that each contribute to risk of disease. Saccharin provides a classic research example of the difference between rats and humans. Several decades ago, the discovery that lifetime exposure to large doses of sodium saccharin caused bladder cancer in male rats nearly led to the removal of this sweetener from the food supply in the United States. Subsequent research showed, however, that the harmful effect of sodium saccharin in the rat bladder is due to a mechanism that does not occur in humans (or even in female rats).

The usual practice of using high doses in animal experiments can also lead to results that may not be applicable to the human situation. High doses are used to increase the likelihood of detecting an effect, if one exists, but they also create the possibility of producing effects that would never be encountered among humans.

In Vitro Studies

In vitro experiments are conducted in cell or tissue cultures or involve isolated chemicals. They are useful for defining and isolating biologic mechanisms but are not directly applicable to humans. Conditions in a living organism are drastically different from those in these simplified experimental models. Thus, the results of *in vitro* studies are useful only as a small part of a larger body of scientific evidence and should never be taken, alone, as evidence of causality.

Observational Epidemiologic Studies

Epidemiology is the study of the occurrence of disease in human populations. Most epidemiologic studies are observational, meaning that study participants determined their own exposures (for example, they chose which foods they ate and whether or not they took dietary supplements), with this information systematically recorded by researchers.

One type of epidemiologic investigation, called an *ecologic study*, compares average measures of exposure and disease for entire population groups. For example, researchers might obtain data on exposures to various environmental contaminants in different countries to see whether any association exists between such exposures and the occurrence of a disease in those countries. Studies of this type are relatively easy and inexpensive and often rely on data collected for other purposes. But they have a critical limitation; they cannot show whether the persons exposed to the factor are the same ones who developed the health effect under investigation. Because of this limitation, ecologic studies derive their greatest value by generating new hypotheses for future study, rather than testing hypotheses about causation.

For example, data from various countries have shown a strong correlation between dietary fat intake and the risk of death from breast cancer, with higher death rates from this disease in countries where fat intakes are higher. These data raised questions about whether dietary fat might play a role in breast cancer causation. However, other types of evidence, including epidemiologic studies that focused on individuals rather than populations, have not supported this hypothesis. In those studies, the dietary fat intakes of individual women were not associated with their later

development of breast cancer. (This does not, of course, mean that there is no correlation between *body* fat and breast cancer.)

One common type of epidemiologic study that focuses on individuals is the *retrospective case-control study*, which uses interviews and medical records to compare the past histories and lifestyles of individuals who have been diagnosed with a disease (cases) with those of otherwise similar individuals (controls) who have not. Case-control studies can be conducted quickly and can be used to investigate any disease — even very rare ones. Studies of this type, however, are prone to bias caused by the different ways in which people who are ill and those who are well recall past events. Collecting valid information for past exposures may differ for cases and controls and may also be difficult because people's recall of the distant past may be poor. In addition, it is challenging to choose control subjects who are truly comparable to the disease cases regarding a multitude of other potentially important characteristics.

In another type of epidemiologic study, called a *prospective cohort study*, information is collected about the lifestyles, exposures, and health of a group of people (the cohort), none of whom, at the start of the study, have the disease under investigation. Follow-up data are then collected from these people for a period of time, often many years. Those who later develop the disease are compared to those who do not to see how their exposures and experiences were different. This type of study design avoids some of the problems inherent in case-control studies. People are not asked to recall events from the distant past, and all subjects are well at the time when they are interviewed regarding exposure. In addition, the researchers don't have to make special efforts to ensure that ill subjects and well subjects are otherwise similar because all participants are drawn from

the same population. Prospective cohort studies have limitations of their own, however. They are time-consuming and expensive, and they require a large number of study participants. Moreover, cohort studies are cost prohibitive for investigation of rare diseases in which follow-up of extraordinarily large cohorts would be required.

In all observational epidemiologic studies, findings of an association between a substance or exposure and a health effect do not necessarily imply causation. For example, a study might show that the habit of carrying matches is associated with an increased likelihood of later developing lung cancer. But this effect is not causal; it is due to a *confounding factor* — a third factor that is associated with both the health effect and a true causative agent — in this instance, cigarette smoking.

The match-carrying example may seem farfetched, but the problem of distinguishing causal relationships from those attributable to confounding factors is real and serious, and it arises in many epidemiologic investigations. For example, prior to the randomized trials of beta-carotene mentioned earlier, numerous observational epidemiologic studies had indicated that people who consumed generous amounts of beta-carotene in their daily diets had lower risks of lung cancer than those who consumed little beta-carotene. Indeed, it was this evidence that prompted scientists to undertake the lengthy and expensive randomized trials that unexpectedly showed that beta-carotene was ineffective at preventing lung cancer.

The debate continues among scientists about reasons for the differences in the findings in the observational studies and randomized trials of beta-carotene. The results of these studies may differ because of differences in the timing and levels of exposure studied.

