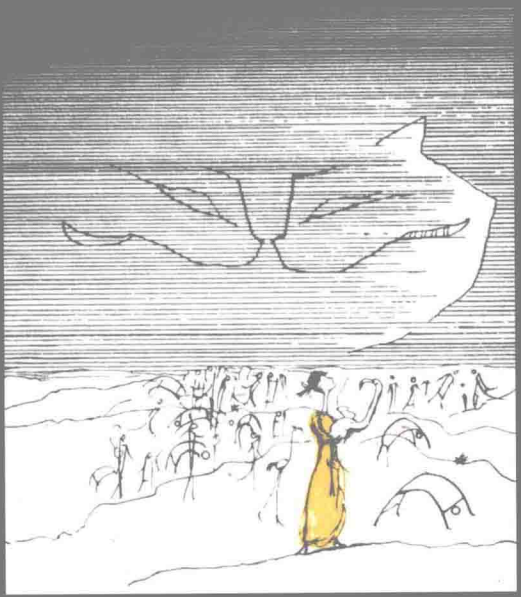
# PHANTOM RISK

## SCIENTIFIC INFERENCE AND THE LAW



EDITED BY

KENNETH R. FOSTER, DAVID E. BERNSTEIN,

AND PETER W. HUBER

# Phantom Risk

## Scientific Inference and the Law

edited by Kenneth R. Foster, David E. Bernstein,
Peter W. Huber

The MIT Press
Cambridge, Massachusetts
London, England

First MIT Press paperback edition, 1999



Published for the Manhattan Institute by The MIT Press.



This book was set in Bembo by Compset Inc and printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Phantom risk : scientific inference and the law / edited by Kenneth R. Foster, David E. Bernstein, Peter W. Huber.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-06156-2 (hc), 0-262-56119-0 (pb)
1. Science and law. 2. Torts. 3. Causation. 4. Risk. I. Foster, Kenneth R. II. Bernstein, David E. III. Huber, Peter W. (Peter William), 1952–
K487.S3P43 1993
346.03—dc20
[342.63]
92-36137
CIP

# Preface

This book examines two intersecting themes: the problems of assessing subtle environmental or occupational risks, and the havoc this creates in the courtroom. In each chapter, a scientist addresses an occupational or environmental health issue that has figured prominently in tort litigation. Each section concludes with a brief summary of the litigation. The final chapter offers guidelines to help the legal system deal with the very real problem of phantom risk.

The editors and authors strive to discuss the scientific and legal issues in a way that will be accessible to the "intelligent lay reader," who—despite occasional claims to the contrary—is not an endangered species.

We gratefully acknowledge the generous support of the Sloan Foundation and the Manhattan Institute for Policy Research. We thank the authors of these chapters and many colleagues for suggestions and helpful discussions, with special thanks to Michael Gough and Isadore Rosenthal. We have also benefited from the excellent secretarial assistance of Marilyn Leeland and our paralegal assistant, Karin Albani.

# Contributors

**Bruce N. Ames** and
**Lois Swirsky Gold**
Division of Biochemistry and
Molecular Biology
401 Barker Hall
University of California
Berkeley, CA 94720

**David E. Bernstein**
Crowell and Moring
1001 Pennsylvania Avenue, N.W.
Washington, DC 20004

**Bernard L. Cohen**
Department of Physics
and Astronomy
University of Pittsburgh
100 Allen Hall
3941 O'Hara Street
Pittsburgh, PA 15260

**Richard S. Cornfeld**
Coburn, Croft & Putzell
One Mercantile Center
Suite 2900
St. Louis, MO 63101

**Kenneth R. Foster**
Department of Bioengineering
University of Pennsylvania
220 South 33rd Street
Philadelphia, PA 19104-6392

**Michael Gough**
Office of Technology Assessment
U.S. Congress
Washington, DC 20510-8025

**Peter W. Huber**
5029 Edgemoor Lane
Bethesda, MD 20814

**Rudolph J. Jaeger** and
**Arlene L. Weiss**
Environmental Medicine, Inc.
263 Center Avenue
Westwood, NJ 07675

**Renate D. Kimbrough**
Institute for Evaluating Health Risks
Suite 608
1101 Vermont Ave., N.W.
Washington, DC 20090-1195

