

Scientific Evidence of Factual Causation
An Educational Module

Prepared by

Steve C. Gold
Professor of Law
Rutgers School of Law-Newark

Michael D. Green
Williams Professor of Law
Wake Forest University School of Law

Joseph Sanders
A.A. White Professor of Law
University of Houston Law Center

For

Committee on Preparing the Next Generation of Policy Makers for Science-Based Decisions

Committee on Science, Technology, and Law

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

October 2016

The copyright in this module is owned by the authors of the module, and may be used subject to the terms of the [Creative Commons Attribution-NonCommercial 4.0 International Public License](https://creativecommons.org/licenses/by-nc/4.0/). By using or further adapting the educational module, you agree to comply with the terms of the license. The educational module is solely the product of the authors and others that have added modifications and is not necessarily endorsed or adopted by the National Academy of Sciences, Engineering, and Medicine or the sponsors of this activity.

The authors are grateful to Kylie Huff, a 2016 graduate of Rutgers Law School, Nicole Regna, Wake Forest School of Law class of 2017, and Kelsey Kolb, a 2016 graduate of Wake Forest School of Law for their superb research assistance, Jennifer Aley, for her tireless copy editing assistance, and Arlene McClannon for the multitude of technical issues she resolved for us.

Contents

<i>I. Introduction</i>	1
<i>II. An Introduction to Causation</i>	3
A. Sine Qua Non and Substantial Factor	3
B. Framing the Causal Inquiry	4
C. The Necessity of Inference in Making Causal Assessments	5
D. The Difficulty of Causal Inference for Toxic Torts	6
E. The Role of <i>Daubert</i> and its Progeny	11
F. General Causation, Specific Causation, and Signature Diseases	13
1. General Causation.....	13
2. Specific Causation	13
3. Signature Diseases	14
G. Causation Standards in Tort and in the Regulatory Arena.....	14
<i>III. Epidemiology</i>	15
A. Introduction	15
B. Experimental and Observational Studies	16
1. Experimental Studies	16
2. Types of Observational Studies.....	18
a. Cohort Studies.....	19
b. Case-control Studies	21

c. Cross-Sectional Studies	22
d. Ecological Studies.....	23
C. Determining Exposure and Measuring Dose	24
D. The Outputs of Studies: Relative Risks, Odds Ratios, Attributable Proportion of Risk, and Standardized Rate Ratios	25
1. Relative Risk	26
2. Odds Ratio.....	27
3. Attributable Risk	30
E. Adjustment for Study Groups That Are Not Comparable: Standardized Mortality and Morbidity Ratios	32
F. Sources of Error.....	35
1. Statistical Methods to Evaluate the Possibility of Sampling Error	36
a. False Positives and Statistical Significance.....	37
b. Confidence Intervals	41
c. False Negatives.....	43
d. Power	43
2. Biases	44
a. Selection Bias	44
b. Information Bias.....	46
c. Other Conceptual Problems.....	48
3. Confounding.....	49

G. Assessing Whether an Association is Causal or Spurious.....	55
1. Temporal Relationship	55
2. The Strength of the Association.....	56
3. Dose-Response Relationship.....	56
4. Consistency of Association.....	57
5. Biologic Plausibility	57
6. Alternative Explanations for the Association.....	57
7. Specificity	57
8. Consistency with Other Information/Ceasing Exposure.....	58
H. Multiple Studies and Meta-analysis	60
I. Reading an Epidemiologic Study.....	62
<i>IV. Toxicology</i>	<i>76</i>
A. The Science of Toxicology.....	76
1. Introduction	76
2. Toxic Effects	143
3. The “Laws” of Toxicology	144
4. Areas of Toxicological Research.....	145
5. Toxicokinetics and Toxicodynamics	145
a. Pharmacokinetics and Toxicokinetics	146
i. Absorption.....	146
ii. Distribution.....	146

iii. Metabolism	147
iv. Elimination	148
b. Toxicodynamics.....	150
6. Types of Research	152
a. Predictive Research.....	152
b. In Vitro Research.....	153
c. In Vivo Research	153
d. Clinical Trials	159
7. Extrapolation.....	162
a. Dose Extrapolation	162
b. Extrapolation across Species.....	167
c. Susceptibility	167
d. The Quality of In Vivo Studies	169
B. Toxicology and the Law	170
1. The Use of Toxicology in Regulation	170
2. Toxicology in Litigation	175
a. Toxicology in Criminal Cases	175
b. Toxicology in Civil Cases.....	185
<i>V. Specific Causation</i>	<i>195</i>
A. Introduction	195
B. The Logic of Relative Risks Greater than 2.0	196

C. Beyond the Basic Logic: The Appropriateness of Applying the Average of a Group Study to a Nonstudy Individual.....	196
D. Heterogeneity: The Effect on Employing the Average Outcome.....	199
E. The (Often Unarticulated) Assumptions in the > 2.0 Threshold.....	201
F. Refining the Probability of Causation Through Subgrouping, Including Genetic Subgrouping and Molecular Epidemiology.....	203
1. The Role of Information about the Individual Plaintiff.....	203
2. Using Genetics to Refine the Probability of Causation: Individual Susceptibility to Toxic Effects.....	213
3. Using Genetics to Refine the Probability of Causation: Individual Inherited Susceptibility to Disease as a Competing Cause.....	221
G. Using Toxicogenomics and Biomarkers as Proof of Specific Causation	231
H. The Sufficiency of Relative Risks ≤ 2.0	235
1. The Basic Logic Recalled.....	235
2. Adjusting the Probability for Specific Individuals	235
3. Adjustments Based on Heterogeneity and Subgrouping.....	236
4. Adjustments through a Differential Etiology	236
5. Accounting for Idiopathic Disease	237
6. Small Relative Risks and General Causation	238
VI. <i>The Role of Consensus Organizations</i>	239

I. INTRODUCTION

Welcome to the module on scientific evidence of factual causation.¹ Causation has always been a requisite element of a tort case (and, indeed, many other legal areas). By contrast with the traumatic torts that burgeoned during the Industrial Revolution and with the invention and widespread ownership of the automobile, in the second half of the twentieth century a new type of tort emerged. Instead of traumatic injury, toxic torts are characterized by victims suffering disease.² Most often the disease is chronic rather than acute, which means that the disease only develops after a lengthy latency period. Cancer due to environmental causes typically develops decades after initial exposure. Birth defects commonly occur six or so months after a mother ingests a teratogen during the first trimester of pregnancy when organs are developing and can be influenced by chemicals that pass through the placental barrier. In addition to these toxic torts, regulation of risk, including environmental risk, food and drug risk, and occupational risk, became staples in the same time period as toxic torts and continues today.

To determine causation in these cases and risk in the regulatory context, for reasons explained in the next section, frequently requires resort to scientific disciplines. Those include epidemiology, toxicology and, the still rapidly developing field of genetics.

This module covers those three scientific areas which provide evidence bearing on causation in the toxic tort or environmental and disease context. These scientific disciplines are used in civil lawsuits and in regulatory proceedings in which causation or risk is an issue. The module is appropriate for non-scientist law students as well as for others interested in learning the science of toxic tort causation, including practicing attorneys, judges, and public policy students in the United States and elsewhere.

The module is designed to provide the following core competencies:

1. An understanding of the concept of factual causation and how evidence bearing on that matter is brought to bear in judicial proceedings and the regulatory context.
2. How the results of an observational epidemiologic study bear on factual causation and the methods and limitations of studies in providing proof of causation.
3. Measures of relationships employed in epidemiologic studies, the role of random chance, bias and confounding in affecting those relationships, and the process of determining whether a correlation is causal or spurious.

¹ Throughout this module, we employ both footnotes (although rarely) and endnotes. Footnotes are included for important explanatory material that would interrupt the main text if placed there. Endnotes contain important source material—we do not generally cite many sources that would support material contained below. Footnotes in reproduced cases reflect notes contained in the cases and their numbering in the case.

² Toxic torts have other characteristics, such as emerging as mass torts involving thousands of victims, but the critical one for this module is the necessity to resort to scientific evidence to determine causation.

4. The difference between epidemiology and toxicology and the latter's focus on whether and how a suspected toxic substance affects genes, cells, tissues, organs, and organisms. Students should acquire an understanding of the mechanisms of how disease is caused and the various areas and types of toxicologic research. Finally, students should obtain an understanding of the basics of extrapolating from various types of toxicology research to questions of causation in humans that are at the heart of legal inquiries.
5. An understanding of the legal issue of specific causation—whether the agent caused the plaintiff's disease. Students should take from this section an understanding of how courts use, and sometimes misuse, epidemiologic outcomes as a way to answer specific causation questions. In addition, genetic information increasingly is used to refine the specific causation question. Students should take from this section an introductory understanding of how toxicogenomics and the use of biomarkers help address specific causation. They should come to understand that currently the results of this area of science help to refine the probability of causation but rarely do they offer a clear answer to specific causation questions.

This module is designed for use in a classroom context with an instructor assisting students to absorb and understand the material in this module. A variety of slides designed to assist in this effort are included in the module along with study questions to be discussed in class. The module contains descriptive material, figures and tables, abstracts and scientific studies, and legal opinions. A comprehensive hypothetical is provided for student assessment at the conclusion of the module.

To provide an overview of the remainder of this module, the next section explains factual causation and how it is understood in law. Here, we distinguish proximate cause and explain that the normative matter of the extent of a defendant's liability is not a matter of scientific inquiry and thus beyond the scope of this module. Section III explains epidemiology, its methods and limitations, and the type of evidence that it can provide bearing on factual causation in toxic tort cases. The section on toxicology does the same for this discipline, explaining how the results of experiments with animals can be extrapolated to humans and the difficulties in attempting to do so. In addition, toxicology's contribution to understanding the underlying mechanisms of how disease is caused is explained. Genetics is outlined in Section F. 2., *infra* and its contribution to facilitating causal inquiry is covered in Sections V. F. 2 and 3. In the section on specific causation, a matter that scientists ordinarily do not examine, we explain how all of the sciences addressed in this module contribute to determining whether an environmental agent was responsible for a particular person's disease.

Throughout this module, we employ both footnotes (although rarely) and endnotes. Footnotes are included for important explanatory material that would interrupt the main text if placed there. Endnotes contain important source material—we do not generally cite many sources that would support material contained below. Footnotes in reproduced cases reflect notes contained in the cases and their numbering in the case.

II. AN INTRODUCTION TO CAUSATION

Causation is the glue that holds virtually all tort cases together. It bonds defendant's misconduct to the plaintiff's injury. Indeed, causation is an element of many legal areas—when a private party seeks to recover for harm, courts must determine if the defendant was a factual cause of that harm. Even in criminal cases, the defendant ordinarily must have caused injury or harm before being found guilty of a crime.

A. Sine Qua Non and Substantial Factor

What do we mean when we say that “X caused Y”?

First, let us briefly address and put aside the commonly used legal phrase “proximate cause,” a term often used to mean different things. Sometimes it is used as a synonym for factual cause. More commonly its distinctive usage is to limit a defendant's liability for harm factually caused (hence, “scope of liability”¹). This concept, which unfortunately uses the word “cause,” is not about whether X caused Y. Rather, proximate cause addresses the normative matter of whether a defendant who, despite tortiously having caused harm, is not held liable for that harm. Although sometimes “proximate cause” is used by courts to mean factual cause, we will not employ that term in this module, instead referring to “cause in fact,” “factual cause,” or just the simpler “cause.”

The core of factual cause is that a specified agent, act, or other stimulus was necessary to an outcome of interest. To be necessary for an outcome, the act (or agent or other stimulus of interest) must be such that if it had not occurred the outcome would also not have occurred. This is sometimes expressed as the “but-for” test: But for the act would the outcome still have occurred? If not, then the act is a factual cause of the harm. If the harm would have occurred regardless of the act, then the act is not a but-for cause of the harm (though, in one circumstance, the law may still view it as a factual cause—more on this later). “*Conditio sine qua non*” or its shorter version, “*sine qua non*” is used synonymously with but-for.

Note that to be a but-for cause, the act must be necessary but does not have to be sufficient for the outcome. For any outcome there are many causes, although most of them, not involving tortious conduct, will not be of interest in a tort suit. Do you see why the decision of a married couple to have a child is a factual cause of harm 20 years later when the child has grown, becomes a pilot, and crashes a small airplane with a passenger aboard?

Similarly, we tend to think that flipping a light switch is *the* cause of a room filling with light. We might appreciate the short sightedness of that assessment when the light bulb in the room burned out, just as we would when a circuit breaker blew because of an overload or when a thief stole all of the copper wiring in the building. Thus, you should appreciate that while we tend to ignore background causes like the existence of electrical wiring or a functioning light bulb, there are many causes (necessary conditions) for an outcome. Rather than being *the* cause of a lit room, turning the light switch is *a* cause of light, along with many others.

Thus, it would be more accurate to think of the cause of an outcome as a set of necessary conditions, all of which must exist for the outcome to occur.² Thus, we sometimes use the terminology of a “causal set” to refer to all of the necessary causes of a specified outcome.

U.S. courts often use another test for causation, the substantial-factor test. Initially adopted in the first Restatement of Torts in the 1930s, it inquires whether an act was a substantial factor in producing the harm of interest. With one exception (again, more on this later), the substantial factor test provided in the Restatement requires that the act be a but-for cause of the harm,³ although sometimes courts have overlooked this requirement.

Thus, for purposes of this module we employ the but-for test for factual causation.

There is one supplement required to the but-for test. When two (or more) acts would each be sufficient to cause the same harm at the same time neither is a but-for cause. Let us consider the classic two-fires hypothetical: both Johnny and Jennifer independently and negligently start campfires that burn out of control. The two fires join together just before the ensuing fire engulfs Bill’s home and destroys it. Jennifer is not a but-for cause of the destruction of Bill’s home because even if she had not started her fire, Johnny’s fire would have destroyed Bill’s home. The same is true of Johnny: Jennifer’s fire would have ruined Bill’s house regardless of what Johnny did. Neither is a but-for cause of the destruction of Bill’s house. In these cases of “overdetermined harm,” or “multiple sufficient causes,” the test for factual causation is supplemented so that each of these multiple sufficient causes is treated as a factual cause of the destruction of Bill’s home.⁴ In the toxic context, this means that if a victim is exposed to two doses of a toxin by different polluters with each being sufficient to cause plaintiff’s disease, each defendant is a factual cause of the victim’s harm. To take a more complex example, if Bill receives six different doses of a toxin from different sources and the threshold for causing Bill’s disease is five doses, each of the six doses is a factual cause of Bill’s disease even though removing any one of the six would not change the outcome.

B. Framing the Causal Inquiry

Before any causal question can be addressed, it must be framed.⁵ This “framing” requires two tasks. First, the act or agent that is to be assessed as the cause of some outcome must be specified. Second, the outcome or effect that is thought to be the result of the identified act or agent must be identified. Only after these two steps are completed can one ask whether the former caused the latter.

In a typical tort case, the act of interest is the tortious conduct of the defendant. The law demands that, before a defendant can be held liable, its negligence (or other tortious conduct) must have caused the harm for which the plaintiff seeks to recover. And that harm is the outcome that completes the framing of the causal inquiry. Thus, in an automobile accident case, the causal inquiry might be: Whether Joe’s speeding (tortious conduct) was a factual cause of the concussion that Steve suffered in the accident. (Note that the inquiry is whether the speeding was *a* cause not whether it was *the* cause.)

In most tort cases, the harm of interest will be a discrete injury, such as the concussion that Steve suffered. However, in other cases, the harm may be the enhanced injury suffered due to tortious conduct. Imagine an ambulance rushing an auto accident victim to the hospital that crashes into a telephone pole due to the negligence of the driver. The driver is not liable for the initial injuries suffered by the victim in the first accident but she is liable for any additional injuries including exacerbation of injuries suffered in the first accident. In other cases, the defendant may only be liable for accelerating the onset of a disease or injury. Consider an individual who, due to a genetic predisposition would develop a particular type of cancer at age 40. That person is exposed to a carcinogen that results in his developing cancer at the age of 35. The carcinogen is responsible for causing the individual to suffer cancer five years earlier than he otherwise would have, which has significant implications for the proper amount of damages to be awarded.⁶ Can you think of another example in which the outcome of interest is not a discrete injury or harm? Do you see why death (sometimes, often, always?) involves an accelerated outcome?

While the causal inquiry in a tort case begins with the tortious conduct of the defendant, our focus in this module is on “agent-disease causation.” That is, the difficult aspect of causation, requiring scientific evidence, is whether a suspected agent was a factual cause of a disease suffered by the plaintiff. To be sure, the defendant must have acted tortiously in exposing the plaintiff to the agent but that inquiry is often more straightforward requiring, for example, assessing whether the defendant failed to warn the plaintiff that respiratory precautions were necessary before working with asbestos or that the defendant was negligent in dumping toxic chemicals into a source of drinking water,¹ or that the defendant knew or should have known that the agent could cause the harm in question.

C. The Necessity of Inference in Making Causal Assessments

Before proceeding, we should appreciate that any assessment of causation involves an inferential process from evidence to the causation conclusion. Causation, unlike, for example, the presence of another human being, is not something we can observe directly with any of our five senses.² Rather, a causal assessment entails an inference based on observed phenomena.⁷ Thus, we observe the sun rising in the east and see that daylight follows. From this evidence, we might infer that the sun’s position in the sky causes daylight. When the effect (daylight) always follows from the cause (sun rising and never occurs in the absence of the sun being overhead), the circumstantial evidence of causation is quite powerful. In other instances the circumstantial evidence of causality may be weak, sometimes so much so that we would think that any statement about causation is more speculation than reasonable inference. Thus, when

¹ An exception is the market-share context in which identifying *the* defendant whose DES caused plaintiff’s disease is not possible.

² We do not mean to suggest that sensory-based testimony is more reliable than circumstantial evidence. As Dean Prosser observed: “there is still no man who would not accept dog tracks in the mud against the sworn testimony of a hundred eye-witnesses that no dog has passed by.” William L. Prosser, *The Law of Torts* § 39, at 212 (4th ed. 1971).

an individual known for clumsiness falls down an unlit stairway, attributing the fall to the lack of light in the stairway—given the alternative cause of the individual’s clumsiness—would border on the speculative end of the strength-of-inference spectrum.

D. The Difficulty of Causal Inference for Toxic Torts

Why is causation more difficult in the toxic tort arena than, say, in the ordinary traumatic tort case? Consider an automobile accident in which a car (without an airbag) crashes into a tree. A passenger in the car emerges with a newly broken arm. Most would readily conclude that the accident caused the passenger’s broken arm. Yet if that same passenger later developed lung cancer, the causal inquiry would be considerably more difficult and might ultimately end inconclusively. Why is this so?

One fundamental concern in making a causal assessment is the existence of “competing causes.” A competing cause is an alternative causal set that *could* have produced the same outcome. Thus, cigarette smoke, radon gas, and asbestos exposure are each causes of lung cancer.¹ Determining the cause of a victim’s lung cancer is more difficult when the victim is at risk due to multiple competing causes.

By contrast with the etiology (cause) of lung cancer, assessing the cause of our passenger’s broken bone is straightforward because there is only one plausible cause of that injury—the automobile crash. Note that simplifying the determination that there is only one potential cause is the latency period for broken bones. Virtually all broken bones result from sudden trauma, and if the passenger entered the car with both arms uninjured, there is only a short period of time in which the cause of the break could have operated. The latency period for lung cancer, on the other hand, is usually 20 years or more, which results in more possibilities for competing causes. Thus, when latency periods are known and brief, causal assessments are more readily made because competing causes can more readily be eliminated.

Contributing to the difficulty or ease of causal attribution is the extent of understanding of the biological mechanism of the causal sequence. When it comes to broken limbs, we know that when an outside force is applied to a bone great enough to displace the bone beyond its “flex” capability, the bone will crack, or if the force is large enough, break or even shatter.⁸ By contrast, although researchers are making great strides in exploring the biological mechanism of disease, understanding lags well behind our understanding of traumatic events.⁹ Indeed, as we will see, epidemiology has developed and grown as a discipline because it provides an alternative (statistical) method of determining causation when biological mechanisms are not well understood.

One final difficulty that exists in determining causation in the toxic tort arena is assessing exposure and dosage. For plaintiff’s disease to be caused by a toxic agent, plaintiff must have been exposed (through any of the known routes, absorption, skin contact, ingestion, inhalation, implantation, irradiation, or injection) to that agent. In some types of toxic tort

¹ For the sake of clarity we put aside that the *combination* of smoking and asbestos is also a competing cause. See Section V. D., *infra*

cases, typically those involving the use of a product such as a drug or chemical in a consumer product, exposure is straightforward. However, in occupational settings, exposure may not be so readily determined. In hazardous waste litigation, there may be a question about whether the waste reached the location where plaintiff resided or worked and, if so, the extent of contamination. Assessments of how pollution and hazardous waste move through ground, water, and air require scientific inquiry in areas beyond the scope of this module. Exacerbating proof problems is that the dose of exposure is often critical. As we discuss below in the toxicology section, one of the fundamental “laws” of toxicology is that the dose makes the poison. Often, little evidence of dose over the period of time plaintiff was exposed occupationally or residentially is available, especially with the passage of decades due to latency periods.

Before proceeding, let us consider a toxic tort case almost a century old that confronted the sorts of causal problems that we have been discussing so far. What we see are primitive efforts to use statistics—epidemiology has become far more sophisticated in the past 50 years—to determine causation.

Stubbs v. City of Rochester
Court of Appeals of New York, 1919.
226 N.Y. 516, 124 N.E. 137.

Hogan, J.

[Defendant City of Rochester supplied Hemlock system water for drinking and Holly system water for firefighting. The evidence revealed that because of the city’s negligence in May, 1910, the systems had become intermingled near the Brown street bridge. Sewage in the Holly water had contaminated the Hemlock water. However, the contamination was not discovered until October. The plaintiff contracted typhoid fever in September and attributed it to the city’s negligence. By a 3-2 vote, without opinion, the Appellate Division affirmed a nonsuit granted by the trial judge.]

* * *

The important question in this case is, Did the plaintiff produce evidence from which inference might reasonably be drawn that the cause of his illness was due to the use of contaminated water furnished by defendant. Counsel for respondent argues that even assuming that the city may be held liable to plaintiff for damages caused by its negligence in furnishing contaminated water for drinking purposes, (a) The evidence adduced by plaintiff fails to disclose that he contracted typhoid fever by drinking contaminated water; (b) that it was incumbent upon the plaintiff to establish that his illness was not due to any other cause to which typhoid fever may be attributed for which defendant is not liable. The evidence does disclose several causes of typhoid fever which is a germ disease, the germ being known as the typhoid bacillus, which causes may be classified as follows:

First. Drinking of polluted water. Second. Raw fruits and vegetables in certain named localities where human excrement is used to fertilize the soil are sometimes sources of typhoid infection. Third. The consumption of shell fish, though not a frequent cause. Fourth. The consumption of infected milk and vegetables. Fifth. The house fly in certain localities. Sixth.

Personal contact with an infected person by one who has a predilection for typhoid infection and is not objectively sick with the disease. Seventh. Ice, if affected with typhoid bacilli. Eighth. Fruits, vegetables, etc., washed in infected water. Ninth. The medical authorities recognize that there are still other causes and means unknown. This fact was developed on cross-examination of physicians called by plaintiff.

[Counsel argues first] that the evidence fails to disclose that plaintiff contracted typhoid fever by drinking contaminated water. The plaintiff, having been nonsuited at the close of his case, is entitled to the most favorable inference deducible from the evidence. That plaintiff, on or about September 6th, 1910, was taken ill, and very soon thereafter typhoid fever developed, is not disputed. That he was employed in a factory located one block distant from the Brown street bridge in which Hemlock lake water was the only supply of water for potable and other purposes, and that the water drawn from faucets in that neighborhood disclosed that the water was roily and of unusual appearance is not questioned. And no doubt prevails that the Holley system water was confined to the main business part of the city for use for fire purposes and sprinkling streets and is not furnished for domestic or drinking purposes.

The evidence of the superintendent of waterworks of the city is to the effect that Hemlock Lake water is a pure wholesome water free from contamination of any sort at the lake, and examinations of the same are made weekly; that the Holley water is not fit for drinking purposes taken as it is from the Genesee river. Further evidence was offered by plaintiff by several witnesses, residents in the locality of Brown street bridge, who discovered the condition of the water at various times during July, August and September and made complaint to the water department of the condition of the same. Dr. Goler, a physician and health officer of the city, was called by plaintiff and testified that in September, when complaint was made to him by a resident of the district, he went to the locality, visited houses in the immediate neighborhood, found that the water drawn from the faucet of the Hemlock supply looked badly and smelled badly. He took a sample of the water to the laboratory and had it examined by a chemist who found that it contained an increase in solids and very many times, that is, 20 to 30 times as much chlorine or common salt as is found in the domestic water supply—the presence of chlorine in excessive quantities indicates contamination in that quantity, bad contamination and usually sewage contamination. Further examination followed in the district. Water was collected from various houses and a large number of samples, perhaps less than 100, but over 25. * * * About the following day, the source of contamination having been discovered, the doctor made an investigation as to the reported cases of typhoid fever in the city in the months of August, September and October, for the purpose of determining the number of cases, where the cases came from, what gave rise to it, and he stated that in his opinion the outbreak of typhoid was due to polluted water, contaminated as he discovered afterwards by sewage. In answer to a hypothetical question embracing generally the facts asserted by plaintiff the witness testified that he had an opinion as to the cause of the infection of plaintiff and such opinion was that it was due to contaminated water.

Dr. Dodge, of the faculty of the University of Rochester, a professor of biology, also bacteriologist of the city of Rochester, about October [first] made an analysis of samples of water * * *. While his examination did not disclose any colon bacillus, it did disclose some evidence of the same. Dr. Brady, the physician who attended the plaintiff, and Dr. Culkin both testified that in their opinion the plaintiff contracted typhoid fever from drinking polluted water.

Plaintiff called a witness who resided on Brown street about two minutes' walk from the bridge, and proved by her that she drank water from the Hemlock mains in the fall of 1910 and was ill with typhoid fever. Thereupon counsel for defendant stipulated that 57 witnesses which the plaintiff proposed to call will testify that they drank water from the Hemlock taps in the vicinity of the district west of the Genesee river and north of Allen street in the summer and fall of 1910, and during said summer and fall suffered from typhoid fever, that in view of the stipulation such witnesses need not be called by plaintiff, and the stipulation shall have the same force and effect as though the witnesses had been called and testified to the facts.

The plaintiff resided with his wife some three miles distant from the factory where he was employed. The water consumed by him at his house outside the infected district was Hemlock water. The only water in the factory was Hemlock water, and he had there an individual cup from which he drank. He was not outside of the city during the summer of 1910. Therefore, the only water he drank was in the city of Rochester.

A table of statistics as to typhoid fever in the city of Rochester for the years 1901-1910, inclusive, was produced by the health officer and received in evidence. * * * The statistics disclose that the number of typhoid cases in the city in 1910 was 223, an excess of 50 cases of any year of the nine years preceding. Recalling that complaints as to water commenced in the summer of 1910 and as shown by the evidence that typhoid fever does not develop until two or three weeks after the bacilli have been taken into the system, in connection with the fact that the source of contamination was not discovered until October, the statistics disclose that of the 223 cases of typhoid in the city in the year 1910, 180 cases appear during the months of August, September, October, and November as against 43 cases during the remaining eight months, 35 of which were prior to August and 8 in the month of December, two months after the source of contamination of the water was discovered.

The evidence on the trial discloses that at least 58 witnesses, residents of the district, drank the contaminated water and suffered from typhoid fever in addition to plaintiff; thus one-third of the 180 cases during the months stated were shown to exist in that district.

Counsel for respondent asserts that there was a failure of proof on the part of plaintiff in that he did not establish that he contracted disease by drinking contaminated water, and in support of his argument cites a rule of law that when there are several possible causes of injury for one or more of which a defendant is not responsible, plaintiff cannot recover without proving that the injury was sustained wholly or in part by a cause for which defendant was responsible. He submits that it was essential for plaintiff to eliminate all other of seven causes from which the disease might have been contracted. If the argument should prevail and the rule of law stated is not subject to any limitation, the present case illustrates the impossibility of a recovery in any case based upon like facts. One cause of the disease is stated by counsel to be "personal contact with typhoid carriers or other persons suffering with the disease, whereby bacilli are received and accidentally transferred by the hands or some other portion of the person or clothes to the mouth." Concededly a person is affected with typhoid some weeks before the disease develops. The plaintiff here resided three miles distant from his place of employment and traveled to and from his work upon the street car. To prove the time when he was attacked with typhoid, then find every individual who traveled on the same car with him, and establish by each one of them that he or she was free from the disease even to his or her clothing is impossible. Again the evidence disclosed that typhoid fever was caused by sources unknown to medical science. If the word of the rule stated is to prevail plaintiff would be required to eliminate sources which had not yet been determined or ascertained. I do not

believe the rule stated to be as inflexible as claimed for. If two or more possible causes exist, for only one of which a defendant may be liable, and a party injured establishes facts from which it can be said with reasonable certainty that the direct cause of the injury was the one for which the defendant was liable, the party has complied with the spirit of the rule.

The plaintiff was employed in the immediate locality where the water was contaminated. He drank the water daily. The consumption of contaminated water is a very frequent cause of typhoid fever. In the locality there were a large number of cases of typhoid fever and near to 60 individuals who drank the water and had suffered from typhoid fever in that neighborhood appeared as witnesses on behalf of plaintiff. The plaintiff gave evidence of his habits, his home surroundings, and his method of living, and the medical testimony indicated that his illness was caused by drinking contaminated water. Without reiteration of the facts disclosed on the trial I do not believe that the case on the part of plaintiff was so lacking in proof as matter of law that his complaint should be dismissed. On the contrary, the most favorable inferences deducible from the plaintiff were such as would justify a submission of the facts to a jury as to the reasonable inferences to be drawn therefrom, and a verdict rendered thereon for either party would rest, not in conjecture but upon reasonable possibilities.

The judgment should be reversed, and a new trial granted, costs to abide the event.

Notes and Questions

1. Context is often important in causation disputes, especially in terms of the state of the relevant science. In the case of typhoid fever, both the state of the science and New York history bore on the decision in *Stubbs*. By 1919, the year of the decision, the cause of typhoid fever was known: Georg Theodor August Gaffky established in 1884 that the bacillus salmonella typhi is the causative agent of typhoid fever, although the carrier can vary, as the *Stubbs* court explained. In addition, local events in the decade before this case were also relevant. In the summer 1906, Charles Henry Warren took his family on vacation to Long Island, hiring Mary Mallon to be their cook. At the end of August, typhoid fever struck one of the Warren daughters, eventually infecting six of the 11 members of the household. Upon investigation, it turned out that Mary Mallon—known now to history as “Typhoid Mary”—was the carrier of the disease. She was passing the disease to family members (and ultimately hundreds and maybe thousands of others) without being symptomatic herself; no public water system was involved.

