Writing about Health Risks

A Handbook for Journalists

AMERICAN COUNCIL
ON SCIENCE AND HEALTH

Writing
about
Health Risks
Challenges and Strategies

A Handbook for Journalists

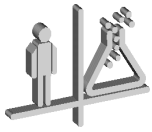
Prepared for the
American Council on Science and Health by
Michael Kamrin, Ph.D., and Marilynn Larkin, M.A.

Project Coordinator and Managing Editor
Alicia M. Lukachko, M.P.H.

Edited by
Jack Raso, M.S., R.D.

Art Director
Yelena Ponirovskaya

February 2000



AMERICAN COUNCIL ON SCIENCE AND HEALTH
1995 Broadway, 2nd Floor, New York, NY 10023-5860
Tel. (212) 362-7044 • Fax (212) 362-4919
URL: http://www.acsh.org • E-mail: acsh@acsh.org

THE AMERICAN COUNCIL ON SCIENCE AND HEALTH (ACSH) APPRECIATES THE CONTRIBUTIONS OF THE REVIEWERS NAMED BELOW.

Thomas S. Allems, M.D., M.P.H.
San Francisco, CA

Joseph F. Borzelleca, Ph.D.
Virginia Commonwealth University

Madelon Finkel, Ph.D.
Cornell University Medical College

Kenneth D. Fisher, Ph.D.
Office of Disease Prevention and Health Promotion
Washington, DC

Charles O. Gallina, Ph.D.
Professional Nuclear Associates
Springfield, IL

Richard M. Hoar, Ph.D.
Williamstown, MA

Rudolph J. Jaeger, Ph.D., DABT
Environmental Medicine Inc.
Westwood, NJ

Ruth Kava, Ph.D., R.D.
American Council on Science and Health

Judith A. Marlett, Ph.D., R.D.
University of Wisconsin–Madison

Gilbert Ross, M.D.
American Council on Science and Health

Ronni Sandroff
New York, NY

Edgar J. Schoen, M.D.
Kaiser Permanente Medical Center
Oakland, CA

Arlene L. Weiss, MD., DABT
Environmental Medicine Inc.
Westwood, NJ

Elizabeth M. Whelan, Sc.D., M.P.H.
American Council on Science and Health

Table of Contents

PREFACE 5

Part One: “HAVE I GOT A STORY FOR YOU!” 6

Introduction 6

Where Do Risk Stories Come From? 6

Should You Cover the Story? 8

Chapter One: Getting Started 11

Approaching Stories about Risk: Basic Steps 11

Uncovering a Risk Story on Your Own: Basic Steps 13

Part Two: REALITY CHECKS 16

Chapter Two: Assessing the Evidence 16

Validity: A Closer Look at the Evidence 16

Chapter Three: Assessing Applicability to the Real World 23

Comparative Dose/Exposure 24

Magnitude of Risk 26

The Precautionary Principle 27

Reverse Onus 28

Putting Risk into Context 32

Part Three: IN THE TRENCHES 32

Chapter Four: Interpreting Risk Based on Laboratory Animal Studies 32

Background 32

The Claim 32

How to Approach the Story 33

Digging Deeper: Assessing the Evidence 34

Digging Deeper: Applicability to the Real World 35

Rounding Out the Story 38

Chapter Five: Interpreting Risk Based on Epidemiological Studies 42

Background 42

The Claim 42

How to Approach the Story 43

Digging Deeper: Assessing the Evidence 44

Digging Deeper: Applicability to the Real World 50

Rounding Out the Story 53
Chapter Six: Interpreting Risk Reduction Claims 57
Background 57
The Claim 58
How to Approach the Story 59
Digging Deeper: Assessing the Evidence 63
Digging Deeper: Applicability to the Real World 67
Part Four: “HERE’S THE REAL STORY!” 71
Conclusion 71
GLOSSARY 73
RECOMMENDED READINGS 77

PREFACE

Reporting about health risks isn't easy. It involves an understanding of the complexities of risk assessment, an ability to distinguish between scientific and pseudoscientific information, the capacity to evaluate and digest complicated material, and the communication skills to portray the risk in the proper context. Simplistic or contradictory messages can leave readers confused and wary; they "tune out"—and you lose your audience.

The aim of this handbook is to facilitate the process of reporting by offering a practical, step-by-step approach to writing about health risks. It is written from the perspectives of a toxicologist concerned about the consequences of misinforming the public—e.g., ineffective actions to protect and improve personal and/or public health—and an equally concerned medical journalist who understands the pressures of producing accurate, balanced stories on tight deadlines.

We provide questions to ask during the development of the story, tools for evaluating the evidence, an annotated glossary of scientific terms relevant to risk assessment, specific examples from real-world risk stories, and advice on how to avoid common pitfalls. These suggestions are intended to streamline the process of writing accurate, balanced risk stories—with the knowledge that clear "take-home" messages will assist readers in making reasonable decisions about their health.

PART 1: "HAVE I GOT A STORY FOR YOU!"

INTRODUCTION

Toxic: Harmful, poisonous. All chemicals or substances are toxic under some conditions and in large enough amounts.

Risk: The possibility of suffering harm or adverse effects. *Acute risk* is the risk from a single or a few closely spaced exposures to an agent; *chronic risk* is the risk from exposure over a significant period of time; *lifetime risk* is the risk from exposure to an agent over an entire lifetime—generally assumed to be 70 years.

Does living near a toxic waste dump increase your chances of getting cancer? Does eating soy reduce cancer risk? Will zinc help a cold? Are pesticides poisoning you?

Health is a hot topic, and stories dealing with risks to health are staples for most print publications. These stories often cry out for simple answers, but the answers are rarely simple. Except for cases of proven risk (e.g., cigarettes cause lung cancer) or proven risk reduction (e.g., wearing a seat belt reduces the risks of injury and death in a car crash), writing about risk is likely to require interviewing more than one expert, sufficient knowledge to understand and assess the evidence, and the ability to place claims and study findings in some kind of context and perspective. It's hard work—but worth the effort. Risk stories are important: They have the potential to influence the health and lives of your readership, and they are apt to be widely read!

WHERE DO RISK STORIES COME FROM?

Ideas for risk stories are generally derived from three main sources: press releases, journal articles, and presentations at scientific meetings. It is important to distinguish among them.

Press Releases

Press releases are of highly variable quality. Some are useful: a well-researched press release may do a good job of explaining the methods by which a

study's investigators arrived at their conclusions, or it may clarify scientific terms or concepts that are difficult to understand. On the other hand, some press releases are filled with errors, especially when the writer attempts to "simplify" a study for lay audiences, or when the release has not been reviewed by the investigator on whose work it is based.

Evidence: Data collected scientifically. In risk assessment, this includes data about exposures and effects. Effect data are generally derived from laboratory studies or observations of human populations.

Press releases often include quotes from a researcher—but when you actually speak with that person, you may find out that he or she was "persuaded" to make certain statements or to express findings in a particular way. Press releases often give a spin to a story aimed at portraying the research institution, the company or organization that sponsored the study, or some other interested party in a favorable light.

So unless you know that the quotes in a press release accurately portray the investigator's thoughts, it's best to use them only to get an idea of what the person might say in a interview or as launch pad for interview questions—*not* as the sole source of quotes for your story. The same is true for other secondary sources—e.g., other news reports.

Journal Articles

Journal articles are also of variable quality, but they usually have the advantage of having undergone at least some form of peer review (see Chapter Two). In the top-tier journals—e.g., *The Lancet, The Journal of the American Medical Association, and The New England Journal of Medicine*—the chances of the article's having undergone rigorous peer and editorial review are better than for some other journals. But the fact that an article has been published in a top journal does not mean that it is exempt from the critical assessment steps described in this handbook, or that a journalist (or anyone) can take its findings as automatically "true."

Peer Review: Careful scientific examination of the data, methodology, results, and conclusions of a study or set of studies by experts in the area of the research.

Meeting Presentations

Meeting presentations can be good sources of news stories because researchers can present preliminary findings from a long-term study, findings from a small (pilot) study, or extended follow-up data from studies that have been published in the scientific literature. On the other hand, even though meeting abstracts are often published in journal supplements, it is important to bear in mind that such abstracts have not undergone peer review: No august body has reviewed the methodology (see Chapter Two) and results of the study to determine if they are scientifically sound, and no experts in the field have reviewed the conclusions drawn from the study.

A Note About "Hot Tips"

Sometimes a source from an earlier story, a member of an advocacy group, or your neighbor down the street will approach you with a "hot tip"—a story idea that hasn't been covered, or allegedly hasn't been covered properly, etc. Such story tips should be assessed in the same way as any other—even if the tip is presented as an exclusive or a scoop. It is vital to have some facts, data, and perspective before forging ahead.

Methodology:
The way in which data are collected. Methodology descriptions include identification of the study subjects or specimens, the sources of the agents being tested, and the techniques used to assess exposure or effect.

Should You Cover the Story?

Sometimes you have no choice about whether or not to cover a story; your editor says "cover this" and you comply. At other times, you may be able to use discretion and decide for yourself whether a story merits covering. You might want to cover a story if it is: something your audience would or should care about; controversial, and either both sides of the issue should be covered or one side is promoting something potentially detrimental to public health; worthy of coverage because of its the public health implications (e.g., the adverse effects of tobacco, the basics of good nutrition); one that has received extensive coverage, but there is new information or critically important information that bears repeating

(e.g., adverse effects of secondhand smoke, unsupported health claims for herbal supplements).

There are other, mainly political or competitive reasons, for covering a risk story: your editor likes the topic; "everyone else" (i.e., your competitors) is covering it; it will make for a good headline (but be aware that alarmist headlines are often inconsistent with good journalism).

A Note About Slanting and Editorializing

Journalists, like everybody else, are likely to have opinions on risk topics. Be aware of your personal views and prejudices, because they may influence your coverage. If you have strong views on a risk topic, for whatever reason, you are unlikely to be able to report objectively on the topic. If this is the case, you are better off refusing the assignment. If you permit your opinions to color your coverage—e.g., by slanting the story in a particular way, only citing sources that support your view, or weaving your opinions into the coverage—you stand to compromise your journalistic integrity. On the other hand, your views may fit perfectly well into your publication's commentary or editorial section, and it would be appropriate to propose an article for that section.

NOTES

Chapter One: GETTING STARTED

Writing a fair, accurate, comprehensible risk story isn't easy—especially when you're on a tight deadline. In fact, the tighter the deadline, the more tempting it is to take shortcuts. This is fine, as long as shortcuts—press releases, statements from other published stories, and similar secondary sources—are used as adjuncts to, not instead of, taking some basic steps to create your own, authoritative story.

Approaching Stories about Risk: Basic Steps

Following these basic steps can help ensure that your risk story—even on a tight deadline—will be accurate and balanced.

- Whenever possible, get a copy of the study report that is the basis for the claim of an increase or decrease in risk. If the claim is based on a meeting presentation, try to get a published abstract—or ask your source whether an article on the subject is "in press" in a journal (in which case, galleys or the original manuscript may be available). Relying solely on a presentation or interview without having corroborating data is a risk in itself—for you.

- Take your time reading the paper and make sure you understand it before doing anything else. If you think it has certain real-world implications, note it, but don't conclude that it does at this point.

- Take note of the affiliation of the study authors. Are they university researchers? Are they advocates with a strong position on the subject being investigated? If the article is about the toxicity of a particular chemical, look at the disclosure statement to see whether the producers of the chemical sponsored the study or whether the author(s) are paid advisors to such companies. This does not necessarily mean the findings are questionable, but

Toxicity: The inherent potential of a substance to have adverse effects in living organisms. "Acute toxicity" refers to adverse effects that occur very soon after a single exposure or small number of closely spaced exposures to a toxic substance; "chronic toxicity" refers to adverse effects that occur after repeated exposures over a long period of time.

it makes good sense to know—so that you are aware of a possible bias.