As already mentioned, randomized trials assess associations between differences in exposure to specific agents, that are frequently quite modest, during a relatively short time-period between exposure and disease assessment. In contrast, observational epidemiologic studies frequently assess exposure differences that are more substantial, with the timing between exposure and disease spanning decades. Just as cigarette smoking is unassociated with risk of lung cancer until two decades after smoking is initiated, difference in timing and exposure dose that are inherent in the two study types are more than sufficient to account for the differences in beta-carotene findings produced in the two study types. As of this writing, then, we still cannot say with certainty if or to what extent beta-carotene is protective against lung cancer.

Some Useful Terminology

To fully understand epidemiologic reports, it is necessary to know the meaning of some terms commonly used by epidemiologists, including the following.

Incidents, Incidence, Risk, Prevalence, and Mortality

The term *incidents* refers to new cases of a disease (e.g., people newly diagnosed with diabetes), while *incidence*, is the risk of disease in a population. Epidemiologists sometimes refer to *incidents* as *incident cases* to ensure distinction between the phonetically similar terms, *incidents* and *incidence*. Incidence is measured as the number of incident (new) cases, divided by the size of the susceptible population in which they developed. Incidence is identical to risk and measures the probability of developing disease within a susceptible population.

In contrast, the number of prevalent cases refers to existing disease cases (e.g., people with diabetes, including both those newly diagnosed and those who have had the condition for some time). Like incidence, prevalence is measured as a rate and represents the number of existing cases (new and old cases), divided by the size of the population in which they are measured. The *mortality rate* refers to the risk of death in a population (e.g., those who died from diabetes, divided by the size of the population in which the deaths were counted).

Data are typically expressed in terms of a population of a particular size. For example, U.S. government statistics for 2004 express the total mortality rate for that year as "816.5 deaths per 100,000 population" and infant mortality as "6.8 deaths per 1,000 live births."[1] Some rates are frequently presented as percentages to display findings in a manner that is intuitive to the audience (e.g., "In 2004, 34% of American adults were obese."[2]).

If the sizes of populations are not taken into account, data may be misinterpreted. For example, consider deaths in 2004 in two of the New England states, Maine and Connecticut.[3] A total of 29,289 people died in Connecticut that year, while 12,405 died in Maine. But even though the total number of deaths was higher in Connecticut, the *mortality rate* for that year, which is a more meaningful piece of information, was higher in Maine (941.7 per 100,000 population vs. 836.0 per 100,000 population in Connecticut).

---

[1] Data from the U.S. Centers for Disease Control and Prevention. Available online at http://www.cdc.gov/nchs/fastats/deaths.htm.
[2] Data from the U.S. government publication *Health, United States, 2006*. Available online at http://www.cdc.gov/nchs/hus.htm.
[3] All Maine/Connecticut data were taken from the Centers for Disease Control and Prevention's National Vital Statistics Reports, Vol. 54, No. 19, June 28, 2006. Available online at http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_19.pdf.

The ages of the members of a population also need to be taken into account when interpreting data. Both total death rates and rates of major chronic diseases, such as heart disease and cancer, are higher in older people than younger ones. To balance for the effects of age, incidence and mortality rates are often *age-adjusted*. The adjusted rate is composed of a weighted average for specific age groups. Age-adjustment or age-standardization allows comparison of disease rates in different populations that are produced by all characteristics other than age.

A researcher who was interested in finding out whether the higher death rate in Maine than in Connecticut results from characteristics other than age differences would want to compare age-adjusted death rates rather than unadjusted ones. The age-adjusted death rates in Connecticut and Maine in 2004 were 705.6 and 803.6, respectively, per 100,000 U.S. standard population. Even after age is taken into account, the death rate in Maine is higher than that in Connecticut, signifying that the difference is attributable to some other reason.

Relative Risk, Risk Ratio, and Odds Ratio

Relative risk (also called *risk ratio* and often abbreviated RR) is a measure that compares the risk of a disease or other event in a group of people exposed to a particular substance or condition to that in a comparison group (typically an unexposed group or one with a low level of exposure). For example, if a study shows a relative risk of disease A of 2.0 in a group of people exposed to a certain factor, as compared to those who were not exposed, that means that the exposed people are twice as likely to develop the disease. A relative risk of 1 indicates no difference in risk between the two groups, and a relative

risk of less than 1 indicates that the exposed group has a lower risk than the comparison group.

Odds ratio (also called relative odds and often abbreviated OR) is a similar though not identical measure that is frequently used in case-control studies. It compares the odds of an event occurring in one group of people to the odds of it occurring in another group. An odds ratio of 1 means that the event is equally likely in both groups. An odds ratio greater than 1 indicates that the event is more frequent in the first group. An odds ratio less than 1 means that it is less frequent in the first group.