Does the Typhoid Mary history provide more factual ammunition for the plaintiff or for the defendant in *Stubbs*?

2. In the last sentence of its opinion, the court states that the evidence provided by the plaintiff would permit a jury to draw a “reasonable inference” of causation. Thus, any such verdict for the plaintiff would not be based on conjecture. When circumstantial evidence (as is all evidence of causation) is introduced by the party with the burden of production, the matter of its sufficiency is often described as whether that evidence is sufficiently supportive of the proposition in question to

permit the factfinder to draw a reasonable inference of the existence of the proposition. By contrast, when the evidence is weak, courts often state that it would require “impermissible speculation” by the factfinder and conclude the party with the burden of production has failed to satisfy it. The Third Restatement of Torts characterizes the line between reasonable inference and impermissible speculation as “one of the more indistinct lines that exists in law and also is one on which reasonable minds can and do differ. Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. b.

3. Note the court’s use of statistics in addressing the likelihood that contamination of the Hemlock water system led to the plaintiff’s typhoid. Today, of course, an entire scientific discipline—epidemiology—has developed to study the causes of human disease. The beginnings of modern epidemiology are traced in part to John Snow, a London general practitioner who confronted a situation not too different from that in *Stubbs*. Dr. Snow developed a hypothesis that cholera was spread through contaminated drinking water and conducted a set of tests to verify his theory. One test entailed comparing the incidence of the disease in a community drawing its water from the Thames upstream from London with the incidence in a community drawing its water downstream of the city. He then attempted to control for other variables by looking at a single London region that was supplied drinking water by two different companies, and observed significant differences in the incidence of the disease in sub-districts supplied by each firm. Finally, he conducted what is known today as a case-control study by going to the homes of each cholera fatality, ascertaining who supplied the deceased’s drinking water, totaling the number of cholera deaths for each water source, and determining the incidence of cholera deaths for each source by dividing by the total number of houses supplied by each company. The results, confirming the connection between the disease and the source of drinking water, were published in 1849.¹⁰

E. The Role of *Daubert* and its Progeny

We have attempted to resist teaching legal matters throughout this module, limiting our discussion of legal matters to those necessary to place science in context and to understand its effect on legal proceedings. Nevertheless, some brief explanation of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹¹ is necessary to explain the way in which the legal system receives and analyzes scientific evidence such as that discussed in this module.

As you probably already appreciate, the way in which causation is proved in toxic tort cases is through the opinion of an expert or experts. Scientific studies are rarely introduced as the *Lindquist* case, Section III. I. *infra*, reveals with the court doing its own extramural research to uncover a scientific study. Jurors are rarely in a position to make any sense of a scientific study if it is introduced (and we may doubt the capacity of judges, at least those who have not

studied from this module, to do so). Experts, however, may identify and testify about scientific studies as support for opinions they express about causation.

In 1976 when the Federal Rules of Evidence were promulgated and approved by Congress, the rules governing expert testimony were liberalized to permit greater use of experts. The rules drafters were motivated by the increasing complexity of modern litigation and the concomitant need for expert evidence. As with many reforms, the liberalization of expert testimony may have overshot the mark. By the late 1970s and early 1980s courts and commentators had become concerned about dubious, speculative, and sometimes outright inaccurate testimony by adversarial experts. That began an effort to reign in what was termed “junk science” by Peter Huber.¹² One court employed the Supreme Court’s decision in *Frye v. United States*,¹³ a criminal case, which required that there be “general acceptance” of new “scientific principle[s] or discover[ies],” in that case a forerunner to the lie detector. *Frye* had previously almost never been invoked in civil cases with regard to the admissibility of an expert’s testimony. Other courts used a variety of techniques to rule inadmissible expert testimony thought to be speculative (because little or no relevant scientific evidence supported it) or inconsistent with existing science. Ultimately, the Supreme Court agreed to review whether *Frye* survived the adoption of the Federal Rules of Evidence in *Daubert*, a Bendectin case.

In an opinion with something for everyone, the Court ruled that the Federal Rules of Evidence superseded the *Frye* general acceptance standard. Nevertheless, the Court found that rule 702, which authorizes expert testimony when “scientific . . . knowledge” would help the jury, imposed a requirement of reliability before scientific expert testimony could be admitted. The Court provided a non-exclusive list of four factors to guide trial judges in assessing the reliability of proposed expert testimony: (1) falsifiability or testing of the theory or technique; (2) peer review and publication; (3) the known or potential rate of error; and 4) general acceptance (now a factor rather than the exclusive criterion).

These factors, however, are not well suited to assessing causation testimony in toxic tort cases, especially with regard to specific causation. Two subsequent opinions by the Supreme Court, *General Electric Co. v. Joiner*, 522 U.S. 136 (1997) and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1997) facilitated a different approach to determining the admissibility of an expert’s testimony about causation that most courts have employed. In this approach, courts consider the scientific evidence identified by the expert and determine whether that evidence is sufficient to support an inference of causation or, alternatively, whether the scientific evidence is inadequate to support such a conclusion or too speculative to permit a reasonable inference. As one court explained itself: “The analytical gap between the evidence presented and the inferences to be drawn on the ultimate issue of human birth defects is too wide. Under such circumstances, a jury should not be asked to speculate on the issue of causation.” *Turpin v. Merrell Dow Pharmaceuticals, Inc.*¹⁴

Thus, as in the PPA case excerpted in Section V. C., *infra* and in the *Johnson* case set forth in Section IV. B. 2. b. of the Toxicology materials, *infra*, judges are digging into the scientific evidence proffered by an expert to support her opinion and assessing whether that body of scientific evidence is sufficient to support the proposed causation opinion. This

emphasizes the critical importance of understanding the sciences explained in this module for lawyers, judges, and others who work in the field of toxic torts.

F. General Causation, Specific Causation, and Signature Diseases

Epidemiology studies the prevalence and incidence of disease in groups as does in vivo animal toxicology. (*Prevalence* is the number of people with disease within a given time period. *Incidence* is the number of individuals who develop disease during a specified period of time and therefore expresses a rate of disease occurrence.) The results of those studies speak directly only to whether an agent increases the incidence of disease in the group. What it does not do is to determine the cause of any individual's disease, that is, whether those in the exposed group who contracted the disease did so because of the studied agent or some other competing cause of the disease. Because of the form of epidemiologic evidence, tort law has developed two concepts to deal with the agent-disease causal inquiry.

1. General Causation

The first, general causation, asks the question of whether the agent in question (or in a more refined concept, the agent in doses to which humans are exposed) is capable of causing the disease in the human population. Courts employ the concept of general causation because that is the issue to which epidemiology and toxicology speak. If there is no evidence of general causation, then agent-disease causation cannot be established in almost all instances.¹ This means that consideration of causation in an individual case can be truncated because of the absence of factual causation. But the existence of general causation, while necessary, is not sufficient to establish agent-disease causation.

2. Specific Causation

Tort law requires the individual claimant to demonstrate that he or she was injured by the alleged cause. Proof that the defendant's agent caused a greater incidence of disease in a group (general causation) is insufficient for tort liability, although it can be the basis for regulatory action. Thus, a claimant must show that exposure to defendant's agent caused that claimant's disease. After coverage of the various sciences, we return, in Section V, to specific causation and explain both how group study results bear on specific causation and other sources of evidence that support a finding of specific causation.

¹ When there is exposure to an agent that is a signature for a disease, i.e., the agent is the only cause of that disease, general causation and specific causation collapse. In addition, in some cases other evidence, mechanism or toxicologic, may be sufficient to establish general causation without epidemiology.

3. Signature Diseases

Some diseases are known as signature diseases (or pathognomonic diseases) because the disease is a signature for a specific causal agent. This occurs when there are no competing causes and, thus, all instances of the disease can be attributed to that specific agent. In the 1960s, an epidemic of vaginal adenocarcinoma developed among young women in their late teens and early twenties. While this disease was previously known, it was only found in women who were middle-aged or older. Researchers determined that the cause of this disease among young women was in utero exposure to the drug DES, a synthetic estrogen that had been prescribed to pregnant women to prevent miscarriage.¹⁵ Thus, young women who suffered this disease were linked to prebirth exposure to DES. Because of the absence of competing causes, general causation and specific causation collapse into a single inquiry.

G. Causation Standards in Tort and in the Regulatory Arena

Courts often state that proof standards for regulation are less stringent than for tort law.¹⁶ There is something to this, but it is worthwhile to unpack the idea. First, it deserves emphasis that, unlike public health regulation, tort law requires proof that an individual defendant was responsible for an individual plaintiff's harm, the reason for specific causation as discussed above. By contrast, in the area of risk regulation, such as that performed by the Environmental Protection Agency or the Food and Drug Administration, risk to a group of individuals or even to the entire population is sufficient for legal action. Thus, unlike tort law, public health regulation is concerned solely with general causation and not specific causation. Risk regulation, unlike tort law, is also concerned with the extent of impact on public health of risk. Adjudication of a tort claim does not depend on whether a risk such as asbestos causes a public health calamity or one unfortunate individual suffers a unique and freakish overdose of a pharmaceutical that causes harm.

While a plaintiff in a civil case must establish causation, including general causation by a preponderance of the evidence, regulators have a lower burden of establishing that there is "sufficient evidence" or in some cases "substantial evidence" to support a determination of general causation.¹⁷

The magnitude of the risk required before regulation is deemed appropriate depends on a variety of factors, including the type of risk, the way in which the risk is a function of dose, that is, the dose-response relationship, the degree of uncertainty about the magnitude of the risk, the extent of public exposure, and the availability of means to avoid or ameliorate it and the costs of such. Most critical is the specific legal standard contained in the regulatory legislation—the "risk trigger"—set by Congress as the threshold for regulatory action.¹ Smaller risks than would likely be adequate to support specific causation may be appropriate for regulation especially when large numbers of persons are exposed to the risk factor. Some

¹ Thus, for example, the Clean Water Act mandates regulation "adequate to protect public health and the environment from any reasonably anticipated adverse effects." Clean Water Act 26 U.S.C. § 1345(d)(2)(D).

statutes specify that regulations must be constructed conservatively so as to provide an adequate margin of safety, often referred to as the “precautionary principle.” Thus, regulatory risk assessments may be relevant to whether general causation exists but rarely have any salience for the matter of specific causation. We discuss the regulatory process of establishing dose levels for toxic substances at greater length in the toxicology section below.

III. EPIDEMIOLOGY

A. Introduction

Epidemiologists study the incidence, distribution, and etiology of human disease and the factors associated with disease incidence and distribution patterns. By contrast with toxicology, which focuses on animals, epidemiologists conduct their investigations on human populations.

The essence of epidemiologic research is to obtain data about exposure to an investigated agent or agents (or occupation or other status) and disease incidence and to examine them for statistical associations between exposure and disease. Then, epidemiologists assess whether those observed associations are causal rather than coincidental or spurious.

Epidemiology focuses on the question of general causation rather than that of specific causation. For example, in the 1950s Doll and Hill and others published articles about the increased risk of lung cancer in cigarette smokers. Doll and Hill’s studies showed that smokers who smoked 10 to 20 cigarettes a day had a lung cancer mortality rate that was about 10 times higher than that for nonsmokers. These studies identified an association between smoking cigarettes and death from lung cancer that contributed to the determination that smoking causes lung cancer. What Doll and Hill’s work did not do was to identify, among smokers with lung cancer, which ones contracted their lung cancer because of smoking and which ones contracted their lung cancer because of other causes.

Although epidemiology has long roots dating back to the Enlightenment,¹⁸ modern epidemiologic methods were developed in the post-World War II period with several important prospective studies¹ undertaken by public health officials, including the Framingham cardiovascular health study and the Salk vaccine trial. In ensuing years, epidemiologists conducting other studies uncovered causal relationships that have played an important role in toxic tort litigation, including smoking and lung cancer, swine flu vaccine and Guillain-Barré Syndrome, asbestos and mesothelioma, lung cancer, and asbestosis, and DES and vaginal adenocarcinoma. Epidemiology continues to play a major role in virtually all significant modern toxic tort cases. Today, epidemiologic evidence is widely accepted, indeed preferred, by courts confronting causal issues in the toxic tort context. As the Third Circuit observed in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*¹⁹: “The reliability of expert testimony founded on reasoning from epidemiologic data is generally a fit subject for judicial notice; epidemiology is a well-

¹ Prospective studies are also known as follow-up studies because the design involves identifying participants and then “following” them to assess exposure and disease states. When the latency period for those studies is lengthy, they can take a long time to complete.

established branch of science and medicine, and epidemiologic evidence has been accepted in numerous cases.”²⁰ Indeed, much more difficult problems arise for courts when there is a paucity of epidemiologic evidence.

Four basic issues arise when epidemiology is used in legal disputes, and the methodological soundness of a study and its implications for resolution of the question of causation must be assessed:

1. Do the results of an epidemiologic study or studies reveal an association between an agent and disease?
2. Could this association have resulted from limitations of the study (bias, confounding, or sampling error), and if so, from which?
3. Based on the analysis of limitations in 2 above and on other evidence, how plausible is a causal interpretation of the association?
4. What are the implications of epidemiologic results and other evidence in determining the cause of a given plaintiff’s disease?

Our focus below is on disease as a dichotomous matter: one either has or does not have lung cancer. Epidemiologists may study diseases and outcomes that are continuous and progressive, such as asbestosis, which is a disease that progresses from having relatively modest clinical symptoms to, in some cases, death. Exposure may be dichotomous—for example an individual either has or does not have a certain genetic anomaly—but more often is matter of degree. Thus, when exposure is continuous, epidemiologists (and toxicologists) may employ a study design with groups who were or are exposed to a high, medium, and low dose of the suspected toxin.

In the materials that follow, we do not focus on how an epidemiologist, interested in a suspected toxic agent, would go about designing a study preparatory to conducting such a study. That topic is addressed in all graduate school introductory Epidemiology texts.²¹ Because your professional role will not be conducting epidemiology studies but using or critiquing them, these materials are designed to make the reader an informed consumer of epidemiologic work that has been completed.

B. Experimental and Observational Studies

1. Experimental Studies

To determine whether an agent increases the risk of developing a certain disease or an adverse health outcome, we might ideally want to conduct an experimental study in which the subjects would be randomly assigned to one of two groups: one group exposed to the agent of interest and the other not exposed.¹ After a period of time, the study participants in both

¹ In clinical trials of new drugs or medical treatments, often another group used for comparison to the new drug or treatment is an existing drug or treatment so that researchers can determine differences in the benefits and risks between the new and existing technology.

groups would be evaluated for the development of the disease being studied. This type of study, called a randomized trial, clinical trial, or experimental study, is considered the gold standard for determining the relationship of an agent to a health outcome or adverse side effect. Such a study design is often used to evaluate new drugs or medical treatments and is the best way to ensure that any observed difference in outcome between the two groups is the result of exposure to the drug or medical treatment, that is, reflects a true causal relationship. Randomization minimizes the likelihood that there are differences in relevant characteristics between those exposed to the agent and those not exposed. Researchers conducting clinical trials attempt to use study designs that are placebo controlled, which means that the group not receiving the active agent or treatment is given an inactive ingredient that appears similar to the active agent under study. These studies also use double blinding, which means that neither the participants nor those conducting the study know which group is receiving the agent or treatment and which group is given the placebo. Note that, unlike animal studies, investigators in human experimental studies cannot control all of the environment in which the experiment is conducted.²² Nevertheless, ethical and practical constraints limit the use of such quasi-experimental methodologies to assessing the agents that are thought to be beneficial to human beings.

When an agent's effects are suspected to be harmful, researchers cannot knowingly expose people to the agent. Thus, it would be unethical for researchers to conduct an experimental study in which they exposed children to different levels of lead paint in their residences to determine the effectiveness of various abatement procedures with regard to the extent of lead contamination in the children's blood (lead poisoning being known to cause cognitive deficits, especially in children).¹ Instead epidemiologic studies typically "observe" a group of individuals who have been exposed to an agent of interest, such as cigarette smoke or an industrial chemical, and compare them with another group of individuals who have not been exposed.² Thus, the investigator identifies a group of subjects who have been exposed and compares their rate of disease or death with that of an unexposed group. In contrast to clinical studies in which potential risk factors can be controlled, epidemiologic investigations generally focus on individuals living in the community for whom differences in characteristics other than the one of interest, such as diet, exercise, exposure to other environmental agents, and genetic background, may distort a study's results. Because these characteristics cannot be controlled directly by the investigator, the investigator addresses their possible role in the relationship being studied by considering them in the design of the study and in the analysis and interpretation of the study results (see Section III. E. 3., *infra*). The Achilles heel of observational studies is the possibility of differences in the two populations being studied with

¹ Nevertheless, a study quite similar to that set out in the text is described in *Grimes v. Kennedy Krieger*, 782 A.2d 807, 816 (Md. 2001), which references other well-known instances of research misconduct of this sort. The court held that researchers conducting non-therapeutic studies that contained risks to participants had a tort-law duty to participants to warn them of the dangers about which the researchers were aware.

² Classifying these studies as observational studies in contrast to randomized trials can be misleading to those who are unfamiliar with the area, because subjects in a randomized trial are observed as well. Nevertheless, the use of the term "observational studies" to distinguish them from experimental studies is widely employed.

regard to risk factors other than exposure to the agent. By contrast, experimental studies, in which subjects are randomized, minimize this problem.

2. Types of Observational Studies

Several different types of observational epidemiologic studies exist.¹ Study designs may be chosen because of suitability for investigating the question of interest, timing constraints, resource limitations, or other considerations.

Most observational studies collect data about both exposure and health outcome in every individual in the study. The two main types of observational studies are cohort studies and case-control studies. Classically, cohort studies were referred to as prospective studies, because participants would be identified and then followed prospectively to determine the incidence of disease in the exposed and unexposed cohorts. Case-control studies have been described as retrospective studies, because once cases and control were identified, researchers would look backward to determine exposure and other factors that might require adjustment between the two groups. However, for several decades researchers have conducted retrospective cohort studies in which they obtain historical information about exposure and disease for an exposed and control cohort. Irving Selikoff's path-breaking work on asbestos consisted of retrospective studies of those employed in the asbestos industry, employing work, industrial, and medical records for asbestos industry workers to assess the health risks of asbestos exposure

A third type of observational study is a cross-sectional study, although cross-sectional studies are rarely useful in identifying toxic agents. A final type of observational study, one in which data about individuals are not gathered, but rather population data about exposure and disease are used, is an ecological study. An example of an ecological study is one conducted by an epidemiologist who testified as an expert witness in the Bendectin litigation (Bendectin, a drug for morning sickness, was alleged to cause birth defects.) The expert compared the incidence of birth defects while Bendectin was on the market with the incidence of birth defects after it was removed from the market and found no difference or a slightly increased incidence of birth defects after Bendectin was no longer sold.²³ The difficulty with these time-trend or secular trend studies is that there may be other things occurring coincidentally at the same time that exposure to the suspected agent changes that may be responsible for the incidence observed after the exposure has changed.

Although not actually a "study," case reports of the existence of disease related to some specific exposure are sometimes obtained or reported in medical journals. Formally, a "cluster" is an unusual aggregation of the same or similar non-infectious diseases that are grouped together in time and space. Note that these reports do not permit calculating an incident rate nor do they have a control group for comparison. Sometimes "clusters" of disease develop that cause concerns that some environmental exposure is responsible. Thus, some years ago there was a cluster of foodborne streptococcal disease in which almost half of the inmates in a

¹ Other epidemiologic studies collect data about the group as a whole, rather than about each individual in the group. These group studies are discussed infra Section II.B.4.

Florida prison developed the (acute) disease. When clusters of disease are found, statistical methods exist to attempt to determine whether the cases are independent or stem from a common source.²⁴

To return to the two types of epidemiologic studies that are most often implicated in toxic tort cases, the difference between cohort studies and case-control studies is that cohort studies measure and compare the incidence of disease in the exposed and unexposed (“control”) groups, while case-control studies measure and compare the frequency of exposure in the group with the disease (the “cases”) and the group without the disease (the “controls”). In a case-control study, the rates of exposure in the cases and the rates in the controls are compared, and the odds of having the disease when exposed to a suspected agent can be compared with the odds when not exposed. The critical difference between cohort studies and case-control studies is that cohort studies begin with exposed people and unexposed people, while case-control studies begin with individuals who are selected based on whether they have the disease or do not have the disease and their exposure to the agent in question is measured. The goal of both types of studies is to determine if there is an association between exposure to an agent and a disease and the strength (magnitude) of that association.

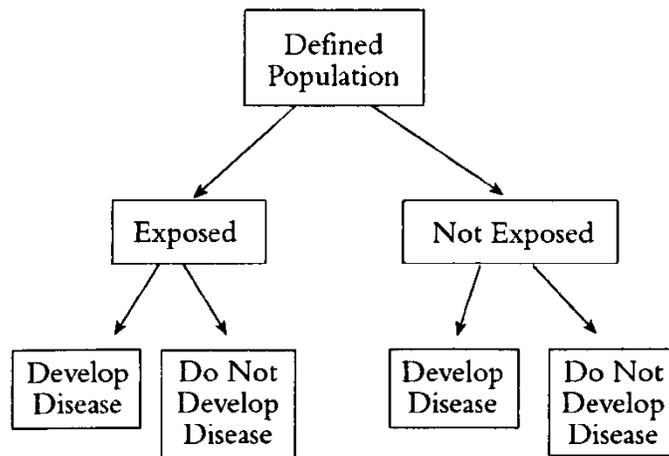
a. Cohort Studies

In cohort studies,¹ researchers define a study population without regard to the participants’ disease status. The cohort may be defined in the present and followed forward into the future (prospectively) or it may be constructed retrospectively as of some time in the past and followed over historical time toward the present. In either case, researchers classify the study participants into groups based on whether they were exposed to the agent of interest (see Figure III-1). In a prospective study, the exposed and unexposed groups are followed for a specified length of time, and the proportion of individuals in each group who develop the disease of interest is compared. In a retrospective study, the researcher will determine the proportion of individuals in the exposed group who developed the disease from available records or evidence and compare the incidence of disease in the exposed group with the incidence in a control group that was not exposed. Thus, as illustrated in Table III-1, a researcher would compare the proportion of unexposed individuals with the disease $c/(a + c)$ with the proportion of exposed individuals with the disease $d/(b + d)$. If the exposure causes the disease, the researcher would expect a greater proportion of the exposed individuals to develop the disease than the unexposed individuals.²

Figure III-1. Design of a Cohort Study

¹ Cohort studies also are referred to as prospective studies, follow-up studies, and, when researchers determine exposure and disease that has already occurred, retrospective cohort studies.

² Researchers often examine the rate of disease or death in the exposed and control groups. The rate of disease or death entails consideration of the number developing disease within a specified period. All smokers and nonsmokers will, if followed for 100 years, die. Smokers will die at a greater rate than nonsmokers in the earlier years.



SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 558. Copyright © 2011 National Academy of Sciences.

Table III-1. Cross-Tabulation of Exposure by Disease Status

	No Disease	Disease	Totals	Incidence Rates of Disease
Not Exposed	a	c	a + c	$c / (a + c)$
Exposed	b	d	b + d	$d / (b + d)$

SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 558. Copyright © 2011 National Academy of Sciences.

One advantage of the prospective cohort study design is that the temporal relationship between exposure and disease can often be established more readily than in other study designs, especially a case-control design, discussed below. By tracking people who are initially not affected by the disease, the researcher can determine the time of disease onset and its relation to exposure. This temporal relationship is critical to the question of causation, because exposure must precede disease onset if it is to be a cause of the disease. Retrospective cohort studies, however, can present temporal relationship difficulties. Retrospective cohort studies are conducted by reviewing historical data of all the exposed and non-exposed persons and then determining the current disease status of each person. Depending on the historical data available, retrospective cohort studies may be unable to establish the temporal relationship between exposure and disease. Imagine a retrospective cohort study in which study participants were asked to fill out a questionnaire regarding cigarette smoking habits (exposure) and whether they developed lung cancer (outcome). Recall bias (the tendency of

those with disease to recall exposures more readily than those without disease, see Section III. E. 2. b., *infra*) may make it difficult to determine a temporal relationship between when exposure occurred and when the disease outcome occurred.

As stated, prospective studies avoid this concern. Typical of prospective cohort studies is one that investigated lung cancer in uranium miners exposed to radon illustrates such studies. Begun in 1950, the study sought to determine whether uranium miners were at increased risk for lung cancer as compared with non-miners. The study group (also referred to as the exposed cohort) consisted of 3,400 white, underground miners. The control group (which need not be the same size as the exposed cohort) comprised white non-miners from the same geographic area. Members of the exposed cohort were examined every three years, and the degree of this cohort's exposure to radon was measured from samples taken in the mines. Ongoing testing for radioactivity and periodic medical monitoring of subjects' lungs permitted the researchers to examine whether lung cancer was linked to prior work exposure to radiation and allowed them to discern the relationship between exposure to radiation and disease. The study found that exposure to radiation was associated with the development of lung cancer in uranium miners.

The cohort design often is used in occupational studies such as the one just discussed. Because the design is not experimental, and the investigator has no control over what other exposures a subject in the study may have had, an increased risk of disease among the exposed group may be caused by agents other than the exposure of interest. A cohort study of workers in a certain industry that pays below-average wages might find a higher risk of cancer in those workers. This may be because they work in that industry, or, among other reasons, it may be because low-wage groups are exposed to other harmful agents, such as environmental toxins present in higher concentrations in their neighborhoods. In the study design, the researcher must attempt to identify factors other than the exposure that may be responsible for the increased risk of the disease being studied. If data are gathered on other possible etiologic factors, the researcher generally uses statistical methods to assess whether a true association exists between working in the industry and disease. Evaluating whether the association is causal involves additional analysis, as discussed in Section III. F, *infra*.

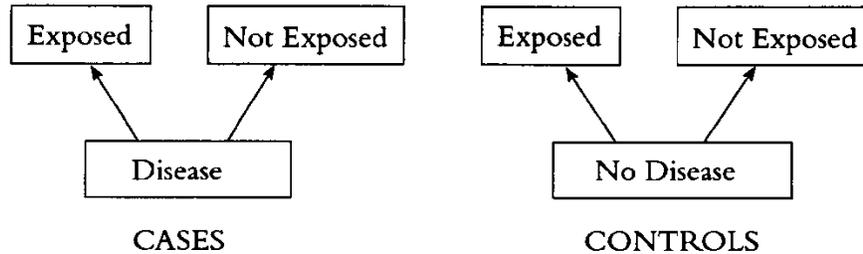
b. Case-control Studies

In case-control studies, the researcher begins with a group of individuals who have a disease (cases) and then selects a similar group of individuals who do not have the disease (controls). Ideally, controls should come from the same source population as the cases. The researcher then compares the groups in terms of past exposures. If a certain exposure is associated with or caused the disease, a higher proportion of past exposure among the cases than among the controls would be expected (see Figure III-2).

Thus, for example, in the late 1960s, doctors in Boston were confronted with an unusual number of young female patients with vaginal adenocarcinoma. Those patients became the "cases" because they had the disease in question in a case-control study and were matched with "controls" who did not have the disease. Controls were selected based on their being born in the same hospitals and at the same time as the cases. The cases and controls were compared

for exposure to agents that might be responsible, and researchers found maternal ingestion of DES (diethylstilbestrol) during pregnancy in all but one of the cases but in none of the controls.²⁵

Figure III-2. Design of a Case-Control Study



SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 560. Copyright © 2011 National Academy of Sciences.

Table III-2. Cross-Tabulation of Disease by Exposure Status

	Exposure	No Exposure	Totals	Exposure Odds
Cases	a	c	a + c	a / c
Controls	b	d	c + d	b / d

SOURCE: Courtesy of the authors.

An advantage of the case-control study is that it usually can be completed in less time and with less expense than a cohort study. Case-control studies are also particularly useful in the study of rare diseases, because if a cohort study were conducted, an extremely large group would have to be studied in order to observe the development of a sufficient number of cases for analysis. Thus, for example, to detect a doubling of disease caused by exposure to an agent where the incidence of disease is 1 in 100 in the unexposed population would require sample sizes of 3,100 for the exposed and nonexposed groups for a cohort study, but only 177 for the case and control groups in a case-control study. But speed and fewer required subjects are not costless: A number of potential problems with case-control studies are discussed in Section III. E. 2, *infra*.

c. Cross-Sectional Studies

A third type of observational study is a cross-sectional study. In this type of study, individuals are interviewed or examined, and the presence of both the exposure of interest and the disease of interest is determined in each individual at a single point in time. Cross-sectional

studies determine the presence (prevalence) of both exposure and disease in the subjects and do not determine the development of disease or risk of disease (incidence). Moreover, because both exposure and disease are determined in an individual at the same point in time, it is not possible to establish the temporal relation between exposure and disease—that is, that the exposure preceded the disease, which would be necessary for drawing any causal inference. Thus, a researcher may use a cross-sectional study to determine the connection between a personal characteristic that does not change over time, such as blood type, and existence of a disease, such as aplastic anemia, by examining individuals and determining their blood types and whether they suffer from aplastic anemia. Cross-sectional studies are infrequently used when the exposure of interest is an environmental agent (current smoking status is a poor measure of an individual's history of smoking), but these studies can provide valuable leads to additional directions for research.

d. Ecological Studies

Up to now, we have discussed studies in which data on both exposure and health outcome are obtained for each individual included in the study. In contrast, studies that collect data only about the group as a whole are called ecological studies. In ecological studies, information about individuals is generally not gathered; instead, overall rates of disease or death for different groups are obtained and compared. The objective is to identify some difference between the two groups, such as diet, genetic makeup, or alcohol consumption that might explain differences in the risk of disease observed in the two groups. Such studies may be useful for identifying associations, but they rarely provide definitive causal answers. The difficulty is illustrated below with an ecological study of the relationship between dietary fat and cancer.

If a researcher were interested in determining whether a high dietary fat intake is associated with breast cancer, he or she could compare different countries in terms of their average fat intakes and their average rates of breast cancer. If a country with a high average fat intake also tends to have a high rate of breast cancer, the finding would suggest an association between dietary fat and breast cancer. However, such a finding would be far from conclusive, because it lacks particularized information about an individual's exposure and disease status (i.e., whether an individual with high fat intake is more likely to have breast cancer). In addition to the lack of information about an individual's intake of fat, the researcher does not know about the individual's exposures to other agents (or other factors, such as a mother's age at first birth) that may also be responsible for the increased risk of breast cancer. This lack of information about each individual's exposure to an agent and disease status detracts from the usefulness of the study and can lead to an erroneous inference about the relationship between fat intake and breast cancer, a problem known as an ecological fallacy. The fallacy is assuming that, on average, the individuals in the study who have suffered from breast cancer consumed more dietary fat than those who have not suffered from the disease. This assumption may not be true. Nevertheless, the study is useful in that it identifies an area for further research: the fat intake of individuals who have breast cancer as compared with the fat intake of those who do

not. Researchers who identify a difference in disease or death in a demographic study may follow up with a study based on gathering data about individuals.