- If stories on the findings have been published (e.g., by a wire service), read them to see what the study authors are saying, to gauge their position, and to see how other experts are responding. Don't assume that other stories are accurate, however, no matter what the source.

- Interview with one or more of the study authors. Ask him or her to summarize the methodology of the study in 100 words or less, so that you can write about it succinctly for your audience (and also to clear up any misunderstandings you may have). You can often get the investigator's perspective, as well as background on the study, by asking this kind of neutral question first.

  Ask for bottom-line messages for your readership—such as implications for the community at large, implications for individuals who might be at particular risk, and clinical messages for physicians—in short, whatever is pertinent to your audience and related audiences. If you had a thought when reading the article and the author doesn't raise this point, don't hesitate to ask. You may be off-base, or conversely, the author may have assumed that the point was clear.

  Ask for context: What are other scientists saying about this work? Is it the first study of its kind? Does it corroborate or contrast with previous findings? What is its overall importance in the field?

- Use additional strategies, if necessary, to determine what the scientific consensus is regarding a risk claim. Contact professional associations or advocacy groups for perspective. Such organizations will generally tell you what they and others think of a particular risk or risk reduction intervention.

- Get at least one independent comment, especially if the topic is controversial (in fact, you may not be aware that a topic is controversial *until* you get an independent comment). Start with the points the study author considers the main messages.

Interviewing Tips

- Tape-record interviews whenever possible (state laws vary on whether or not you must get permission from the interviewee). This will help ensure that your quotes are accurate (it's amazing how quickly our memories can fail us or distort what a source says). Replaying the tape can also remind you of terms or points you may not have understood during the interview—but which came up when you didn't want to break the "flow" to ask for clarification (and you forgot to ask afterwards!).
- Don't be afraid to ask questions when you don't understand what is being said. Some journalists feel they must come across as experts in the area they're writing about, but if you cover a wide range of topics, it's impossible to know it all. It's better to tell your source that you need context or background or clarification and go from there.
- If your source uses an acronym, ask what it stands for. The same acronyms mean different things to different people (e.g., "ACS" stands for "American Cancer Society" and also "American College of Surgeons"). Do the same for technical terms that you and your readers may not understand (e.g., "snips," which are really SNPs—single nucleotide polymorphisms).

Uncovering a Risk Story on Your Own: Basic Steps

Investigative journalism has its own special skills and requirements, which are beyond the scope of this handbook. But even if you are not an investigative journalist, there may be times when you are approached with a risk story tip from a respected source, or you read a murky-

sounding story and decide to investigate further on your own. For example, you may hear that scientists have not been able to duplicate the results of a well-publicized study that reported a significant risk associated with a chemical exposure.

How do you know whether a story deserves covering? The steps are similar to those described above, but the order in which you take them is apt to differ.

Exposure: Coming into contact with an agent. *Acute exposure* occurs when a person comes into contact with a single dose or a small number of doses of the agent over a short period of time, such as one day. *Chronic exposure* occurs when a person receives repeated doses over a long period of time, e.g., years.

- Get context. Before going any further, see what other people in the field are saying about the tip. If, for example, other scientists have also heard about the reproducibility problem, the story is probably worth pursuing.

- If the story still seems plausible, get more details from your main source. If you're starting fresh on the trail of a story because the subject interests you, track down an appropriate source. For example, if you hear of a suspicious-sounding alleged cure for cancer, such as BioResonance Therapy (BRT), which purportedly enlivens defective genes through oscillations, you might contact the American Cancer Society for comment.

- Ask yourself again whether you are pursuing a viable story. In the BRTexample, if everyone is saying it is worthless and there is no suggestion that anyone is promoting it, there is nothing to worry about for now and thus no story. But if the BRT promoters have set up a new web site, for instance, you may have a reason to investigate—and a worthwhile story.

- Finally, take the steps outlined in the section on working on assignment—read any published studies critically, get a sense of the scientific consensus, and interview independent sources.

NOTES

PART 2: REALITY CHECKS

Chapter Two: ASSESSING THE EVIDENCE

Not everything represented as science is science—and not all decisions by scientific or regulatory bodies are based on scientific evidence. Risk management considerations are based only partially on science. For example, a maximum allowable level of 1.0 ppm (parts per million) of an agent in drinking water sounds scientific because it is a precise number, but it is essentially a regulatory value that has been set partly on the basis of scientific findings and partly on the basis of assumptions about managing risk. Two such assumptions are that any effect seen in animals will also be seen in humans and that high dose studies in animals are relevant to low dose environmental exposures.

Risk Management: The steps undertaken to reduce or eliminate risks. This may be accomplished either by controlling the source of the risk (e.g., limiting emissions from an industrial source) or changing behavior (e.g., issuing a fish consumption advisory).

To distinguish good from bad or mediocre evidence or studies, journalists must make judgments about the validity of research findings; that is, they must evaluate how well the scientific work has been done and whether the findings have been interpreted properly. In most cases, this does not mean you have to be a scientific expert; often, you can make reasonable qualitative judgments with a modest understanding of how scientists assess validity (the confidence they have that the study's findings are accurate and free of distortion) and with a list of pertinent questions to ask.

Validity: Degree of confidence that the conclusions from a scientific study are accurate (i.e., are not due to chance, bias, confounders, and other causes of distortion).

VALIDITY: A CLOSER LOOK AT THE EVIDENCE

An evaluation of a study's validity depends on several factors, including how and where the study was done; whether the authors accounted for common sources of distortion—bias (on the part of the investigators or the subjects), confounders (variables associated with the factor under investigation that can

influence the outcome of interest), chance; and whether the design is such that it can be duplicated independently by other scientists.

Type of Evidence

Your evaluation of validity depends at least in part on the type of study on which the risk claim is based. The type of evidence largely determines how the findings may be generalized (to a human population, for example). Three main types of studies are used to assess risk. Laboratory animal studies assess the risks and/or benefits of a given agent in animal models; epidemiological studies show trends or possible associations between a given agent and health outcomes in human populations; and clinical trials assess the risks and/or benefits of a drug or another intervention in people. *In vitro* (test tube) studies are also used to evaluate the effects of exposure to a suspect agent on human or animal cells.

Quality of Evidence

Generally, publication in a peer-reviewed journal indicates that the evidence in a study meets scientific standards. In effect, by accepting the article for publication, the reviewers and journal editors have done much of the work for you in assuring the validity of the study. But peer review is not infallible. Not all journals have high standards; sometimes the best reviewers are not available; and sometimes the reviewers overlook important problems with a study because of time constraints.

The viewpoints of peer reviewers may also reflect the bias of those who select them. For example, the Natural Resources Defense Council (NRDC) report on the agricultural chemical Alar ("Intolerable Risk: Pesticides in our Children's Food," NRDC, February 1989) was said to have been "peer-reviewed," but it was never subjected to independent scientific review and has never been published in a reputable scientific journal. The authors claimed that the use of Alar

Bias: A factor or process that systematically distorts the collection or analysis of study results. An example is "recall bias"—people who have diseases may be more likely to report real or imagined exposures to the study agent than are those who are disease-free.

Confounder: A factor that can distort the results of the study because it is related both to the exposure and to the effects—e.g., secondhand cigarette smoke may be a confounder in studies of the effects of an environmental contaminant on childhood asthma.

Laboratory Study: Astudy done in test tubes *(in vitro)* or in animals *(in vivo).*

Epidemiological Study: A study of the patterns of human disease and the factors that are associated with specific diseases.

on apples was causing cancer in children, but the group's conclusions were later discredited by the broader scientific community. Also, advocacy groups are not bound to accept the conclusions of peer reviewers and may cite only those opinions that support their conclusions. Thus, all peer reviews are not created equal.

Dose Response

A fundamental principle of toxicology is that there must be a clear relationship between the amount of exposure to a given substance and the degree of effect. Simply put, this means that if a low exposure to an agent produces slight effects, a higher exposure will produce greater effects. For example, one aspirin may only partially relieve a headache while two aspirins (twice the dose) will relieve it completely. Or the fumes from paint may not bother you until you have been exposed for several hours, at which point they may make you feel lightheaded; in this case, the total dose in your body increases over time because your body rids itself of the chemicals in the fumes more slowly than you breathe them in.

Dose Response: The relationship between exposure to an agent and the occurrence of an effect. Generally, the higher the dose, the greater the response.

Clinical Trial: A study in humans that assesses the efficacy and safety of a specific intervention or treatment on a disease or other health-related outcome. Generally, patients are randomly assigned to receive the study treatment or another treatment, or a placebo, and are followed over a period of time.

Since this concept is so basic, you might think it would be a requirement for the evaluation of a toxicity study. Regrettably, this is not the case. In some instances, an agent causes no effect as exposure increases and only produces an effect when a very high level of exposure is reached. An example is the sweetener saccharin, which showed no toxicity until a very large amount was administered to experimental animals. Nonetheless, the high-dose results were used to attempt to discredit the safety of this sweetener.

Why should dose response be considered in an evaluation of the evidence? For one thing, it may suggest that there is no solid basis for a claim that exposure to an agent is related to toxic effects. Suppose, for example, that at very high exposures, a contaminant present in tiny amounts in a particular agent reaches

significant levels and causes toxicity.

Or perhaps something happens to change the chemical or physical properties of a substance at very high doses, and this change—rather than the agent itself—is responsible for the effect. In the case of saccharin, studies suggest that at high doses, the sweetener formed crystals instead of staying dissolved. The crystals may have caused the toxicity. At the very least, you should question the generalizability of high exposure results to lower exposures if no dose response has been shown. There is generally a threshold of exposure below which effects do not occur.

Weight of Evidence

An important principle to keep in mind is that the validity of individual scientific studies cannot be considered in isolation. Each piece of research is part of a larger picture that includes other studies—some of which duplicate part of the current work, show effects of the same agent on other species, demonstrate the effects in test tube experiments, and so on. Generally, the pieces fit before scientists are confident of their conclusions. Thus, the validity of one study result is intertwined with the validity of others. For example, the conclusion that the organophosphate (OP) pesticide fonofos affects the nervous system in particular ways is based on studies of fonofos in a variety of laboratory animals, studies of other OPs in laboratory animals, and human experience with fonofos and other OPs.

Generally, in the introduction or the background section of published reports, researchers will describe the context of their work by discussing earlier studies that led to theirs and by showing how their study fits into the larger picture. Areas of agreement and controversy are often covered in the article's discussion section. It helps to read these sections for context, and to obtain selected references for more details.

Clinical Trial ***(continued)*****:** Findings from the different study groups are then compared to assess the effects of the intervention.

Association: Relationship between an exposure and a health effect—e.g., between exposure to sunlight and skin cancer. Association is not the same as causation. Associations can be positive or negative. For example, a positive association exists between cigarette smoking and lung cancer. Anegative association exists between the use of seat belts and deaths from car crashes.

When results of individual studies reinforce, or are at least consistent with, others, it is easier to have confidence in the validity of the individual studies. On the other hand, if 10 well-designed studies give one result, and one study gives the opposite result, it is likely the lone study has not been done properly or is in some way fundamentally different from the others.

A single study that gives results contrary to a larger body of evidence should not in itself significantly shake your confidence in the validity of the other studies. While it is always tempting to champion the underdog—i.e., the person who has contrary results—especially if that individual is persuasive or claims to be persecuted by the scientific community, it is better to report the consensus opinion unless a body of published work supports an alternative view.

Mechanistic Studies: Investigations to determine how a particular effect is produced. These include studies of how chemicals are absorbed into the body, changed in the body, distributed throughout the body, excreted from the body, and stored in the body, and how they have toxic or beneficial effects.

Of course, there are cases in which scientists obtain varying and conflicting results and no clear pattern emerges. For example, recent findings call into question studies whose findings suggested that dietary fiber reduces the risk of colon cancer. To report the current picture, it is important to talk to scientists on both sides and to present their views in support of one conclusion or the other.

How Does the Agent Produce an Effect?