Criteria for Distinguishing Association from Causation

Since no study, regardless of type, yields perfectly valid findings, how can association be distinguished from causation? Scientists have long wrestled with this question and have established criteria that have proven helpful in making this distinction. One of the best known is a set of criteria proposed by British epidemiologist Austin Bradford Hill in 1965. In the subsequent decades, five of the these criteria, as summarized in the table and discussed below, have proven particularly useful.

Most Useful Criteria for Deciding That an Association Is Causal

Criterion	Comments
Temporality	Cause precedes effect
Strength	Large relative risk
Dose-response	Larger exposures associated with higher frequency of effect
Consistency	Repeatedly observed by different investigators, in different places, circumstances, and times
Biological plausibility	Causal interpretation is congruent with knowledge of the natural history/biology of the effect

- *Temporality.* For a relationship to be causal, the cause must precede the effect. Considerations of temporality are especially noteworthy for diseases that take a long time to develop, such as cancer. Thus, a change in cancer rates in the year 2000 could not have been caused by a change in exposure to an environmental chemical during the same year.
- *Strength.* Scientists can be more confident in the causality of strong associations (those with a large relative risk) than weak ones. When an association is strong, it is more likely to be causal (as is true for the relationship between cigarette smoking and lung cancer, which has a relative risk of at least 10) or due to a readily identifiable confounding factor (as is true for the relationship between the carrying of matches and lung cancer). The causality of weaker relationships is more difficult to establish because such relationships could easily be due to subtle confounding factors that are hard to identify.

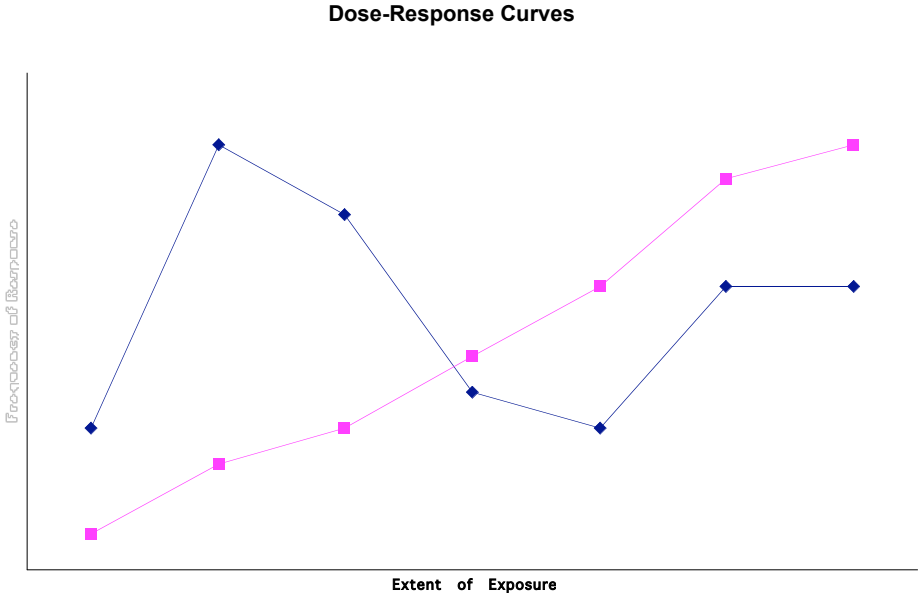
  Epidemiologists would look at increases in relative risk of less than 2 as being no better than moderate and would be reluctant to label them as causal unless there is a great deal of additional evidence in support of the relationship. Weak relationships of this type could easily be produced by confounding that might never be detected. Small increases in relative risk, however, may be of great public health importance, especially if they pertain to very common diseases, such as heart disease.

  Approximately 700,000 people die of heart disease in the United States each year.[4] If an epidemiologic study indicated that some factor was associated

[4] Data from the U.S. Centers for Disease Control and Prevention. Available online at http://www.cdc.gov/heartdisease/facts.htm#facts.

with a 50% increase in the heart disease death rate (that is, a relative risk of 1.5), that would mean an additional 350,000 deaths per year — a very meaningful change in terms of public health. It would be important — but also very challenging — to determine whether this association is causal. To investigate causality, scientists would need to replicate the original finding, study the association in other populations and using other types of studies, and investigate its biological plausibility and whether or not a dose-response relationship exists, as described below. The dilemma that surrounds weak associations that are potentially important but are marked by uncertainty about their causal nature makes the interpretation of such research findings challenging.