Another epidemiologic approach is to compare disease rates over time and focus on disease rates before and after a point in time when some event of interest took place. For example, thalidomide's teratogenicity (capacity to cause birth defects) was discovered after Dr. Widukind Lenz found a dramatic increase in the incidence of limb reduction birth defects in Germany beginning in 1960, shortly after the introduction of thalidomide, which was heralded as a new and safer sedative. Yet, other than with such powerful agents as thalidomide, which increased the incidence of limb reduction defects by several orders of magnitude, these secular-trend studies (also known as time-line studies) are less reliable and less able to detect modest causal effects than the observational studies described above. Other factors that affect the measurement or existence of the disease, such as improved diagnostic techniques and changes in lifestyle or age demographics, may change over time. If those factors can be identified and measured, it may be possible to control for them with statistical methods. Of course, unknown factors cannot be controlled for in these or any other kind of epidemiologic studies.

C. Determining Exposure and Measuring Dose

Epidemiologists must determine if the subjects of their studies were exposed to the agent or factor they are interested in studying. Exposures of interest vary widely and include environmental agents, physical characteristics, genetics, and lifestyle choices. Sometimes the exposure is dichotomous and readily assessed. Suppose that a researcher is interested in whether mammography screening by the age of 50 has an effect on mortality due to breast cancer. Determining whether women were exposed or not should be relatively straightforward based on each subject's medical records. Different assessment methods are employed based on the exposure of interest. Because the interest of tort law is in human or product interventions, we focus on environmental agents, drugs, chemicals, and hazardous waste in this discussion.

Often, unlike mammography, exposure to an agent is a matter of degree—the intensity of an environmental factor like dust concentration, radiation, or industrial chemicals. Dose, then, consists of the concentration or intensity of the agent and the duration of exposure. (To be a bit more precise, often the relevant measure is the dose absorbed by the individual but frequently the environmental dose serves as a surrogate measure of the former for practical, privacy, technological, or ethical constraints.) Risk may depend on short-term intense exposure, a time-weighted average of intensity of exposure, or the cumulative life-time dose. The relevant exposures may be many years or decades before the manifestation of disease, as in the case of cancer where latency periods are lengthy. Researchers must attempt to understand the biology of the disease they are studying to determine the appropriate measure of dose to employ in a study. Consider this explanation of the difficulties of determining exposure to cigarette smoke for purposes of investigating its relationship to lung cancer:

Assume for discussion purposes that . . . the inhaled amount of benzo[*a*]pyrene best predicts

lung cancer risk. Even in a cohort study and certainly in a case-control study, one cannot hope to measure the inhaled amount of benzo[*a*]pyrene. What can be measured? Perhaps the daily consumption of cigarettes. But then one needs to know what type of tobacco is used, how far down each cigarette is smoked, whether there is a filter on the cigarette, and how deeply the individual inhales, among other things. Generally, none of this can be determined with any reasonable accuracy. Even if it could be, the ideal measure of exposure must integrate this information over a period of time and allow for a reasonable but usually unknown induction period. . . . [O]ne would theoretically need accurate cigarette-smoking information for some period of time long before the lung cancer occurs or might occur. Since the relevant time is uncertain, in principle one needs accurate exposure information for a period covering many decades, including the details of how the exposure varied by time during this period. Because historical information of such accuracy is not attainable, some misclassification of relevant exposure is unavoidable.¹

With hazardous waste, different disciplines are involved in assessing the migration of waste to proximity of humans who are exposed. Dispersal models may be employed with air pollution while geologists employ hydrogeologic methods to understand how waterborne waste is carried through rivers or other bodies of water. Hydrogeologic methods may also be used to examine how pollutants migrate through soil and into groundwater.

Biomarkers sometimes are available. They classically consisted of internal physiologic changes that reflect exposure to an agent and which can be identified and measured. These biomarkers are potentially helpful as evidence of exposure to an agent. We discuss biomarkers, including genetic characteristics that may reflect diseases associations and thus individual susceptibility below in Section V. G, *infra*. This rapidly developing field is known as molecular or genetic epidemiology.²⁶

In occupational epidemiology studies, researchers may undertake environmental exposure measurements of the agents being studied to determine the intensity of the agent in the workplace. For retrospective studies in which the work site changed over time, exposure measurement may not be possible, and researchers can only use the existence of the agent and duration of exposure to it as the measure of dose. Often occupational epidemiology examines the risks of specific occupations rather than exposure to individual chemical or other toxic agents.

D. The Outputs of Studies: Relative Risks, Odds Ratios, Attributable Proportion of Risk, and Standardized Rate Ratios

Epidemiologists are ultimately interested in whether a causal relationship exists between an agent and a disease. However, the first question an epidemiologist addresses is whether an association exists between exposure to the agent and disease. An association between exposure to an agent and disease exists when they occur together more frequently than one would expect by chance. For example, approximately 3–5% of births involve some

¹ Kenneth J. Rothman & Sander Greenland, *Modern Epidemiology* 142 (2d ed. 1999).

form of a birth defect. If the incidence of birth defects in women taking a suspected drug is also 3–5%, there is no association because the incidence is exactly what we would expect in those women if they had not taken the drug.

A causal relationship is one possible explanation for an observed association between an exposure and a disease. However, we emphasize that, *an association does not necessarily mean that there is a cause-effect relationship*. “Correlation does not imply causation,” is a phrase frequently seen in statistics texts. Although it emphasizes the lack of congruence between causation and association, it is somewhat inaccurate as an association *may* be, but is not necessarily, causal. Interpreting the meaning of an observed association is discussed below in Sections III E. & F, *infra*.

The strength of an association between exposure and disease can be stated in various ways, including as a relative risk, an odds ratio, or an attributable risk. Each of these measurements of association examines the degree to which the risk of disease increases when individuals are exposed to an agent.

1. Relative Risk

A commonly used approach for expressing the association between an agent and disease is relative risk (RR). It is defined as the ratio of the incidence rate (often referred to as incidence) of disease in exposed individuals to the incidence rate in unexposed individuals and was previously shown in Table III-1 in Section III B. 2. a.:

$$RR = \frac{\text{Incidence rate in the exposed } (\frac{c}{a+c})}{\text{Incidence rate in the unexposed } (\frac{d}{b+d})}$$

The incidence rate of disease is defined as the number of cases of disease that develop during a specified period of time divided by the number of persons in the cohort under study. Thus, the incidence rate expresses the risk that a member of the population will develop the disease within a specified period of time. Epidemiologists also use the concept of prevalence, which measures the existence of disease in a population at a given point in time, regardless of when the disease developed. Prevalence is expressed as the proportion of the population with the disease at the chosen time.

To illustrate the determination of a relative risk from data obtained during a study, assume a researcher studies 100 individuals who are exposed to an agent and 200 who are not exposed. After one year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals also are diagnosed as having the disease. The relative risk of contracting the disease is calculated as follows:

- The incidence rate of disease in the exposed individuals is 40 cases per year per 100 persons (40/100), or 0.4.
- The incidence rate of disease in the unexposed individuals is 20 cases per year per

- 200 persons (20/200), or 0.1.
- The relative risk is calculated as the incidence rate in the exposed group (0.4) divided by the incidence rate in the unexposed group (0.1), or 4.0.

A relative risk of 4.0 means that the risk of disease in the exposed group is 4 times as high as the risk of disease in the unexposed group.

In general, the relative risk can be interpreted as follows:

- If the relative risk equals 1.0, the risk in exposed individuals is the same as the risk in unexposed individuals. There is no association between exposure to the agent and disease. When there is no true association, causation does not exist.
- If the relative risk is greater than 1.0, the risk in exposed individuals is greater than the risk in unexposed individuals. There is a positive association between exposure to the agent and the disease, which could be causal.
- If the relative risk is less than 1.0, the risk in exposed individuals is less than the risk in unexposed individuals. There is a negative association, which could reflect a protective or curative effect of the agent on the risk of disease. For example, immunizations lower the risk of disease.

Although relative risk is a straightforward concept, care must be taken in interpreting it. Whenever an association is uncovered, further analysis must be conducted to assess whether the association reflects a causal relationship or whether there is another explanation, such as sampling error, confounding, or bias, which we address below. These same sources of error may mask a true association, resulting in a study that erroneously finds no association.

2. Odds Ratio

The odds ratio¹ (OR) is similar to a relative risk in that it expresses in quantitative terms the association between exposure to an agent and a disease. It is a convenient way to estimate the relative risk in a case-control study when the disease under investigation is rare.² The odds ratio approximates the relative risk when the disease is rare.

A relative risk cannot be calculated for a case-control study, because a case-control study begins by examining a group of persons who already have the disease. Consider Table 2, which shows the data that would be obtained in a case-control study. For both sets of cases (those with the disease), there is no denominator available to determine the incidence of

¹ The odds of an event, such as that a sports team will win a contest, is the ratio of the probability it will win the contest to the probability it will lose the contest. Thus, if the betting odds for the Boston Red Sox to win the World Series is 3-2, this means that bookmakers think that 3 times out of 5, the Red Sox will win and 2 times out of 5, they will lose, corresponding to a probability of 60% they will win, which makes one coauthor of this module, who is a Red Sox fan, very happy.

² If the disease is not rare, the odds ratio is still valid to determine whether an association exists, but interpretation of its magnitude is less intuitive.

disease (recall that the incidence of disease is the proportion of exposed persons who contract the disease in a given time period). Without a rate or incidence of disease, a researcher cannot calculate a relative risk.

In a case-control study, the odds ratio is the ratio of the odds that a case (one with the disease) was exposed to the odds that a control (one without the disease) was exposed. In a cohort study, the odds ratio is the ratio of the odds of developing a disease when exposed to a suspected agent to the odds of developing the disease when not exposed.

Consider a case-control study, with results as shown schematically in a 2 x 2 table (Table III-3):

Table III-3. Cross-Tabulation of Cases and Controls by Exposure Status

	Cases (with disease)	Controls (no disease)
Exposed	a	b
Not Exposed	c	d

SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 568. Copyright © 2011 National Academy of Sciences.

In a case-control study:

$$\text{Odds Ratio (OR)} = \frac{\text{the odds that a case was exposed}}{\text{the odds that a control was exposed}}$$

Looking at the above 2 x 2 table, this ratio can be calculated as:

$$\frac{a/c}{b/d} = \frac{ad}{bc}$$

Because we are multiplying two diagonal cells in the table and dividing by the product of the other two diagonal cells, the odds ratio is also called the cross-product ratio.

Consider the following hypothetical study: A researcher identifies 100 individuals with a disease who serve as “cases” and 100 people without the disease who serve as “controls” for her case-control study. Forty of the 100 cases were exposed to the agent and 60 were not. Among the control group, 20 people were exposed and 80 were not. The data can be presented in a 2 x 2 table (Table III-4):

Table III-4. Case-Control Study Outcome

	Cases (with disease)	Controls (no disease)
Exposed	40	20
Not Exposed	60	80

SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 569. Copyright © 2011 National Academy of Sciences.

The calculation of the odds ratio would be:

$$OR = \frac{40/60}{20/80} = 2.67$$

If the disease is relatively rare in the general population (about 5 percent or less), the odds ratio is a good approximation of the relative risk, which means that there is almost a tripling of the disease in those exposed to the agent. The odds ratio is usually marginally greater than the relative risk. As the disease in question becomes more common, the difference between the odds ratio and the relative risk grows.

That the relative risk and the odds ratio roughly approximate each other when the incidence of disease is small can be demonstrated by returning to Table III-1, which contains the results of a cohort study:

Table III-1. Cross-Tabulation of Exposure by Disease Status

	No Disease	Disease	Totals	Incidence Rates of Disease
Not Exposed	a	c	a + c	c / (a + c)
Exposed	b	d	b + d	d / (b + d)

SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 558. Copyright © 2011 National Academy of Sciences.

When a disease is rare, *c* (the number of people with disease who were not exposed to the agent) is small compared to *a* (the number of people without disease who also were not exposed to the agent). Thus, $(a + c)$ approximates *a*. Similarly, when the disease is rare, *d* (the number of people with the disease who were exposed to the agent) is small compared to *b* (the number of people without the disease who were exposed to an agent). Thus, $(b + d)$ approximates *d*.

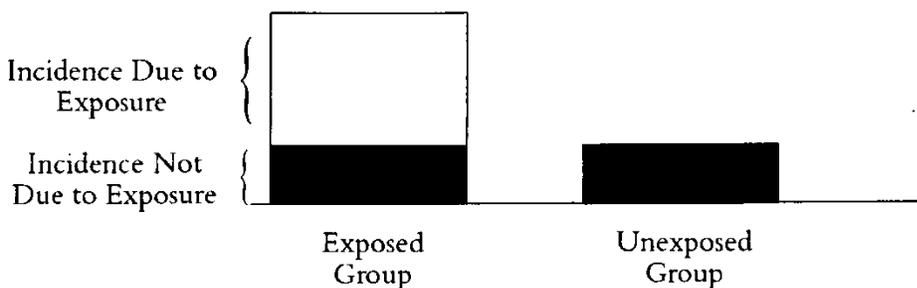
Recall from the discussion of relative risk that it is the ratio of the incidence rate of disease in the exposed cohort, $d / (b + d)$, divided by the incidence rate of those in the control cohort $c / (a + c)$ or $\frac{d/(b+d)}{c/(a+c)}$. When c and d are small, the relative risk approximates $\frac{d/b}{c/a}$. The OR, recall, is the rate of exposure of the cases to the rate of exposure of the controls or $\frac{d/b}{c/a}$.

3. Attributable Risk

A frequently used measurement of risk is the attributable risk (AR), sometimes also referred to as the attributable proportion of risk (APR). The attributable risk represents the amount of disease among all of the exposed individuals that can be attributed to the exposure as opposed to other causes of the disease. Can you explain why the AR for a signature disease is 100%?

It also can be expressed as the proportion of the disease among exposed individuals that is associated with the exposure (also called the “etiologic fraction,” or the “attributable risk percent”). The attributable risk reflects the proportion of the disease that can be attributed to exposure to an agent and consequently the proportion of disease that could be potentially prevented by blocking the effect of the exposure or by eliminating the exposure. In other words, if the association is causal, the attributable risk is the proportion of disease in an exposed population that might be caused by the agent and that might be prevented by eliminating exposure to that agent (see Figure 3). It also represents the average probability that an exposed study subject’s disease was caused by exposure.¹

Figure III-3. Risks in Exposed and Unexposed Groups



SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 570. Copyright © 2011 National Academy of Sciences.

To determine the proportion of a disease that is attributable to an exposure, a researcher would need to know the incidence of the disease in the exposed group and the

¹ We defer to Subsections V.B.-D. & F. discussion of the issues that arise in attempting to determine the probability that a given individual’s disease was caused by an agent based on the attributable proportion of risk

incidence of disease in the unexposed group. The attributable risk for a cohort study is:

$$\text{Attributable Risk (AR)} = \frac{(\text{incidence in the exposed}) - (\text{incidence in the unexposed})}{\text{incidence in the unexposed}}$$

The attributable risk can also be calculated from the relative risk found in a study:

$$\text{Attributable Risk (AR)} = \frac{RR - 1}{RR}$$

The equivalence of the two equations for attributable risk above can be understood by appreciating that when 1.0 is subtracted from the relative risk as in the second equation, it reflects the excess risk in the exposed cohort over and above the risk in the unexposed—the same numerator as in the first equation. In the second equation, the excess relative risk attributable to exposure is divided by the entirety of the risk in the exposed group equivalent to the denominator in the first group.¹

The attributable risk can be calculated using the example described in Section III A, *supra*. Suppose a researcher studies 100 individuals who are exposed to a substance and 200 who are not exposed. After one year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals are also diagnosed as having the disease.

- The incidence of disease in the exposed group is 40 persons out of 100 who contract the disease in a year.
- The incidence of disease in the unexposed group is 20 persons out of 200 (or 10 out of 100) who contract the disease in a year.
- The proportion of disease that is attributable to the exposure is 30 persons out of 40, or 75 percent. This is so because among the 40 in the exposed group, 10 would have contracted the disease despite the exposure (see Figure 3).

This means that 75 percent of the disease in the exposed group is attributable to the exposure. We should emphasize here that “attributable” does not necessarily mean “caused by.” Up to this point, we have only addressed associations. Inferring causation from an association is addressed in Section III E, *infra*.

To calculate the attributable proportion of risk for a case-control study, we cannot use the first equations above, because, remember, we don’t know the incidence rates. However, because we know that the odds ratio approximates the relative risk when the incidence of disease is low, we can use the second equation (substituting the odds ratio for the relative risk) to calculate an attributable proportion of risk from a case-control study.

¹ The equivalence of this equation to the first can be demonstrated algebraically by taking the first equation and multiplying numerator and denominator by I_c , which yields $\frac{I_e - I_c}{\frac{I_c - I_c}{I_e}}$. Substituting RR for I_e/I_c yields $(RR - 1)/RR$.

E. Adjustment for Study Groups That Are Not Comparable: Standardized Mortality and Morbidity Ratios

Comparing mortality and morbidity rates in two or more populations can be useful in assessing disease or other health outcomes. Morbidity refers to the disease state of an individual. The morbidity rate looks at disease incidence across a population during a specified period of time. Mortality refers to death, and the mortality rate is the number of deaths in a population. Populations often differ in characteristics that relate to disease risk, such as age, sex, and race. Most diseases occur at different rates in different age groups. For example, those who live in Florida have a much higher death rate than those who live in Alaska.¹ Is sunshine and warmth dangerous? Perhaps, but the Florida population is much older than the Alaska population, and some adjustment must be made for the differences in age distribution in the two states in order to compare disease or death rates between the two populations. The technique used to accomplish this is called adjustment, and two types of adjustment are used—direct and indirect. Both direct and indirect adjustment may be used during data analysis to minimize the effect of extraneous sources of variation that may affect the study's results.

The “crude rate” is a weighted average calculated by dividing the total number of cases by the total number of people in the population. Crude death rates measure the magnitude of mortality in a population. It may be tempting to compare the crude rates of two populations; however, this can be misleading. Adjustment allows us to take into account differences in confounding factors, like age, to provide more accurate comparisons. Let us first examine the crude death rate (C) in two different populations, the first of which might resemble Florida and the second Alaska:

Population 1			
Group	Age (years)	Total Population	Deaths
1	0-24	150	5
2	25-49	100	9
3	50-74	100	14
4	75+	170	24
Total		520	52

$C_1 = 52 \text{ deaths} / 520 \text{ people} = 0.10 = 10 \text{ deaths per } 100 \text{ people in population.}$

SOURCE: Courtesy of the authors.

¹ See David E. Lilienfeld & Paul D. Stolley, *Foundations of Epidemiology* 68–70 (3d ed. 1994) (mortality rate in Florida is approximately three times what it is in Alaska).

Population 2			
Group	Age (years)	Total Population	Deaths
1	0-24	230	13
2	25-49	125	13
3	50-74	85	13
4	75+	70	13
Total		520	52

$C_2 = 52 \text{ deaths}/520 \text{ people} = 0.10 = 10 \text{ deaths per } 100 \text{ people in population.}$

SOURCE: Courtesy of the authors.

These two populations have the same crude death rate (0.10). Is the risk of dying the same in both populations 1 and 2? Why is the comparison of the crude death rates of populations 1 and 2 not accurate? Population 1 has a greater percentage of older people and old age is associated with increased mortality. When the crude death rate is calculated for each age sub-group (e.g., subgroup 1 of those aged 0–24 years) this is called an age-specific death rate. Now let us examine the age-specific death rate in each population and compare the age-specific death rates from the two populations.

Population 1				
Group	Age (years)	Total Population	Deaths	Death Rate per 100
1	0-24	150	5	3
2	25-49	100	9	9
3	50-74	100	14	14
4	75+	170	24	12
Total		520	52	10

SOURCE: Courtesy of the authors.

Population 2				
Group	Age (years)	Total Population	Deaths	Death Rate per 100
1	0-24	230	13	6
2	25-49	125	13	10
3	50-74	85	13	15
4	75+	70	13	19
Total		520	52	10

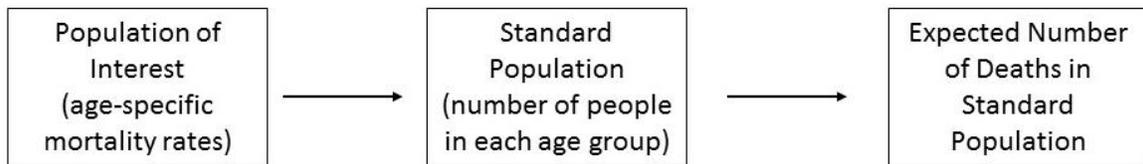
SOURCE: Courtesy of the authors.

Here we see that despite each population having the exact same crude death rate, Population 2 has a higher age-specific death rate in each identified age sub-group. Comparison of crude death rates may be confounded by differing population distributions. That is, the

number of people in each age subgroup varies greatly between the two populations. One way to account for this confounding is through adjustment.

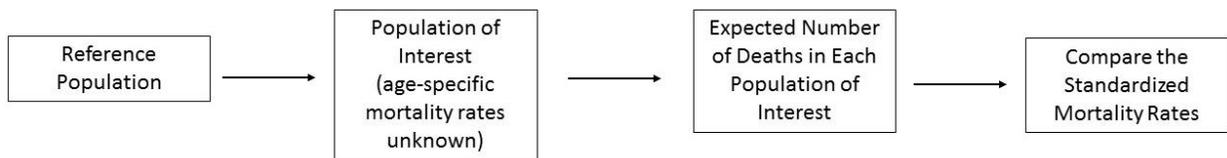
There are two methods of adjustment that can be used to calculate standardized rates. When using the direct method of adjustment, the rates observed in a population of interest are applied to a standard population so that the number of deaths expected in the standard population can be calculated. The ratio of actual deaths to expected number of deaths provides a measure of the increased (or decreased) rate of morbidity or mortality in the studied population. With the indirect method of adjustment, the rates from the reference population are applied to the population of interest to determine the expected number of deaths in each population.

*Direct Adjustment*²⁷



SOURCE: Courtesy of the authors.

*Indirect Adjustment*²⁸



SOURCE: Courtesy of the authors.

Direct adjustment allows researchers to compare populations that have different distributions of characteristics (e.g., age distribution). Direct age adjustment analyses can be used to determine what the comparable death rate would be in each population if both populations had the same age distributions. In direct age adjustment, overall disease/death rates are calculated for the population as though the population had the age distribution of another standard or reference population. We can then compare these overall rates, called age-adjusted rates, knowing that any difference between these rates cannot be attributed to differences in age distributions in the populations. The standard population may be created by combining two or more populations or by using an existing population (e.g., North America population).

Indirect adjustment is used when the age-specific rates for a study population are not known or the numbers are too small. In indirect adjustment, the age-specific rates are taken

from the standard population and then applied to each study population.

The standardized mortality ratio (SMR) is the simplest form of indirect adjustment. The mortality rate of the reference population is applied to the observed population of interest. When the outcome of interest is disease rather than death, it is called the standardized morbidity ratio. If the ratio equals 1.0, the observed number of deaths equals the expected number of deaths, and the mortality rate of the population of interest is no different from that of the reference population. If the SMR is greater than 1.0, the population of interest has a higher mortality risk than that of the reference population, and if the SMR is less than 1.0, the population of interest has a lower mortality rate than that of the reference population.

Thus, age adjustment provides a way to compare populations while in effect holding age constant. Adjustment is used not only for comparing mortality rates in different populations but also for comparing morbidity (disease) rates in different groups of subjects selected for study in epidemiologic investigations. Although this discussion has focused on adjusting for age, it is also possible to adjust for any number of other variables, such as gender, race, occupation, and socioeconomic status that might affect the risk of the diseases being studied. It is also possible to adjust for several factors simultaneously.

F. Sources of Error

Incorrect study results can occur in a variety of ways. A study may find a positive association (relative risk or odds ratio greater than 1.0) when there is no true association. Or a study may erroneously find that there is no association when in reality there is. A study may also find an association when one truly exists, but the association found may be greater or less than the true association.

Three general categories of phenomena can cause a study's results to be erroneous: chance, bias, and confounding. Before any inferences about causation are drawn from a study, the possibility of these phenomena must be examined to determine if they are responsible for the association found in the study rather than a causal relationship. Epidemiologists, aware of these sources of error, attempt to minimize them in advance when designing a study.

The findings of a study may be the result of (or affected by) chance (or random error). In designing a study, the size of the sample can be increased to reduce (but not eliminate) the likelihood of random error. Once a study has been completed, statistical methods (discussed in the next subsection) permit an assessment of the extent to which the results of a study may be due to random error.

The methods for assessing random error are statistical significance and confidence intervals. A study that is statistically significant has results that are unlikely to be the result of random error, although any criterion for "significance" is somewhat arbitrary. A confidence interval provides both the relative risk (or other risk measure) found in the study and a range (interval) within which the outcome risk likely would fall if the study were repeated numerous times. These two techniques (which are closely related) are explained in subsection III.E.I, *infra*. We should emphasize a matter that those unfamiliar with statistical methodology frequently find confusing. That a study's results are statistically significant says nothing about the

magnitude of any association (i.e., the relative risk or odds ratio) found in a study or about the biologic or clinical importance of the finding. “Significant,” as used with the adjective “statistically,” does not mean either large or important. A study may find a statistically significant relationship that is quite modest—perhaps it increases the risk only by 5%, which is equivalent to a relative risk of 1.05.¹ An association may be quite large—the exposed cohort might be 10 times more likely to develop disease than the control group—but the association is not statistically significant because of the potential for random error given a small sample size. In short, statistical significance is not about the magnitude of the risk found in a study.

Bias² (or systematic error) can also produce error in the outcome of a study. Epidemiologists attempt to minimize bias through their study design, including data collection protocols. However, even the best designed and conducted studies have biases, which may be subtle. Consequently, after data collection is completed, analytic tools are often used to evaluate potential sources of bias. Sometimes, after bias is identified, the epidemiologist can determine whether that bias would tend to inflate or dilute any true association that may exist. Identification of the bias may permit the researcher to make an assessment of whether the study’s conclusions are valid. Epidemiologists may reanalyze a study’s data to correct for a bias identified in a completed study or to validate the analytic methods used. Common biases and how they may produce invalid results are described in Section III E. 2.

Finally, a study may reach incorrect conclusions about causation because, although the agent and disease are associated, the agent is not a true causal factor. Rather, the agent may be associated with another agent that is the true causal factor, and this factor confounds the relationship being examined in the study. Confounding is explained in Section IV C. A bald head does not cause death, but because bald heads are differentially associated with old age, there is an association between bald heads and death. That association is a product of confounding rather than causation.

1. Statistical Methods to Evaluate the Possibility of Sampling Error

Before detailing the statistical methods used to assess random error (which we use as synonymous with sampling error), we explain two concepts that are central to epidemiology and statistical analysis. Understanding these concepts will facilitate comprehension of the statistical methods.

Epidemiologists often refer to the true association (also called “real association”), which is the association that really exists between an agent and a disease and that might be found by a perfect (but nonexistent) study. The true association is a concept that is used in evaluating the results of a given study even though its value is (and remains) unknown. By contrast, a study’s outcome will produce an observed association, which is known.

¹ In general, small risks, such as the one discussed in the text, require more subjects in order to obtain a statistically significant result. When effects are larger, fewer subjects are needed to obtain statistical significance.

² Scientists use bias in a different way from lawyers. To scientists, bias is a source of error that can produce spurious results in a study. Systematically incorrectly classifying some subjects as unexposed when they were partially exposed would be a source of bias.

Formal procedures for statistical testing often begin with the null hypothesis, which posits that there is no true association (i.e., a relative risk of 1.0) between the agent and disease under study. Data are gathered and analyzed to see whether they disprove the null hypothesis. The data are subjected to statistical testing to assess the plausibility that any association found is a result of random error or, alternatively, whether it supports rejection of the null hypothesis. The use of the null hypothesis for this testing should not be understood as the a priori belief of the investigator. When epidemiologists investigate an agent, it is usually because they think that the agent is a cause of some outcome. There is little incentive to conduct a study that finds no effects. Nevertheless, epidemiologists prepare their study designs and test the plausibility that any association found in a study was the result of random error by using the null hypothesis.

a. False Positives and Statistical Significance

When a study outcome is positive (i.e., a relative risk greater than 1.0), epidemiologists try to determine whether that outcome represents a true association or is the result of random error. Random error can be illustrated by thinking about flipping a fair coin (i.e., not modified to produce more heads than tails (or vice versa)). On average, for example, we would expect that coin tosses would yield half heads and half tails. But sometimes a set of coin tosses might yield an unusual result, for example, six heads out of six tosses, an occurrence that would result, purely by chance, in about 1.5 percent of a series of six tosses.¹ In the world of epidemiology, sometimes the study findings, merely by chance, do not reflect the true relationships between an agent and outcome. Any single study—even a clinical trial—is in some ways analogous to a set of coin tosses, being subject to the play of chance. Thus, for example, even though the true relative risk (in the total population) is 1.0, an epidemiologic study of a population may find a relative risk greater than (or less than) 1.0 because of random error or chance. A researcher may then conclude that the study, incorrectly, reveals a true association and, therefore, the null hypothesis is incorrect. Such an erroneous conclusion that the null hypothesis is false (i.e., a conclusion that there is a difference in risk between the two study groups when no difference actually exists) owing to random error is called a false-positive error (also type I error or alpha (α) error). By contrast, an incorrect conclusion that there is no difference in risk (i.e., that the null hypothesis is true) is known as false-negative error (also type II error or beta (β) error). These two forms of error are reflected in Table III-5:

¹ $\frac{1}{2}$ to the 5th power, as the odds of another heads after the first one is 1 out of 2.

Table III-5. False Positive and False Negative Error

		Null Hypothesis is Actually:	
		True	False
Conclusion Drawn	Accept the Null Hypothesis	Correct Conclusion!	Type I Error (α) False Positive
	Reject the Null Hypothesis	Type II Error (β) False Negative	Correct Conclusion!

SOURCE: Courtesy of the authors.

Common sense leads one to believe that a large enough sample of individuals must be studied if the study is to identify a relationship between exposure to an agent and disease that truly exists. Common sense also suggests that by enlarging the size of the study groups, researchers can form a more accurate conclusion and reduce the chance of random error in their results. Both statements are correct and can be illustrated by a test to determine if a coin is fair. A test in which a fair coin is tossed 1,000 times is more likely to produce close to 50 percent heads than a test in which the coin is tossed only 10 times. It is far more likely that a test of a fair coin with 10 tosses will come up, for example, with 80% heads than will a test with 1,000 tosses. With large numbers, the outcome of the test is less likely to be influenced by random error—the results are statistically stable—and the researcher would have greater confidence in the inferences drawn from the data. In probability theory, this is known as the law of large numbers.