In general, you can have more confidence in a study's results if there is an accepted explanation of how the results came about (i.e., mechanism of action). For example, mechanistic studies of fonofos show that the pesticide interferes with an enzyme critical for the proper conduction of nerve impulses. This finding would reinforce the validity of a study whose claim is that fonofos adversely affects the nervous system. But if other research had shown that fonofos is rapidly broken down into its metabolites and is quickly eliminated from the body, you might be more skeptical about the validity of the study

showing that the pesticide is harmful.

It is important to note that information about what happens to a particular agent once it enters living beings is often not available. However, information on a similar chemical could provide some indication of whether the hypothesis for a mechanism of action makes sense. For example, studies on the organophosphate pesticide chlorpyrifos could help explain the presumed mechanism of action of a related pesticide—but not necessarily; in this situation, an expert's evaluation can be helpful.

Also be aware that a given agent may not work the same way—quantitatively or qualitatively—in different species (see Chapter Four for details). This is a critical consideration when assessing the generalizability to humans of studies of a particular agent based solely on animal experiments, and it has an impact on risk claims. Sometimes scientists are well aware of differences between an animal species and humans and might take advantage of them. For example, because insects break down certain OP insecticides in a way different from that of humans, these insecticides can be used to kill insects without harming people.

NOTES

Chapter Three: ASSESSING APPLICABILITY TO THE REAL WORLD

Once you've assessed the evidence, you will want to tell your readers what the results mean in a real-life situation—e.g., when persons are exposed to a suspect agent in their food. This requires an understanding of how relevance—the applicability of scientific findings to the real world—is assessed, and knowledge of the right questions to ask scientists.

The assessment of applicability is often overlooked because the idea that something is "toxic" is equated with the conclusion that the substance is a "risk" to people. An example is the oft-repeated phrase that "dioxin is the most toxic manmade chemical," implying that it carries the most risk to people. The fallacy in this logic stems from the most fundamental principle of toxicology: *Everything is toxic under certain conditions.* All agents will cause toxic effects if exposures are high enough—including substances essential to life, such as oxygen and water. Even the most toxic agent will cause no harm if exposure levels are low enough. So while dioxin may be a highly toxic chemical, if your exposure to it is minimal, it is unlikely to pose a risk to your health.

In assessing the applicability of a study's findings to the real world, it is important to identify the situation in which the study's findings are being applied. If a chemical in paint is said to be a health risk, who is being exposed? Artists who use paints daily? Workers in a factory where appropriate health precautions are not being taken? Inner city children living in dilapidated housing conditions? Such information is often absent from risk claims.

Relevance: The extent to which toxicity findings from laboratory, epidemiological, or clinical studies are applicable to people living in the real world. Studies done under unrealistic conditions—e.g., in which animals are exposed to extremely high doses of an agent—may not be relevant to people exposed to low environmental levels of the same agent.

Comparative Dose/Exposure

How do the amount, duration, and frequency of exposure to people in a real risk situation compare with the exposures in the scientific studies on which the risk evaluation is based? In animal studies, such exposures generally are much greater for all parameters—amount, duration, and frequency—than exposures in people. For example, in the case of saccharin, it is estimated that a human would have to drink hundreds of cans of diet soda daily for a lifetime to obtain the amount of saccharin harmful to rodents. Even in epidemiological studies (see Chapter Five), the study population often is not representative of the general population but instead consists of selected individuals—e.g., those who were highly exposed to a chemical while working in a factory. For example, estimates of the risk of lung cancer from indoor radon exposure are based on studies of miners whose exposure was much greater than that of people living in their homes.

Risk Assessment: The process that estimates the type and magnitude of risk to human health, wildlife, or the environment posed by exposures to hazardous agents.

A Note about the U.S. EPA's Approach to Risk Assessment

The U.S. Environmental Protection Agency (EPA) approaches risk assessment by assuming that, while experimental and real exposures are very different, the results of experimental studies can be used to estimate effects in people exposed to lower levels of a chemical. This assumption makes it especially important to determine (1) the relevance of animal studies to humans and/or the relevance of epidemiological studies to people in the real situation; and (2) the relevance of high, lengthy, and frequent exposures in animals to the real-life situation of exposed people in the community.

Magnitude of Risk: A quantitative estimate of the degree of risk associated with a specific exposure—e.g., one additional case of cancer in one million exposed individuals.

Magnitude of Risk

Risks generally are described in two different ways. Cancer risks are typically reported as the number of

additional cancers that would occur per million exposed individuals. By contrast, exposures to an agent associated with a non-cancer health risk are usually described as safe or unsafe—more of a value judgment than an actual risk value. Value judgments, of "acceptable" or "unacceptable," are also used to describe a cancer risk. For example, if exposure to a given agent leads to an additional cancer risk of more than one in a million, the risk might be labeled unacceptable. The "safe" and "acceptable" levels are established by government agencies; in the United States, the EPA is the source of most of these numbers.

Using value judgments such as "safe" or "unsafe" and "acceptable" or "unacceptable" allows regulators to state that specific actions must be taken when "safe" or "acceptable" exposure levels are exceeded. But this approach cannot provide absolute risk levels. For example, if your story involves a comparison of measured ozone levels and ambient air standards, it is logical to ask how much risk is incurred when the "safe" level is exceeded. But that question has no answer because a "safe" level is not a risk value. The best you can get is a comparative answer—that is, the more the concentration of ozone or some other agent exceeds the "safe" value, the more likely it is to lead to adverse effects. This is a reasonable response when the story deals with a specific agent.

"Safe" Value: A quantitative estimate of the amount of a non-carcinogenic substance a person can consume without incurring appreciable risk; exposures above the "safe" value do not necessarily result in adverse effects.

It is important to bear in mind, however, that "safe" levels are determined uniquely for each agent and are not generalizable to other agents—even for the same type of agent. Thus, if the safe level of agent A is exceeded by a factor of three, this does not necessarily mean that it causes more harm than does agent B, whose concentrations are only twice the "safe" level. A recent Consumers Union report ("Do You Know What You Are Eating?" Consumers Union, February 1999) on pesticides in foods ignores this problem. Its authors assess relative risk on the basis of compar-

isons among "safe" levels of a large number of pesticides in a variety of foods—which is akin to comparing apples and oranges.

With respect to cancer, you can get an answer to your question about the amount of increased risk because the assessment usually provides a numerical risk estimate—e.g., one in a million. Assuming that this is a valid number (and we will see shortly that this is not very likely), one could reasonably claim that, because the measured environmental level of a given agent is twice the acceptable risk level, two additional persons in a million will get cancer from the exposure—persons who would not have developed cancer otherwise.

But before you can truly gauge the impact of an exposure, you need to ask a simple question: How many people have been or are likely to be exposed? If, for example, only 100 people are exposed, then it is extremely unlikely there will be one additional case of cancer, much less two, since one in one million people is the same as 0.0001 person in a hundred. This distorted type of claim is often associated with risk claims about hazardous waste sites, which generally are located in areas of low population density.

Hazard: The potential of an agent to cause an adverse effect under specific circumstances.

Knowing how many people are likely to be exposed is especially important when comparing two different risk situations in which different numbers of people are exposed. For example, quadrupling the risk in a small population living near a contaminant source such as a hazardous waste site may not affect anyone; but increasing the risk from an indoor contaminant by a much smaller amount—say, 25 percent—may affect a huge number of people if most of the population is exposed to the contaminant.

THE PRECAUTIONARY PRINCIPLE

When interviewing people about a risk claim, you may be told of the importance of the "Precautionary

Principle" in interpreting evidence. The principle basically says that when information about risk is uncertain, it is prudent to assume the worst and aim for zero risk. The implication is that following the principle will always lead to greater protection of the public's health. While such "prudence" sounds good on the surface, closer examination reveals that simple-minded application of the principle may lead to adverse public health consequences.

Following the precautionary principle may, for example, lead communities to act on information that has not been scientifically validated (if it were validated, the information wouldn't be "uncertain"—and the principle would not be needed). A possible negative impact of such action could be the unnecessary removal from use of a product beneficial to public health. A tragic example occurred in Peru in 1991, when chlorination of drinking water was discontinued because of fear of cancer from chlorination by-products. This resulted in a widespread cholera epidemic.

Another potential negative impact is the expenditure of large amounts of public health resources to correct a problem of limited or negligible impact, with a concomitant reduction of funding for more valuable public health measures. For example, vast amounts of resources have been devoted to cleaning up lead at hazardous waste sites, while much more significant sources of lead exposure—apartment paint and soil in urban areas—have received less attention.

The Precautionary Principle: The premise that actions should be taken to reduce or eliminate a hazard even if there is little credible evidence that a significant risk exists, as a precaution against potential risk.

Reverse Onus: Placing on those who claim an agent is safe the onus of proving that it will not have adverse effects under any circumstances.

Reverse Onus

A complementary argument often put forth by those who support the Precautionary Principle is that it is not the responsibility of those making risk claims to prove the validity of the claims; rather, it is up to those who deny the claims to prove that they could not be true. The "reverse onus" argument has popular appeal because often the individuals making risk

claims are members of environmental safety groups, while those who produce the suspect agent are industrialists. Yet the fact remains that products generally undergo rigorous testing before coming to market and must meet at least the minimum safety standards set by regulatory bodies. This does not ensure that they will be absolutely safe for every person in every situation, but it would be impossible to hold manufacturers to absolute standards and expect that new products would be produced. In fact, it is impossible to prove absolute safety for any agent, natural or synthetic, since all may cause harm under some circumstances.

Natural: Occurring in nature. Natural substances may show a wide range of toxicities.

Putting Risk into Context

Context is critical in situations where individuals may be exposed to a potentially toxic agent from more than one source. Say, people living near an incinerator (or near the prospective site of one) are concerned about their risk of cancer from exposure to benzene, which the incinerator emits. This may be only part of the story. The same agent may be found in indoor air, and indoor exposure levels could be much greater than the outdoor exposure, especially if people are exposed all day, every day.

Synthetic: Not natural; formed through human intervention. Generally refers to compounds made in the laboratory. Synthetic substances may show a wide range of toxicities.

Benzene is also a component of tobacco smoke, and in homes where someone smokes cigarettes, this is the primary source of indoor exposure. Petroleum products also emit benzene; thus, household products that contain petroleum distillates as a solvent, or gasoline or automotive products in an attached garage, can also increase indoor exposures to benzene. Since the risk from indoor exposure may be many times greater than the risk posed by the incinerator, reducing incinerator emissions would have very little impact on risk, while reducing exposure in the home could have a substantial impact. It might be important to point that out.

Comparisons to other cancer risks might also give the reader more perspective on what the risk values associated with the incinerator mean. For example, how does the increase in risk from exposure to benzene compare to the risk from exposure to second-hand smoke or to diesel-exhaust particles?

A Note about Natural Versus Synthetic Agents

A common misperception among laypersons is that natural agents are less toxic than synthetic agents. One argument put forth in support of this contention is that natural chemicals, such as those found in plants and food, have been around for a long time and that human beings have adapted to them. Synthetic chemicals, on the other hand, are new, and thus more dangerous. This argument may appear sensible, but it is ill-founded, because it does not take into account at least two important considerations.

The first consideration is the principle that "the dose makes the poison." Sufficiently large doses of any substance can prove to be harmful. The second consideration is the scientific evidence that has been collected. The results of standard toxicological tests demonstrate that natural and synthetic agents show the same range of toxicities. For example, botulinum toxin, produced naturally by a bacterium, is generally considered the most acutely toxic agent known. Even tiny doses can be deadly. In contrast, high doses of sugar, another natural agent, are required to produce adverse effects. Similarly, dioxin, a contaminant produced during certain industrial operations, is acutely toxic. But another industrial byproduct, sodium sulfate, has low toxicity. The comparative toxicities of natural and synthetic agents hold for both carcinogenic and noncarcinogenic effects.

Carcinogenic: Capable of causing cancer.