**Ralph E. Lapp**
7215 Park Terrace Road
Alexandria, VA 22307

**Louis Lasagna** and
**Sheila R. Shulman**
Center for the Study of
Drug Development
Tufts University
192 South Street
Boston, MA 02111

**Michael I. Luster,**
**Gary J. Rosenthal,** and
**Dori R. Germolec**
Systems Toxicity Branch
National Toxicology Program
National Institute of Environmental
Health Sciences
Research Triangle Park, NC 27709

**James L. Mills**
Epidemiology Branch
Prevention Research Program
National Institute of Child Health and
Human Development
National Institutes of Health
6100 Building
Room 7B03
Bethesda, MD 20892

**Marvin M. Romsdahl**
University of Texas
M.D. Anderson Cancer Center
1515 Holcombe Boulevard
Houston, TX 77030

**Stuart F. Schlossman**
Dana-Farber Cancer Institute
Harvard Medical School
44 Binney Street
Boston, MA 02115

**George K. Tokuhata**
410 Rupley Road
Camp Hill, PA 17011

**Ralph D'Agostino, Jr.**
and **Richard Wilson**
Lyman Laboratory of Physics
Harvard University
Cambridge, MA 02138

*Cognitive dissonance is all but unavoidable when the data are ambiguous and the stakes are high.*

Weinberg 1985, p. 67

*We would know nothing at all about chronic risk attributable to most toxic substances if scientists had not detected and evaluated them. Our response to such risks, therefore, must be based on a set of scientific findings. Science, however, is hardly ever unambiguous or unanimous, especially when the data on which definitive science must be founded scarcely exist.*

Ruckelshaus 1985, p. 26

# Contents

# 1

# A Scientific Perspective

Phantom risk is not lung cancer from smoking, or pelvic inflammatory disease (PID) from one intrauterine contraceptive device, the Dalkon Shield. It is not mesothelioma from asbestos; it is surely not AIDS from unprotected sex.

Phantom risk is, arguably: miscarriage resulting from work at video display terminals, birth defects in children caused by the mother's use of the drug Bendectin, cancer from low-intensity magnetic fields, lung cancer produced by the very slight levels of airborne asbestos in buildings with intact asbestos insulation, or cancer resulting from a slip and fall in a grocery store. By *phantom risk* we mean cause-and-effect relationships whose very existence is unproven and perhaps unprovable.

In using the term *phantom risk* we do not question the existence of human tragedies or imply lack of sympathy for the affected individuals. A family that has a child with a major birth defect has suffered real tragedy, whatever caused it. The problem we address is whether its cause can be identified and (inevitably in our litigious society) how to resolve the resulting legal claims. Phantom risk is a problem at the interface of science and the law.

We are interested, for present purposes, in the problems that arise in tort litigation in proving that an injury did (or did not) arise from low-level exposure to a potentially hazardous substance. Phantom risk also causes problems in the regulatory process, but that is not of principal concern in this book. The issues we discuss aroused much public anxiety and litigation, yet the hazards were never proven, or at least not proven to be serious at the levels of exposure that litigants experienced.

This book explores two intersecting problems. The first is the great disparity between the ease with which a controversy about a suspected hazard can begin and the difficulty in resolving the nature of the connection, if any, between the suspected hazard and a health effect. The second is the havoc the resulting confusion wreaks in the courts. The two prob-

lems cannot be separated: litigation may reflect scientific controversy, but it may also help to create it.

Phantom risk can arise from chance observations in everyday life. Breast cancer developing after serious trauma to the chest (chapter 15), the birth of a child with a major defect following use of the drug Bendectin by the mother (chapter 5), miscarriage in a user of a video display terminal (chapter 6), the illness of a child after a vaccination, or the appearance of essentially any cancer in a person exposed to a pesticide might understandably appear as causal sequences to the people concerned. Such incidents, which in retrospect (and in the light of accepted scientific theory and clinical practice) might be regarded as chance occurrences, helped to fuel the scientific and legal controversies discussed in this book.

Phantom risk may also arise from errors and ambiguity in science itself. Often in science, the first studies in a new line of investigation are preliminary in design, crude in execution, or flawed in concept. Follow-up studies are often better designed, more focused, better executed, and less ambiguous in interpretation. Phantom risk is a manifestation of the confusion and uncertainty that invariably accompany risk research, a science that in all respects is in its early stages of development. This confusion is particularly troublesome in epidemiology, which is an observational (rather than experimental) science and therefore more subject to ambiguity in interpretation than an "exact" science such as physics.