- *Dose-response.* The term "dose-response" (Bradford Hill used the term "biological gradient") means that the likelihood or intensity of a biological effect is greater in people or animals with greater exposures to an agent than in those with lesser exposures. The presence of a dose-response relationship tends to support causality. In the graph below, the association indicated by the pink line, where the frequency of response increases as exposure increases, is more convincingly supportive of a causal relationship than the one indicated by the blue line, where response does not vary predictably with the extent of exposure. It is important to note, though, that a dose-response relationship might also be due to a confounding factor that varies in intensity along with the factor under investigation (for example, people who carry matches more frequently might be heavier smokers than those who occasionally carry matches).



- *Consistency.* An association is more likely to be causal if it is observed by different researchers, in different places, circumstances, and times. It is premature and inappropriate to regard the results of any single study as causal because there is no basis upon which to judge consistency. Although associations produced by confounding factors are expected to vary between studies, the force of a common biologic effect should be consistent across studies.
- *Biological plausibility.* An association is more likely to be causal if it makes sense in terms of scientific understanding of the biology of the disease or health effect under investigation. (An association between cigarette smoking and lung cancer makes biological sense; an association between match-carrying and lung cancer does not.)

Statistical Analysis and Peer Review

Two other key factors that should be taken into account when evaluating a research study are statistical analysis and peer review. Scientists pay close attention to these considerations when evaluating a study's credibility.

Studies that use appropriate methods to determine whether a finding is "statistically significant" (i.e., likely to be due to a real association between two factors rather than to mere chance) are more reliable than those that lack or misuse statistical analysis. It is important to remember, though, that statistical significance does not imply that an association reflects cause and effect and is not a criterion for causation. If researchers actually analyzed data on match-carrying and lung cancer risk, they might find that the association between the two is statistically significant, but it is certainly not causal.

Peer review is the process of subjecting the report of a scientific study to the scrutiny of other experts before publication. The reviewers examine the work for possible flaws or weaknesses, and if any are present, the report may be rejected for publication or the authors may be required to revise their report or conduct additional research before it can be published. Because scientific reports published in peer-reviewed journals have withstood detailed scrutiny by experts, they have a much higher degree of credibility than other types of reports on research — such as presentations at scientific meetings, press releases, announcements on Web sites, or self-published reports. It is not completely unreasonable for journalists to report on these other types of scientific announcements, but their tentative nature must be explained.

Reporting on Studies: Some Helpful Pointers

The following points may help journalists cover health news stories that require distinctions to be made between association and causation.

- *Focus on the study design, not just the conclusions.* What kind of study was it? Human? Animal? In vitro? Epidemiologic? Some study designs are more reliable than others, and findings derived from better-designed studies should carry more weight.
- *Ask about possible confounding.* Written reports on research studies may or may not discuss possible confounding. Researchers should be prompted to discuss whether there are potential confounders that may have influenced the results. Confounding, like bias (such as the bias in case-control studies that may result from well people and ill people remembering past events in different ways), is a major weak spot in epidemiologic research.
- *Scrutinize animal tests with care.* Were there appropriate controls? Were the results statistically significant? Did the study use well-accepted methodology? Is this animal a good model for possible reactions in humans? Do effects occur only at high doses unlike those to which humans are subjected? To what extent, if any, can the results be applied to the human situation?
- *Check out the bona fides of a study and its authors.* Completed studies published in peer-reviewed scientific journals should carry much more weight than other types of reports. Studies that include appropriate statistical analysis of the data should carry much more weight than those that do not. Studies produced by

authors not affiliated with a university, medical center, or other established research organization should receive particularly careful scrutiny.

- *Provide context and analysis.* New research results need to be interpreted in the context of related previous research, especially with regard to the criteria of consistency and biological plausibility, as discussed above. Checking such data with reputable health professionals and organizations who can contribute expertise and balance to developing stories can be very helpful in putting findings into the appropriate context.
- *Beware of overinterpretation of study results by scientists themselves.* Because researchers tend to be enthusiastic about their own work, some may overinterpret their findings, sometimes suggesting the possibility of causation when the data only support an association. Other scientists who work in the same general area but who were not involved in a particular study — and are not intimately connected to or predisposed to support the study's authors — can often spot such overenthusiasm and put new findings into perspective. They may also be able to point out related work by researchers other than those who performed the new study that is making news. An analysis of the work of multiple research groups may provide a more balanced perspective than a story exclusively devoted to the work of a single group of researchers.
- *Fight the temptation to fill explanatory vacuums.* Human beings dislike uncertainty. We are unsettled when the reason for an occurrence cannot readily be found. It is natural, therefore, to embrace any explanation, however unlikely, for an unexplained phenomenon. It is important that claims of causation are not made

lightly, so that the public isn't encouraged to take actions that are not beneficial to their health, instead of taking actions that are supported by sound science.

- *Use your wits.* The first response to an incredible finding should be to question its credibility.

For more information on distinguishing association from causation, consult the subsequent ACSH white paper "How to Distinguish Association from Causation: A Guide for Journalists," by William P. Kucewicz, John W. Morgan, and Diana M. Torres.