Any agent, whether natural or synthetic, must be

evaluated individually, and on the basic of appropriate laboratory and human investigations, to determine how great a risk, if any, it poses to human health. It can be dangerous to assume that natural substances are inherently safer than synthetic ones. For example, some people have risked their health—and lives—by abandoning established therapies based on synthetic drugs and relying instead on alleged "natural" remedies. Others have poisoned themselves by ingesting large amounts of "natural" supplements in the mistaken belief that natural substances are safe in any amount.

NOTES

PART THREE: IN THE TRENCHES

Chapter Four: INTERPRETING RISK BASED ON LABORATORY ANIMAL STUDIES

Background

Claims of adverse health effects from an environmental agent usually are based on risk assessment numbers calculated from studies done on rodents. As noted in Chapter Two, such studies generally involve daily exposures over a long period of time to a single agent at levels much higher than those found in the environment (e.g., food or water) or in products made for human consumption.

Two types of toxicity values are generated from such studies: (1) a potency value for agents that cause cancer and (2) a "safe" value for noncarcinogenic agents. These toxicity values are used to manage the risk—for example, by establishing a maximum contaminant level (MCL) in drinking water. Because the aim is to be protective, these values do not represent scientists' best estimates of cancer potencies or safe levels; rather, they are "prudent" values that include margins of safety. Thus, risk values are based both on science and on policy considerations.

The Claim

You receive a press release stating that agent X, a grain contaminant also found in children's breakfast cereal, has been shown to cause cancer in laboratory animals. The release further claims that a risk assessment, conducted according to EPA guidelines, shows that the presence of agent X will result in a doubling of the incidence of childhood cancers; thus, agent

X—and cereals that contain it—should be removed from the market. As is often the case, the actual cancer potency value is not presented in the press release; instead, the risk—a combined measure of the potency of an agent and the exposure to that agent—is provided.

Cancer Potency Value: A quantitative measure of cancer risk derived from applying the linearized multistage (LMS) model (see page 35). It describes how cancer risk varies according to dose.

How to Approach the Story

How do you reach a conclusion regarding the risk of agent X and the applicability of the study findings to children who eat commercial breakfast cereals? As you follow the basic steps outlined in Chapter One—reading the article, interviewing a study author, getting independent comments—keep these questions and points in mind:

- What type(s) of cancer were found in the animals and what type(s) of cancers do the authors suggest are increasing in the children? There are many different types of cancer, each with multiple causes involving genes, lifestyle, and other environmental factors.

- Who would be affected? All children, or only those who eat a particular type or amount of cereal?

- Was the study published in a peer-reviewed journal—or in a peer-reviewed report generated by a special interest group?

- How was the study conducted? How many species were tested? How often were animals given agent X? What doses were used? What effects were seen, and was there a dose response? Were results the same for male and female animals? Were several species affected or only one?

- How long has the substance been in commercial use? Have any human studies suggested a

risk of disease?

Digging Deeper: Assessing the Evidence

Let's assume that agent X was given daily by injection to an animal over its lifetime, and that the agent was given at doses one thousand times higher than the doses humans typically get. This is a likely scenario, since EPA guidelines suggest evaluation of toxicity at doses near the maximum that an animal can tolerate—known as the maximum tolerated dose (MTD). Let's also assume that the results showed a dose response for bone cancer in male rats, but not in female rats, suggesting that the applicability of the study may be limited.

- How well does the study fit with related scientific research? What other studies have investigated (1) the cancer-causing potential of agent X or related compounds; and (2) the mechanism by which it causes cancer?

It may well be that no other studies of agent X have been done. If so, the evidence is quite weak; only one species has been studied, and the effects were gender-specific. Perhaps studies have also been done in mice—but these did not show any cancer in either sex. Such an inconsistency of findings would also suggest weakness in the evidence, and it would raise serious doubts about the study's generalizability to humans. On the other hand, if results in rats and mice are similar, they reinforce each other and add to the overall validity of the individual studies.

- What happens at lower doses?

If studies have shown that agent X is easily eliminated from the body at low but not at high doses, this suggests that the validity of the rodent studies may be limited to high dose exposures.

- How does agent X cause cancer?

Suppose it works by interacting with calcium accumulation in cells and that there is a threshold for this effect—i.e., no changes result from increasing exposures until a certain dose level is reached. This would support the validity of the rat study in one way—namely, it would be plausible that effects on these cells could be linked to occurrence of bone cancer. But the threshold phenomenon suggests that the validity of the results is limited to high dose exposures.

If, on the other hand, agent X causes bone cancer indirectly, by a mechanism that requires repeated exposures over long periods of time, the validity of the study may be limited to situations of continual, long-term exposure.

✓ A Note about the Linear Multistage (LMS) Model

A mathematical model, the linearized multistage (LMS) model, is used to extrapolate from high dose data what might happen at lower doses. The formula assumes that high dose data alone can be used to predict what will happen at low doses and also that cancer can result from exposure to even a single molecule of the agent. But if studies have shown that high and low doses of agent X do not have similar effects, or that there is a threshold below which effects will not occur, or that many exposures (rather than exposure to a single molecule) are required to cause cancer, this standard formula will not give a valid estimate of the potency of agent X; it may underestimate or overestimate its potency, depending on what happens in real-life situations.

Linearized Multistage (LMS) Model: A mathematical model used to estimate the carcinogenic potency of specific agents. The LMS model is based on data collected from giving animals very high doses of an agent over their lifetime. This is not a biological model; it does not include any consideration of the mechanisms involved in cancer formation.

Digging Deeper: Applicability to the Real World

How do exposures to agent X in the experimental animals compare with those of children? Two different types of information are needed to answer

this question: (1) how much exposure do children have to agent X and how does the magnitude of this exposure compare to the magnitude of laboratory animals' exposures? (2) How applicable is the information gathered in animal experiments to children?

Since not all cereals have been tested for agent X and the dietary habits of children are quite variable, the study authors' conclusions about exposures will be based on several assumptions. You can ferret out these assumptions in the article and by asking appropriate questions in the interview with the author.

The authors may have assumed that each child: (1) ate cereals containing the highest level of agent X measured in any sample of any cereal; (2) ate cereals at the highest consumption rate (amount of cereal per day) of any child surveyed; and/or (3) ate cereals for the highest number of years of any child surveyed. If this is the case, ask how the conclusions might differ if other assumptions were made—i.e., an average contamination level and average amounts and durations of consumption. It would be useful to find out whether contaminant levels or cereal consumption have been increasing or decreasing in recent years. Since cancer in the animals resulted from exposures over a lifetime—not just during childhood—it is expected that, for humans, high levels of exposures over a number of years would also be required to cause this effect.

The route of exposure should also be considered. For example, suppose the animals were exposed to agent X by injection? How does this compare with children *ingesting* agent X? Different routes of exposure can lead to very different body levels and thus, different severities or types of effects. If agent X is ingested, not only is it unlikely that all of it will be absorbed into the bloodstream; it is also unlikely that absorption will be immediate. If it is injected into the blood-

stream, all of it gets there immediately. In general, assuming that these two routes of exposure are equivalent would lead to an overestimation of toxicity.

Other important questions include the following.

- In reaching their conclusions, how did the authors account for the differences between their assumptions and real-world conditions?

As noted earlier, not only the amount, but also the frequency and duration of exposure, are likely to be much lower in people than in experimental animals. The animals received the same amount of agent X every day for a lifetime; this is unlikely to be the case for people. And, since cereal comes from grains grown in various places, contaminant levels are unlikely to be uniform from box to box, much less from brand to brand.

Even in the extremely unlikely event that a particular child always eats the same brand of cereal made from grain from just one location, contaminant levels would undoubtedly vary over time. In fact, children don't eat exactly the same amount of cereal every day; they do not eat cereal every day, much less the same cereal every day throughout childhood.

- Is what happens to agent X in an animal's body the same as what happens in a child's body?

The standard EPA assumption is that anything that happens in experimental animals will also happen in humans. But this is not necessarily so. Experimental animals may respond differently from humans, especially at varying doses. The chemical breakdown pathway may lead to different byproducts in rodents compared with people. If the rodent byproduct alone is linked to cancer, the experimental studies are not

directly relevant to children eating cereal. Interviews with scientists can help you decide how much weight to give to this type of evidence.

ROUNDING OUT THE STORY

- How serious is the risk?

Once you've done your groundwork, it's important to put the childhood cancer risk from agent X in perspective. For example, how many cases of bone cancer in children occur from all causes? If the total number of cancers is small, even a large increase in the percentage of such cases may reflect additional cancers in very few children. Or the alleged increased risk from exposure to agent X may be small. Either way, agent X is unlikely to have a major impact on overall cancer rates for children.

- What is the public health message?

Certainly, any increased risk of cancer in children, no matter how small, should not be ignored; but the costs involved in gaining a small public health benefit must also be taken into account. Could the steps needed to reduce or eliminate agent X from cereal also have public health implications? For example, if agent X is a pesticide, and the decision is to ban it—might another pesticide, with risks of its own, replace it? Even if the other pesticide, or a non-chemical alternative, is not directly associated with health risks, might it be less effective than agent X, and thus lead to an increase in the amounts of natural toxicants remaining in the cereal? Would these bring additional risks?

Even if the alternatives are risk-free and equally effective, they might be more expensive. Would the cost of cereal increase to the point that many families couldn't afford it? Would breakfast then become less nutritious? How would this impact the public's health?

Furthermore, it is vital to determine whether agent X also occurs elsewhere—e.g., other foods, environmental tobacco smoke—and whether exposures from these other sources are greater or less than exposure from the cereal. Even if other exposures exist, it doesn't mean that a cancer risk from cereal is acceptable—but it should be factored into information used by parents and health officials in determining priorities for cancer protection for children.

By now, it should be clear that risk assessment stories and their implications are quite complicated and could have far-reaching ramifications. Although the issues may seem simple on the surface, they rarely are—if they were, there would be no "story," since everyone would be in agreement about what the studies mean and what to do about them.

Applying this Approach to Noncarcinogenic Chemicals

Because so many people fear getting cancer, many of the risk stories you are called upon to cover pertain to risks of cancer. But non-cancer effects—e.g., birth defects or nerve damage—are also of concern.

Certain differences between cancer-related and non–cancer-related claims make the detailed evaluation of evidence and applicability somewhat different. As noted in Chapter Three, claims about noncarcinogenic chemicals are based on "safe" daily exposures rather than on potency numbers. Thus, this type of claim would not describe the number of additional people who would be harmed by the effect; rather, the claim would be that the "safe" exposure level for a given agent had been exceeded, and so an untoward effect might be expected to occur in some unquantifiable number of people.

"Safe" exposure levels for various substances have been developed by the EPA and other government agencies. The EPA refers to the safe exposure level

for a substance as the "reference dose" (RfD), while most other agencies use a more descriptive term, "Acceptable Daily Intake" (ADI). The highest dose that causes no adverse effect in experimental animals is used as the starting point for the development of the "safe" exposure level. This "no-effect" value is divided by at least 100, on the basis of two assumptions: (1) that humans are always more sensitive than experimental animals to environmental agents; and (2) that sensitivities vary much more among humans than among laboratory animals.

Here again, information about how a particular agent produces its effects can help you determine the validity of the first assumption and the degree of confidence you can have in the validity of the "safe" exposure level. Similarly, information about real-life exposures compared with doses used in the animal experiments will help in evaluating the relevance of the results.

NOTES

Chapter Five: INTERPRETING RISK BASED ON EPIDEMIOLOGICAL STUDIES

BACKGROUND

Epidemiology is the study of patterns of determinants of human disease. By studying these patterns—e.g., how variations in workplace exposures to a given agent correspond to variations in the occurrence of skin lesions—it is sometimes possible to identify an association between a particular agent and a particular outcome.