## Setting the Stage

*Risk* in technical usage is the probability that a person will experience an adverse effect from some activity or exposure. *Risk assessment* is the process of quantifying and evaluating risk. Risk assessment occupies the time of many scientists and physicians in academia, industry, and government, and a vast literature exists on the subject. [Good introductions are the collection of readings by Glickman and Gough (1990) and the book by Rodricks (1992)].

A broad perspective of the process of risk assessment is given in an influential 1983 report by the National Research Council. According to this report, risk assessment can be broken down into four stages:

*Hazard identification* is the process of determining whether exposure to an agent can cause an adverse health effect, such as an increase in incidence of cancer or miscarriage. The question "Can ionizing radiation cause cancer?" is one of hazard identification. (The answer is an unqualified yes.)

*Dose-response assessment* is the process of quantifying the relation between the dose of an agent and the increase in incidence of the health effect. The question "How much exposure to ionizing radiation is needed to produce a 50 percent increase in the incidence of leukemia?" is one of dose-response assessment. (A great deal of exposure is needed, roughly comparable with what many people experienced at Hiroshima when the atomic bomb exploded.) Many of the controversies described in this book involve claims of injury from far smaller exposures.

*Exposure assessment* is the process of measuring how much exposure people actually receive from a potentially hazardous agent. The question "How much exposure did people receive from radioactive fallout from atmospheric weapons testing?" is one of exposure assessment. (The answer, for most people, is very little, compared with other sources of exposure; see chapter 13).

Finally, *risk characterization* involves the estimate of the magnitude of the public health problem involved with exposure to a potentially hazardous agent and other study of the problem.

Thus, at least three factors must be considered when assessing the likelihood that a person was harmed by exposure to an agent (which is, after all, the central question in personal injury suits). The first is the existence of a hazard, that is, the potential health effect. The second is the exposure the person received, and the third is the level of risk associated with that exposure. Public discourse (and media coverage) about subtle environmental hazards usually focuses on the first of these factors (e.g., whether radiation causes cancer); the other two need to be considered as well.

## Association and Causation

The most direct way to identify a hazard is to observe human populations. Epidemiology, as defined by one authority, is "the branch of applied statistics that studies the determinates and correlates of human disease" (Bailar 1989). Epidemiology has a strong medical component also; perhaps the most famous collaboration in epidemiology in recent decades has been between Richard Peto (a biostatistician) and Richard Doll (a physician/epidemiologist).

By observing human populations, epidemiologists try to identify statistical associations between exposure (use of a drug or exposure to an environmental chemical, for example) and incidence of a disease.

For example, consider the birth of a child with a major defect to a woman who had used the drug Bendectin during pregnancy. This event,

certainly tragic to the family, is not necessarily an adverse effect of the drug. Major birth defects occur in roughly 3 percent of all pregnancies, including users of Bendectin even if the drug had no effect on the outcome of the pregnancy. As Lasagna and Shulman point out (chapter 5), the interpretation of such an event depends on whether exposure to the drug is associated with an increase over the usual incidence of such outcomes. This problem is best approached by means of controlled observations on populations of users and nonusers of the drug.

For such purposes, a useful concept is *relative risk,* in this case the incidence of serious birth defects in a population of Bendectin users, divided by that in a population of nonusers. A relative risk of 1 indicates lack of association between use of Bendectin and birth defects in the child; progressively higher values indicate progressively stronger associations. As Lasagna and Shulman point out, epidemiologic studies failed to make a strong case linking birth defects and the mother's use of Bendectin, despite heart-rending cases that were well publicized in the media.

Consider, by contrast, smoking and lung cancer. Epidemiologic studies by Doll and Hill (1952) conducted in the early 1950s strongly indicated that a pack-a-day smoker has a tenfold higher chance of developing lung cancer than a nonuser. These studies (which have been repeatedly confirmed by other investigators) stand as classics in their field; some of their results are shown in chapter 3. Moreover, lung cancer is a major public health problem, and one that is largely avoidable. The disease, primarily a result of smoking, now accounts for the largest number of new cancer cases in the United States, about 157,000 cases in 1990 (Henderson et al. 1991).