One way of assessing the possibility that an observed association could have occurred as a result of random error is by calculating a p-value. A p-value represents the probability that an observed positive association (or one even greater) could result from random error even if no association were in fact present, that is, even if the null hypothesis is correct. The qualification *“even if no association were in fact present”* is crucial. Because the p-value is calculated based on the assumption that the null hypothesis is true for the population, the p-value cannot and does not reflect the probability that the null hypothesis is true or false (conventional statistical methods do not permit such a calculation). Thus, a p-value of 0.1 is correctly interpreted as saying, assuming that the null hypothesis is true, you would obtain the observed difference (or a greater one) in 10% of identical studies due to random sampling error. A p-value of 0.1 is often incorrectly interpreted to mean that if the null hypothesis is rejected there is a 10% chance that, on the contrary, the null hypothesis is correct. The probability of incorrectly rejecting a true null hypothesis is actually a lot higher than the p-value. Or, to put it another way, the complement of α , the risk of false positive error, is not β , the risk of false negative error, i.e., $1 - \alpha \neq \beta$. Since the null hypothesis is assumed to be true, there is no way to determine whether the null hypothesis is actually true or, because of sampling error the null hypothesis is actually false.

To minimize false positives, epidemiologists use a convention that the p-value must fall below some selected level known as alpha or the significance level for the results of the study to be statistically significant. Thus, an outcome is statistically significant when the observed p-value for the study falls below the preselected significance level. The most common significance level, or alpha, used in epidemiologic research (and other empirical sciences) is .05. A .05 value means that the probability is 5% of observing an association at least as large as that found in the study when in truth there is no association. Although .05 is often the significance level selected, other levels can and have been used. Thus, in its study of the effects of secondhand smoke, the Environmental Protection Agency (EPA) used a .10 standard for significance testing.

Because the burden of persuasion in civil cases is a preponderance of the evidence, that is, more probable than not, legal commentators, courts, and others, including scientists, sometimes incorrectly attempt to equate this legal test to tests of statistical significance used in epidemiologic research. The most egregious and yet common examples occurs with cases and commentators arguing that in civil cases courts should simply disregard the typical p-value of tests of significance in the legal setting, .05, because requiring 95% certainty is far higher than the preponderance of evidence burden.

Such statements are based on a misunderstanding of what tests of significance accomplish and their difference from standards of proof. Significance testing states the probability that the study outcome (or an even more extreme outcome) would occur if the null hypothesis is correct. It does not permit us to conclude that the null hypothesis has only a 5% chance of being correct. The p-value tells us what is likely to happen when the null hypothesis is correct, that is, the probability of a false positive. It does not tell us the probability that the study result is true. Recall also, that beta (the probability that we will incorrectly conclude there is no effect, a false negative is not the complement of the chosen alpha (p-value), i.e., $\alpha \neq 1 - \beta$, unlike the case the preponderance standard, that is, the probability of incorrectly ruling for the defendant is the complement of the probability the probability of ruling incorrectly for the plaintiff. Thus, to employ a .50 significance level standard, the result would be to greatly increase the ratio of false positive (Type I) errors to false negative (Type II) error. To paraphrase Professor David Kaye, analogizing significance standards to the legal standard of proof is like trying to find one's way from New York to California by consulting a map of Beijing.¹

Do you see why a better solution to accommodating the legal preponderance standard into significance testing would be to choose a test of significance that equalizes the probability of a false positive and a false negative? Why not, then, use a p-value of .5?

¹ David H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54, 66 (1987).

In re Ephedra Products Liab. Litig.
United States District Court for the District of Southern New York, 2005.
393 F. Supp. 2d 181, 192-93.

Rakoff, J.

Daubert was designed to exclude “junk science.” It was never intended to keep from the jury the kind of evidence scientists regularly rely on in forming opinions of causality simply because such evidence is not definitive. The legal standard, after all, is preponderance of the evidence, i.e., more-probable-than-not, and that applies to causality as to any other element of a tort cause of action. Rule 702, a rule of threshold admissibility, should not be transformed into a rule for imposing a more exacting standard of causality than more-probable-than-not simply because scientific issues are involved. It is one thing to prohibit an expert witness from testifying that causality has been established “to a reasonable degree of scientific certainty” when the very exacting standards for determining scientific certainty have not been met. But it by no means follows that a scientific expert may not testify to the scientific plausibility of a particular hypothesis of causality or even to the fact that a confluence of suggestive, though non-definitive, scientific studies make it more-probable-than-not that a particular substance (such as ephedra) contributed to a particular result (such as a seizure).

The difference between statistical significance and preponderance of the evidence is well illustrated by an examination of the one study that attempted to measure a possible association between ephedra and hemorrhagic stroke (one of the five listed injuries), namely, Morgenstern LB et al., Use of ephedra-containing products and risk for hemorrhagic stroke (*Neurology*, 2003, 60:132–135), DCC Exh. A404 (hereinafter Morgenstern). Morgenstern found a fivefold increased risk of hemorrhagic stroke in participants who had taken more than 32 mg of ephedra alkaloids on the day before a stroke. The result, however, was not statistically significant because of the small number of participants found to have taken ephedra at this dose, even though a typical recommended dose of the products in the instant cases is 96 mg/day. * * *

* * * [T]he DCC says the Morgenstern results cannot support opinion testimony on general causation because they fail to meet science's conventional test for statistical significance.

In motion papers and through argument and examination of experts at the hearings, the DCC repeatedly showed the Court how epidemiological studies quantify statistical significance in two ways—the “P-value” and the “confidence interval.” Generally accepted scientific convention treats a result as statistically significant if the P-value is not greater than .05. The expression “P=.05” means that there is one chance in twenty that a result showing increased risk was caused by a sampling error—i.e., that the randomly selected sample accidentally turned out to be so unrepresentative that it falsely indicates an elevated risk. “Confidence interval” measures the same risk of sampling error in a form that is less easy for a layman to picture. Morgenstern reports that the fivefold increased rate of hemorrhagic stroke among study participants who took more than 32 mg of ephedra on the day before the case's stroke has a “95% confidence interval of 0.84 to 41.33.” Because this interval includes the value 1.0 (which would mean no increased risk), the result is not considered statistically significant. This necessarily means that the P-value is greater than .05, though Morgenstern does not state the precise P-value.

The reason why Morgenstern's P-value is greater than .05 is that too few of the 702 stroke cases and 1,376 matching controls turned out to have taken more than 32 mg of ephedra within 24 hours before the case's stroke. In particular, only three controls had done so, even though the study design provided for two matching controls for each case. So the eight-year process described above, with 43 participating hospitals in six states, was insufficient to find enough ephedra users.

* * *

Scientific convention defines statistical significance as " $P \leq .05$," *i.e.*, no more than one chance in twenty of a finding a false association due to sampling error. Plaintiffs, however, need only prove that causation is more-probable-than-not. Although this legal standard may lead to what some scientists might consider an unacceptably high error rate in jury verdicts, the law has tolerated the jury error rate for centuries because it has not yet found a better way of adjudicating disputes. This Court will be guided by *Daubert's* "general observations" about scientific knowledge in its determination to keep junk science out of the courtroom. At the same time, it will not treat *Daubert's* dictum about scientific validity as authority for increasing the burden of proof imposed by substantive law.

Notes and Questions

1. Critique the court's discussion above with regard to statistical significance, p-values, and the civil standard of proof.

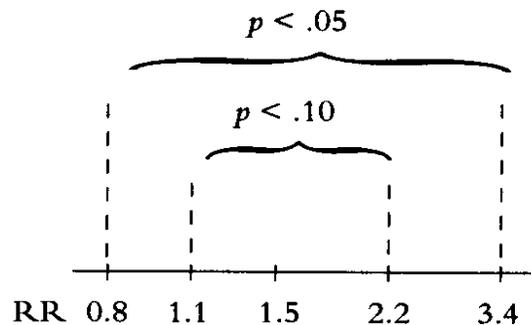
There is some controversy among epidemiologists and biostatisticians about the appropriate role of significance testing. To the strictest significance testers, any study whose p-value is greater than the level chosen for statistical significance should be rejected as inadequate to disprove the null hypothesis. Others are critical of using strict significance testing, which rejects all studies with an observed p-value above that specified level. Epidemiologists have become increasingly sophisticated in addressing the issue of random error and examining the data from a study to ascertain what information it may provide about the relationship between an agent and a disease, without the necessity of rejecting all studies that are not statistically significant. Reporting on the actual p-value, rather than merely whether a study's results were or were not statistically significant is helpful in understanding the extent of statistical instability in the outcome. Meta-analysis, as well, a method for pooling the results of multiple studies sometimes can ameliorate concerns about random error.

b. Confidence Intervals

Calculation of a confidence interval permits a more refined assessment of appropriate inferences about the association found in an epidemiologic study in light of the risk of sampling error. A confidence interval is a range of possible values calculated from the results of a study. If a 95% confidence interval is specified, the range encompasses the results we would expect 95% of the time if samples for new studies were repeatedly drawn from the same population. Thus, the width of the interval, see Figure 4 below, reflects a range of potential random error.

The narrower the confidence interval, the more statistically stable the results of the study. The advantage of a confidence interval is that it displays more than simply reporting that a study result is or is not statistically significant or providing a p-value. “Statistically significant” does not convey the magnitude of the association found in the study or indicate how statistically stable that association is. A confidence interval shows the boundaries of the relative risk based on selected levels of alpha or statistical significance. Just as the p-value does not provide the probability that the risk estimate found in a study is correct, the confidence interval does not provide the range within which the true risk must lie. Rather, the confidence interval reveals the likely range of risk estimates consistent with the null hypothesis being correct. An example of two confidence intervals that might be calculated for a given relative risk found in a study is displayed in Figure III-4.

Figure III-4. Confidence Intervals



SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 580. Copyright © 2011 National Academy of Sciences.

The confidence intervals shown in Figure 4 are for a study that found a relative risk of 1.5, with boundaries of 0.8 to 3.4 when alpha is set at 0.05 (equivalently, a confidence level of .95), and with boundaries of 1.1 to 2.2 when alpha is set at 0.10 (equivalently, a confidence level of .90). The confidence interval for alpha set at .05 might be reported in a study as “The relative risk found was 1.5 (95% CI, .8 to 3.4).”

The confidence interval for alpha equal to .10 is narrower because it encompasses only 90% of the expected test results (if the null hypothesis is correct). By contrast, the confidence interval for alpha equal to .05 includes the expected outcomes for 95 percent of the tests. To generalize this point, the lower the alpha chosen (and therefore the more stringent the exclusion of possible random error) the wider the confidence interval. At a given alpha, the width of the confidence interval is determined by sample size. All other things being equal, the larger the sample size, the narrower the confidence boundaries (indicating greater numerical stability). For a given risk estimate, a narrower confidence interval reflects a decreased likelihood that the association found in the study would occur by chance if the true association is 1.0.

For the example in Figure 4, the boundaries of the confidence interval with alpha set at 0.05 encompass a relative risk of 1.0, and therefore the result is not statistically significant at

the 0.05 level. Alternatively, if the confidence boundaries are set with an alpha equal to 0.10, then the confidence interval no longer includes a relative risk of 1.0, and the 1.5 relative risk result would be described as statistically significant at the 0.10 level.

c. False Negatives

As Figure 4 illustrates, false positives can be reduced by adopting more stringent values for alpha. Using an alpha of 0.05 will result in fewer false positives than using an alpha of 0.10, and an alpha of 0.01 or 0.001 would produce even fewer false positives. The trade-off for reducing false positives is an increase in false negative errors (or beta (β) errors or type II errors). This concept reflects the possibility that a study will be interpreted as “negative” (not disproving the null hypothesis), when in fact there is a true association of a specified magnitude. The beta for any study can be calculated only based on a specific alternative hypothesis about a specified positive relative risk and a specific level of alpha selected. Importantly, beta is not the complement of alpha, that is, $1 - \alpha$. Thus, the beta for a study when alpha is set at .10 might be .7, which reflects a high likelihood of a false negative due to random error.

d. Power

A study may fail to find a statistically significant association not because one doesn't exist, but because the study was insufficiently powered to find an association of that magnitude. The power of a study is the probability of finding a statistically significant association of a given magnitude (if it exists) in light of the sample sizes used in the study. The power of a study depends on several factors: the sample size; the level of alpha (or statistical significance) specified; the background incidence of disease; and the specified relative risk that the researcher would like to detect if it exists. When designing a study power can be analyzed to reveal the likelihood of finding any given relative risk or odds ratio in light of these factors. Often power curves are used in the design of a study to determine what size the study populations should be. If you were advising a researcher who was designing a study that might be used in future litigation what magnitude of relative risk would you recommend the researcher use in conducting power analyses?

Here is one researcher's report on the power calculation performed in the study design of an epidemiologic investigation:

The power of this study to detect a significant odds ratio of >3.00 for maternal involvement in agricultural work during the acute risk period and all selected congenital malformations was a relatively low 0.33 (significance level, 0.05). For fathers, there was a similar power to detect a significant odds ratio of >1.50 between handling pesticides during the acute risk period and all selected congenital malformations. For fathers who had ever handled pesticides, this same power was 0.55. The analysis by maternal handling of pesticides and by groups of congenital malformations was very limited by small numbers. In this study, significant associations were observed for maternal involvement in agricultural activities during the acute risk period. The risk

for reported paternal handling of pesticides was increased as well, although the increase was not statistically significant.

Ana M. Garcia et al., *Parental Agricultural Work and Selected Congenital Malformations*, 149 Am. J. Epidemiol. 64 (1999).

The power of a study is the complement of beta ($1 - \beta$). Thus, a study with a likelihood of .25 of failing to detect a true relative risk of 2.0 or greater has a power of .75. This means the study has a 75% chance of detecting a true relative risk of 2.0 that is statistically significant.

When the disease rarely occurs, finding an effect is more difficult. The same is true when the incidence of exposure is low. In situations of low exposure rates, case-control studies can be particularly valuable because of their relatively greater power.

After a study is completed, the confidence interval reveals which outcomes were not statistically compatible with the study's results. Thus, in Figure 4 the range of relative risks from 1.1 to 2.4 (for a p-value of .10) are all statistically compatible with the results of the study while a relative risk of 4.0 is not compatible with the study outcome.

2. Biases

The second major reason for an invalid outcome in epidemiologic studies is systematic error or bias. Bias may arise in the design or conduct of a study, data collection, or data analysis. The meaning of scientific bias differs from conventional (and legal) usage, in which bias refers to a partisan point of view. When scientists use the term bias, they refer to anything that results in a systematic (nonrandom) error in a study result and thereby compromises its validity.

Most observational studies have some degree of bias that may affect the outcome. If major bias is present it may imply the study results are invalid. Determining the existence of bias, however, can be difficult if not impossible. In examining the validity of an epidemiologic study, the reviewer must identify potential biases and analyze the amount or kind of error that might have been induced by the bias. Often the direction of error can be determined; depending on the specific type of bias, it may exaggerate the real association, dilute it, or even completely mask it. Two important categories of bias are selection bias (inappropriate methodology for selection of study subjects) and information bias (a flaw in measuring exposure or disease in the study groups).

a. Selection Bias

Selection bias¹ refers to the error in an observed association that results from the method of selection of the exposed and unexposed individuals (in a cohort study) and the cases and controls (in a case-control study). Ideally, controls should be drawn from the same

¹ Selection bias is defined as “[e]rror due to systematic differences in characteristics between those who are selected for study and those who are not.” *A Dictionary of Epidemiology* 153 (John M. Last ed., 3d ed. 1995).

population that produced the cases. Selecting control participants becomes problematic if the control participants are selected for reasons that are related to their having the exposure being studied.

Hospital-based studies, which are relatively common among researchers located in medical centers, illustrate the problem. Suppose an association is found between coffee drinking and coronary heart disease in a study using hospital patients as controls. The problem is that the hospitalized control group may include individuals who had been advised against drinking coffee for medical reasons, such as to prevent the aggravation of a peptic ulcer. In other words, the controls may become eligible for the study because of their medical condition, which in turn is related to their exposure status—their likelihood of avoiding coffee. If this is true, the amount of coffee drinking in the control group would understate the extent of coffee drinking expected in people who do not have the disease and thus bias upwardly (i.e., exaggerate) any odds ratio observed. Bias in hospital studies may also understate the true odds ratio when the exposures at issue led to the cases' hospitalizations and also contributed to the controls' chances of hospitalization.

Just as cases and controls in case-control studies should be selected independently of their exposure status, so the exposed and unexposed participants in cohort studies should be selected independently of their disease risk. For example, if women with hysterectomies are over-represented among exposed women in a cohort study of cervical cancer, this could result in understating the association between the exposure and the disease. Without a cervix, a woman would not be at risk of cervical cancer and thus the exposed group would have fewer cancers than if women with hysterectomies were proportionally the same as those in the control group.

A further source of selection bias occurs when those selected for the study decline to participate or drop out before the study is completed. Many studies have shown that individuals who participate in studies differ significantly from those who do not. If a significant portion of either study group declines to participate, the researcher should investigate whether those who declined are different from those who agreed. The researcher can compare relevant characteristics of those who participate with those who don't to demonstrate the extent to which the two groups are comparable. Similarly, if a significant number of subjects drop out of a study before completion, the remaining subjects may not be representative of the original study populations.

The fact that a study may suffer from selection bias does not necessarily invalidate its results. A number of factors may suggest that a bias, if present, had only limited effect. If the association is particularly strong, for example, bias is less likely to account for all of the increased risk. In addition, a consistent association across different control groups suggests that possible biases applicable to a particular control group are not invalidating. Similarly, a dose-response relationship (see Section III F. 3. *infra*) found among multiple groups exposed to different doses of the agent would provide additional evidence that biases applicable to the exposed group are not a major problem.

b. Information Bias

Information bias is a result of inaccurate information about either the disease or the exposure status of the study participants or a result of confounding. In a case-control study, potential information bias is an important consideration because the researcher depends on information from the past to determine exposure and disease and their temporal relationship. In some situations, researchers may be required to interview subjects about past exposures, thus relying on the subjects' memories. Research has shown that individuals with disease (cases) tend to recall past exposures more readily than individuals with no disease (controls); this creates a potential for bias called recall bias.

For example, consider a case-control study conducted to examine the cause of congenital malformations. The epidemiologist is interested in whether the malformations were caused by an infection during the mother's pregnancy. A group of mothers of malformed infants (cases) and a group of mothers of infants with no malformation (controls) are interviewed regarding infections during pregnancy. Mothers of children with malformations may recall an inconsequential fever or runny nose during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if in reality the infection rate in mothers of malformed children is no different from the rate in mothers of normal children, the result in this study would be an apparently higher rate of infection in the mothers of the children with the malformations solely on the basis of recall differences between the two groups. The issue of recall bias can sometimes be evaluated by finding a second source of data to validate the subject's response (e.g., blood test results from prenatal visits or medical records that document symptoms of infection). When the exposure being studied is a prescription drug, pharmacy records may provide a more accurate measure of exposure. Alternatively, the mothers' responses to questions about other exposures may shed light on the presence of a bias affecting the recall of the relevant exposures. Thus, if mothers of cases do not recall greater exposure than controls' mothers to pesticides, children with German measles, and so forth, then one can have greater confidence recall bias is unlikely for the exposure of interest.

Bias may also result from reliance on interviews with surrogates who are individuals other than the study subjects. This is often necessary when, for example, a subject (in a case-control study) has died of the disease under investigation or may be too ill to be interviewed.

There are many sources of information bias that affect the measure of exposure, including its intensity and duration. Exposure to the agent can be measured directly or indirectly. Sometimes researchers use biomarkers as a measure of exposure to an agent—an alteration in tissue or body fluids that occurs as a result of an exposure and that can be detected in the laboratory (see Section V. G., *infra*). Biological markers, however, are only available for a small number of toxins and usually only reveal whether a person was exposed. Biological markers rarely help determine the intensity or duration of exposure.

Monitoring devices also can be used to measure exposure directly but often are not available for exposures that have occurred in the past. For past exposures, epidemiologists often use indirect measures of exposure, such as interviewing workers and reviewing employment records. Thus, all those employed to install asbestos insulation may be treated as

having been exposed to asbestos during the period that they were employed. However, there may be a wide variation of exposure within any job, and these measures may have limited applicability to a given individual. If the agent of interest is a drug, medical or hospital records can be used to determine past exposure. Thus, retrospective studies, which are often used for occupational or environmental investigations, entail measurements of exposure that are usually less accurate than prospective studies or follow-up studies, including ones in which a drug or medical intervention is the independent variable being measured.

The route (e.g., inhalation or absorption), duration, and intensity of exposure are important factors in assessing disease causation. Even with environmental monitoring, the dose measured in the environment generally is not the same as the dose that reaches internal target organs. If the researcher has calculated the internal dose of exposure, the scientific basis for this calculation should be examined for soundness. This topic is addressed at greater length in Section IV. A. 5., *infra*.

In assessing whether the data may reflect inaccurate information, one must determine whether the data were collected from objective and reliable sources. Medical records, government documents, employment records, death certificates, and interviews are examples of data sources that are used by epidemiologists to measure both exposure and disease status. The accuracy of a particular source may affect the validity of a research finding. If different data sources are used to collect information about a study group, differences in the accuracy of those sources may affect the validity of the findings. For example, using employment records to gather information about exposure to narcotics probably would lead to inaccurate results, because employees tend to keep such information private. If the researcher uses an unreliable source of data, the study may not be useful.

The kinds of quality-control procedures used may affect the accuracy of the data. For data collected by interview, quality-control procedures should probe the reliability of the individual and whether the information is verified by other sources. For data collected and analyzed in the laboratory, quality-control procedures should probe the validity and reliability of the laboratory test.

Information bias may also result from inaccurate measurement of disease status. The quality and sophistication of the diagnostic methods used to detect a disease should be assessed. The proportion of subjects who were examined also should be questioned. If, for example, many of the subjects refused to be tested, the fact that the test used was of high quality would be of relatively little value.

The scientific validity of a research finding is influenced by the reliability of the diagnosis of disease or health status under study. The disease must be one that is recognized and defined to enable accurate diagnoses. Thus, for example identification of mesothelioma's being caused by asbestos exposure was impeded by inaccurate death certificates prepared by medical examiners who were unfamiliar with mesothelioma and instead recorded the cause of death as lung cancer. If someone had been conducting a study of the relationship between asbestos exposure and lung cancer, what effect would this information bias have? Subjects' health status may be essential to the hypothesis under investigation. For example, a researcher interested in studying spontaneous abortion in the first trimester must determine that study

subjects are pregnant. Diagnostic criteria that are accepted by the medical community should be used to make the diagnosis. If a diagnosis had been made at a time when home pregnancy kits were known to have a high rate of false positive results (indicating pregnancy when the woman is not pregnant), the study will overestimate the number of spontaneous abortions.

Misclassification bias is a consequence of information bias in which, because of problems with the information available, individuals in the study may be misclassified with regard to exposure status or disease status. Bias due to exposure misclassification can be differential or nondifferential. In nondifferential misclassification, the inaccuracies in determining exposure are independent of disease status, or the inaccuracies in diagnoses are independent of exposure status—in other words, the data are crude, with a great deal of random error. This is a common problem. Generally, nondifferential misclassification bias leads to a shift in the odds ratio toward one, or, in other words, toward a finding of no effect. Thus, if the errors are nondifferential, it is generally misguided to criticize an apparent association between an exposure and disease on the grounds that data were inaccurately classified. Instead, nondifferential misclassification generally underestimates the true size of the association.

Differential misclassification is systematic error in determining exposure in cases as compared with controls or disease status in unexposed cohorts relative to exposed cohorts. In a case-control study this would occur, for example, if, in the process of anguishing over the possible causes of the disease, parents of ill children recalled more exposures to a particular agent than actually occurred, or if parents of the controls, for whom the issue was less emotionally charged, recalled fewer. This can also occur in a cohort study in which, for example, birth control users (the exposed cohort) are monitored more closely for potential side effects, leading to a higher rate of disease identification in that cohort than in the unexposed cohort. Depending on how the misclassification occurs, a differential bias can produce an error in either direction—the exaggeration or understatement of a true association. If researchers have a means to estimate the size and direction of differential misclassification, statistical tools can be employed to determine a more accurate relative risk than the one found originally in the study.

c. Other Conceptual Problems

There are dozens of other potential biases that can occur in observational studies. Sometimes studies are limited by flawed definitions or premises. For example, if the researcher defines the disease of interest as all birth defects, rather than a specific birth defect, there should be a scientific basis to hypothesize that the effects of the agent being investigated could be so broad. If the effect is in fact more limited, the result of this conceptualization error would be to dilute or mask any real effect that the agent might have on a specific type of birth defect. Thus, in *Brock v. Merrell Dow Pharmaceuticals, Inc.*,²⁹ (reproduced below in Section III E. 3.), the court discussed a reanalysis of a study in which the effect was narrowed from all congenital malformations to limb reduction defects. The magnitude of the association increased by 50 percent when the effect was defined in this narrower fashion.

Some biases go beyond errors in individual studies and affect the overall body of

available evidence in a way that skews what appears to be the universe of evidence. Publication bias is the tendency for medical journals to prefer studies that find an effect. If negative studies are never published, the published literature will be biased, a matter discussed further in Section IV. A. 7. d., *infra*. Financial conflicts of interest by researchers and the source of funding of studies have been shown to have an effect on the outcomes of such studies. Clinical trials run by researchers hired by drug companies find greater efficacy of the drug than do studies by independent researchers.¹

Examining a study for potential sources of bias is an important task that helps determine the accuracy of a study's conclusions. In addition, when a source of bias is identified, it may be possible to determine whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, additional considerations remain. As repeatedly noted, an association does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert must consider the possibility of confounding factors. The expert must also evaluate several criteria to determine whether an inference of causation is appropriate. These matters are discussed below.

3. Confounding

The third major reason for error in epidemiologic studies is confounding. Confounding occurs when another causal factor (the confounder) confuses the relationship between the agent of interest and the outcome of interest. Thus, one instance of confounding is when a confounder is both a risk factor for the disease and a factor associated with the exposure of interest. For example, researchers may conduct a study that finds individuals with gray hair have a higher rate of death than those with hair of another color. Instead of hair color having an impact on death, the results might be explained by the confounding factor of age. If old age is associated differentially with the gray-haired group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death. Researchers must separate the relationship between gray hair and risk of death from that of old age and risk of death. When researchers find an association between an agent and a disease, it is critical to determine whether the association might be the result of confounding. Do you understand why clinical trials are much less likely to suffer from confounding?

Confounding can be illustrated by a hypothetical prospective cohort study of the role of alcohol consumption and emphysema. The study is designed to investigate whether drinking

¹ Richard Smith, the editor in chief of the British Medical Journal, wrote on this subject:

The major determinant of whether reviews of passive smoking concluded it was harmful was whether the authors had financial ties with tobacco manufacturers. In the disputed topic of whether third-generation contraceptive pills cause an increase in thromboembolic disease, studies funded by the pharmaceutical industry find that they don't and studies funded by public money find that they do.

alcohol is associated with emphysema. Participants are followed for a period of 20 years and the incidence of emphysema in the “exposed” (participants who consume more than 15 drinks per week) and the unexposed is compared. At the conclusion of the study, the relative risk of emphysema in the drinking group is found to be 2.0, with a p-value of .027, an association that suggests a possible effect. But does this association reflect a true causal relationship or might it be the product of confounding?

One possibility for a confounding factor is smoking, a known causal risk factor for emphysema. If those who drink alcohol are more likely to be smokers than those who do not drink, then smoking may be responsible for some or all of the higher level of emphysema among those who do not drink.

A serious problem in observational studies such as this hypothetical study is that the individuals are not assigned randomly to the groups being compared. As discussed above, randomization maximizes the possibility that exposures other than the one under study are evenly distributed between the exposed and the control cohorts. In observational studies, by contrast, other forces, including self-selection, determine who is exposed to other (possibly causal) factors. The lack of randomization leads to the potential problem of confounding. Thus, for example, the exposed cohort might consist of those who are exposed at work to an agent suspected of being an industrial toxin. The members of this cohort may, however, differ from unexposed controls by residence, socioeconomic or health status, age, or other extraneous factors. These other factors may be causing (or protecting against) the disease, but because of potential confounding, an apparent (yet false) association of the disease with exposure to the agent may appear. Confounders, like smoking in the alcohol drinking study, do not reflect an error made by the investigators; rather, they reflect the inherently “uncontrolled” nature of exposure designations in observational studies.

In designing a study, researchers sometimes make assumptions that cannot be validated or evaluated empirically. Thus, researchers may assume that a missing potential confounder is not needed for the analysis or that a variable used was adequately classified. Researchers employ a sensitivity analysis to assess the effect of those assumptions should they be incorrect. Conducting a sensitivity analysis entails repeating the analysis using different assumptions (e.g., alternative corrections for missing data or for classifying data) to see if the results are sensitive to the varying assumptions. Such analyses can show that the assumptions are not likely to affect the findings or that alternative explanations cannot be ruled out.

Choices in the design of a research project (e.g., methods for selecting the subjects) can prevent or limit confounding. In designing a study, researchers must determine other risk factors for the disease under study. Researchers familiar with the disease and the mechanisms of action are best able to identify these risk factors, which are potential confounders. When a factor or factors, such as age, sex, or even smoking status, are risk factors and potential confounders in a study, investigators can limit the differential distribution of these factors in the study groups by selecting controls to “match” cases (or the exposed group) in terms of these variables. If the two groups are matched, for example, by age, then any association observed in the study cannot be due to age, the matched variable.

Restricting the persons who are permitted as subjects in a study is another method to

control for confounders. If age or sex is suspected as a confounder, then the subjects enrolled in a study can be limited to those of one sex and those who are within a specified age range. When there is no variance among subjects in a study with regard to a potential confounder, confounding as a result of that variable is eliminated.

For remaining potential confounding factors, data should be gathered about the existence of those confounders, which enables statistical analysis as explained below. Unanticipated confounding factors that are suspected after data collection can sometimes be controlled during data analysis, if data have been gathered about them.

If researchers have good data on potential confounders, they can control for those confounders in the data analysis. There are several analytic approaches to account for the distorting effects of a confounder, including stratification and multivariate analysis. Stratification permits an investigator to evaluate the effect of a suspected confounder by subdividing the study groups based on a confounding factor.