But assessing the evidence and real-world applicability of associations found in epidemiological studies can be difficult, because human beings are exposed to many different agents every single day and the extent of these exposures to particular agents varies from day to day. The multitude of agents in the environment makes teasing out the effects of a single agent particularly challenging. Thus, epidemiological studies are most successful in establishing associations involving large effects of a particular agent and unique exposures and/or rare disorders—for example, the discovery of an association of vinyl chloride and a rare type of liver cancer in factory workers.

It has been estimated that the incidence of a common cancer caused by a particular agent would have to increase by about 10 percent before it would become detectable epidemiologically; it is certainly not possible to identify factors that increase risks by one in a million. Nevertheless, epidemiological studies are valuable because they can provide direct human evidence of the relationship between an agent and disease and don't require extrapolations from animals to humans.

THE CLAIM

You receive a press release touting an article in the current issue of *American Journal of Toxic Agents,* a

peer-reviewed journal, in which the investigators conclude that leukemia risk is doubled in people who live near dry cleaning establishments.

How to Approach the Story

Because of the difficulties inherent in determining risk from epidemiological studies—which are not done under controlled conditions (where only one agent is administered in exactly the same amounts every day for a specific period of time)—extensive information must be gathered to root out and examine confounders (factors that could affect or account for the association that is claimed). As you follow the basic steps outlined in Chapter One—reading the article, interviewing a study author, getting independent comments—keep these questions and points in mind:

- How large is the sample size? How many people were studied in the exposed and unexposed groups? How was the exposed group defined?
- Were levels of specific agents emitted from dry cleaning establishments measured? What was found?
- How long have these agents been in use?
- Were other possible exposures studied—e.g., environmental tobacco smoke, emissions from nearby commercial or industrial facilities, and occupational exposures?
- How were leukemia cases identified? Was the time of diagnosis determined?
- How was the possible association between dry cleaning establishments and the increase in leukemia risk established? Were other health effects studied, and if so, were any trends or associations found?

Sample Size:
The number of participants or subjects being examined in a study. The larger the sample size, the likelier it is that the study will be able to detect an effect, if there is one.

Digging Deeper: Assessing the Evidence

To assess appropriately the quality of the evidence in this study, you need to look at how the exposed and comparison groups were selected, the type and duration of the exposures, the magnitude of the effect, and the time course of the study.

Magnitude of the Effect: A quantitative measure of the strength of an association between an exposure and an effect, often measured using the risk ratio.

- Was the exposed population defined simply by drawing concentric circles around the dry cleaners and arbitrarily saying that people living within each of these defined areas had similar exposures?

While this might be the simplest approach, it may not take into account factors such as prevailing winds, which could drive the dry cleaning vapors preferentially in one direction rather than another, or barriers to movement such as tall buildings, which would block flow in another direction.

- How was the comparison population defined? Were comparisons made to all people living in the city, state, or country of the exposed community—that is, people with average exposure—or to a population in a similar community, with similar demographics, but with no exposure to emissions from dry cleaning establishments?

The more similar the exposed and unexposed populations, the likelier it is that: (1) both groups will be exposed to the same mix of background agents; and (2) both will have similar demographic and health characteristics. Since health status and exposure to other agents could potentially affect the likelihood of the health outcome—in this case, leukemia—a comparison of similar populations would increase the likelihood that the perceived association is real.

- What are the potential confounders?

Even when study populations are similar, potential confounders must be taken into consideration. Even if demographics and overall health status are similar, lifestyle habits such as alcohol and tobacco consumption, genes, family history of cancer, and other factors possibly related to cancer may still have effects that skew the study findings. If the exposed and unexposed populations are different at the outset of the study, there are likely to be many more confounding variables.

- When did the exposure occur? What was the duration of the exposure? How were these pieces of information used in calculating total exposures?

Studies have shown that clothes that are dry-cleaned retain residues of the dry cleaning solvents. These residues vaporize over time, and this results in a certain amount of indoor exposure. Such solvents may also be ingredients of household products, thus increasing indoor exposures. If exposures to emissions from dry cleaning establishments are small compared with exposures in the home, it is unlikely that these emissions are causing leukemia.

- Was a dose response relationship demonstrated in the exposed population?

Did the most highly exposed groups—perhaps those living nearest the dry cleaning establishments—have the highest leukemia incidence, the next most highly exposed a lower rate of leukemia, and so on? The absence of a dose response should significantly decrease your confidence in the validity of the claim.

- What was the magnitude of the effect of exposure?

The magnitude of the effect is generally described as

Risk Ratio (RR): The ratio of the risk (or incidence) of disease or death in the exposed population to the risk in the unexposed or comparison population. RR is used in epidemiology to indicate how much more or less likely individuals with a particular characteristic (e.g., obesity, cigarette smoking) are to show the effects under investigation (e.g., heart disease or lung cancer) than are those who do not have that characteristic. Risk ratios above 1.0 indicate a positive relationship—i.e., the risk is higher when the agent is present—while risk ratios below 1.0 indicate that the risk is lower when the agent is present (i.e., the agent is protective).

a risk ratio—the ratio of incidence between the exposed and comparison (unexposed) populations. Small risk ratios—those less than 2.0, for example—are likely to be explained by the effects of chance, bias, or confounders. In our example, if leukemia incidence in the exposed group was tenfold rather than twofold greater than that in the unexposed group, you would have more confidence in the results.

Results are not absolute, however, and another measure—the confidence interval—must also be taken into consideration. The confidence interval describes the statistical variability of the result, on the basis of uncertainties in the study. Let's say, for example, that the risk ratio is 3.0 but the 95% confidence interval is 0.8 to 4.6. (The 95% interval means there is only a 5% chance that the true risk ratio is higher or lower than the confidence limits.) Since the confidence interval includes the value 1.0, there may in fact be no association at all between the exposure and the outcome. The lower limit of the confidence interval suggests the possibility of a slight protective effect of the exposure, while the upper limit indicates that the risk in the exposed population may be up to 4.6 times greater than that in the unexposed population. Most often, a small sample size is the greatest contributor to uncertainty. The key is to look for confidence intervals that span the value 1.0. This indicates that the findings are questionable.

- How was the association established? Was the incidence of other health effects—e.g., other cancers—examined in the exposed populations? Were associations found? How was each case of leukemia identified—by self-report or with actual medical records? Was the time of a leukemia diagnosis determined?

- What was the time course of the study?

Confidence Interval: The margin of error calculated for a given risk estimate. Generally, 95% confidence intervals are calculated; this means that there is a probability of 95 percent that the true risk is no higher or lower than the range of values included in the interval. For example, a risk ratio (RR) of 2.0 might have a 95% confidence interval ranging from 0.5 to 4.0. A RR of 1.0 suggests that there is no association between an exposure to an agent and the outcome being measured. In a case like this, where the confidence interval spans the value of 1.0, it is not

Most environmental epidemiology studies are retrospective—that is, they are based on something that happened in the past. Prospective studies involving a potentially toxic agent are of value but often are not feasible or ethical, since they would require individuals to be exposed to an agent that might be harmful.

The most common type of retrospective study is the "case-control" study, which compares past exposures of persons who currently show an effect (cases) with exposures of persons who don't show the effect (controls). Exposures are described on the basis of individuals'recollections, along with some environmental measurements. In our example, estimates of exposures to a dry cleaning agent depend on the individuals' recollections of where they lived, when and for how long they lived there, and possibly even how much they spent outdoors versus how much time they spent indoors.

Exposure estimates could also depend on the dry cleaning proprietors' memories or records of the types and amounts of chemicals they used each year and the kinds of ventilation systems that were in place from year to year. The reliability of such memories and the availability of actual measures of exposure can give you clues as to validity. For example, you would expect that memories would be better for more recent events than earlier ones. Thus, the validity of studies that include self-reported exposures from the distant past would probably be lower. Epidemiologists know that studies based on self-reports and recollections are subject to "recall bias"—that is, memories can be influenced by what participants think the investigators want them to say or by knowing that some effect is expected to have resulted from the exposure.

- What is the relationship between the time the exposures occurred and the time the effects were diagnosed?

Confidence Interval

(continued):
clear whether there is a positive (RR more than 1.0) relationship or a negative (RR less than 1.0) relationship. Thus, the results are not considered significant.

Controls:
Depending on the type of study, the population that has not experienced the exposure or that does not show the effect.

It may seem obvious that the exposure should precede the effect. But this is not always the case in epidemiological studies, and this fact may be buried in the methodology. For example, to have a large enough number of cases of leukemia to study, data may be combined over several decades. But the dry cleaning business in question may not have opened until 25 years ago. Thus, some cancers included in the study would have resulted from exposures that occurred before the business opened—and the occurence of such exposures would decrease the likelihood that exposure to dry cleaning agents caused the illnesses. While the timing of exposures and outcomes is often taken into account, there may be situations in which the authors implicitly assume a short latency period between exposure and effect; if this is the case, it should be stated explicitly, and the basis for this assumption should be explored with the authors and perhaps independent experts.

- Are the study findings consistent with those of other published research?

Have other epidemiological studies shown that leukemia occurs only after long-term exposures? If so, an increase in the incidence of leukemia shortly after the dry cleaning establishment began operations is not likely to have resulted from exposures to dry cleaning emissions. Therefore, even if the effect does follow exposure, other evidence may suggest that the interval between the alleged cause and the observed effect is not sufficient to show a true association.

- Do the findings jibe with those of others similar studies—e.g., studies of occupational exposures or of workers exposed to related solvents? Occupational studies are particularly valuable, because the exposures are generally higher than those in the environment, and therefore any effects should show up more strongly in workers.

- Are the findings consistent with general-population exposures to dry cleaning vapors and/or industrial emissions of the related chemicals? The greater the consistency of results, the more confidence you can have in the validity of the study of interest.

The fact that other cancers may be associated with the same exposure would not enhance the validity of the leukemia study. As noted in Chapter Four, cancers are a heterogeneous group of diseases, each with a variety of potential causes. It is very unlikely that a single agent or even a mixture of agents, acting alone, will cause different kinds of cancer. In assessing the evidence, consistency should be narrowly defined as the same exposure showing the same specific effect.

A misunderstanding of the importance of such specificity may be behind concerns that an environmental agent is causing cancer in a particular neighborhood. People may claim that three members of the family next door and two members of the family down the street have cancer, and so on. On investigation, it may turn out that, for the most part, the cancers differ and could not have been caused by exposure to a single agent.

- Have animal studies shown similar effects?

Results from cancer assays in animals may also contribute to an evaluation of the evidence. If these show that even high doses of dry cleaning emissions do not result in cancer, the claim that these emissions are causing leukemia in humans is seriously weakened. If the emissions cause a different kind of cancer in animals, you have a mixed message. Toxic agents don't always cause the same cancers in animals as in humans. Thus, this result would add somewhat to the evidence but not as strongly as would the result of animal studies showing the same cancer.

Animal studies may also provide information about the levels of exposure and the length of time necessary for potential carcinogens to cause cancer, which could be compared with the magnitude and time course of exposures demonstrated in the epidemiological study.

DIGGING DEEPER: ASSESSING APPLICABILITY TO THE REAL WORLD

Since the results of epidemiological studies reflect effects in humans, there is no need to be concerned, as you should be in the case of *in vitro* or animal studies, about the relevance of the findings to humans. But there may be questions about the relevance to groups other than the study population. For example, the results of a study of workers in dry cleaning establishments may not be directly relevant to environmental exposures. In addition to differences in the magnitude of exposures, other characteristics—genetic susceptibility, other health behaviors, diet, and lifestyle factors—may differ between workers and the general population, making comparisons questionable.

The more similar the study population is to the population reading the story, the more relevant the results are likely to be. If you are writing for a very broad audience, it might be useful to include mention of the uncertainties in applying the study findings to all people. Depending on the available information, you can be more or less specific about these uncertainties; e.g., including information about known ethnic differences in susceptibility to various agents can make your article more useful to your readers.

Making Sense of the Numbers

Risks can be reported quantitatively in various ways. These include the incidence of the risk—e.g., 1 in 10,000 over a lifetime; the number of people affected—e.g., an additional 200 people will contract cancer; and comparison—e.g., this will lead to a dou-

bling of the risk (or the risk will be 200% of background).