Epidemiology provides information that is directly relevant to health and disease in humans. This must be weighted against the frequent difficulty in interpreting the evidence, and the frequent inconsistency of epidemiologic findings. Our authors emphasize the need to examine the whole body of evidence about a suspected hazard, and judge what can be concluded reliably in spite of inconsistency and confusion.

## Interpreting the Evidence

The authors in this book are clearly optimistic about the use of science to detect and evaluate risks. But they point to the need to weigh conflicting scientific evidence and to debate interpretations of data. In so doing, they often return to two major themes. The first is the difficulty of inferring cause-and-effect relationships from epidemiologic evidence. The

second is the difficulty of inferring risks to humans from high-dose animal experiments.

In sorting through the evidence, our authors repeatedly evaluate epidemiologic studies along several dimensions. The first is the strength of association between exposure and a disease, that is, the relative risk. Many arguments about phantom risk revolve about small relative risks, that is, about weak statistical associations. A relative risk close to 1 implies that the exposure is, at best, only one of perhaps many factors that contribute to the development of a disease.

The second dimension is *significance of association*. Because of statistical uncertainties (such as sampling error), the results of an epidemiologic study are surrounded by a penumbra of uncertainty. The size of a study is an important factor; sampling error declines as study size increases. A study might report a relative risk greater than 1 (indicating an increase in risk), but with a margin of sampling error that includes a relative risk of 1. Thus, the increase is not statistically significant, which means that it has an uncomfortably high probability of arising from sampling error. This creates obvious problems for laymen on juries who must interpret such evidence, as happened in the Bendectin litigation.

A related problem, mentioned by several authors, is that some risks are reported on the basis of very limited data. Foster notes (chapter 3) that the claim, often repeated in the lay media, that 10–15 percent of all childhood cancers result from exposure to magnetic fields, derives from data from a group of 27 households in Denver. In statistical jargon, such results are unstable, that is, highly uncertain because of the small numbers of subjects.

The final dimension is *validity of a study*. Epidemiologic studies are susceptible to many potential errors, which are particularly difficult to control in early stages of research on a problem whose dimensions and nature are unknown. For example, Mills suggests (chapter 4) that early studies linking spermicides and birth defects were flawed and seriously misleading. Kimbrough describes major flaws in early studies that reported adverse health effects of polychlorinated biphenyls (PCBs) (chapter 9). Gough (chapter 11) describes studies that are inconsistent with (and probably invalidate) the earliest studies that linked low-level exposure to dioxin and cancer. In all of these cases, the early positive findings were not confirmed by later investigation and are presumably wrong.

Recall bias is one potential difficulty that several of our authors mention. This problem arises if some individuals in a study are more likely than others to report using a drug or remember an exposure. Epidemiologists have learned, through hard experience, that external factors

such as bad publicity about a drug will affect the likelihood that a person will report having used it. Recall bias may have been a factor in studies of miscarriage and video display terminals (chapter 6) or birth defects and spermicides (chapter 4) or Bendectin (chapter 5). Again, these early studies were not confirmed by later investigations and are presumably in error.

Another problem that our authors frequently mention is the multiple comparison effect. An investigator who makes enough comparisons in a study will surely find many "statistically significant" associations in the data, merely because of chance. For example, there are many different kinds of birth defects, and an investigator fishing for some connection between use of a drug and birth defects can make many comparisons, each of which (under the statistical rules most investigators adopt for interpreting such studies) has 1 chance in 20 of yielding a "statistically significant" difference, even though there may be no real effect of the drug. There are many occupations and many forms of cancer, and an investigator who performs enough comparisons will surely find associations between the two.

This problem is obvious enough, but it is a recurrent and hard-to-prove source of error. According to our authors, initial reports of associations between Bendectin and birth defects (chapter 5), or dioxin and cancer (chapter 11), or "electrical" occupation and cancer (chapter 3), or PCBs and diverse health effects (chapter 9) may have been false-positive errors that arose from the many comparisons the investigators performed. The initially reported "associations" were not confirmed by later studies, which were more focused and less susceptible to the multiple comparison problem.

Some errors might lead an investigator to *underestimate* a risk. One common problem is the misclassification of subjects according to their exposure. A study that randomly grouped its subjects as "exposed" or "unexposed" to a suspected hazardous agent would detect no effect, even if one existed. Smaller misclassification errors will tend to diminish the size of any detected effect.

For this and other reasons, a scientist might view the borderline results of a study as suggesting a large effect. If so, additional studies with better exposure assessment will show the effect more clearly. For the scientific issues discussed in these pages, the trend has been in the opposite direction.