To return to the hypothetical emphysema study, to evaluate whether smoking is a confounding factor, the researcher would stratify each of the exposed and control groups into smoking and nonsmoking subgroups to examine whether subjects' smoking status affects the study results. If the relationship between alcohol drinking and emphysema in the smoking subgroups is the same as that in the all-subjects group, smoking is not a confounding factor. If the subjects' smoking status affects the relationship between drinking and emphysema, then smoking is a confounder, for which adjustment is required. If the association between drinking and emphysema completely disappears when the subjects' smoking status is considered, then smoking is a confounder that fully accounts for the association with drinking observed. Table III-6 reveals our hypothetical study's results, with smoking being a confounding factor, which, when accounted for, eliminates the association. Thus, in the full cohort, drinkers have twice the risk of emphysema compared to nondrinkers. When the relationship between drinking and emphysema is examined separately in smokers and in nonsmokers, the risk of emphysema in drinkers compared to nondrinkers is not elevated in smokers or in nonsmokers. This is because smokers are disproportionately drinkers and have a higher rate of emphysema than nonsmokers. Thus, the relationship between drinking and emphysema in the full cohort is distorted by failing to take into account the relationship between being a drinker and a smoker.

Table III-6. Hypothetical Emphysema Study Data

Drinking Status	Total Cohort				Smokers				Nonsmokers			
	Total	Cases	Incidence	RRR	Total	Cases	Incidence	RRR	Total	Cases	Incidence	RR
Nondrinkers	4471	116	.034	11.0*	111	9	0.081	1.0*	360	7	0.019	1.0*
Drinkers	7739	551	.069	22.0	592	48	0.081	1.0	147	3	0.020	1.0

* “RR” in Table 6 is the relative risk. The relative risk for each of the cohorts is determined based on reference to the risk among nondrinkers, that is, the incidence of disease among drinkers is compared with nondrinkers for each of the three cohorts separately.

SOURCE: Courtesy of the authors.

Even after accounting for the effect of smoking, there is always a risk that an undiscovered or unrecognized confounding factor may contribute to a study’s findings, by either magnifying or reducing the observed association. It is, however, necessary to keep that risk in perspective. Often the mere possibility of uncontrolled confounding is used to call into question the results of a study. This was certainly the strategy of some seeking, or unwittingly helping, to undermine the implications of studies persuasively linking cigarette smoking to lung cancer. The critical question is whether it is plausible, based on the best information available, including mechanism evidence, that the findings of a given study could indeed be due to unrecognized confounders.

Because stratification can result in reducing the statistical stability of a study’s results (by requiring more comparisons as in Table 6, each with fewer numbers of subjects), it works best when there are very few potential confounders that require stratification. When there are a number of potential confounders, multivariate analysis³⁰ is often employed. Through statistical methodology that can employ several mathematical models, multivariate analysis allows a researcher to neutralize the effects of confounding and isolate the role of the studied agent on the outcome. When confounding is found to have had some effect, multivariate analysis can provide an “adjusted” outcome that eliminates the influence of confounding.

Both of these methods allow for adjustment of the effect of confounders.³¹ They both modify an observed association to take into account the effect of risk factors that are not the subject of the study and that may distort the association between the exposure being studied and the disease outcomes. If the association between exposure and disease remains after the researcher completes the assessment and adjustment for confounding factors, the researcher must then assess whether an inference of causation is justified. This entails consideration of the Hill factors explained in Section III F. below.

Brock v. Merrell Dow Pharmaceuticals, Inc.
United States Court of Appeals, Fifth Circuit, 1989.
874 F.2d 307.

GARZA, J.

Mr. & Mrs. Floyd Brock filed suit in federal district court on behalf of their minor child, Rachel Brock, to recover damages for birth defects that allegedly resulted from Mrs. Brock's ingestion during her pregnancy of the anti-nausea drug Bendectin, which is manufactured by Merrell-Dow Pharmaceuticals, Inc. ("Merrell-Dow"). The Brocks obtained a jury verdict in the amount of \$550,000 against Merrell-Dow, representing \$240,000 in compensatory damages and \$310,000 in punitive damages. Merrell-Dow appeals that verdict here, arguing that the Brocks did not present sufficient evidence to allow the jury to conclude that Bendectin caused Rachel Brock's birth defect. After reviewing the record and decisions of other courts confronted with similar suits regarding Bendectin, we hold that Merrell-Dow was entitled to judgment notwithstanding the verdict, and the judgment in favor of the Brocks is therefore reversed and the case will be dismissed.

Background

Mrs. Brock conceived Rachel Brock on or around July 2, 1981. On July 28, 1981, Mrs. Brock began to experience morning sickness, and she began to take Bendectin, a prescription drug manufactured by Defendant, Merrell-Dow. Rachel Brock was born on March 19, 1982 with a limb reduction defect known as Poland's Syndrome, which is recognized by a shortening or absence of fingers with a decrease in the corresponding pectoralis muscle on one side.

Mr. and Mrs. Brock filed a diversity suit against Merrell-Dow on behalf of their daughter in the U.S. District Court for the Eastern District of Texas. The complaint alleged theories of improper inspection, design defect, and failure to warn. Causation was a hotly contested issue, with both sides presenting expert testimony and studies regarding the possible teratogenicity of Bendectin. At the end of trial, Merrell-Dow moved for a directed verdict, arguing that there was no credible evidence tending to show that Bendectin causes birth defects. Merrell-Dow's motion was denied, and the issue of whether Bendectin caused Rachel Brock's birth defect was given to the jury. The jury found for the Brocks, and awarded both compensatory and punitive damages. Merrell-Dow then moved for judgment notwithstanding the verdict, and that motion was denied. Merrell-Dow here appeals the denial of its motions for directed verdict and for judgment notwithstanding the verdict.

* * *

* * * Ultimately, the "correctness" of our decision that there was insufficient evidence presented by plaintiff on the issue of whether Bendectin caused Rachel Brock's limb reduction defect to enable a jury to draw a reasonable inference may be just a matter of opinion, but hopefully the reasoning below will persuade others of the insights of our perspective.

* * *

Sufficiency of the Evidence Presented

Undoubtedly, the most useful and conclusive type of evidence in a case such as this is epidemiological studies. Epidemiology attempts to define a relationship between a disease and

a factor suspected of causing it—in this case, ingestion of Bendectin during pregnancy. To define that relationship, the epidemiologist examines the general population, comparing the incidence of the disease among those people exposed to the factor in question to those not exposed. The epidemiologist then uses statistical methods and reasoning to allow her to draw a biological inference between the factor being studied and the disease's etiology.

One difficulty with epidemiologic studies is that often several factors can cause the same disease. Birth defects are known to be caused by mercury, nicotine, alcohol, radiation, and viruses, among other factors. When epidemiologists compare the birth defect rates for women who took Bendectin during pregnancy against those who did not take Bendectin during pregnancy, there is a chance that the distribution of the other causal factors may not be even between the two groups. Usually, the larger the size of the sample, the more likely that random chance will lead to an even distribution of these factors among the two comparison groups, unless there is a dependence between some of the other factors and the factor being studied. For example, there would be a dependence between variables if women who took Bendectin during pregnancy were more or less likely to smoke than women who did not take Bendectin. Another source of error in epidemiological studies is selective recall—i.e., women who have children with birth defects may be more likely to remember taking Bendectin during pregnancy than those women with normal children. Fortunately, we do not have to resolve any of the above questions, since the studies presented to us incorporate the possibility of these factors by use of a *confidence interval*. The purpose of our mentioning these sources of error is to provide some background regarding the importance of confidence intervals.

In this case, the parties described the results of epidemiologic studies in terms of two numbers: a relative risk and a confidence interval. The relative risk is a number which describes the increased or decreased incidence of the disease in question in the population exposed to the factor as compared to the control population not exposed to the factor. In this case, the relative risk describes the increased or decreased incidence of birth defects in the group of women who took Bendectin versus women who did not take Bendectin. A relative risk of 1.0 means that the incidence of birth defects in the two groups were the same. A relative risk greater than 1.0 means that there were more birth defects in the group of women who took Bendectin.

Just because an epidemiological study concludes that a relative risk is greater than 1.0 does not establish that the factor caused the disease. If the confidence interval is so great that it includes the number 1.0, then the study will be said to show no statistically significant association between the factor and the disease. For example, if a study concluded that the relative risk for Bendectin was 1.30, which is consistent with a 30% elevated risk of harm, but the confidence interval was from 0.95 to 1.82, then no statistically significant conclusions could be drawn from this study because the relative risk, when adjusted by the confidence interval, includes 1.0. Again, it is important to remember that the confidence interval attempts to express mathematically the magnitude of possible error, due to the above mentioned sources as well as others, and therefore a study with a relative risk of greater than 1.0 must always be considered in light of its confidence interval before one can draw conclusions from it.

[The court explained that none of the studies relied on by plaintiffs' experts had a confidence interval that excluded the null, i.e., a relative risk of 1.0.]

We find, in this case, the lack of statistically significant epidemiological proof to be fatal to the Brock's case. While we do not hold that epidemiologic proof is a necessary element in all toxic tort cases, it is certainly a very important element. This is especially true when the only

other evidence is in the form of animal studies of questionable applicability to humans. We are not the first court to emphasize the importance of epidemiologic analysis.

[The court proceeded to examine the animal toxicology evidence used by plaintiff's experts and explained why that evidence was inadequate to the task of providing the basis for a reasonable inference of causation.]

Accordingly, the judgment below is REVERSED and RENDERED, and the court below will enter an order of dismissal.

G. Assessing Whether an Association is Causal or Spurious

Throughout these materials, we emphasize that association is not the same as causation. Associations may occur for a number of non-causal reasons. As you should appreciate, the discussion of sources of error above are all potential reasons for an association that is spurious rather than evidencing a real causal relationship. To assess whether an association is causal, epidemiologists have developed a number of guidelines to inform their judgment. These guidelines, known as the "Hill factors" because they were proposed in a published speech by Sir Austin Bradford Hill,¹ provide structure to assessing whether an observed association reflects a causal relation. Application of these considerations in a particular case requires an informed exercise of scientific judgment rather than completion of a mandatory checklist. As Hill explained: "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non." Hill also cautioned his audience that his factors needed to be supplemented with consideration of the role of random chance and small-magnitude associations. Perhaps needless to say, no algorithm exists for toting up the results of consideration of these factors and reaching a conclusion about causation.³²

1. Temporal Relationship

A temporal, or chronological, relationship must exist for causation to exist. If an exposure causes disease, the exposure must occur before the disease develops. If the exposure occurs after the disease develops, it cannot have caused the disease. Although temporal relationship is often listed as one of many Hill factors for consideration in assessing whether an inference of causation is justified, this aspect of a temporal relationship is a necessary factor: Without exposure before the disease, causation cannot exist.

With regard to specific causation, a subject dealt with in detail in Section V. below, there may be circumstances in which a temporal relationship supports or negates the existence of a causal relationship. If the latency period between exposure and outcome is known, then exposure consistent with that information lends credence to a causal relationship. This is

¹ See Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. ROYAL SOC'Y MED. 295 (1965). The year before Hill's speech was published, the U.S. Surgeon General proposed similar guidelines. U.S. Dep't of Health, Educ. & Welfare, Public Health Serv., *Smoking and Health: Report of the Advisory Committee to the Surgeon General* (1964).

particularly true when the latency period is short and competing causes are known and can be ruled out. Thus, if an individual suffers an acute respiratory response shortly after exposure to a suspected agent and other causes of that respiratory problem are known and can be ruled out, the temporal relationship involved supports the conclusion that a causal relationship exists. Similarly, exposure outside a known latency period constitutes evidence, perhaps conclusive evidence, against the existence of causation. Thus, in *In re Swine Flu Immunization Products Liab. Litig.*,³³ the court, sitting as factfinder, found against the plaintiff whose Guillain-Barré disease occurred more than 10 weeks after his swine flu vaccination because epidemiologic studies had found an increased incidence of disease only in those vaccinated within 10 weeks of the onset of disease. On the other hand, when latency periods are lengthy, variable, or not known and a substantial proportion of the disease is due to unknown causes, temporal relationship provides little beyond satisfying the requirement that cause precede effect.

2. The Strength of the Association

The magnitude of the relative risk is one of the cornerstones for causal inferences. The higher the relative risk, the greater the likelihood that the relationship is causal. For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10. That is, the risk of lung cancer in smokers is approximately ten times the risk in nonsmokers.

A relative risk of 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any bias or confounding factor that might account for it. The higher the relative risk, the stronger the association and the lower the chance that the effect is spurious. Although lower relative risks can reflect causality, epidemiologists scrutinize such associations more closely because there is a greater chance that they are the result of uncontrolled confounding or biases.

3. Dose-Response Relationship

A dose-response relationship means that the greater the exposure, the greater the risk of disease. Generally, higher exposures should increase the incidence (or severity) of disease. However, some causal agents do not exhibit a dose-response relationship when, for example, there is a threshold phenomenon (i.e., an exposure may not cause disease until the exposure exceeds a certain dose).¹ Further discussion of dose-response relationships is at Section IV. A. 7. a., *infra*. A dose-response relationship is strong, but not essential, evidence that the relationship between an agent and disease is causal.

¹ Determining whether a no-effect threshold exists is very difficult because at low doses the incidence of disease will typically be very small and thus hard to detect. The question whether there is a no-effect threshold dose is a controversial one in a variety of toxic substances areas. See National Research Council, *Science and Decisions: Advancing Risk Assessment* 8-9, 127- 43 (2009). We return to this issue in the toxicology materials § IV. A. 7. a., *infra*.

4. Consistency of Association

Rarely, if ever, does a single study persuasively demonstrate a cause-effect relationship. Epidemiologists and other scientists prefer that a study be replicated in different populations and by different investigators before coming to a firm causal assessment about an association found in a given study. The need to replicate research findings permeates most fields of science. Consistency in these findings is an important factor in making a judgment about causation. Different studies that examine the same exposure–disease relationship generally should yield similar results. While inconsistent results do not necessarily rule out a causal nexus, any inconsistencies signal a need to explore whether different results can be reconciled with causality.

5. Biologic Plausibility

Biologic plausibility is not an easy criterion to use and depends upon existing knowledge about the mechanisms by which the disease develops. When biological plausibility exists, it lends credence to an inference of causality. For example, the conclusion that high cholesterol is a cause of coronary heart disease is plausible because cholesterol is found in atherosclerotic plaques. However, observations have been made in epidemiologic studies that were not biologically plausible at the time but subsequently were shown to be correct. When an observation is inconsistent with current biological knowledge, it should not be discarded, but the observation should be confirmed before significance is attached to it. The saliency of this factor varies depending on the extent of scientific evidence about the cellular and subcellular mechanisms through which the disease process works. The mechanisms of some diseases are understood quite well based on evidence, including from toxicologic research, while other mechanism explanations are merely hypothesized—although plausible hypotheses are sometimes accepted under this factor.³⁴

Biologic plausibility is always specific to the specific disease process under investigation. Thus, different agent-disease relationships require consideration of what is known about the mechanism involved in the development of that particular disease and whether the agent and its biologic action is consistent with the disease process.

6. Alternative Explanations for the Association

The most prominent alternative explanations for an association, other than a causal relationship, are biases and confounding, which are addressed above. Of course, random error is another possibility that is assessed through significance testing and confidence intervals.

7. Specificity

An association exhibits specificity if the exposure is associated only with a single disease or type of disease. The vast majority of agents do not cause a wide variety of effects. For

example, asbestos causes mesothelioma and lung cancer and may cause one or two other cancers, but there is no evidence that it causes any other types of cancers. Thus, a study that finds that an agent is associated with many different diseases should be examined skeptically. Nevertheless, there may be causal relationships in which this guideline is not satisfied. Cigarette smoking causes a wide variety of harms, including lung cancer, emphysema, bladder cancer, heart disease, pancreatic cancer. This may be due, at least in part, to the fact that cigarette smoke has many components and thus cigarette smokers are exposed to numerous toxic agents, with multiple possible effects. Thus, while evidence of specificity may strengthen the case for causation, lack of specificity does not necessarily undermine it where there is a good biological explanation for its absence.

8. Consistency with Other Information/Ceasing Exposure

If an agent is a cause of a disease one would expect that cessation of exposure to that agent ordinarily would reduce the risk of the disease. This was the case, for example, with the drug Bendectin, a suspected teratogen. In 1981, use dropped precipitously because of concerns about its teratogenicity, and by 1983 the manufacturer withdrew the drug from the market. In the Bendectin multi-district litigation, Steven Lamm, a neutral expert epidemiologic consultant, prepared a preliminary version of the study abstracted and excerpted below, comparing the withdrawal of Bendectin with the incidence rates of various birth defects. Virtually no correlation was revealed. The federal judge overseeing the multi-district litigation remarked that “[t]he most telling single piece of evidence I have ever seen after 23 years on the Federal Bench is the [Lamm] exhibit.”³⁵

In many situations, however, relevant data are simply not available regarding the possible effects of ending the exposure. But when such data are available and eliminating exposure reduces the incidence of disease, this factor supports a causal relationship, just as the independence of disease rates and exposure supports a lack of causation.

Bendectin and Birth Defects II: Ecological Analyses

Jeffrey S. Kutcher,¹ Arnold Engle,² Jacqueline Firth,³ and Steven H. Lamm^{2*}

¹University of Michigan, Department of Neurology, Ann Arbor, Michigan 48109

²Consultants in Epidemiology and Occupational Health, Inc., Washington, DC 20007

³Tulane University School of Medicine, New Orleans, Louisiana 70112

Received 18 May 2001; Accepted 5 December 2001

BACKGROUND: Bendectin was the primary pharmaceutical treatment of nausea and vomiting of pregnancy (NVP) in the United States until the early 1980s. Its manufacture was then discontinued after public allegations that it was causing birth defects. Subsequently, meta-analyses of the many epidemiological cohort and case/control studies used to examine that hypothesis have demonstrated the absence of a detectable teratogenic effect. This study presents an ecological analysis of the same hypothesis that examines specific malformations. **METHODS:** Annual birth defect prevalence data for the 1970s to the 1990s have been obtained for specific birth defects from the Center for Disease Control's nationwide Birth Defect Monitoring Program. These data for the US have been compared graphically to the annual US Bendectin sales for the treatment of NVP. Data have also been obtained for annual US rates for hospitalization for NVP. The three data sets have been temporally compared in graphic analysis. **RESULTS:** The temporal trends in prevalence rates for specific birth defects examined from 1970 through 1992 did not show changes that reflected the cessation of Bendectin use over the 1980–84 period. Further, the NVP hospitalization rate doubled when Bendectin use ceased. **CONCLUSIONS:** The population results of the ecological analyses complement the person-specific results of the epidemiological analyses in finding no evidence of a teratogenic effect from the use of Bendectin. *Birth Defects Research (Part A)* 67:88–97, 2003. © 2003 Wiley-Liss, Inc.

Limb Malformations

Figure demonstrates that, of the limb malformations analyzed, only clubfoot without CNS defects showed a decrease (2%) in annual incidence between 1970–92. Polydactyly (25%), syndactyly (45%), and limb reduction (18%) each showed a notable increase. Hip dislocation and subluxation showed a jagged increase in between 1974–79 that declined sharply before the drop in Bendectin sales and began to climb again after the cessation of its use; overall hip dislocation and subluxation showed a dramatic 291% increase between 1970–92. None of these trends indicate a pattern similar to Bendectin sales.

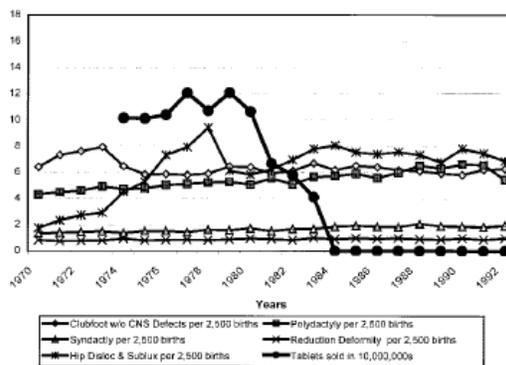


Figure 7. Incidence of limb malformations (BDMP) and Bendectin tablet sales in the United States, 1970–92.

H. Multiple Studies and Meta-analysis

Not infrequently, the scientific record may include a number of epidemiologic studies whose findings differ. These may be studies in which one or more finds an association and others do not, or studies that report associations, but of different magnitude. In view of the fact that epidemiologic studies may disagree and that sometimes studies are small and lack the statistical power needed to give confidence about the statistical stability of the results, the technique of meta-analysis was developed. Before explaining meta-analysis, we pause to explain that traditionally multiple studies of the same causal relationship were assessed qualitatively in a review article. The author would review all studies, make an assessment of their strengths and weaknesses, and provide a synthesis of what the studies found.

Abstract from Chengqian Li et al., *Main air pollutants and diabetes-associated mortality: a systematic review and meta-analysis*, 171 *Mechanisms in Endocrinology* 183 (2014). The abstract is Copyright © 2014, Bioscientifica, Ltd.

Abstract

Objective: Exposure to high levels of air pollutants may be linked to diabetes-associated mortality, but the associations remain unclear. To assess the associations between main air pollutants and diabetes-associated mortality, a systematic review and meta-analysis was performed.

Methods: PubMed, Embase and Web of Science were searched for studies investigating the associations between increments in gaseous (nitrogen dioxide (NO₂), sulphur dioxide, ozone (O₃) and carbon monoxide) and particulate matter (PM; diameter <2.5 μm (PM_{2.5}) or <10 μm (PM₁₀)) air pollutants and diabetes-associated mortality. Using a random-effects model, relative risks (RRs) and 95% CIs were calculated per interquartile range (IQR) increment or per 10 μg/m³ increment in pollutant concentrations.

Results: Out of 925 identified articles, 36 were reviewed in depth and 12 studies from 13 articles satisfying the inclusion criteria (five time-series, five case-crossovers and two cohorts) were finally included. Increased risk of diabetes-associated mortality was associated with higher levels of PM_{2.5} (per 10 μg/m³: RR=1.123, 95% CI 1.036–1.217, *P*=0.005, *I*²=96.1%), PM₁₀ (per 10 μg/m³: RR=1.008, 95% CI 1.004–1.013, *P*<0.001, *I*²=0%), NO₂ (per 10 μg/m³: RR=1.024, 95% CI 1.007–1.041, *P*=0.006, *I*²=49.7%) and O₃ (per IQR increment: RR=1.065, 95% CI 1.017–1.115, *P*=0.007, *I*²=0.0%). No obvious risk of publication bias was observed.

Conclusions: Exposure to high levels of air pollutants is significantly associated with an increased risk of diabetes-associated mortality.

Notes and Questions

1. Li et al. conducted a meta-analysis of 12 studies including 5 time-series studies, 5 case-crossover studies, and 2 cohort studies. Time-series studies are cohort studies during which data points are measured over a specific time interval. Case-crossovers are retrospective studies where an individual acts as his own control for the purpose of comparison in order to decrease confounding. Cases are identified and then their exposure status prior to the time when they became a case is assessed. This is then compared to the individual's own previous exposure at a time when he did not become a case.

Meta-analysis is a quantitative method of pooling study results to arrive at a single figure to represent the totality of the studies reviewed. It is a way of systematizing the time-honored approach of reviewing the literature and placing it in a standardized framework with quantitative methods for estimating risk. In a meta-analysis, studies are given different weights in proportion to the sizes of their study populations and other characteristics.

Meta-analysis is most appropriate when used in pooling randomized clinical trials, because the studies included in the meta-analysis share the most significant methodological characteristics, in particular, use of randomized assignment of subjects to different exposure groups. However, often one is confronted with non-randomized observational studies of the effects of possible toxic substances or agents. A method for summarizing such studies is greatly needed, but when meta-analysis is applied to observational studies—either case-control or cohort—it becomes more difficult because of methodological differences among studies. Hence, one must approach the results of these meta-analyses with skepticism and caution.

Among the problems and issues that arise in conducting and understanding meta-analyses are:

- Should only published papers be included in the meta-analysis, or should any available studies be used, even if they have not been peer reviewed?
- How can the problem of differences in the quality of the studies reviewed be taken into account?
- Can the results of the meta-analysis itself be reproduced by other analysts?
- When there are several meta-analyses of a given relationship, why do the results of different meta-analyses often disagree?

Another consideration is that often the differences among the individual studies included in a meta-analysis and the reasons for the differences are important in themselves and need to be understood; however, those matters are often masked in a meta-analysis. A final problem with meta-analyses is that they generate a single estimate of risk and may lead to a false sense of security regarding the certainty of the estimate. People often tend to have an inordinate belief in the validity of the findings when a single number is attached to them, and many of the difficulties that may arise in conducting a meta-analysis, especially of observational

studies like epidemiologic ones, may consequently be overlooked.

I. Reading an Epidemiologic Study

To conclude our treatment of epidemiology, we include a study of firefighters and mortality.¹ This case affords an opportunity to consolidate our understanding as consumers of studies, the role in which lawyers and judges typically confront such work. After reading and analyzing the study, we include a court opinion in which the New Jersey Supreme Court read and relied on the study in the course of deciding a workers compensation case in which causation was at issue.

¹ Paul A. Demers et al., Mortality Among Firefighters from Three Northwestern United States Cities, 49(9) BRITISH J. OF INDUS. MED. 664-670 (1992). Reproduced with permission.

Mortality among firefighters from three northwestern United States cities

Paul A. Demers, Nicholas J Heyer, Linda Rosenstock

Abstract

To explore whether exposure among fire-fighters to fire smoke could lead to an increased risk of cancer, lung disease, and heart disease, the mortality of 4546 firefighters who were employed by the cities of Seattle and Tacoma, WA and Portland, OR for at least one year between 1944 and 1979 were compared with United States national mortalities and with mortality of police officers from the same cities. Between 1945 and 1989, 1169 deaths occurred in the study population and 1162 death certificates (99%) were collected. Mortality due to all causes, ischemic heart disease, and most other non-malignant diseases was less than expected based upon United States rates for white men. There was no excess risk of overall mortality from cancer but excesses of brain tumours (standardised mortality ratio (SMR) = 2.09, 95% confidence interval (95% CI) 1.3-3.2) and lymphatic and haematopoietic cancers (SMR = 1.31, 95% CI = 0.9-1.8) were found. Younger firefighters (<40 years of age) appeared to have an excess risk of cancer (SMR = 1.45, 95% CI 0.8-2.39), primarily due to brain cancer (SMR = 3.75, 95% CI 1.2-8.7). The risk of lymphatic and haematopoietic cancers was greatest for men with at least 30 years of exposed employment (SMR = 2.05, 95% CI 1.1-3.6), especially for leukaemia (SMR = 2.60, 95% CI 1.0-5.4).

Department of Environmental Health

PA Demers, NJ Heyer, L Rosenstock

Department of Epidemiology

PA Demers

Department of Medicine, University of Washington, Seattle, Washington, USA

L Rosenstock

Since the end of the second world war the use of synthetic materials for both the structures and interiors of buildings has increased the complexity and toxicity of the smoke generated when these buildings catch fire.¹⁻² The potential exposure to suspected or known carcinogens has raised the concern that firefighters may be at excess risk of cancer. Benzene and polycyclic aromatic hydro-carbons are likely encountered at most fires and other, less common, exposures may include asbestos, aromatic amines, chlorinated dioxins, and other potential carcinogens.³⁻⁸ Excesses of brain cancer, cancers of the colon or rectum, malignant melanoma or skin cancer, bladder cancer, leukaemia, and multiple myeloma have been found,⁹⁻¹⁵ although the results have been far from consistent. Perhaps surprisingly, given a priori suspicions, only one cohort study has noted an excess of lung cancer in firefighters.¹⁶

It is plausible that firefighters could also be at excess risk of death due to heart and respiratory disease. Many respiratory irritants, such as hydrogen chloride, nitrogen dioxide, isocyanates, and acrolein, are commonly present in smoke.^{3-5 8} Evidence exists for respiratory dysfunction after acute high exposures¹⁷⁻¹⁹ although studies designed to look at chronic effects have produced mixed results.²⁰⁻²³ An increased risk of cardiovascular disease due to intense physical and psychological stress after periods of inactivity or exposure to carbon monoxide and other toxic gases is also plausible.²⁴⁻²⁵ Most cohort mortality studies, however, have found firefighters to be at the same or lower risk than the general population for both heart and lung disease.

Death rates for the general population have been used as the reference in most mortality studies of occupational cohorts. A major bias introduced by using general population rates has been termed the healthy worker effect.²⁶⁻²⁸ In many ways firefighters, with their strict physical entry requirements and good employment benefits, typify a population in which a particularly strong healthy worker effect would be expected. This may in part account for the low risk of death due to heart and respiratory disease noted in these studies; however, when police, an occupation with similar entrance criteria, have been used as a reference population^{14 29} evidence that firefighters are at increased risk of respiratory disease has been found. Also, a previously reported study of a sub-population of this same cohort found that the risk of heart disease increased with duration of employment.³⁰

In 1984 we began a retrospective cohort study of Seattle firefighters to explore the relation between exposure to fire smoke and mortality.³⁰ Later we expanded the study to include two other major cities in the region and to collect data on police from the same cities as a comparison group. This is a report of the results of the expanded mortality study with follow up to the end of 1989.

Methods

The study population consists of all men who were employed as firefighters for at least one year between 1944 and 1979 by the cities of Seattle and Tacoma, Washington and Portland, Oregon. Women were excluded from the study because they first began employment as firefighters in the 1970s and their numbers continue to be comparatively small. Years of active duty in positions involving fire combat was used as a surrogate measure of exposure to smoke. Records of the Seattle and Portland fire departments were reviewed and no time for exposure to fire smoke was assigned for years spent in administration, fire prevention, or support services. Because Tacoma lacked the necessary records to make this distinction, fire smoke exposure time was assigned for all years of firefighter employment. A cohort of police from the same cities was also identified for use as a comparison group.

The follow up period was from 1 January 1945 to 31 December 1989. Follow up for vital status and collection of death certificates were performed for both the firefighter and police cohorts using information from pension board and department records, the death records of Washington and Oregon, the records of the Washington and Oregon motor vehicle departments, and the National Death Index. Those who were lost to follow up were only considered at risk until the date on which they were last known to be alive. Persons lost to follow up subsequent to 1978 were assumed to be alive if no death was identified through the National Death Index. Underlying cause of death was coded by a former Washington state nosologist after information identifying the deceased person as either a former firefighter or police officer was removed from the death certificate.

Standardised mortality ratios (SMRs) compared with United States white men were calculated using the microcomputer version of the Occupational Mortality Analysis Program.³¹ Reference rates for United States white men were obtained from the National Institute for Occupational Safety and Health. White male rates were used because most firefighters from the cities studied were Caucasian and department records did not include information on race. Confidence intervals were calculated using a Poisson distribution. Incidence density ratios (IDRs) and 95% confidence intervals (95% CIs) for firefighters relative to police were calculated using Mantel-Haenszel methods with standardisation by five year age groups and time periods and test based confidence intervals.³² Mortality was examined in stratified analyses by years of fire combat exposure, years since first employment as a firefighter, and age at risk.

Results

Complete follow up was achieved for 98% of the 4401 firefighters (table 1). Between 1945 and 1989, 1169 deaths occurred and 1162 death certificates (99%) were collected. The comparison cohort consisted of 3676 police officers and complete follow up information was attained for 3599 (98%). During the follow up period 714 police deaths were identified and 703 death certificates (98%) were collected.