Cancer risk values are generally reported by the government as incidence risk—e.g., the lifetime risk of lung cancer from a certain amount of exposure to asbestos is 10 per million. Because incidence values are based on lifetime studies of animals, and cancer is assessed only at the end of the study (i.e., when the animal has reached "old age"), it is not possible to assess what the cancer risk is at earlier ages. When risk values are reported to the public, confusion may result—people often assume that the risk value refers to current risk, rather than the risk over an entire lifetime. Such misinterpretation can have a significant impact on the priorities people give to risk reduction and preventive activities.

A good example of the confusion this can cause is the oft-quoted estimate that women have a one in eight risk of developing breast cancer. This does not mean that any woman has this risk in any year, or that one in eight women will develop breast cancer every year. It means that if a woman lives to an advanced age (about 85) and doesn't die from other causes, her overall lifetime risk of developing breast cancer is one in eight, or about 12%. At any age below about 85, the risk for the total female population is lower than one in eight.

As noted earlier, ignoring the number of people who are potentially affected by a given health risk can also distort the picture. From a public health standpoint, a one in a million risk of developing nerve damage from a given agent for workers in a specialized field that employs 5,000 people is not nearly as serious as a one in a million risk from an air pollutant that is present in all large cities.

Using ratios to describe risk—e.g., a doubling of the cancer rate—can cause the same confusion if the

affected population is not considered. In addition, when using ratios, it is important to discuss the background risk—how much risk there is in the general population, at average exposures. A doubling of a cancer with a background lifetime population risk of 1 in 100,000 (e.g., liver cancer in women) will have a different impact from doubling a cancer for which the lifetime background risk is about 100 in 100,000 (e.g., breast cancer in women). The same holds true for percentages. It is important to remember that the true impact of a potential health risk is dependent upon both the size of the exposed population and the background risk.

Statistical Significance: The value used as the basis for deciding that two factors under investigation are related to each other rather than associated by chance. Scientists generally are willing to accept a finding as "statistically significant" if there is a probability of five percent or less that the results are due to chance. Usually, a *p-value* of 0.05 is the cut-off point; study results with p-values at or below 0.05 (5 percent) are considered statistically significant.

✓ A Note about Cancer Clusters

Health authorities are routinely called upon to investigate the presence of an alleged "cancer cluster"—that is, what people think is an unusual number of cancers in a small area. Often, those who believe they are affected live near an obvious contamination source, such as a hazardous waste site. In the vast majority of cases, closer investigation shows that what appeared to be a serious problem is probably the result of coincidence.

Why? First, people tend to underestimate the true incidence of cancer. The American Cancer Society estimates that between 25–35 percent of the population will have some type of cancer at some point in their lives; thus, it is not unusual to find people with cancer. Second, people think of cancer as one disease when in fact, as noted above, cancer is a group of many different diseases with different—and often multiple—causes. The presence in a particular geographic region of individuals with different types of cancer does not in and of itself suggest an unusual event attributable to a specific agent.

Furthermore, the public does not scientifically select the boundaries of the region in which a cancer clus-

ter has purportedy occurred. Instead, people are likely to decide selectively that a cancer cluster "ends" where an area of lower cancer incidence begins. Using this approach, it is easy to take a large area that has an average cancer incidence and divide it in such a way that some subsection has a higher incidence.

Finally, most purported cancer clusters are based on relatively small populations—e.g., one or two hundred people living in a particular neighborhood. With such a small population, variation can be very large and it would be difficult to show statistical significance for an outcome as common as cancer.

Rounding Out the Story

Keeping all of the above issues in mind, ask yourself the following questions.

p-Value: Probability value; a measure of chance. The closer the p-value is to zero (values range from 0.0 to 1.0), the likelier it is that the study results show a real effect. A p-value of 0.05 or less is generally accepted as evidence that there is a real effect (see also "statistical significance," above).

- How serious is the risk?

Four factors can be used to assess the strength of the association between exposure to a dry cleaning emissions and leukemia and help put the study in perspective: the risk ratio, the background incidence of leukemia, the size of the affected population, and the health impact. Each piece of information is a part of the puzzle. For example, you know the rate of leukemia has doubled. But what is the usual (background) rate? It may be that a background rate of one in 100,000 is being doubled in a population of 200 people. Or that a rate of one in 1,000 is being doubled in the entire U.S. population. These are two very different effects—the latter possibly quite serious, the former unlikely to yield even a single additional case in the exposed group.

- What is the public health message?

For the dry cleaners story, it would be helpful to get an estimate of the number of people likely to be

affected if the claim is valid. The availability of technologically and economically feasible strategies for reducing dry cleaning emissions and the possible health impact of using alternative agents would also help put the story in perspective.

A Note about Meta-analyses

As noted above, the wide variations both in people and in exposures can make it quite difficult to perform epidemiological studies that provide convincing evidence of an association between a particular agent and a particular effect. Thus, it is not surprising that epidemiological studies done at different times and in different places often have conflicting results. Examples include the studies of the possible association of power lines with childhood leukemia, secondhand smoke with lung cancer, and hormone replacement therapy with breast cancer. Another difficulty in assessing the evidence is that negative studies—those that show no effects from exposures—may not be published; thus, a review of the published literature might not show the entire picture.

Weighing the strengths and weaknesses of individual studies can help scientists interpret conflicting findings, but often such analyses do not yield a single, unambiguous answer. One way epidemiologists have attempted to address this situation is by performing meta-analyses. Simply put, this involves combining data from all or a subgroup of available relevant studies and determining the outcome. This approach is helpful because it results in a larger study population, and larger numbers afford an opportunity to establish statistical significance for subtle effects.

This approach has drawbacks, however. There may not be agreement on which studies to include and which to exclude from analysis. Also, methodology often varies from study to study, and combining data that have been collected and analyzed using varying

techniques can introduce confounders. Measurements relating to power lines, for example, may include electrical and/or magnetic fields and various surrogate measures. Or effects data may be gathered by various means—self-reports, hospital records, and death records, for example—that are not equivalent.

Finally, the studies included in a meta-analysis may have been performed on different populations; if these populations have different sensitivities, results from higher-risk populations may be diluted out or may lead to an overestimation of risk for populations with lower sensitivity. The more the studies in a meta-analysis differ, the less scientifically acceptable is the rationale for combining them. If you are faced with reporting a story based on a meta-analysis, it is critical to sort out these factors in assessing the validity and relevance of the findings to the health of a particular population.

NOTES

Chapter Six: INTERPRETING RISK REDUCTION CLAIMS

Does taking vitamin C reduce the risk of getting a cold? Despite a spate of studies conducted over many years, the answer remains unclear. Does calcium reduce the risk of osteoporosis? It seems so. But how best to get that calcium—from food, one type of supplement or another—is an open question. Can flaxseed reduce the risk of breast cancer? Possibly, according to studies in rodents. But that finding says nothing about the effects of human consumption.

The bottom line: Claims of risk reduction must be evaluated just as critically as claims of increased risk, and with a similar approach.

BACKGROUND

Claims of benefit made about specific foods, micronutrients, or other substances are particularly attractive to many publications. Such claims often serve as fodder for "feel good" stories—stories editors may insist on covering because they give readers something "positive" to do, or because, on a more "practical" level, such stories create product-friendly editorial environments for advertisers, which in turn may generate more revenue for the publication.

Companies that might benefit from such claims—juice, dairy-product, or cereal manufacturers, for examples—have a stake in prevention/risk reduction stories, which can be used directly or indirectly as promotional tools. Finally, risk reduction stories also benefit the medical journal that publishes research on which the claims are based; the articles are often highlighted in press releases to encourage media coverage, and this in turn enhances the recognition or "impact factor" of the journal.

Prevention:
Reducing or eliminating the occurrence of adverse health effects. *Primary prevention* is the preservation of health in susceptible individuals by personal and population-wide efforts (e.g., immunization of susceptible populations). *Secondary prevention* consists of early detection and prompt treatment of diseases in all populations (e.g. mammography to screen for breast cancer).

Because the agendas of diverse entities could be well-served by risk reduction stories, and because the pressures to cover such stories can be immense, it is particularly important to follow the basic tenets in this handbook in assessing the research, putting it in context, and presenting a balanced view of the findings.

As noted in Chapter Two, critically reading the article itself—not just a press release, even if it is issued by the journal—is particularly important, as are interviews with one of the authors and, when time permits, an independent source. If a story idea comes from a meeting presentation, see whether there is a published abstract. Even if there is, remember that it generally represents preliminary findings and is not subject to the same rigors of peer review as is a full journal article.

If the evidence is strictly anecdotal—that is, no scientifically acceptable studies have been performed—be careful. Anecdotal evidence alone, no matter how appealing, is not sufficient to substantiate any kind of claim.

Anecdotal Evidence: Data produced from nonscientific observations and experiences, e.g., testimonials.

The Claim

Claims of benefit made for products and/or interventions may include not only risk reduction claims, but also claims of prevention or cure. In fact, some authors use "risk reduction" and "prevention" interchangeably—and thus wrongly.

First, for either claim, you must know what population is involved. Interventions to reduce the risk of heart disease or a particular type of cancer often are tested in high-risk populations; thus, an intervention may be used to reduce the risk of heart disease in people with a family history of the disease, or to reduce the risk of a heart attack in people who already have documented atherosclerosis. Similarly, a prevention claim may mean preventing a cancer

recurrence in people who have had a documented occurrence, or prevention of cancer, such as breast cancer, in those with a strong family history (mother and sister have breast cancer). Findings relating to high-risk populations may not be applicable to the public.

Second, reducing the risk of developing an illness is not the same as preventing the illness. Even when risk is reduced, some risk remains; to most laypersons, the word "prevention" implies that the risk of occurrence is negligible or nonexistent.

The misuse of these terms is particularly likely to occur when scientific findings are overzealously interpreted for consumers. For example, the physician author of a consumer book who purported to present a breast cancer prevention diet later admitted that "risk reduction" would have been a better term to use. It also turned out that many of the book's claims—whether for risk reduction or prevention—were based on *in vitro* or animal studies. The author made a leap that the original investigators did not, by claiming the interventions were good for humans. Thus, the book gave readers a false impression of the evidence.

Separating science from exaggeration, and presenting the differences clearly, are among the most important responsibilities in risk reduction reporting.

How to Approach the Story

The process of handling the story is similar to the one described in the preceding chapters: reading the article critically, interviewing a study author, getting independent comments, and assessing the evidence and the applicability of the findings in real-life situations. Here are some key questions and points to keep in mind:

- What is the quality of the evidence?

In this handbook, we deal mainly with story ideas generated from scientific sources, such as a journal article, a meeting-presentation abstract, or a consensus panel, with appropriate supporting evidence in the form of other peer-reviewed articles.

But what if someone outside of the scientific establishment is making the benefit claim? Consider what is being offered in the way of supporting information: Is the intervention in use only outside the U.S.? Has it been described as effective in a language other than English? If so, it may make sense to have such material translated. Remember, however, that given the ease of global communication today, there is little reason to think that an intervention that truly reduces the risk of a given disease or disorder would be sanctioned only in a small part of the world. At the least, such findings are likely to have been presented at an international meeting. If no one knows about it, why not?

- Is the claim specific?

Most risk reduction claims pertain to reducing a person's risk of succumbing to a disease; to improving the chances of living a healthy, disease-free life; or to increasing longevity—all very attractive endpoints, but also very general and difficult to prove. Most diseases are complex and, except for certain inherited disorders, rarely have a single cause. Thus, a single intervention, no matter how beneficial, is unlikely to have a large impact. An exception is the halting of a proven detrimental behavior, such as cigarette smoking or having unprotected sex. Stopping a behavior that increases risk can reduce risk. Beyond this, showing cause and effect is no simple matter.