Much of the legal controversy in these pages turns on different interpretations of epidemiologic evidence. In epidemiology, as in any science, the search for small effects is uncertain and confusing; and epidemiology

is hardly an exact science in any event. Unfortunately, the irreducible uncertainties in epidemiology are frequently large enough to be legally significant. The usual test in civil litigation is whether the suspected agent "more likely than not" caused a claimant's disease. This might be interpreted as whether the exposure doubled the risk. Many epidemiologists, if pressed, would admit that twofold increases in risk are difficult to measure reliably, except under special circumstances.

## Assessing Causation

Epidemiology relies on statistics, and it is a cardinal rule of science that statistics cannot prove causation. For this reason scientists are usually reluctant to speak of exposure "causing" a disease on the basis of epidemiologic data alone. We consider different criteria for medicine and epidemiology that scientists have developed to weigh inferences about causation.

**Henle-Koch-Evans (HKE) postulates.** HKE postulates were first proposed by the famous bacteriologist and Nobel Prize (1905) winner Robert Koch (1843–1910) during the nineteenth century, when medicine was struggling with cholera, typhus, typhoid fever, and tuberculosis. Some of these postulates were enunciated as early as 1840 by Koch's teacher Henle. They were restated in 1976 by Evans, and are best referred to now as the Henle-Koch-Evans (HKE) postulates (Evans 1976, 1977). They are listed below.

1. The disease is significantly more prevalent in those exposed to the hypothesized cause than in unexposed controls.
2. Exposure is more frequent among those with the disease than in controls without the disease.
3. Prospective studies (which begin with an entirely healthy population) show a significantly higher incidence of disease in those exposed than in those not exposed.
4. Disease should follow exposure after an incubation period that tracks a log-normal curve.
5. Responses follow exposure along a logical biologic gradient from mild to severe.
6. Exposure triggers a measurable response (e.g., antibodies, cancer cells), with a high probability after exposure, or increases the number of responses if already present before exposure. This response pattern occurs infrequently or never in persons not exposed.
7. Experimental reproduction of the disease, in volunteers or laboratory experiments, or by controlled regulation of natural exposure, occurs more frequently in exposed animals or humans than in those not exposed.

8. Elimination of the suspected agent (e.g., control of polluted water, removal of tar from cigarettes) decreases the incidence of the disease.
9. Modification of the host's response on exposure (e.g., by immunization, or drugs to lower cholesterol) should decrease or eliminate the disease.
10. All of the relationships and findings should make scientific sense.

Most scientists would agree that evidence that satisfies the HKE postulates would make a compelling case for causation. They are most easily satisfied for infectious diseases, when the investigator believes that he or she has isolated the infectious agent. The postulates played a role in the scientific debates about the cause of AIDS. Virtually all scientists now agree that the HIV virus causes AIDS; what appears to be the last remaining holdout of high stature based his argument on the denial that the evidence satisfies these postulates (Duesberg 1989). Duesberg's views have been hotly contested by other scientists, and remain very much a minority position.

**Hill's Criteria for Assessing Causation From Epidemiologic Studies.** Most chronic diseases have an unknown cause, and some perhaps have multiple causes. For them, the scientific evidence does not—and may never—satisfy the HKE postulates. Nevertheless, epidemiologic studies can identify factors that increase a person's risk for a disease and allow effective public health measures to be developed to reduce the problem.

Some diseases rarely occur except in association with exposure to some hazardous substance. Mesothelioma rarely occurs except following exposure to asbestos (although other fibers may also produce it; see chapter 8); vaginal adenocarcinoma is unequivocally associated with in utero exposure to the drug DES; fetal alcohol syndrome arises from alcohol consumption by the mother. These "signature diseases" are characterized by very high values of relative risk, and for most purposes the question of causation is moot.

Other diseases are strongly but not uniquely associated with some risky behavior: smoking and lung cancer. Some associations may be weak (passive smoking and lung cancer). Most reported associations between passive smoking and cancer, in fact, are not statistically significant, but are accepted as real by many scientists because of the much stronger data from smokers.

In his celebrated lecture of 1965, Sir Austin Bradford Hill proposed nine criteria to help decide whether a reported association is causal or spurious. The criteria are summarized below, with Hill's comments (Hill 1965, 295–300).