The risk of death due to any cause among firefighters was less than expected (SMR = 0.81, 95% CI 0.77--0.86) due to a lower than expected risk of most types of non-malignant diseases (table 2). Twofold excess of brain tumours was seen (SMR = 2.09, 95% CI 1.31-3.17). The death certificates listed seven of the tumours as glioblastoma multiforme, three as astrocytoma, three as other gliomas, five as other or unspecified malignant brain tumours, and four as unspecified brain tumours. Smaller excesses were found for cancers of the lymphatic and haematopoietic tissues (SMR = 1.31, 95% CI 0.92-1.81)

and prostate (SMR = 1.34, 95% CI 0.90-1.91). The number of observed cases of most other cancers, including lung cancer, was similar to expected with only cancers of the bladder (SMR = 0.23, 95% CI 0.03--0.83) and kidney (SMR = 0.27, 95% CI 0.03--0.97) significantly lower than expected

Table 1 Employment and vital status and years of follow up at 1 January 1990

Status	Seattle	Portland	Tacoma	Total	(%)
Currently employed	610	458	217	1285	(29)
Retired	782	396	239	1417	(32)
Other alive	318	95	22	435	(10)
Deceased	516	509	144	1169	(27)
Certificates collected	510	508	144	1162	(99)*
Unknown status	55	24	16	95	(2)
Total	2281	1482	638	4401	
Years of follow up	64388	41085	17379	122852	

*Per cent of death certificates collected.

Table 2 Seattle, Portland, and Tacoma firefighter mortality: 1945-89

Cause of death (ICD 9 codes)	Deaths	SMR	(95%)
All causes (001-999)	1169	0.81	(0.77 - 0.86)
All cancers (140-152.2, 156.9-165.9, 170-175, 179-208)	291	0.95	(0.85 - 1.07)
Oral and pharyngeal cancers (140-149)	7	0.81	(0.33 - 1.66)
Oesophageal cancer (150)	6	0.83	(0.30 - 1.80)
Stomach cancer (151)	16	1.07	(0.61 - 1.73)
Colon cancer (152, 153)	24	0.85	(0.54 - 1.26)
Rectal cancer (154)	8	0.95	(0.41 - 1.87)
Biliary passages and liver cancer (155.0 - 155.1, 156)	6	1.19	(0.44 - 2.59)
Pancreatic cancer (157)	14	0.89	(0.49 - 1.49)
Laryngeal cancer (161)	2	0.47	(0.06 - 1.70)
Lung cancer (162)	95	0.96	(0.77 - 1.17)
Prostate cancer (185)	30	1.34	(0.90 - 1.91)
Kidney cancer (189.0 - 189.2)	2	0.27	(0.03 - 0.97)
Bladder and other urinary cancers (188, 189.3-189.9)	2	0.23	(0.03 - 0.83)
Skin cancer (172, 173)	6	0.98	(0.36 - 2.13)
Brain/nervous system tumours (191, 192, 237.5-237.9, 239.6-239.7)	22	2.09	(1.31 - 3.17)
Brain and nervous system cancers (191, 192)	18	2.07	(1.23 - 3.28)
Unspecified nervous system tumours (237.5-237.9, 239.6-239.7)	4	2.20	(0.60 - 5.62)
Lymphatic/haematopoietic cancers (200-208)	37	1.31	(0.92 - 1.81)
Lymphosarcoma and reticulosarcoma (200)	7	1.42	(0.57 - 2.93)
Hodgkin's disease (201)	3	1.05	(0.22 - 3.08)
Leukaemia(204-208)	15	1.27	(0.71 - 2.09)
Other lymphatic/haematopoietic (202, 203)	12	1.40	(0.72 - 2.44)
Heart disease (390-398, 402, 404, 410-414, 420-429)	461	0.79	(0.72 - 0.87)
Ischaemic heart disease (410-414)	394	0.82	(0.74 - 0.90)
Other circulatory disease (401, 403, 405, 415-417, 430-438, 440-459)	131	0.96	(0.80 - 1.14)
Cerebrovascular disease (430-438)	79	0.85	(0.67 - 1.06)
Diseases of arteries, veins and pulmonary circulation (415-417, 440-459)	48	1.24	(0.91 - 1.64)
Respiratory disease (460-466, 470-478, 480-487, 490-519)	81	0.89	(0.71 - 1.10)
Acute upper respiratory infection (460-466)	2	3.57	(0.43 - 12.9)
Pneumonia (480-486)	22	0.67	(0.42 - 1.01)
Chronic respiratory diseases (470-478, 490-519)	56	1.00	(0.76 - 1.30)
Emphysema (492)	20	1.19	(0.72 - 1.83)
Asthma (493)	3	1.05	(0.22 - 3.08)
COPD and other respiratory disease (470-478, 494-519)	32	0.98	(0.67 - 1.38)

COPD = Chronic obstructive pulmonary disease.

The risks for death due to heart and circulatory disease were similar to or lower than expected with the exception of diseases of the arteries, veins, and pulmonary circulation, which were somewhat increased (SMR = 1.24, 95% CI 0.91-1.64).

Table 3 presents firefighter mortality relative to that of police and police mortality relative to that of United States white men for causes of death of a priori interest and those found to be in excess as shown in table 2. Hodgkin's

disease, asthma, and acute respiratory infections were not included in the table because no deaths due to these causes were found among police. Although the confidence limits were wide, firefighters appear to have a higher risk than police of colon cancer, prostate cancer, brain tumours, "other" lymphatic and haematopoietic cancers, and emphysema. The category of "other" lymphatic and haematopoietic cancer includes multiple myeloma (seven out of 12 firefighter and two out of five police deaths were in this

category). Although national rates for the study period were not available, the risk of multiple myeloma for firefighters relative to police was 1.91 (95% CI 0.4-8.4). Of the brain tumours among police, five were listed on the death certificates as glioblastoma multiforme, two as astrocytomas, and one as a malignant neuroblastoma.

Firefighters were at somewhat lower risk than police for deaths due to all causes and circulatory disease and at much lower risk of bladder cancer.

The causes of death that were found to be in excess were further analysed by duration of exposed employment (table 4). The risks for lymphatic and haematopoietic cancer, especially leukaemia, and diseases of the arteries, veins, and pulmonary circulation were highest for firefighters with at least 30 years of exposure, although the risks do not increase consistently with duration of exposed employment. The risk of leukaemia in firefighters with 30 years of exposed employment remained increased (IDR = 1.80, 95% CI 0.6-5.4) when comparisons were made with police, whereas the risk of all lymphatic and haematopoietic cancers did not (IDR = 1.14, 95% CI 0.5-2.6). The risk of mortality from all chronic respiratory disease peaked among firefighters with 20 to 29 years of exposure; the excess risk of emphysema was highest among those with 10 to 19 years of exposure.

Table 3 Seattle, Portland, and Tacoma firefighter mortality compared with police and police mortality compared with United States white male rates: 1945-89

Cause of death	Firefighters v police			Police v United States white men		
	Deaths	IDR	(95% CI)	Deaths	SMR	(95% CI)
All causes	1169	0.87	(0.79-0.95)	714	0.87	(0.81-0.93)
All cancers	291	0.97	(0.80-1.17)	169	0.95	(0.81-1.11)
Colon cancer	24	1.58	(0.73-3.43)	8	0.50	(0.22-0.99)
Rectal cancer	8	0.89	(0.30-2.66)	5	1.11	(0.36-2.59)
Biliary passages and liver cancer	6	0.71	(0.19-2.71)	4	1.40	(0.38-3.59)
Trachea, bronchus, and lung cancer	95	0.95	(0.67-1.33)	55	0.92	(0.69-1.19)
Prostate cancer	30	1.43	(0.71-2.85)	11	1.02	(0.51-1.82)
Bladder cancer	2	0.16	(0.02-1.24)	4	0.91	(0.25-2.34)
Skin cancer	6	1.12	(0.27-4.76)	4	0.94	(0.26-2.41)
Brain and nervous system tumours	22	1.88	(0.82-4.31)	8	1.14	(0.49-2.25)
Brain and nervous system cancer	18	1.63	(0.70-3.79)	8	1.36	(0.59-2.69)
Lymphatic/haematopoietic cancers	37	1.03	(0.62-1.73)	21	1.22	(0.75-1.86)
Lymphosarcoma and reticulosarcoma	7	0.81	(0.30-2.22)	5	1.72	(0.56-4.02)
Leukaemia	15	0.80	(0.38-1.70)	11	1.56	(0.78-2.80)
Other lymphatic/haematopoietic	12	1.40	(0.48-4.07)	5	0.93	(0.30-2.17)
Heart diseases	461	0.86	(0.74-1.00)	269	0.85	(0.75-0.96)
Ischaemic heart disease	394	0.88	(0.74-1.04)	223	0.86	(0.75-0.98)
Other circulatory disease	131	0.72	(0.54-0.96)	86	1.25	(1.00-1.55)
Cerebrovascular disease	79	0.65	(0.45-0.92)	59	1.28	(0.98-1.65)
Diseases of the arteries, veins, and pulmonary circulation	48	0.91	(0.54-1.52)	25	1.24	(0.70-2.04)
Respiratory disease	81	1.11	(0.71-1.73)	30	0.64	(0.43-0.91)
Pneumonia	22	1.04	(0.46-2.36)	10	0.60	(0.29-1.11)
Chronic respiratory diseases	56	1.11	(0.65-1.89)	20	0.68	(0.42-1.06)
Emphysema	20	1.45	(0.54-3.88)	5	0.63	(0.20-1.46)
COPD and miscellaneous lung disease	32	0.89	(0.47-1.69)	15	0.83	(0.47-1.37)

of the arteries, veins, and pulmonary circulation (SMR = 2.55, 95% CI 1.43-3.38), and colon cancer (SMR = 1.69, 95% CI 0.77-3.20). Lagging also further accentuated the risks for emphysema among firefighters with 20 to 29 years of exposed employment (SMR = 1.49, 95% CI 0.82-56).

Firefighters with at least 30 years since their first employment had increased risks for brain tumours (SMR = 2.63), lymphatic and haematopoietic malignancies (SMR = 1.48), prostate cancer (SMR = 1.42), diseases of the arteries, veins, and pulmonary circulation (SMR = 1.33), and emphysema (SMR = 1.39) (table 5). These firefighters also had an increased risk for brain tumours (IDR = 3.62, 95% CI 1.2-11.2), prostate cancer (IDR = 1.58, 95% CI 0.8-3.2), and emphysema (IDR = 1.48, 95% CI 0.6-3.9) compared with police.

In general, the risk for mortality from most causes was highest among firefighters 65 years of age or older (table 6). Firefighters under the age of 40, however, had an

SMR for all cancers of 145 (95% CI 0.81-2.39) due primarily to a greater than expected number of brain tumours (SMR = 3.75) and lymphatic and haematopoietic malignancies (SMR = 1.74). The excess observed for cancer is by contrast with the deficits found for all non-cancer causes of death (SMR = 0.47). The excess of cancer among firefighters under the age of 40 persisted when the comparison was made with police (IDR = 1.51, 95% CI 0.7-3.5).

Discussion

We found an excess of brain tumours among firefighters compared with United States white men and police. Previous studies of workers exposed to vinyl chloride, acrylonitrile, and polycyclic aromatic hydrocarbons have noted excesses of brain cancer.³⁴

Although it is difficult to quantify, it is likely that

Table 4 Seattle, Portland, and Tacoma firefighter mortality by duration of exposed employment: 1945-89

Cause of death	<10 years			10-19 years			20-29 years			>30 years		
	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)
Colon cancer	2	0.54	(0.1-2.0)	9	0.62	(0.3-1.2)	9	1.21	(0.6-2.3)	4	1.40	(0.4-3.6)
Prostate cancer	3	2.42	(0.5-7.1)	2	1.12	(0.1-4.1)	14	1.23	(0.7-2.1)	11	1.36	(0.7-2.4)
Brain and nervous system tumours	5	2.57	(0.8-6.0)	8	3.53	(1.5-7.0)	6	1.24	(0.5-2.7)	3	2.04	(0.4-5.9)
Lymphatic/haematopoietic cancers	4	0.91	(0.2-2.3)	7	1.46	(0.06-3.0)	14	1.06	(0.6-1.8)	12	2.05	(1.1-3.6)
Leukaemia	2	1.13	(0.1-4.1)	2	1.04	(0.1-3.7)	4	0.73	(0.2-1.9)	7	2.60	(1.0-5.4)
Diseases of the arteries, veins, and pulmonary circulation	4	1.36	(0.4-3.5)	4	0.94	(0.3-2.4)	15	0.79	(0.4-1.3)	25	1.99	(1.3-2.9)
Chronic respiratory diseases	2	0.42	(0.1-1.5)	5	0.82	(0.3-1.9)	34	1.15	(0.8-1.6)	15	0.97	(0.5-1.6)
Emphysema	1	0.92	(0.1-5.1)	3	1.83	(0.4-5.3)	12	1.35	(0.7-2.4)	4	0.76	(0.2-1.9)

Table 5 Seattle, Portland, and Tacoma firefighter mortality by years since first employment: 1945-89

Cause of Death	<20 years			20-29 years			>30 years		
	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)
Colon cancer	1	0.51	(0.1-2.9)	3	0.66	(0.1-1.9)	20	0.91	(0.6-1.4)
Prostate cancer	0	0.00	(0.0-26.6)	0	0.00	(0.0-3.1)	30	1.42	(1.0-2.0)
Brain and nervous system tumours	6	2.45	(0.9-5.3)	2	0.73	(0.1-2.6)	14	2.63	(1.4-4.4)
Lymphatic/haematopoietic cancers	8	1.65	(0.7-3.2)	2	0.39	(0.1-1.4)	27	1.48	(1.0-2.2)
Leukaemia	3	1.50	(0.3-4.4)	1	0.50	(0.1-2.8)	11	1.40	(0.7-2.5)
Diseases of the arteries, veins, and pulmonary circulation	1	0.51	(0.1-2.8)	4	0.91	(0.2-2.3)	43	1.33	(1.0-1.8)
Chronic respiratory diseases	1	0.45	(0.1-2.5)	2	0.32	(0.1-1.1)	53	1.12	(0.8-1.5)
Emphysema	0	0.00	(0.0-7.9)	0	0.00	(0.0-1.8)	20	1.39	(0.9-2.2)

exposure to polycyclic aromatic hydrocarbons at fires is common whereas exposure to vinyl chloride and acrylonitrile may happen only under certain conditions. If the excess of brain cancer were due to exposures that were not necessarily present at most fires, this might at least in part explain our finding that duration of exposed employment was not associated with increasing risk.

We also found an excess risk of leukaemia, which was highest among persons employed 30 or more years in fire combat positions, confirming our earlier finding of an increased risk among Seattle firefighters.³⁰ A twofold excess of multiple myeloma relative to police was also found. Other studies have noted an excess of lymphatic and haematopoietic cancers of various histologies^{11 12 14 15} and an excess of these malignancies is plausible given the exposure of firefighters to benzene.³⁻⁵ Although exposure to benzene is likely to be short term, measurements have been taken in excess of 100ppm.^{3,4} Our ability to conclude with certainty an association with exposure to fire smoke is limited by our finding of a similar excess in police. We are unable to assess whether the excess among police is due to factors held in common between the two occupational groups, to some exposure unique to police, or to chance. Of interest, two other studies that have examined cancer in firefighters v police found firefighters to be at higher risk for leukaemia.^{12 14}

We also found an excess of prostate cancer, an effect of uncertain significance and not seen in other studies.

The persistence of this excess compared with police makes a diagnostic bias an unlikely explanation. We did not find excess cancers of the skin, bladder, or lung, which have been noted in some other studies of firefighters. Limited support was found for previously described excesses of colon cancer when the comparison was made with police, but not with the United States reference group. The inconsistency with previous studies may be due to the small number of deaths found for some sites or to the different methods used and varying time periods examined.

As anticipated, many of the results of this study are consistent with the healthy worker effect. One exception was deaths due to diseases of the arteries, veins, and pulmonary circulation, which were increased among firefighters with at least 30 years of exposed employment compared with both United States white men and police. This result is difficult to interpret given the heterogeneous nature of conditions in this category.

In analyses of this cohort performed with follow up through 1983 we found an excess of non-malignant respiratory disease compared with police (IDR = 1.59), as opposed to a deficit when compared with United States rates (SMR = 0.88).²⁹ One other study that compared deaths from lung disease in firefighters with those for police officers found a similar

result.¹⁴ Although this disparity was also found in the current analysis, the magnitude of the effect was much reduced. This

may be in part accounted for by the increasing availability

Table 6 Seattle, Portland, and Tacoma firefighter mortality by age at risk: 1945-89

Cause of Death	<18-39 years old			40-64 years old			>65 years old		
	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)
Colon cancer	1	0.51	(0.1-2.9)	3	0.66	(0.1-1.9)	20	0.91	(0.6-1.4)
Prostate cancer	0	0.00	(0.0-26.6)	0	0.00	(0.0-3.1)	30	1.42	(1.0-2.0)
Brain and nervous system tumours	6	2.45	(0.9-5.3)	2	0.73	(0.1-2.6)	14	2.63	(1.4-4.4)
Lymphatic/haematopoietic cancers	8	1.65	(0.7-3.2)	2	0.39	(0.1-1.4)	27	1.48	(1.0-2.2)
Leukaemia	3	1.50	(0.3-4.4)	1	0.50	(0.1-2.8)	11	1.40	(0.7-2.5)
Diseases of the arteries, veins, and pulmonary circulation	1	0.51	(0.1-2.8)	4	0.91	(0.2-2.3)	43	1.33	(1.0-1.8)
Chronic respiratory diseases	1	0.45	(0.1-2.5)	2	0.32	(0.1-1.1)	53	1.12	(0.8-1.5)
Emphysema	0	0.00	(0.0-7.9)	0	0.00	(0.0-1.8)	20	1.39	(0.9-2.2)

and use of respiratory protection since the 1970s. Also, the risk of death due to non-malignant respiratory disease among police was higher in the current (SMR = 0.64) than in the earlier analysis (SMR = 0.48).

Nonetheless, a raised risk of emphysema was found among firefighters compared with both United States white men and police. All of these deaths occurred among subjects at least 30 years after first employment and was highest among those with 10 to 29 years of exposed employment. If a relation does exist between exposure to firesmoke and emphysema, the fact that the risk was reduced among firefighters with 30 or more years of exposed employment might be due to those most susceptible to disease leaving employment early due to disability. Attempts to draw conclusions should be tempered by the fact that the specificity of death certificates is low for differentiating between different types of obstructive lung diseases. Whereas the results for all chronic respiratory diseases combined roughly parallel those for emphysema, the risks were of lesser magnitude.

Some limitations should be borne in mind when interpreting the results of this study. Firstly, duration of fire combat employment, although an improvement over total duration of employment, may still be an inadequate measure of exposure, particularly for substances that may not be present at all fires. Exposure may vary substantially between and within fires due to the composition of the material being burned, the temperature of the fire, and availability of oxygen.^{3 5} Thus the lack of association seen between duration or fire combat employment and various outcomes in this study may in part be due to the use of a poor surrogate for exposure.

Another limitation of this study is the lack of accuracy and specificity of information on cause of death on death certificates. In the case of heart and lung disease it may be difficult to assign a specific cause of death without a postmortem examination. Information about cancer on death certificates usually lacks detail and only rarely includes anatomical subsite or histological information. To the

extent that a cause of death category contains a wide range of etiologically unrelated diseases, the relation between the exposure and any one specific disease will be obscured.

Police were chosen as an alternative reference population because they have a similar socioeconomic state, health benefits, and strict physical injury requirements, and are generally free from any major fire smoke inhalation. Two studies of smoking habits by occupation show that police and firefighters are similar,^{35 36} although a somewhat greater percentage of firefighters reported having never smoked. Because of the small number of police deaths, however, the risk estimates based upon them lack statistical stability and their confidence limits are correspondingly wide. Also, police have rarely been studied and their occupational exposures and risks for death due to various causes have not been well characterised. An excess or deficit of deaths among police could be due to their own unique exposures or characteristics and thus lead to spurious conclusions about firefighters. Potential police exposures include psychological stress and motor vehicle exhausts. The magnitude and health effects of these exposures are not fully known and their potential for introducing bias should be borne in mind.

In conclusion, this study found excesses of brain cancer and leukaemia among city firefighters from the northwest United States and suggests that they may be at excess risk of dying from emphysema. Exposures to known carcinogens and respiratory irritants are likely to explain these findings; future efforts should be directed towards reducing and eliminating these exposures.

Lindquist v. City of Jersey City Fire Department

New Jersey Supreme Court, 2003.

175 N.J. 244, 814 A.2d 1069.

Coleman, J.

The issue raised in this appeal is whether petitioner's employment as a fireman for approximately twenty-three years caused or contributed to his development of pulmonary emphysema within the meaning of the occupational disease provisions of the Workers' Compensation Act. Resolution of that issue requires us to decide how much workplace contribution is enough to trigger employer responsibility. The Judge of Compensation found that petitioner's occupational exposure materially contributed to the development of emphysema. The Appellate Division reversed, finding that the evidence was insufficient to establish medical causation between the employment and the emphysema. We disagree and reverse.

I.

Petitioner Richard Lindquist was employed as a full-time paid fireman with the City of Jersey City Fire Department from July 1972 until his retirement in January 1995. He was promoted to the rank of captain in 1979. Petitioner testified that during the first ten years of his employment, he responded to "30 to 60 large fires per year," "small one-room" fires, car fires, and "dump" fires. When he began his job in 1972, each firefighter was given a self-contained breathing apparatus, "but it was just very new and people didn't seem to use it until 1982." Although petitioner was exposed to "heavy smoke" for up to forty-five minutes to an hour and a half during larger fires, he frequently did not use the apparatus. * * *

Some of the fires involved burning chemicals, plastics, household cleaners, and propane.

* * *

From 1986 to 1992, petitioner was assigned to supervise the Hazardous Materials Unit of the fire department. During that time, petitioner responded to both residential and industrial fires. The burning items consisted of plastics and chemicals, causing much more toxic smoke than the &70s and &60s. After 1992, petitioner returned to his position as captain.

* * *

Petitioner retired in 1995 at the age of forty-seven, due in part to an early buyout offer and in part to health considerations. At the time of his retirement, petitioner was less able to perform his responsibilities as a firefighter, and in particular as captain, because his energy and normal breathing capacity gradually had diminished. * * * He also suffers from dry eyes and shortness of breath and is no longer able to play basketball with his son or take long walks with his wife. He is able to walk only one quarter to one half of a mile "before [he begins] breathing heavily." He cannot perform

yard work or house work, such as “building sheds, [and] putting [together] decks,” without some difficulty. * * *

Petitioner smoked approximately three-fourths of a pack of cigarettes per day for twenty-two years, stopping in 1992 or 1994. * * *

Dr. Bernard Eisenstein testified on petitioner’s behalf. Dr. Eisenstein specializes in heart and lung medicine and is Board Certified in internal medicine. He performed a complete examination of petitioner on January 16, 1995, to evaluate his pulmonary disability. In addition to the physical examination, Dr. Eisenstein performed a chest x-ray, and pulmonary function studies. * * *

Based on those tests, Dr. Eisenstein concluded that petitioner suffered from “chronic obstructive pulmonary disease in the form of emphysema.” He attributed petitioner’s condition primarily to occupational exposure as a firefighter to fire, smoke, hazardous waste, combustion, and secondarily to cigarette smoking. However, he was unable to allocate an exact percentage to each cause.

* * *

The doctor concluded that, “based upon a reasonable degree of medical probability,” petitioner suffered “30 percent of partial total” permanent disability. On cross-examination, Dr. Eisenstein admitted that he could not cite any studies in which non-smoking firefighters developed emphysema.

In response to Dr. Eisenstein’s testimony, respondent presented the testimony of Dr. Douglas Hutt. Dr. Hutt is Board Certified in internal, pulmonary, and critical care medicine. During his examination of petitioner on December 19, 1996, petitioner informed Dr. Hutt that his primary symptom was a post-nasal drip that began one year after he retired from the fire department. * * * [Dr. Hutt] noted that petitioner “did not remember any long term symptoms that he had after any of the * * * exposures to any of the [] bad fires” * * * He * * * told Dr. Hutt that * * * his grandfather died from emphysema.

Based on the physical examination and the diagnostic testing, Dr. Hutt concluded that petitioner suffers from emphysema caused by petitioner’s cigarette smoking. According to the doctor, “even though only [twenty percent] of people that smoke cigarettes actually get emphysema, that number is [between seventy and eighty percent] higher if you have relatives that smoke cigarettes and get emphysema which is true in this patient’s family in his grandfather.” * * *

* * *

According to the doctor, out of approximately “a hundred” studies concerning firefighters and lung disease in general, none address emphysema but rather deal with air flow obstruction, chronic bronchitis, and other “more serious diseases.” He stated that he had not “seen [studies] that specifically mention emphysema as an increased risk when you factor out cigarette smoking in firefighters.” * * *

The Judge of Compensation concluded that “petitioners occupation[al] disease is due in a material degree to the occupational exposures described” during the trial. The judge also determined that petitioner had suffered an “appreciable impairment of [his] ability to carry on the ordinary pursuits of his retirement lifestyle.” The judge awarded petitioner a disability of thirty percent for emphysema.

On appeal, the Appellate Division reversed in an unpublished opinion, concluding that “the evidence of the causal connection between petitioner’s employment and his emphysema is insufficient to sustain the award.” * * *

* * *

We granted petitioner’s petition for certification * * * and now reverse.

II.

Petitioner argues that the Appellate Division exceeded the scope of its appellate review and ignored testimony in the record that provided an evidentiary basis to support medical causation. * * *

III.

* * *

[A] successful petitioner in workers compensation generally must prove * * * causation.

* * *

The controlling test to be applied in this case is whether the work exposure substantially contributed to the development or aggravation of emphysema. Petitioner had the burden to demonstrate by a preponderance of the evidence that his environmental exposure while fighting fires was a substantial contributing cause or aggravation of his emphysema. To satisfy that obligation, he was not required to prove that his work exposure exceeded the exposure caused by smoking cigarettes. * * *

In a case such as this one in which petitioner concedes that his personal risk factor played a significant role in developing emphysema, the Legislature has provided some relief to employers. When there are dual causes of an injury or disease, such as cigarette smoking and employment exposure, a 1979 amendment to the Act, L. 1979, c. 283, effective January 10, 1980, codified as *N.J.S.A. 34:15-12(d)*, requires a credit to “be given [to] the employer or the employer’s insurance carrier for the previous loss of function and the burden of proof in such matter shall rest on the employer.” *Ibid*. The purpose of that amendment was to ameliorate the effect of prior law that an employer takes an employee as he finds the employee. Although that theory still pertains, the amendment permits a credit, regardless of whether or not the previous loss was work-related, “to encourage [the] hiring [of] workers with pre-existing disabilities.”

* * *

VI.

A.

We now consider whether petitioner’s emphysema is medically related to his work exposure. * * * Emphysema is a “[c]hronic obstructive pulmonary disease (COPD), also called chronic obstructive lung disease[. It] is a term that is used for two closely related diseases of the respiratory system: chronic bronchitis and emphysema. In many

patients these diseases occur together. . . .” Div. of Lung Diseases & Office of Prevention, Educ. & Control, Natl Insts. of Health, Pub. No. 95-2020, *Chronic Obstructive Pulmonary Disease 1* (3d prtg.1995) * * *.

In the general population, emphysema usually develops in older individuals with a long smoking history. However, there is also a form of emphysema that runs in families. People with familial emphysema have a hereditary deficiency of a blood component, alpha-1-protease inhibitor, also called alpha-1-antitrypsin (AAT). The number of Americans with this genetic deficiency is quite small, probably no more than 70,000. It is estimated that 1 in 3,000 newborns have a genetic deficiency of AAT, and 1 to 3 percent of all cases of emphysema are due to AAT deficiency.

* * *

Although “[c]igarette smoking is the most important risk factor for COPD . . . [o]ther risk factors include age, heredity, *exposure to air pollution at work* and in the environment. . . .” [Div. of Lung Diseases & Office of Prevention, Educ. & Control, Natl Insts. of Health, Pub. No. 95-2020, *Chronic Obstructive Pulmonary Disease 1* (3d prtg.1995)] That means the National Institutes of Health has recognized that exposure to air pollutants at work can cause both chronic bronchitis and emphysema. Furthermore, “[s]cientists believe that, in addition to smoke-related processes, there must be other factors that cause emphysema in the general population since only 15 to 20 percent of smokers develop emphysema.” *Id.* at 4.

* * *

B.

* * *

Dr. Eisenstein agreed that emphysema can be caused exclusively by smoking cigarettes, from fighting fires and inhaling the smoke, fumes, gases, and heat alone or a combination of smoking cigarettes and occupational exposure. * * * Based on his experience in examining firefighters, Dr. Eisenstein concluded that petitioner’s emphysema is due to his work plus his smoking. * * *

* * * He stated that he could point to no study done on firefighters who are non-smokers and who had emphysema.

In contrast, Dr. Hutt testified that the emphysema was caused by petitioner’s cigarette smoking and family history that revealed that his grandfather died of emphysema. * * *

Dr. Hutt also testified that although he has read many unspecified studies on lung diseases that included firefighters, none dealt with firefighting and emphysema. * * *
* He was unaware of any studies linking emphysema to any smoke except cigarette smoke.

* * *

C.

When, as in this case, studies of firefighters and other groups have been utilized to assist experts with the medical causation issue * * *, consideration of some or all of

those studies would be useful to a reviewing court. Although the numerous studies Dr. Hutt stated that he utilized in arriving at his opinion in this matter were never identified in the record and have not been made part of the appellate record before us, our independent research has uncovered many studies in this field. We have examined some of the articles presumably reviewed by Dr. Hutt. In any event, we take judicial notice of the studies uncovered in our research. * * *

* * * The “healthy worker effect,” whereby sick workers leave employment and are not included in studies, complicates most studies of disease in firefighters. To reduce that effect, two studies were performed comparing mortality in firefighters and police officers. Because the socioeconomic background, smoking habits, and health requirements of these groups are similar, any increase in lung disease among firefighters is likely to have been caused by their employment. Paul A. Demers et al., *Mortality Among Firefighters From Three Northwestern United States Cities*, 49 *British J. Indus. Med.* 664, 668-69 (1992); * * * Linda Rosenstock et al., *Respiratory Mortality Among Firefighters*, 47 *British J. Indus. Med.* 462, 464 (1990).