Even if a claim seems specific at first glance, a little reflection or investigation may reveal other-

wise. For example, the claim that eating a certain food such as broccoli will reduce the risk of cancer is specific only in that it is specifying broccoli as the beneficial substance. Again, cancer is a class of heterogeneous diseases, with many types and many, often interacting causes. Even a claim of risk reduction for a specific cancer, such as breast cancer, would be difficult to prove; a woman's risk of developing breast cancer depends on multiple factors, including genes, age, obesity, family history, and numerous lifestyle factors. Can eating one particular food or substance make a significant dent in overall risk? It's possible, but unlikely. Thus, to assess claims of benefit, we must know something about the disease, disorder, or condition to which the claim pertains.

- Is the claim based on the practitioner's clinical experience or on scientific studies?

If the claim is based on clinical experience, find out how many patients have been treated and what results were seen.

If the claim is based on scientific studies, find out what types of study are offered as evidence. Like claims of increased risk, claims of reduced risk may be based on a study on "test-tube," animal, or human studies. If the study is on animals, is the animal an appropriate model for the disease or disorder? Certain animal models are accepted as responding similarly to humans to agents given for a specific condition—e.g., the woodchuck for hepatitis B. Common laboratory animals, such as rats or mice, may be used in studies mainly because they are relatively inexpensive, have been extensively studied previously, or have a short life span—not because they have proven similarities to humans. Of course, even an accepted animal model is not a human being; thus, a risk reduction claim based solely on animal studies in a single species may be made only for that animal, not for people.

Regarding "test-tube" or laboratory studies, the same caveats presented for studies attributing a risk increase to a given substance (Chapter Four) hold for studies attributiong a risk reduction to a substance: How large was the dose? Is it realistic to think that a human could tolerate a large dose? Or could the substance be given in a sufficiently large dose to provide a benefit? For example, an agent may show powerful activity against the human immunodeficiency virus (HIV) *in vitro* at doses not obtainable by common routes of administration in people.

Epidemiological studies, as we saw in Chapter Five, may show an association between an intervention and risk reduction, but not cause and effect. One of the axioms of epidemiological studies is that their findings may be relevant to populations but cannot be extrapolated to individuals. Simply put, this means that you cannot draw conclusions about individual risk based on epidemiological studies, though you can make broad recommendations. For example, if after accounting for alternative explanations, several large epidemiological studies have shown that postmenopausal women who undergo hormone replacement therapy (HRT) are less likely to develop osteoporosis (a loss of bone density that increases the risk of fractures) than post-menopausal women who do not undergo HRT, a public health message that "HRT may reduce your risk of osteoporosis" may be considered.

But this is not the same as demonstrating the benefits in a randomized, controlled clinical trial, in which a particular intervention is compared with standard care or with a placebo. For the example above, a clinical trial might involve one group of postmenopausal women who are given HRT—an estrogen/progesterone combination pill once a day. Another group of postmenopausal women would receive a placebo once a day. Various assessments of bone density would be made at the onset of the trial, at several

points during the course of the trial, and at the endpoint—five years after the onset. If, after the completion of the trial, the group using HRT had a lower incidence of osteoporosis compared with the placebo group, the claim that HRT is useful in reducing the risk of osteoporosis has been supported by this trial. Still, other questions must be asked before accepting this or any other finding. For example, have other large, well-designed clinical trials reproduced the findings of this trial? And is there a known or accepted biological mechanism by which the intervention might result in the benefits achieved?

Certainly when reporting on the results of clinical trials, it is important to provide readers with the proper context for evaluating possible benefits and risks. For example, it is important to tell readers whether risks for other diseases, such as breast or endometrial cancer, increase with HRT. If so, public health messages about HRT need to include consideration of the balance between risks and benefits.

Digging Deeper: Assessing the Evidence

- How does the finding fit with other published studies?

As discussed in previous sections, one of the important contributors to validity is a pattern of evidence that supports a claim. This is especially important when dealing with studies involving humans, since such studies cannot be strictly controlled, as animal studies can, and human responses are quite variable. Thus, an article about a new risk reduction finding should include cautions that this single study is at best suggestive, and that it is possible that future studies will lead to different—even totally opposite—conclusions.

- How long was the study period?

For a claim of long-term benefit or risk reduction,

one needs to do a long-term clinical trial. For example, a Phase 3 trial of a drug thought to be neuroprotective in Parkinson's disease must run two years before a claim of neuroprotection—a long-term effect—can be made.

- How large is the study population?

Cost is a major factor limiting the size of scientific studies; the more study participants, the greater the expense. But if a population is too small, the results may not have enough "power" to show an effect or may not be generalizable to the population at large. Larger trials mean that patient compliance with the study protocol may be more difficult to control, which can also affect results.

- What are the side effects or long-term consequences of taking a substance for which a risk reduction claim is made?

For instance, taking a prescription weight loss pill for 3 months or 6 months may help an individual lose weight and thus reduce his or her risk of developing certain diseases. But if the same pill puts the person at risk of liver disease, then the risk–benefit equation changes. Sometimes, when the long-term consequences of taking a new drug aren't known, the drug is approved for use only for a specific period of time. Some people may nonetheless use it indefinitely—at their peril.

- What mechanism of action is proposed for the effect?

Very often, the mechanism by which a substance is said to produce beneficial effects simply isn't known, and if this is the case, the study author should say so.

When unsupported claims of benefit are made, the promulgators rarely say that anything about the inter-

Power: The ability of a study to detect an effect if there is one. Power is expressed as a percentage. The larger the sample size used in a study, the likelier it is that the study will find an effect, if there is one. If a study has a low power, it may only be able to detect large effects (such as a doubling of risk). Thus, a study with a low power might not "find" a small or moderate effect.

vention is unknown; part of the selling of the substance involves giving at least the appearance of a scientifically based mechanism. An extreme example is BioResonance Therapy, touted as a cure for all types of cancer. A long-winded explanation is presented, in which the dubious therapy is said to work through electromagnetic waves that "enliven" the p53 gene. This is nonsense, but it might sound plausible to someone who does not understand that genetic mutations cannot be undone by vibrations.

- What is the magnitude of the effect?

Even if the claimed percentage of risk reduction is enormous, in real life it may translate into a very small benefit compared with the risks involved. For example, a recent study assessed 639 women deemed at high risk of developing breast cancer who had undergone bilateral prophylactic mastectomy (i.e., they had had both breasts removed, even though they had showed no evidence of breast cancer). The authors calculated that, at 14 years follow-up, the procedure reduced the risk of developing breast cancer in these women by at least 90 percent, compared with those at similar risk who did not undergo the procedure.

A look at the actual numbers revealed that, instead of an expected 37 cases of breast cancer, only four developed in women who had undergone the procedure; thus, the 90% risk reduction claim was accurate. But what of the 602 other women who had undergone the procedure? These women either had not been destined to get breast cancer, or if they had gotten it, might have been cured without having to undergo bilateral mastectomy. Thus, 602 women had unnecessary surgery to spare 33 others. Of course, the women who underwent the procedure could not have known in advance that they were among those who did not need it. Nevertheless, it puts the 90% risk reduction claim in a different perspective.

- How does the intervention compare with standard therapy?

How does the intervention compare with the benefits of standard therapy, and how have such comparisons been made? Comparisons based on objective measures are more reliable than those based on subjective assessments. For example, a study that compares a new type of cholesterol-lowering treatment with standard treatment may assess the effect of each intervention on the actual lowering, or the percentage of lowering, of cholesterol over the course of the study period. This is an objective measurement. On the other hand, a study that evaluates the ability of a treatment to relieve a symptom that is difficult to assess objectively—such as nighttime cough in asthma—must usually rely on subjective patient reports, which may be unreliable.

As noted in Chapter Four, when people are asked to compare how they felt before and after an intervention, recall bias must be considered. When an intervention is said to reduce the risk of pain (e.g., head pain in migraine) the criteria used to assess the reduction cannot be truly objective, even if the same measurement tool (e.g., a scale of 1 to 5) is used before and after the intervention.

- How does the intervention compare with a placebo?

Sometimes an intervention is not compared with standard care, but rather with a placebo—an inactive medication or procedure. Using a placebo as a comparison is not the same as comparing the effectiveness of one drug with another in reducing the risk of disease, or comparing a drug intervention to a dietary intervention. Thus, when looking at risk reduction claims for interventions such as drugs, supplements, or diet, it is important to know what comparisons have been made.

The placebo effect itself can be very powerful. Numerous studies have shown that when people participate in a clinical trial, improvement is reported by some who receive the placebo as well as by some who get the active treatment. The care and attention received during the study period and the participant's motivation to get well are important factors in the positive effects seen with placebos.

In fact, the placebo effect may be as high as 40 percent in drug trials for some conditions such as migraine. That means that 40 percent of people might feel better simply because they believe they are helping themselves and that others are helping them. If the active treatment reduces pain in 45 percent of people, that is only 5 percent more than might be expected without the treatment.

The placebo effect must be considered when assessing all claims of risk reduction, but particularly for symptoms such as pain, which often has psychological components and for which benefit is difficult to assess using objective measures.

Digging Deeper: Applicability to the Real World

- What is the study population?

Several large, ongoing epidemiological studies involve individuals in specific communities, such as those in Framingham, Massachusetts, or in Rochester, Minnesota—who have been participants for many years. When drawing conclusions about interventions that might benefit the public's health, it is important to remember that these populations are not representative of the U.S. population as a whole, which is much more diverse both in demographics and in health risk factors. Recently, for example, the U.S. federal food guidelines'recommendation of two or more daily servings of dairy products has been questioned by activist groups, who note that many minorities have lactose intolerance. It has been sug-

gested that the dairy food group be renamed the "calcium-containing" food group and expanded to include soy and tofu products.

The characteristics of study populations must also be assessed. Some risk reduction studies use people who already have a particular disease, such as coronary artery disease, and attempt to show how an intervention—taking a cholesterol-lowering drug, for example—reduces the risk of a consequence of the disease, e.g., a heart attack. As noted earlier, participants may be people who are already hospitalized, or who have had previous heart attacks, rather than people in the community at large without these risk factors.

A Note About Dietary Supplements

Many herbal products, hormones such as dehydroepiandrosterone (DHEA), and other agents sold in health food stores (and, increasingly, in other types of retail stores), are legally and popularly considered "dietary supplements" rather than drugs. In the U.S., the regulation of dietary supplements is much less strict than is the regulation of drugs. Manufacturers may make claims for supplements—e.g., that they reduce stress and fatigue or make you feel years younger—without providing the extensive evidence that would be required to substantiate a similar claim for a drug. Unlike drugs, dietary supplements do not necessarily undergo rigorous study concerning benefits and risks. Claims are likely to be supported by anecdotal evidence; physicians may say they work "based on my experience." Claims of this sort are totally unscientific.

To promote such products, manufacturers and vendors may claim they are safe because they are "natural"—implying that natural chemicals are somehow inherently safe, despite that they may have exactly the same properties as their synthetic counterparts.

Such claims may completely misrepresent a product's ingredients, setting the stage for possible harm. For example, a manufacturer of a "natural" so-called energy booster claimed the product contained no estrogenic compounds. But a little research on the plant extracts in the product revealed that they contained potent phytoestrogens.

NOTES

PART 4: "HERE'S THE REAL STORY!"

CONCLUSION

After you have gathered and assessed the evidence, interviewed appropriate sources, gained some insight into the issues surrounding a particular risk claim, and reached some conclusions about its impact on the public's health, you are ready to write the story. Very often you will have to distill a wealth of information into a relatively small space. Here are some tips for telling the "real story" in a succinct, accurate, and lively way:

- Tell the reader in the first couple of sentences what the risk claim is, who is making it, and why the reader should care about this issue.

- Write the story as you understand it; don't just parrot what a source says unless you are quoting the person directly.

- Write non-technically even if you are writing for a professional readership. You may use more technical terms for a professional readership, but don't make the mistake of thinking that stories written for medical professionals or scientists should be filled with jargon. Boring stories are unlikely to be read through to the end by anyone.

- Include qualifiers when appropriate (i.e., when the data support the claim) rather than making blanket statements (e.g., say "Exposure to this agent may cause respiratory failure if you have severe asthma," not "Exposure to this agent causes respiratory failure if you have asthma").

- Remember that there are many ways of saying the same thing. If you describe a risk in a potentially alarmist way (e.g., "One in eight women will get breast cancer at some point in their lives"; see Chapter Five), explain what the risk means in real-world terms (e.g., "This means that if a woman lives to the age of about 85 and doesn't die from other causes, her overall lifetime risk of developing breast cancer is one in eight, or about 12 percent; at any age below about 85, the risk is lower than one in eight").

Also remember that reporting risk using percentages can be problematic because the reader may not get the proper message. For example, an increase of 50 percent in the risk may be misinterpreted by the reader as equivalent to a doubling of the risk rather than an increase to one and one-half times the risk. Thus, percentages should be carefully explained or accompanied by the corresponding ratio.

Ultimately, readers must decide for themselves how dangerous a risk is and whether they will take action to reduce it. But journalists can play a key role in encouraging sound decision-making by providing a balanced view of the data and presenting relevant information to help their readership evaluate the risk. When you take on a risk story, you also take on the responsibility for telling the "real" story. This can be a rewarding challenge!

GLOSSARY

anecdotal evidence: data from nonscientific observations and experiences, e.g., testimonials. (See page 58.)

association: relationship between an exposure and a health effect—e.g., between exposure to sunlight and skin cancer. Association is not the same as causation. Associations can be positive or negative. For example, a positive association exists between cigarette smoking and lung cancer. A negative association exists between the use of seat belts and deaths from car crashes. (See page 19.)

bias: a factor or process that systematically distorts the collection or analysis of study results. An example is "recall bias"—people who have diseases may be more likely to report real or imagined exposures to the study agent than those who are disease-free. (See page 17.)

cancer potency value: a quantitative measure of cancer risk derived from applying the linearized multistage (LMS) model. It describes how cancer risk varies according to dose. (See page 33.)

carcinogenic: capable of causing cancer. (See page 29.)

clinical trial: a study in humans that assesses the efficacy and safety of a specific intervention or treatment on a disease or other health-related outcome. Generally, patients are randomly assigned to receive the study treatment or another treatment, or a placebo, and are followed over a period of time. Findings from the different study groups are then compared to assess the effects of the intervention. (See pages 18–19.)

confidence interval: the margin of error calculated for a given risk estimate. Generally, 95% confidence intervals are calculated; this means that there is a 95% probability that the true risk is no higher or lower than the range of values included in the interval. For example, a risk ratio (RR) of 2.0 might have a 95% confidence interval ranging from 0.5 to 4.0. A RR of 1.0 suggests that there is no association between an exposure to an agent and the outcome being measured. In a case like this, where the confidence interval spans the value of 1.0, it is not clear whether there is a positive (RR more than 1.0) relationship or a negative (RR less than 1.0) relationship. Thus, the results are not considered significant. (See pages 46–47.)

confounder: a factor that can distort the results of the study because it is related to both the exposure and the effects—e.g., secondhand cigarette smoke may be a confounder in studies of the effects of an environmental contaminant on childhood asthma. (See page 17.)

controls: depending on the type of study, the population that has not experienced the exposure or that does not show the effect. (See page 47.)

dose response: the relationship between exposure to an agent and the occurrence of an effect. Generally, the higher the dose, the greater the response. (See page 18.)

epidemiological study: a study of the patterns of human disease and the factors that are associated with specific diseases. (See page 18.)

evidence: data collected scientifically. In risk assessment, this can include data about exposures as well as effects. Effect data are generally derived from laboratory studies or observations of human populations. (See page 7.)

exposure: coming into contact with an agent. *Acute exposure* occurs when a person comes into contact with a single dose or small number of doses of the agent over a short period of time, such as one day. *Chronic exposure* occurs when a person receives repeated doses over a long period of time, e.g., years. (See page 14.)

hazard: the potential of an agent to cause an adverse effect under specific circumstances. (See page 26.)

laboratory study: a study done in test tubes *(in vitro)* or in animals *(in vivo).* (See page 17.)

linearized multistage (LMS) model: a mathematical model used to estimate the carcinogenic potency of specific agents. The LMS model is based on data collected from giving animals very high doses of an agent over their lifetime. This is not a biological model; it does not include any consideration of the mechanisms involved in cancer formation. (See page 35.)

magnitude of the effect: a quantitative measure of the strength of an association between an exposure and an effect; often measured using the risk ratio. (See page 44.)

magnitude of risk: a quantitative estimate of the degree of risk associated with a specific exposure—e.g., one additional cancer in one million exposed individuals. (See page 24.)

mechanistic studies: investigations to determine how a particular effect is produced. These include studies of how chemicals are absorbed into the body, changed in the body, distributed throughout the body, excreted from the body, and stored in the body, and how they have toxic or beneficial effects. (See page 20.)

methodology: the way in which data are collected. Methodology descriptions include identification of the study subjects or specimens, sources of the agents being tested, and the techniques used to assess exposure or effect. (See page 8.)

natural: occurring in nature. Natural substances may show a wide range of toxicities. (See page 28.)

p-value: probability value; a measure of chance. The closer the p-value is to zero (values range from 0.0 to 1.0), the likelier it is that the study results show a real effect. A p-value of 0.05 or less is generally accepted as evidence that there is a real effect (see also "statistical significance," below). (See page 53.)

peer review: careful scientific examination of data, methodology, results and conclusions of a study or set of studies by experts in the area of the research. (See page 7.)

power: the ability of a study to detect an effect if there is one. Power is expressed as a percentage. The larger the sample size used in a study, the likelier it is that the study will find an effect, if there is one. If a study has a low power, it may only be able to detect large effects (such as a doubling of risk). Thus, a study with a low power might not "find" a small or moderate effect. (See page 64.)

precautionary principle: the premise that actions should be taken to reduce or eliminate a hazard even if there is little credible evidence that a significant risk exists, as a precaution against potential risk. (See page 27.)

prevention: reducing or eliminating the occurrence of adverse health effects. (See page 57.) *Primary prevention* is the preservation of health in susceptible individuals by personal and population-wide efforts (e.g., immunization of susceptible populations). *Secondary prevention* consists of early detection and prompt treatment of diseases in all populations (e.g., mammography to screen for breast cancer).

relevance: the extent to which toxicity findings from laboratory, epidemiological or clinical studies are applicable to people living in the real world. Studies done under unrealistic conditions—e.g., when animals are exposed to extremely high doses of an agent—may not be relevant to people exposed to low environmental levels of the same agent. (See page 23.)

reverse onus: placing on those who claim an agent is safe the onus of proving that it will not have adverse effects under any circumstances. (See page 27.)

risk: the possibility of suffering harm or adverse effects. *Acute risk* is the risk from a single or a few closely spaced exposures to an agent*; chronic risk* is the risk from exposure over a significant period of time*; lifetime risk* is the risk from exposure to an agent over an entire lifetime—generally assumed to be 70 years. (See page 6.)

risk assessment: the process that estimates the type and magnitude of risk to human health, wildlife, or the environment posed by exposures to hazardous agents. (See page 24.)

risk management: the steps undertaken to reduce or eliminate risks. This may be accomplished either by controlling the source of the risk (e.g., limiting emissions from an industrial source) or changing behavior (e.g., issuing a fish consumption advisory). (See page 16.)

risk ratio (RR): the ratio of the risk (or incidence) of disease or death in the exposed population to the risk in the unexposed or comparison population. RR is used in epidemiology to indicate how much more or less likely individuals with a particular characteristic—e.g., obesity, cigarette smoking—are to show the effects under investigation (e.g., heart disease, lung cancer) than those who do not have that characteristic. Risk ratios above 1.0 indicate a positive relationship—i.e., the risk is more likely when the agent is present—while risk ratios below 1.0 indicate that the risk is less likely when the agent is present (i.e., the agent is protective). (See page 45.)

"safe" value: a quantitative estimate of the amount of a noncarcinogenic substance a person can consume without incurring appreciable risk; exposures above the "safe" value do not necessarily result in adverse effects. (See page 25.)

sample size: the number of participants or subjects being examined in a study. The larger the sample size, the more likely it is that the study will be able to detect an effect, if there is one. (See page 43.)

statistical significance: the value used as the basis for deciding that two factors under investigation are related to each other, rather than being associated by chance. Scientists generally are willing to accept a finding as "statistically significant" if there is a probability of five percent or less that the results are due to chance. Usually, a *p-value* of 0.05 is the cutoff point; study results with p-values at or below 0.05 (5%) are considered statistically significant. (See page 52.)

synthetic: not natural; formed through human intervention. Generally refers to compounds made in the laboratory. Synthetic substances may show a wide range of toxicities. (See page 28.)

toxic: harmful, poisonous. All chemicals or substances are toxic under some conditions, and in large enough amounts. (See page 6.)

toxicity: the inherent potential of a substance to cause adverse effects in living organisms. "Acute toxicity" refers to adverse effects that occur very soon after a single exposure or small number of closely spaced exposures to a toxic substance; "chronic toxicity" refers to adverse effects that occur after repeated exposures over a long period of time. (See page 11.)

validity: degree of confidence that the conclusions from a scientific study are accurate (i.e. not due to chance, bias, confounders and other causes of distortion). (See page 16.)

RECOMMENDED READING

American Chemical Society, *Understanding Risk Analysis: A Short Guide for Health, Safety and Environmental Policy Making,* Washington, DC, 1998.

Cohn, V., *News and Numbers: A Guide to Reporting Statistical Claims and Controversies in Health and Other Fields,* Iowa State University Press, Ames, IA, 1989.

Glickman, T.S., and Gough, M., eds., *Readings in Risk,* Resources for the Future, Washington, DC, 1990.

Kamrin, M., *Toxicology: A Primer on Toxicology Principles and Applications,* Lewis Publishers, Chelsea, MI, 1988.

Kamrin, M., Katz, D., and Walter, M.L., *Reporting on Risk: A Journalist's Handbook on Environmental Risk Assessment,* National SeaGrant College Program, Ann Arbor, MI, 1995.

Lauden, L., *The Book of Risks,* Wiley & Sons, New York, 1994.

Ottoboni, A.M., *The Dose Makes the Poison: A Plain Language Guide to Toxicology,* J. Wiley & Sons, New York, 1997.

Ward, B., *Chemicals, the Press and the Public: A Journalist's Guide to Reporting on Chemicals in the Community,* Environmental Health Center, National Safety Council, Washington, DC, 1989.

West, B., Sandman, P.M., and Greenberg, M.R., *The Reporter's Environmental Handbook,* Rutgers University Press, Piscataway, NJ, 1995.

RECOMMENDED READING FROM THE AMERICAN COUNCIL ON SCIENCE AND HEALTH:

American Council on Science and Health, *Cigarettes: What the Warning Label Doesn't Tell You*, Prometheus Books, Amherst, NY, 1997.

"Cancer Clusters," American Council on Science and Health, New York, 1989.

"Holiday Dinner Menu," American Council on Science and Health, New York, 1999.

Juberg, D., "Traces of Environmental Chemicals in the Human Body: Are They a Risk to Health?" American Council on Science and Health, New York, 1999.

Of Mice and Mandates: Animal Experiments, Human Cancer Risk, and Regulatory Policies, American Council on Science and Health. New York, 1997.

Whelan, E.M., *Toxic Terror: The Truth Behind the Cancer Scares,* Prometheus Books, Buffalo, NY, 1993.

AMERICAN COUNCIL
ON SCIENCE AND HEALTH
1995 BROADWAY, 2ND FLOOR
NEW YORK, NY 10023-5860 USA

Non-Profit Org.
U.S. Postage
PAID
Permit No. 9513
New York, NY