The Demers study is a follow-up of the Rosenstock study, published two years later. It found a smaller increase in the risk of non-malignant respiratory disease for firefighters than previously thought, but nonetheless concluded that “a raised risk of emphysema was found among firefighters compared with both United States white men and police.” Demers, *supra*, at 668-69. Those studies contain a predictable list of limitations, such as small sample size, difficulty in tracking subjects after retirement, vague death certificates, and inability to determine the amount and chemical content of smoke exposure. * * * Those studies comparing populations of healthy workers, similar in all relevant respects except fire smoke exposure, present the strongest scientific support for the proposition that firefighting is a significant cause of lung disease. Additional studies support that conclusion.

[The court discussed four other studies of firefighters that examined acute responses in firefighters shortly after fighting a fire and found decrements in breathing ability. One follow up study found no residual impairment several years afterwards. A study of firefighters involved in 9-11 examined them within six months of the event and found: “Intense, short-term exposure to materials generated during the collapse of the World Trade Center was associated with bronchial responsiveness and the development of cough.”]

* * *

D.

* * *

This Court has recognized for many years that the Act is “humane social legislation designed to place the cost of work-connected injury upon the employer who may readily provide for it as an operating expense.” *Tocci v. Tessler & Weiss, Inc.*, 28 N.J. 582, 586, 147 A.2d 783 (1959). * * * Similarly, this Court should be solicitous of firefighters who have demonstrated a substantial likelihood that their fire suppression duties have contributed to the development of emphysema.

* * *

More than a possibility of causal connection exists in this case. Although we do not relax the requirement that petitioner must prove his case by a preponderance of the evidence, and that his evidence must be scientifically reliable, we must examine the evidence in light of science's inability to provide conclusive answers to every question of causation. . . .

In this case, it is true that petitioner's expert did not cite any scientific studies to support his conclusion. Respondent's expert, Dr. Hutt, testified that he had read about one hundred unspecified studies concerning firefighters and lung disease, none of which established a causal link between firefighting and emphysema. However, our independent review of articles addressing firefighting and lung disease confirmed that some evidence to the contrary exists. * * *

Dr. Hutt suggested that petitioner's family history could account for his emphysema, and studies do indicate that "familial factors" can increase the risk. * * * Alpha-1-antitrypsin deficiency is detectable by a blood test that apparently was not performed on petitioner. We therefore do not know the extent to which petitioner's family history contributed to his emphysema.

We find that enough scientific data exists in support of petitioner's case to allow a Judge of Compensation to find in petitioner's favor. * * * That conclusion is compelled by the principles that the Act represents social legislation, and is to be interpreted to expand rather than limit coverage, and that under the social compromise theory it is intended that a petitioner's burden of proof be lighter than in a common-law tort action. The conclusion is further compelled by the fact that the studies reveal that although smoking is the most significant risk factor, some other causal factors must exist because no more than twenty percent of smokers contract emphysema. Nat'l Insts. of Health, *supra*, at 4. Both experts testified that industrial exposure can cause emphysema and that the signs and symptoms have the same manifestation regardless of whether they are caused by cigarette smoking, industrial exposure, or a combination of exposures. We reemphasize that it is not necessary for petitioner to prove that firefighting was the most significant cause of his disease. Rather, he need only show that his employment exposure contributed in a material degree to the development of his emphysema. We hold that there is sufficient scientific evidence to support the Judge of Compensation's conclusion that petitioner sustained his burden of proof. * * *

VII.

The judgment of the Appellate Division is reversed, and the judgment of the Division of Workers' Compensation is reinstated.

Notes and Questions

1. Although workers' compensation requires factual causation, the issue is different from tort cases in which it is the defendant's tortious conduct that must be the cause of harm. Instead, in workers'

compensation the harm must result from employment. Thus, the court's statement that plaintiff had to prove that "environmental exposure while fighting fires" caused or aggravated his emphysema.

2. The court adverts to "family history" and the role that it might play as a risk factor for disease. We defer discussion of the role of genetics in assessing causation until Sections V. E. and F., *infra*.

IV. TOXICOLOGY

A. The Science of Toxicology

1. Introduction

Toxicology is a field that builds on knowledge of chemistry and biology to identify and understand the adverse effects of chemical and physical agents on biological systems. Because of this focus on adverse effects, toxicology is sometimes called the science of poisons, but as we shall see, it deals with many agents, including prescription drugs, food, tobacco and other consumer products, and pollutants that may be toxic to humans.

Toxicity is the production of adverse effects on the structure or functioning of any organ or system of organs of the body. Chemicals vary greatly in the types of toxicities they can cause and in the doses and durations of exposure at which they cause toxicity. Toxicity is studied in animals and other types of experimental systems and can also be identified in observational epidemiology studies. A process called risk assessment is applied to data developed from such studies to identify doses below which adverse effects are likely to be avoided in humans. As we discuss below, these doses are used in the regulatory and public health policy arenas to establish health protective standards for food, drinking water, air, the workplace, etc. They are also used to establish safe starting doses for the study of drugs in clinical trials.

We begin this introduction to toxicology with an excerpt from one type of toxicological research, a study of the effect of a substance on animals other than humans. As the study suggests, toxicology and epidemiology often are complementary methods of detecting whether a chemical causes harm to humans.

The complimentary nature of this relationship is demonstrated in the research undertaken to ascertain whether the morning sickness medication Bendectin is a teratogen (Recall this drug is the topic of the *Brock* opinion in the epidemiology section and the Supreme Court's *Daubert* decision). Early animal

studies generally failed to find effects but a few studies did find a correlation between the drug and birth defects. For example, one study found that fetuses of monkeys given Bendectin at higher than human dosage suffered from a ventricular septal defect (however this was observed in fetuses, not when monkeys exposed in utero were followed postnatally). Concerns about the safety of the drug caused the FDA's Bureau of drugs to request the National Toxicology Program (<http://ntp.niehs.nih.gov/>) to conduct a study to evaluate the teratogenic potential of Bendectin in rats. The study exposed pregnant dams with 0, 200, 500, and 800 mg/kg of Bendectin per day during the relevant gestation period (days 6-13). At gestation day 20 the mothers and fetuses were killed and examined. Among other things, the fetuses were examined for external, visceral, and skeletal malformations. The examinations were done by individuals without knowledge of the dose group of the dams or fetuses. Following is the abstract of the paper followed by two tables constructed from tables found in the full article. The tables are, in turn, followed by the authors' discussion of these findings.

Rochelle W. Tyl, et al., Developmental Toxicity Evaluation of Bendectin in CD Rats¹

ABSTRACT Bendectin, composed of doxylamine succinate and pyridoxine HCl (1:1), is an anti-nauseant previously prescribed for nausea and vomiting during pregnancy. The present study examined the maternal and developmental effects of Bendectin (0, 200, 500, or 800 mg/kg/day, po) administered to timed-pregnant CD rats (36-41/group) during organogenesis (gestational days [gd] 6-15). At death (gd 20), all live fetuses were examined for external, visceral, and skeletal abnormalities. At 500 and 800 mg/kg/day, maternal toxicity included reduced food consumption during treatment and for the gestation period, increased water consumption in the posttreatment period, reduced weight gain during treatment, and sedation; water consumption was reduced during treatment and for the gestation period, and maternal mortality (17.1%) was observed only at the high dose. Developmental toxicity included reduced prenatal viability (800 mg/kg/day) and reduced fetal body weight/litter (500 and 800 mg/kg/day). In addition, reduced ossification of metacarpals (800 mg/kg/day), phalanges of the forelimbs (500 and 800 mg/kg/day), and of caudal vertebral centra (all doses) was observed. No increase in percent malformed live fetuses/litter was observed. The proportion of litters with one or more malformed fetuses was higher than vehicle controls only at 800 mg/kg/day, with short 13th rib (to which the test species is predisposed) as the predominant observation. * * * In conclusion, the incidence of litters with one or more malformed fetuses was increased only at a dose of Bendectin which produced maternal mortality (17.1%) and other indices of maternal and developmental toxicity.

¹ Rochelle W. Tyl, et al., "Developmental Toxicity Evaluation of Bendectin in CD Rats" 37 *Teratology* 539 (1988). Copyright © 2005, John Wiley and Sons. Reprinted with permission.

Reproductive Parameters Following Exposure to Bendectin on Gestational days 6-15

	Bendectin (mg/kg/day, po)			
	0	200	500	800
Number of Fetuses Examined ¹	510	495	445	383
Number of Litters Examined ²	37	36	36	30
Fetal body weight (g) per litter ^{§§§}	3.55 ± 0.04	3.48 ± 0.05	3.14 ± 0.04**	2.49 ± 0.06**
Anterior limbs ^{3,4,5,}				
No. ossified metacarpals ^{§§}	3.62 ± 0.06	3.51 ± 0.08	3.46 ± 0.07	3.31 ± 0.08*:
No. ossified phalanges ^{§§}	0.08 ± 0.03	0.03 ± 0.02	0.01 ± 0.01	0.00 ± 0.00*:
Tail ^{3, 5}				
No. ossified caudal vertebral centra ^{§§§}	4.24 ± 0.12	3.86 ± 0.12**	3.81 ± 0.08*:	3.26 ± 0.12*:

1 Only live fetuses were examined for malformations

2 Includes only litters with live fetuses

3 Reported as a mean ± Standard Error of the Mean for all live fetuses

4 Includes only dams with live fetuses; litter size = No. live fetuses per dam

5 Data are reported as the No of ossified sites per fetus per litter.

6 Litters with one or more malformed fetuses

* P < .05, Fisher Exact Probability Test vs. controls

** p < .01 vs. controls

§ p < .05 Test for Linear Trend

§§ p < .01 Test for Linear Trend

§§§ p < .001 Test for Linear Trend

Note: po is a toxicological shorthand for "by mouth" from the latin per os

Specific Teratological Skeletal Defects Observed
Following Exposure to Bendectin on Gestational days 6-151

	Bendectin (mg/kg/day, po)			
	0	200	500	800
Number of Fetuses Examined ²	510	495	445	383
Number of Litters Examined ³	37	36	36	30
Skeletal malformations				
No. Fetuses with Malformations ⁴	4	8	4	17
No. Litters with Malformations ⁵	3	3	3	9*
Short 13th rib	4	8	3	14
Lumbar centra off center				2
Missing lumbar arch				2
Thoratic centra off center				2
Fused thoratic arches				1
Missing centra				1
Missing 13th rib			1	

1 A single fetus may be represented more than once in listing individual defects

2 Only live fetuses were examined for malformations

3 Includes only litters with live fetuses

4 Fetuses with one or more malformations

5 Litters with one or more malformations

* P < .05, Fisher Exact Probability Test

The present study has demonstrated that administration of Bendectin, po, during organogenesis in CD rats, resulted in maternal and developmental toxicity including an increase in the proportion of litters with one or more malformed fetuses, the last at a dose which caused maternal lethality. These effects were observed at doses which are two to three orders of magnitude above the human therapeutic dose range, estimated at 1-2 mg/kg/day. Bendectin tablets as formulated by Merrell-Dow, contain 20 MG of Bendectin and the usual prescription calls for one to four tablets daily, irrespective of maternal weight. In the present study, the high dose (800 mg/kg/day) resulted in the death of 17.1% (7/41) of the treated dams. * * * Maternal toxicity at 500-800 mg/kg/day was also expressed as reduced body weight and weight gain [and] reduced food intake during the treatment period. * * * The pattern of food and water intake was consistent with a toxic response during the treatment period and compensatory

overeating and drinking in the posttreatment period (gd 15-20). * * *

Ebryolethality was observed at 800 mg/kg/day due to an increase in reabsorptions. * * * Developmental toxicity was observed at 500 and 800 mg/kg/day expressed as a reduction in fetal body weight per litter and as a reduction in fetal ossification in the anterior distal limb bones (metacarpals at 800 mg/kg/day) and phalanges at 500 and 800 mg/kg/day). Increased incidence of poorly ossified skull plates and pubic bones and misaligned sternbrae at 800 mg/kg/day also indicates developmental toxicity. These findings are most likely due to the compromised status of the dams (reduced food and water intake and reduced weight and weight gain).

The increased incidence of malformations observed in CD rat fetuses exposed to 800 mg/kg/day Bendectin during

organogenesis in the present study, occurred only in the presence of maternal toxicity and other indications of developmental toxicity. The parameters which were statistically increased at 800 mg/kg/day were numbers of litters with one or more malformed fetuses and the number of litters with one or more skeletally malformed fetuses. The predominant malformation, short 13th rib, occurred in 4/510 (0.78%), 8/495 (1.62%), 3/433 (0.68%), and 14/383 (3.66%) fetuses examined from the vehicle control through high-dose groups, respectively. Short 13th rib was observed in 34 fetuses out of 1,898 corn oil control CD fetuses (1.79%) and in 12/2,048 (0.59%) distilled water controls for an overall incidence of 46/3946 (1.17%) in historical control data collected in this laboratory and was the most common skeletal anomaly observed in these data. Therefore, the increased incidence of this commonly occurring malformation at a dose which was also associated with maternal toxicity may be a reflect of the stress on the test system. * * *

The ossified elements of the fetal appendicular skeleton * * * were counted. The skeletal districts examined were suggested by Aliverti et al., who proposed that reduced fetal body weight and evidence of retardation of skeletal ossification provided a reliable index of delayed development (i.e. fetotoxicity) in teratogenicity studies. Reduced numbers of ossification centers (with no associated changes in gross limb or digit morphology) were seen in the fetal anterior limb metacarpals at 800 mg/kg/day and phalanges at 500 and 800 mg/kg/day and these effects paralleled the reduction in fetal body weight in the same dose groups. The number of ossified caudal vertebral centra was slightly but statistically significantly reduced in all Bendectin-exposed groups relative to the vehicle controls; however, the biological relevance of an average reduction at 200 mg/kg/day of approximately one-third of an ossification site per fetus in the absence of any of the indications of maternal or developmental toxicity is questionable. This is especially questionable since visualization of these ossification sites depends on the visual

acuity of the observer even with magnification using a dissecting microscope, and since these observations were performed by a number of staff members. The difference between the high-dose and control group was only approximately one ossification site per fetus. No indications of abnormal ossification were seen in any districts of the fetal skeleton.

In order to further evaluate the interrelationships of Bendectin exposure and fetal skeletal development (specifically the number of ossified caudal centra), fetal body weight and litter size were analyzed by using an analysis of covariance (ANCOVA), a posteriori, with dose as the class variable and the other parameters as the dependent variables or covariates. The relationship between fetal body weight and the number of ossified caudal centra was significantly positive, consistent with [other findings] that heavier fetuses exhibit more advance ossification, including the caudal region, than do lighter weight fetuses. [Another study] also found that fetal body weight and the number of ossified caudal centra were both linearly related to fetal age and thus presumably linearly related to each other. In the present study, the slope of fetal weight vs. ossification with dose appeared to indicate that fetal weight and the number of ossified caudal centra co-varied and that Bendectin affected fetal weight more than ossification – i.e. the slopes differed.

The data for this study are consistent with a threshold for maternal and developmental toxicity at 200 mg/kg/day, above which homeostatic mechanisms are apparently overwhelmed and toxicity to the dams and conceptuses is observed. The extent to which the presence of maternal toxicity affected (caused) the observed developmental toxicity, including the short 13th rib, is not known.

In conclusion, Bendectin produced minimal evidence of developmental toxicity in CD rats when administered orally during organogenesis except at high dose levels, which also produced profound maternal toxicity.

Notes and Questions

1. What are the competing hypotheses for the malformations exhibited by fetuses whose mothers were exposed to 800 mg/kg/day? How did the researchers attempt to resolve which of the hypotheses is correct?
2. As the Bendectin study suggests, one role of toxicology is to confirm or contradict the results of epidemiological research. However, there is an important difference between the two fields. Unlike epidemiology, which focuses primarily on whether or not a substance causes harm, the strength of toxicology is its focus on *how* the substance causes harm.

The remainder of Part I of the toxicology section provides an overview of the science of toxicology. Part II turns to a discussion of the use of toxicological evidence in the legal system.

2. Toxic Effects

Toxic substances have the potential to affect all organs/systems, for example, the liver, kidneys, immune, dermal, and respiratory systems. Toxic substances can cause harm in several distinct ways. A number of substances may affect the central nervous system. Examples include addictive substances such as alcohol, heroin and cocaine but also other chemicals such as benzene, metals such as mercury and lead, and bacteria such as clostridium botulinum, responsible for botulism, one of the most toxic substances known.

Still other toxins cause damage to the cardio-vascular system. Although the disease most closely associated with cigarette smoking in the public mind is lung cancer, tobacco use actually causes more deaths through damage to the lungs and heart than all tobacco related cancers combined. Some drugs may also cause heart problems as an unwanted side effect. This is true of nonsteroidal anti-inflammatory drugs (NSAIDs).

A third general type of injury is fibrosis, a process of laying down fibers outside cells. The result is loss of tissue resiliency and intercellular communication. Typical examples are silicosis, asbestosis and emphysema, illnesses associated with, in turn, exposure to silica dust, asbestos and tobacco smoke. Cirrhosis of the liver is another form of fibrosis that may be caused by, among other things, chronic exposure to ethanol. As these examples demonstrate, substances often produce adverse effects on specific organ systems such as the lungs, the liver and the kidneys.

Other toxicants affect reproduction and act as teratogens: agents that disturb the development of an embryo or fetus may cause a birth defect in the child or simply halt a pregnancy. Radiation, chemicals and drugs may produce this outcome. Thalidomide is one such drug. Another teratogen that has produced substantial litigation is diethylstilbestrol (DES). The timing of exposure to the offending substance is often critical. Exposure at the very earliest stages is likely to produce embryonic death. Exposure during the middle stage of organogenesis is most likely to lead to structural defects. If the exposure occurs during a critical

period of organ system formation it may lead to defects to that system. For some systems such as the nervous system, the critical period may extend throughout development. For example, consumption of ethanol by pregnant women throughout the pregnancy poses a risk of fetal alcohol syndrome. Other risks may be of limited duration. Thalidomide ingestion poses a significant risk of limb-reduction defects if taken during the critical time of limb bud development but if the exposure occurs either before that or after, the substance apparently has no adverse effect on the fetus. Substances may cause harm either by directly altering DNA sequences or through epigenetic control systems that alter what genes get expressed at what point in time.

Perhaps most significantly, numerous toxicants are carcinogens. Cancer is not a single disease, but rather a term used to describe many types of malignant growths that invade adjoining tissues. Perhaps the cancer most strongly associated with an external agent is lung cancer associated with tobacco use, especially cigarette smoking, closely followed by lung cancer and mesothelioma (a cancer of the pleural lining of the lung) related to exposure to airborne asbestos fibers. The prevailing theory of carcinogenesis is mutational. Carcinogenesis begins with an initiation phase in which DNA is damaged. Whether this results in a tumor turns on a number of factors. For example, genetic factors also affect the development of cancers, in part due to defects in DNA repair systems. There are several competing theories of carcinogenesis but all agree that alterations in gene expression are central to the carcinogenic process. Many substances are known to have the potential to cause cancer. Cigarette smoke alone is known to contain between 60 and 70 carcinogens.¹

3. The “Laws” of Toxicology

There are three central tenets of toxicology, sometimes called the three “laws” of toxicology. The first concerns dosage. The phrase, “the dose makes the poison” reflects the fundamental idea that the higher the dose of a substance the greater the response. Perhaps all chemical agents are hazardous if consumed in large enough doses. Water, sugar, salt and other substances we consume every day are hazardous in sufficient dosage. Dosage is a function of both exposure concentration at a given point in time and a cumulative dose received over a relevant period of time, a fact reflected in government exposure regulations. For example, OSHA’s permissible exposure limit (PEL) for benzene in the workplace is 1 part per million of air as a time-weighted average (TWA) for an 8-hour work shift¹ and a short-term exposure limit of 5 parts per million in any 15-minute sampling period.

The second central tenet is that chemicals and other toxicants produce specific patterns of biological effects due to the unique chemical structure of the agent and the laws of biology

¹ A particularly valuable web site on the toxic effects of various substances is <http://toxnet.nlm.nih.gov/> and the embedded database <http://toxnet.nlm.nih.gov/newtoxnet/toxline.htm>

that govern the organism's response. Knowing the structural determinants of the activity of a chemical and the way in which it interacts with an organism may permit us to identify the substance to which an individual has been exposed. However, as we discuss in greater detail below, the structural consequences of an exposure may depend on the dose of the substance and subtle changes in chemical structure can produce very different biological effects.

The third "law" is simply that humans are members of the animal kingdom. The importance of this tenet is that by studying the effects of chemicals on other animals we may better understand and predict their effect on humans. Indeed, without animal testing we would know much less about the effect of many substances on humans.

4. Areas of Toxicological Research

Toxicology proceeds on several tracks. Forensic toxicology traditionally used laboratory techniques to determine the cause of death, but has expanded to support police and the courts in developing evidence suitable for use in criminal cases, the widespread use of DNA evidence being one example. Below we present a drunk driving case that turns on questions in forensic toxicology.

Environmental toxicology constitutes a second track. This field has grown rapidly, propelled in part by the many new occupational and environmental laws passed in the past few decades. Assessing the toxicology of chemicals and other substances in the workplace and elsewhere is one component. Another component of environmental toxicology is concern with the adverse effects of both water and air pollution. Water pollutants include organic waste; organic compounds such as petroleum products and pesticides; inorganic substances including phosphates and nitrates used in fertilizers; and biological agents such as viruses, bacteria and parasites that may cause disease. Air pollutants include carbon oxides, sulfur oxides, volatile hydrocarbons such as benzene and methane as well as particulates such as silica, asbestos, dust, pollen and the byproducts of the combustion of fossil fuels and other organic materials.

A third track of toxicology is pharmaceutical toxicology, a field that has grown apace with the development of prescription drugs. As with other areas of toxicology, a good deal of effort goes into measuring the risks and benefits of drugs before they are allowed to go on the market.

5. Toxicokinetics and Toxicodynamics

Therapeutic drugs and toxins such as poisons, industrial chemicals and environmental pollutants are substances that are foreign to the organism. Toxicologists generally refer to such substances as *xenobiotics*. The interaction of xenobiotics with an organism involves three components, exposure, toxicokinetics and toxicodynamics. We discuss exposure issues, i.e. dose and differences between species, and susceptibility below. In this section we outline the basics of toxicokinetics and toxicodynamics.

a. Pharmacokinetics and Toxicokinetics

Pharmacokinetics is the study of how drugs enter, reside, and exit an organism. Toxicokinetics is the study of the same processes as it relates to toxic substances. Each involves four primary processes: absorption, distribution, metabolism, and excretion (ADME).

i. Absorption

Absorption, the uptake of an internal agent into the body, occurs through three primary pathways; ingestion, inhalation and through the skin, although some drugs may be injected directly into the body. A central question with respect to absorption is the dose actually received by an organism. The answer to this question is usually straightforward with respect to prescription drugs, but is quite difficult to assess with respect to other exposures. Here it is important to distinguish between exposure and dose. One may be exposed to X amount of some chemical spilled on the skin, but only some of the chemical is absorbed into organism and this amount is the dose. Many factors influence the rate of absorption through different pathways. For example, barriers through diffusion in the lungs are far less than through the skin because cells that line the lungs are thin and located very close to blood vessels. To give another example, the size of particles influences their absorption through respiration; larger particles are more likely to be intercepted in the nose or upper airways and diverted through swallowing into the digestive tract where they may be evacuated rather than absorbed. Once ingested, the ratio of absorption versus evacuation is measured in terms of the substance's bioavailability. The bioavailability of an ingested substance is the fraction of a dose that reaches systemic circulation. A substance administered intravenously is, by definition, 100% bioavailable. The bioavailability of a substance may be influenced by the matrix of materials enclosing the agent. For example, more than half the dioxins that contaminated the soil in Times Beach, Missouri was bioavailable when fed to laboratory animals, while the bioavailability of the dioxins that contaminated soil in Newark, New Jersey was hundreds-fold less.

ii. Distribution

Once a chemical enters the bloodstream, its distribution within the body is affected by several factors. They include: a) whether a substance binds to red blood cells or plasma proteins. Arsenic, for example, has a high affinity for red blood cells. Substances that are so bound are less available for filtration or diffusion into organs that metabolize and excrete toxins (liver, kidneys) and therefore remain longer in the organism but their uptake by other organs is decreased; b) the rate at which blood is supplied to various tissues. The heart, liver, and kidneys experience a relatively high rate of blood flow; c) the ability of the chemical to cross cell membranes, which in turn is influenced by its solubility in water or fats (primarily lipids). Organic chemicals are more soluble in lipids and, therefore, generally pass through lipid-rich

cell membranes more readily than can inorganic chemicals. Consequently, organic chemicals are generally absorbed more extensively into cells than are inorganic chemicals. This ability is enhanced if the toxicant is neutrally charged; and d) the existence of anatomical and physiological attributes that limit the passage of some chemicals. The most well-known of these attributes is the blood-brain barrier that prevents the penetration of some toxicants into the central nervous system. Toxicants that affect the central nervous system tend to be small, highly lipid soluble, nonpolar molecules. Similarly, the placental barrier, a semipermeable layer of tissue in the placenta that serves as a selective membrane to substances passing from maternal to fetal blood, may prevent some toxins absorbed by the mother from entering the bloodstream of a fetus. With respect to all suspected teratogens, the first question to ask is whether the chemical can cross the placental barrier.

Other xenobiotics may do their damage without ever entering the bloodstream. For example, asbestos fibers produce their adverse effects by damaging cells in the lungs and pleura and acids cause skin damage by directly harming skin cells.

iii. Metabolism

Understanding metabolism is a critical component of toxicology. As exogenous agents enter the body their chemical or physical status is altered by the organism. Metabolites are the product of enzyme-catalyzed chemical changes in the initial drug or toxicant. First stage metabolic processes include hydrolysis and oxidation. These processes make the substance available for use by the organism, and facilitate elimination. Organs that play a major role in metabolism such as the liver may themselves be injured in the process. Primary metabolites may undergo further metabolic changes. Metabolomics is the study of an organism's metabolome, which is the collection of metabolites found in particular organism.

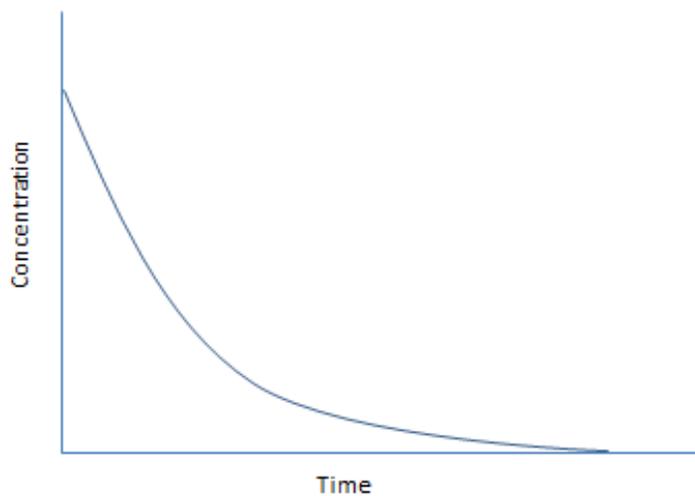
Often, metabolic processes render an otherwise harmful substance harmless in the body. The process of hydrolysis makes the metabolites of lipid soluble xenobiotics more hydrophilic and oxidation makes them more highly charged and thus less able to cross cell membranes and more easily eliminated. However, for some chemicals, metabolism converts a relatively benign substance into a toxic agent. For example, metabolites of benzene are toxic to bone marrow and may lead to aplastic anemia and some forms of leukemia, most clearly acute myeloid leukemia (AML).

The way in which different species metabolize a chemical may significantly affect its toxic effects. As noted above, thalidomide is much more toxic to human fetuses than the fetuses of rats and mice a result, apparently due to the different ways and rates of metabolism of the drug in the organism. This often relates to the existence of species specific proteins involved in the metabolic process. Adverse effects are also affected by the rate at which metabolic processes occur. Both the young and the elderly have slower rates of xenobiotic metabolism with multiple consequences.

iv. Elimination

Some substances that are ingested are never absorbed and are eliminated through the gastrointestinal tract, that is, their bioavailability is zero. The primary route of excretion of most drugs and toxic substances that are absorbed is through the kidneys and excreted as urine. Other drugs and chemicals and their metabolites may be removed by the liver, excreted into the bile. Other routes include the respiratory system and secretion such as sweating and lactation. The rate of elimination of many toxicants and drugs is related to the concentration of the substance in the body: higher concentrations lead to more rapid elimination, and elimination slows as concentrations fall—what is commonly called a first-order process of elimination (see Figure IV-1).

Figure IV-1.

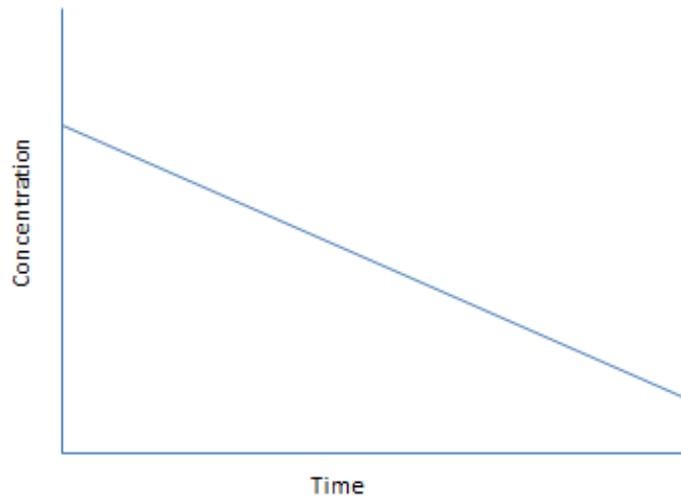


Elimination Rate affected by Concentration

SOURCE: Courtesy of the authors.

However, some toxicants are eliminated at a steady state regardless of concentration: what is commonly called a zero-order process of elimination. Unfortunately, perhaps, ethanol is one such substance (for reasons that are discussed below). Thus the concentration of alcohol, or other drugs and substances introduced orally, the concentration in the blood at any point in time is the sum of the absorption rate from the gut and the elimination rate (see Figure IV-2).

Figure IV-2.

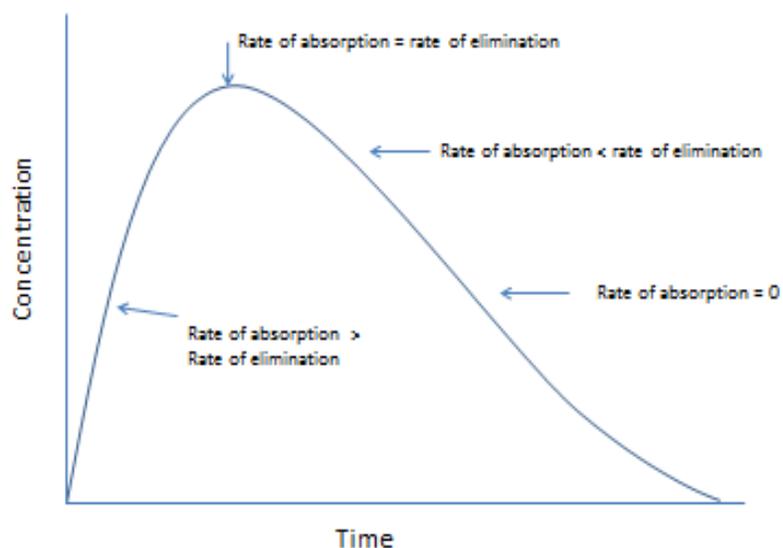


Elimination is a Constant Over Time

SOURCE: Courtesy of the authors.

Peak concentration occurs when the absorption rate equals the elimination rate (see Figure IV-3).

Figure IV-3.



Absorption and Elimination of an Oral Dose

SOURCE: Republished with permission of Taylor and Francis Group LLC Books, from Karen E. Stein and Thomas M. Brown, *Principles of Toxicology*, Third Edition, 2015; permission conveyed through Copyright Clearance Center, Inc.

Obviously when the elimination rate is unaffected by concentration, peak concentration is primarily affected by changes in the absorption of a substance. One passes the peak concentration of alcohol only when the absorption of alcohol into the bloodstream falls below the rate of elimination.

The speed of elimination is influenced by whether the substance is lipophilic (soluble in lipids). Substances with high levels of lipophilicity and which degrade slowly may remain in the body for substantial periods of time. For example dichlorodiphenyldichloroethylene (DDE) a metabolite of the now-banned insecticide DDT accumulates in the fatty tissues of fish and other wildlife. Although, most pharmaceuticals are rather quickly eliminated from the body, some newer products are more lipophilic by design and therefore remain in the body longer. This is true of nonsteroidal analgesics such as Ibuprofen, and COX-2 inhibitors, giving these products a greater opportunity to produce side effects.

b. Toxicodynamics

While toxicokinetics describes the changes in the concentration of a substance in an organism over time, toxicodynamics describes the dynamic interactions of a substance with

biological targets and the resulting biological effects. A principal target of toxicants is receptors on the cell's surface. Signals are sent from receptors into the cell. Receptors are specific biochemical sites that recognize and bind to chemicals of a specific size and shape (ligands). Many xenobiotics produce their effect by mimicking a biological ligand. For example, heroin and morphine mimic endorphins.

Because many toxic substances begin their adverse effect by binding to or reacting with a particular biological molecule such as a protein or lipid within a cell, mechanistic toxicology, a component of all subdisciplines of toxicology, focuses on the molecular basis of how these agents affect biological targets.

Improving our understanding of the mechanisms by which substances produce their toxic or therapeutic effects should increase our ability to design drugs and predict toxicity.

Some interactions may not cause molecular changes but nevertheless may cause harm to the organism by disrupting the ordinary function of proteins. For example, the primary protein that transports oxygen in the body is hemoglobin, a molecule in red blood cells. The four amino acid chains that make up the molecule possess an iron-containing structure (a heme group) that can carry a molecule of oxygen. Carbon monoxide interferes with this process because hemoglobin has a higher affinity for carbon monoxide than it does for oxygen so that even small amounts of carbon monoxide can block oxygen binding and, in sufficient concentrations, may lead to death.

More frequently, xenobiotics or their metabolites actually alter or damage their target. In general, the adverse effect is related to the dose of the substance. Toxic effects occur in several general ways. For example, xenobiotics may create oxidative stress, by causing cell death, and through processes of cell proliferation and tissue repair.

Comparing toxic effects across species may provide a better understanding of this dynamic. For example, it is known that chronic exposure to unleaded gasoline, a complex mixture of hydrocarbons and other organic chemicals, increases the number of kidney tumors in male rats, but not humans or mice of either sex. Research discovered that one of the chemicals in gasoline, *tert*-butyl alcohol, binds to a specific plasma protein, α_2 u-globulin. This protein is filtered in the kidney and mostly reabsorbed in epithelial tubular cells in the kidney. There they are normally degraded by enzymes in the lysosomes in these cells. However, the chemically altered proteins are degraded more slowly by lysosomal enzymes and, therefore, accumulate in the tubular cells, causing them lethal damage. The repeating cycle of cell damage and regeneration substantially increases the rate of cell proliferation in the kidney and this rapid proliferation is an established mechanism of carcinogenesis. α_2 u-globulin is a male rat-specific protein, not expressed, in female rats, mice, or humans, strongly suggesting that the human risk of kidney cancer from chronic gasoline exposure is small.

As this example, suggests, although some toxins cause problems at the site where they come in contact with an organism, for example, acid burns on the skin, and others cause general systemic effects, for example, carbon monoxide poisoning, many are absorbed and distributed throughout the body where their effects are occur primarily in specific target organs.

Our level of understanding of the mechanistic processes by which unleaded gasoline produces its toxic effect in rats is far better than our understanding of how many other known toxic substances produce their ill effects. Thalidomide offers one such example. The drug was developed as a sedative and proved to be useful as an antiemetic and was widely prescribed in the 1950s as a morning sickness drug in Germany, England and Canada, but not the United States, primarily due to an FDA employee Dr. Frances Kelsey, who repeatedly stalled its approval due to inadequate safety data. As is now well known, the drug turned out to be one of the most serious known teratogens: approximately 20% of the children whose mothers took the drug during their pregnancy suffered from some birth defect. The most frequent injuries were limb reduction defects. Thalidomide was withdrawn from the market in 1961, but this disaster caused the United States and other countries to develop systematic toxicity testing protocols for all new drugs, protocols which we discuss below.

The disaster also generated a substantial amount of research on exactly how the drug produced its adverse effect. As is the case with gasoline, thalidomide does not equally affect all species. Indeed, at the time the awareness of the drug's adverse effects was unfolding, some well-respected scientists believed that it was not a teratogen because animal studies on rats had not resulted in fetal defects similar to those found in humans. Ultimately, researchers discovered that the drug did produce similar effects in rabbits, causing the FDA to begin requiring that new drugs be tested on two animal species, only one of which could be a rodent.

Even with insight in the 50 years since the drug was removed from the market as a treatment for morning sickness the exact mechanism of its effect is not fully understood. Over 30 different hypotheses have been advanced, but to date there is no consensus about the relevant mechanism, in part perhaps because the drug produces its harm through several pathways.

6. Types of Research

a. Predictive Research

The relatively new subdiscipline of predictive toxicology uses computer models (and thus is sometimes called *in silico* toxicology) to assist in predicting the effect of agents on biological systems. The field tends to focus on, among other things, molecular modeling demonstrating how drugs and other substances interact with the nuclear receptors of cells and whole cell simulations. In the case of prescription drugs, these methods may help to reduce the very high cost of bringing new products to market.

Predictive toxicology and mechanistic toxicology are part of the process of examining the xenobiotic's chemical structure and comparing it to other compounds of similar structure for which there is existing toxicity information. The *structure-activity relationship* can expedite the identification of potentially beneficial or harmful substances. However, as noted earlier, very modest differences in chemical structure can lead to different levels of toxicity, in part because of the way organisms metabolize the substance. For example, benzene and the alkyl

benzenes, for example, toluene, xylene and ethylbenzene, share a similar structure. This similarity is reflected in the fact that acute exposure to each of them produces similar central nervous system anesthetic-like effects. However, only benzene causes leukemia and damage to bone marrow. The damage is caused not by benzene itself but a toxic metabolite of benzene.

b. In Vitro Research

In vitro testing is conducted on bacteria, human or animal cells, isolated tissues, embryos or organs. Such testing takes place outside a living organism. Cultures in a test tube or petri dish are exposed to potential toxicants (or drugs). Toxicologists study the perfusion of the substance through the culture and for physiological responses. These methods are useful to mechanist toxicologists because they can provide insight to the mechanisms of toxicity such as the way potentially cancer causing substances may damage DNA or cause changes in the nucleus of a cell.

The advantages of in vitro testing are considerable. One important consideration is cost when compared to testing on living organisms. A second advantage is that because living organisms are very complex, it is often difficult to identify particular interactions and processes of interest. Because in vitro systems are simpler, it is easier to isolate and study these interactions such as the particular way immune system proteins attach themselves to antigens. Still a third advantage is that in vitro research can use human cells in situations where it would be unethical to do research on living people. Because cellular responses and whole organism responses are often species-specific, using human cells eliminates the need to make a cross-species extrapolation. For example, one study found that across 16 chemicals tested, in vitro tests using human cultures predicted human adverse skin reaction better than in vivo tests using rabbits, primarily because the tests using rabbits over-predicted skin effects in humans.

Shortcomings of in vitro tests include the fact that for many tests there are not established protocols that permit easy comparison across studies, although this is an area where published protocols are becoming more commonplace. Even more important, to date there is limited ability to extrapolate in vitro findings to effects of the substance on living organisms. With respect to potential new drugs, effectiveness in vitro more often than not does not translate into effectiveness in living animals.

c. In Vivo Research

In vivo testing is done on living organisms other than humans. Such research may be designed to understand the impact of a substance on the species itself, but more frequently its purpose is to tell us something about the substance's potential impact on humans. The Tyl et al. study at the beginning of this section is one example of animal research. At the end of the toxicology section we include another animal study of the effects of chlorine exposure on rats. Although both of these studies use rats, the choice of which animal models to employ is a complex one, influenced by the end point of interest (e.g., cancer, birth defects, etc.),

similarities between the animal and human systems of interest, and, at a practical level, the costs involved in testing on different species. Because the choice of the appropriate animal model is not perfect, the FDA requires the use of two species for each drug. As noted above, the wisdom of this requirement first became apparent during the thalidomide disaster. The drug is not a teratogen in rats or mice but had it been tested on rabbits, its adverse effects would have been discovered.

As toxicologists develop better mechanistic models, the choice of appropriate animal species is becoming less and less of a hit-and-miss proposition. Moreover, due to advances in genomics, toxicologists have begun using transgenic models, that is, animals whose genome has been modified so as to increase their similarity to humans in some particular way. One example is the use of “knockout animals” in which a particular gene has been disabled.

The primary strength of well conducted animal studies is that they can be randomized clinical experiments. Ideally, animals are randomly assigned to treatment or control and therefore with a sufficient sample size we can be reasonably certain that any observed effect is due to the treatment. Moreover, the best studies are blinded so that the researcher doing the experiment does not know whether a given animal received a treatment or a placebo. As we discuss below, however, not all studies conform to this ideal.

The dose to which an animal is subjected is a central issue in animal studies. In some animal studies the goal is to assess the acute toxicity of a chemical, that is, the results from a single dose of the substance. However, in other studies the goal is to assess the toxicity of repeated or continuous exposure over a longer or shorter period of time. With respect to acute response studies, animal researchers often employ a LD₅₀ dose experiment to ascertain the dosage (usually measured as milligrams per kilogram weight of the animal) needed to kill half the animals in a study within a relatively short period of time. A substance with a median lethal dose of 1 mg/kg is generally considered highly toxic. A substance with a median lethal dose of greater than 500 mg/kg is considered to be slightly toxic (see Table IV-1).

Table IV-1
Acute Oral LD₅₀ In Rats: Various Substances

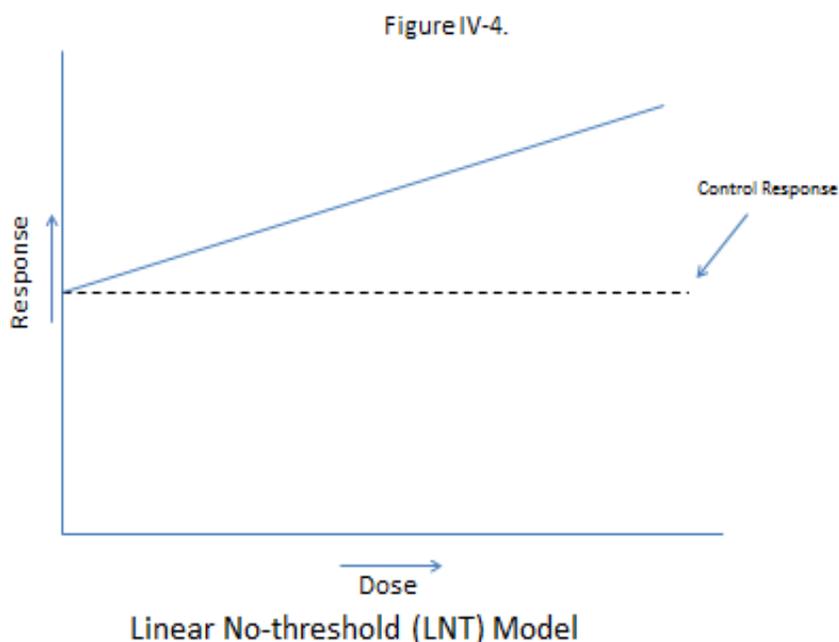
• Ethanol	7,000	mg/kg
• Sodium chloride	3,000	
• Salt	3,000	
• Aspirin	1,000	
• Caffeine	200	
• Nicotine	50	
• Vitamin D	10	
• Cyanide	10	
• Botulin toxin	0.00001	

SOURCE: Republished with permission of National Science Teachers Association, from Nancy M. Trautmann, *Assessing Toxic Risk*, Student's Edition, 2001; permission conveyed through Copyright Clearance Center, Inc.

In studies not focused on acute poisoning, where the question is whether the drug causes cancer or similar adverse outcomes, animals may be given the maximum tolerated dose (MTD); that is the dose just below a dose that may cause premature mortality due to short-term toxic effects. The maximum dose is used, rather than a lower dose, to reduce the number of test animals—and thus the high cost of *in vivo* research—and to detect significant differences between animals exposed to a substance and controls. For example if an adverse outcome occurs naturally, that is, in the absence of the xenobiotic under investigation, in 2 of 100 cases and exposure to the toxin at approximately the dose humans would experience increases the incidents of the adverse outcome to 3 in 100 (a 50% increase), a study using 200 animals (100 exposed and 100 controls) that resulted this outcome would not be statistically significant. Indeed, if these findings were replicated using a thousand animals (i.e., 15 of 500 sick in the exposed group and 10 of 500 in the control group) the results would still not be statistically significant. If, on the other hand, we increased the dose to 10 times what humans were likely to be exposed to and at this greater dose 8 in 100 animals suffered the adverse effect under investigation, a study producing this result with 200 animals would reach statistical significance. And even this example is misleading because actual effects for animals and humans may not be 3 in 100, but rather 3 in 1,000 or even less frequent. Of course, we are now confronted with the problem of extrapolating this result back to effects and human dose rates, a topic addressed in Section IV.G., *infra*.

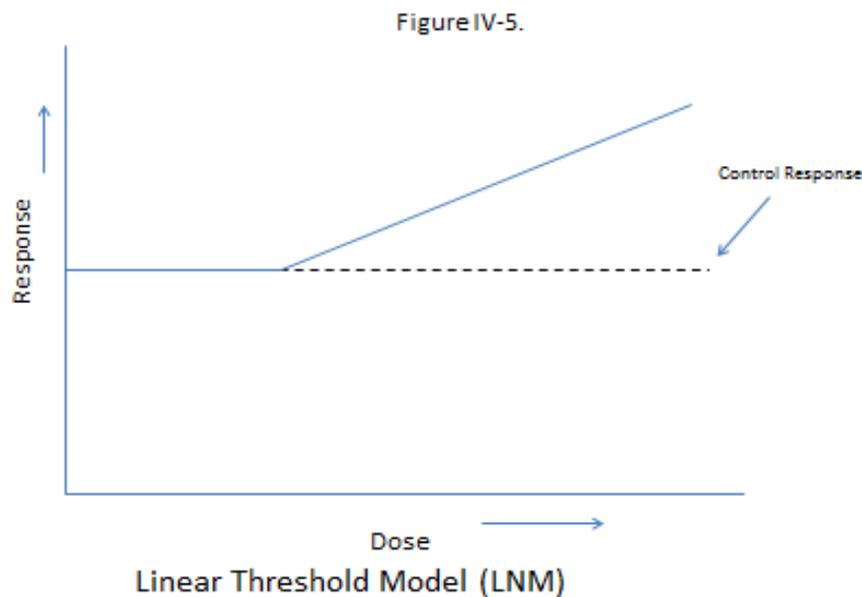
With respect to new drugs, a compound that survives acute tests will be given to animals in repeated doses, with attention paid to the organs that are most likely to suffer toxic effects. In quantitative studies, animals may be subjected to increasing dosages of a drug to assess the effect of dose on outcomes. Information gained from these studies is valuable for estimating starting doses in clinical trials.

Within the fields of regulatory and environmental toxicology, once a hazardous chemical has been identified next steps involve dose-response assessment, exposure assessment, and risk characterization. Toxicologists assume that with respect to cancer risks there is no low-dose level at which a substance has no adverse effects. Therefore, assessing risks a standard procedure is to fit mathematical models to the observed tumor incidence in animal studies some of which may have used high doses as such as MTD and extrapolate risk in a linear fashion, that is, a linear no-threshold (LNT) model (see Figure IV-4).



SOURCE: Courtesy of the authors.

However, with respect to noncancer injuries, toxicologists recognize that there are likely to be threshold doses below which there is no adverse effect (see Figure IV-5).



SOURCE: Courtesy of the authors.

To use a trivial example, sulfuric acid spilled on the skin in high concentrations may produce a significant injury, but sufficiently diluted, it causes no injury at all. In this situation, it is important to estimate a threshold below which an injury does not occur. As a practical matter, this means finding an exposure level that fails to produce a statistically significant difference between exposed and unexposed animals.

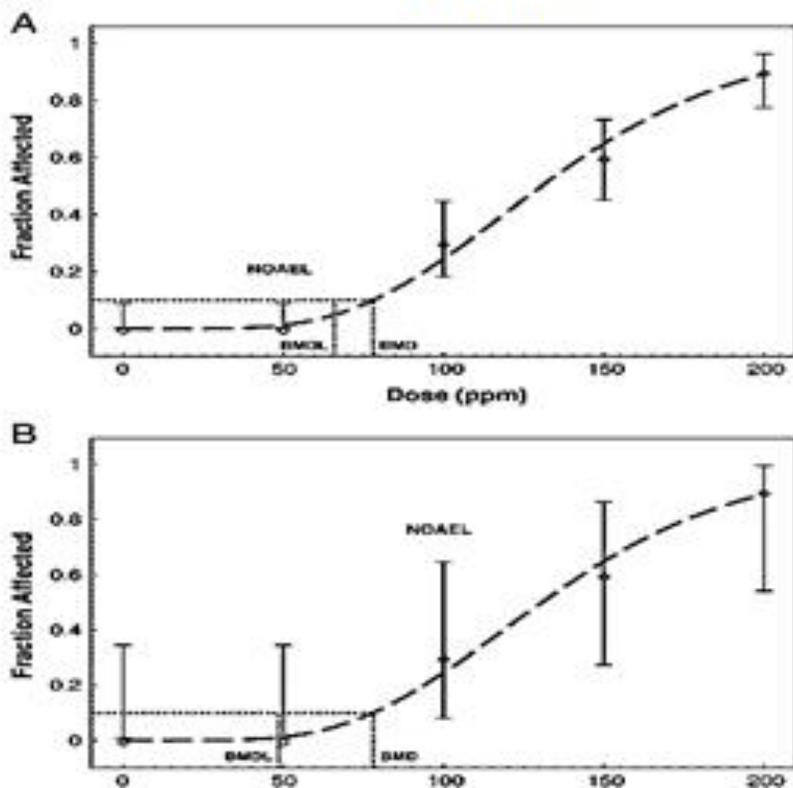
Traditionally, the method used to do this was to undertake tests designed to uncover a no observed adverse effect level (NOAEL), which is defined as the highest dose at which there is not statistically significant difference between subject animals and controls, or in the alternative to test for the lowest observed adverse effect level (LOAEL). NOAEL and LOAEL models may be important in litigation when an individual has been exposed to a dose substantially below a NOAEL level. We return to this in Part II, below.

NOAEL models have some well-known limitations. They are dependent on the dose selection, dose spacing, and sample size of the study from which the critical effect has been identified. And the NOAEL approach does not take into consideration the shape of the dose-response curve. Therefore, in regulatory toxicology the NOAEL approach is being supplemented and sometimes replaced with an approach known as the Benchmark Dose (BMD).

The BMD is now the Environmental Protection Agency's preferred approach to estimating thresholds. Using this method, one collects all known studies of adequate quality

and if they are of sufficient number and quality, one fits various mathematical models to the observed data to estimate the BMD, which is the central estimate of the dose or concentration that produces a predetermined change in the response rate of an adverse effect. The change is called a benchmark response (BMR). An example would be the dose estimated to cause a 10% increase in the number of animals developing a particular disease. The models produce an estimated dose-response curve across the entire dose range of the studies involved and thus are not limited to experimental doses as is the case for the NOAEL approach. A second benefit is that this method, as used by the EPA calculates a 95% confidence level lower bound for the BMD (called the BMDL). Unlike the NOAEL approach, where if a study has low statistical power the NOAEL is higher, here, if the studies have small sample sizes or if there is a high background rate the BMD method calculates a lower, more conservative BMDL (see Figure IV-6 and Table IV-2).

Figure IV-6.



SOURCE: Reprinted from *Toxicology and Applied Pharmacology* 254(2), J. Allen Davis, Jeffrey S. Gifft, Q. Jay Zhaob, "Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1," pp. 181-91, (2011), with permission from Elsevier.

Table IV-2.

Determination of NOAEL and BMDL dependent on dose group sample size.

Animals per dose group	Dose (ppm)	Incidence	Fisher's exact p-value	NOAEL*	BMD	BMDL
50	0	0	1.00	50	78.05	65.85
	50	0	1.00			
	100	15	< 0.001			
	150	30	< 0.001			
	200	45	< 0.001			
10	0	0	1.00	100	78.05	48.40
	50	0	1.00			
	100	3	0.105			
	150	6	0.005			
	200	9	< 0.001			

*NOAEL determined based on highest dose with Fisher's exact p-value of ≤ 0.1 .

SOURCE: Reprinted from *Toxicology and Applied Pharmacology* 254(2), J. Allen Davis, Jeffrey S. Gifft, Q. Jay Zhaob, "Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1," pp. 181-91, (2011), with permission from Elsevier.

Whether one uses a NOAEL or a BMD approach, as we discuss further below in the extrapolation section, regulators use these results to establish "safe" levels of exposure to substances. Recall, for example, OSHA's permissible exposure limits for exposure to benzene in the workplace.

d. Clinical Trials

The chemicals that are most thoroughly tested are prescription drugs. Drugs are designed to have biological effects. As such, they are often a two-edged sword. It may be impossible to retain a drug's therapeutic effect while eliminating all unwanted side effects. For example, the toxicological effect of causing shock through dilation of blood vessels can, with appropriate manipulation of chemical structure and dose, lead to a valuable drug to treat high

blood pressure, but the dilation of blood vessels may carry its own risks.

Obviously, with respect to substances believed to be toxic but with no therapeutic potential, testing ends with animal studies and any research on human effects largely falls within the domain of epidemiology. However, in the case of potential new drugs, tests continue using human subjects. These clinical trials come at the intersection of toxicology and epidemiology.

Before they can be approved by the government, drugs must pass through first-, second-, and third-phase human clinical trials. Prior tests on animals are used to establish a safe starting dose for Phase I clinical trials. In Phase I, a new drug is tested on 20 to 80 healthy volunteers. However, if a new drug is intended for use in cancer patients, researchers conduct the studies in patients with that type of cancer, usually individuals for which other treatments have proven unsuccessful. The object is to observe how the drug interacts with the human body and to adjust dosing schemes to assess the best way to administer the drug to limit risks and maximize possible benefits. Approximately 70% of drugs that enter Phase I move on to Phase II.

In Phase II studies, researchers administer the drug to a group of patients with the disease or condition for which the drug is being developed. Typically involving a few hundred patients, these studies aren't large enough to show whether the drug will be beneficial. Phase II studies do provide researchers with additional safety data. Researchers use these data to refine research questions, develop research methods, and design new Phase 3 research protocols. Phase II studies may last several months up to 2 years. Approximately one-third of the drugs that enter Phase II clinical trials move on to Phase III.

Phase III studies are the final stage before a new drug request is submitted to the FDA. This phase may include up to several thousand individuals with the disease or condition. Because of the larger number of people in Phase 3 studies, there is a greater chance of detecting less common side effects that went undetected in earlier Phases. Phase 3 trials may last 1 to 4 years. A majority of drugs that enter Phase III studies successfully complete this phase of testing. Overall, however, one study published in 2014 reported that 12% of drugs making it to the clinical trial stage are ultimately approved by the FDA for human use.

As noted in the epidemiology materials, most Phase II studies and all or nearly all Phase III studies involve randomized trials where one group receives the experimental drug while a control group receives a placebo or a standard treatment. Often in Phase II and uniformly in Phase III trials, the trials are "double blinded." That is, neither the doctor nor the patient knows whether the patient received the experimental drug.

It is worth stating once again why these trials and in vivo studies should use random assignment and why they should be blinded.

The problem that haunts nonrandomized designs is the possibility that some important variable that affects the relationship of the purported cause (the drug) and the effect (the clinical outcome derived from using the drug or the adverse outcome caused by the drug) has not been measured and, therefore we may misestimate or misinterpret a correlation between the two variables of interest (the drug and the outcome) as causal. How does randomization

help? It does so because, ideally, by randomly choosing who gets a drug and who gets the placebo, we have controlled for the effect of unmeasured variables. This is so because, in a large enough sample, the distributions of unmeasured variables should appear approximately equally among the two groups, thus canceling out any effect it might have in mediating the relationship between the drug and the outcome of interest.

One could conduct trials, of course, simply by randomly assigning people to either receive the experimental drug or to receive nothing at all, as would normally occur in in vivo studies. Why give the human control group a placebo? The answer is to guard against a “placebo effect.” This effect is both psychological and physical and without giving the control group a placebo it is difficult to disentangle the effect of an experimental drug from the effect that would occur if the patient were given a placebo.

Finally, why is it useful to employ a double-blind design? Recall that “double blind” means that neither the patient nor the treating physician knows whether the patient is receiving the experimental drug or a placebo (or a standard treatment). The advantage of employing a double-blind design is two-fold. Research on the placebo effect indicates that when patients know that they do not know whether they are getting the placebo or the experimental drug, the placebo effect is somewhat reduced. And blinding the treating physician eliminates the possibility that she would provide the patient with some unintended clue as to whether the patient is receiving the experimental drug.

Some adverse events may be too rare to be detected in even relatively large clinical trials. For example, even if a clinical trial includes 2,000 subjects, if an adverse event occurs in 1 in 5,000 subjects, perhaps because of some unique susceptibility in a subgroup of the population, the clinical trial lacks sufficient power to detect the adverse event. There is a second way as well in which Phase II and phase III trials may fail to detect an adverse event. Increasingly, new drugs are intended to treat chronic rather than acute diseases. Unless clinical trials ran for many years, they may be unable to detect adverse effects from long-term use. Because of these limitations, the FDCA has been amended to call for what are now called Phase IV studies, often called Postmarketing Surveillance Trials. These, once again, are epidemiological studies designed to (1) compare a drug with other drugs already in the market; (2) monitor a drug’s long-term effectiveness and impact on a patient’s quality of life; and (3) determine the cost-effectiveness of a drug therapy relative to other traditional and new therapies. Phase IV studies can result in a drug or device being taken off the market or restrictions of use could be placed on the product depending on the findings in the studies.

Cyclooxygenase (COX) inhibitors are an instructive example of the complexities surrounding both beneficial and harmful long term drug effects. Cyclooxygenase is an enzyme that facilitates the production of compounds called prostaglandins. Prostaglandin-mediated physiological effects can be blocked by COX inhibitors. Aspirin is the best known COX inhibitor. The therapeutic benefits of this and other COX inhibitors include a lowering of elevated body temperature, pain relief, anti-inflammatory effects, and reduced platelet aggregation. However, aspirin therapy also produces gastric ulcers due to its effect on stomach acid secretion, prolonged bleeding time, and, in some cases, the onset of Reye’s syndrome in children.

In the late 1980s, researchers discovered that there is a family of COX enzymes. Importantly, prostaglandins whose synthesis involves the cyclooxygenase-I enzyme, or COX-1, are responsible for maintenance and protection of the gastrointestinal tract. Prostaglandins whose synthesis involves the cyclooxygenase-II enzyme, or COX-2, are responsible for inflammation and pain. Thus, inhibiting COX-2 could produce anti-inflammatory effects associated with diseases such as arthritis, without the adverse gastrointestinal tract effects. In relatively short order, pharmaceutical companies developed several COX-2 inhibitors, including Celebrex, Vioxx, and Mobic, all of which were approved by the FDA in the late 1990s, having passed through all stages of new drug approval. Only in 2004, in a new clinical study designed to investigate whether Vioxx would be beneficial in deterring the development of colon polyps, did it become fully clear that the drug increased the risk of heart attack and stroke. Although another study of Celebrex using a lower dosage did not find a significant increase of these adverse side effects, it appears that all COX-2 inhibitors, and indeed all nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen, do increase cardiovascular risks. The maker of Vioxx voluntarily removed it from the market place and ultimately reached a settlement of a class action for \$4.85 billion. Both Celebrex and Mobic remain on the market. The push toward more formal Phase IV studies is motivated in large part by the hope that with such studies in place we can more quickly detect adverse effects of drugs taken for chronic problems such as arthritis.

7. Extrapolation

As noted above, a primary limitation of both in vitro and in vivo animal studies is one of extrapolation. This section focuses on extrapolation from animal studies. Extrapolation is a question of external validity. External validity involves the ability to generalize conclusions from one setting to another, in this case, from the results of an animal study to humans. External validity threats are particularly significant in all non-epidemiological research.

a. Dose Extrapolation

Recall from the earlier discussion that through the calculation of a NOAEL or a BMD, toxicologists may establish a “safe” dose for an animal model, that is, a dose that does not cause a selected increase in the number of animals developing a particular disease. From this starting place, one still must extrapolate this dose to an appropriate human dose.

Dose extrapolation involves two related dose issues: dose extrapolations across animals of different size and extrapolations that are necessary because animal tests expose the animals to dose levels greater than expected human doses. As to the first extrapolation, the question is, if one gives a laboratory animal an X-mg dose, what is the human equivalent? The question falls within the field of allometry: the study of size and its consequences. Unfortunately, there is no uniformly accepted formula for the extrapolation across species. Traditionally, three methods of extrapolation have been used; body mass equivalence, caloric scaling, and surface

area equivalence. However, advances in pharmacokinetics suggest other considerations as well, such as bioavailability.

The second dosage extrapolation relates to the relative dose given an animal compared to the relative dose experienced by the average human who is exposed to the substance. As discussed above, because many substances produce an adverse effect in only a small percentage of organisms when ingested at a rate similar to that encountered in the environment or prescribed by a physician, it takes a large number of animal subjects to detect a substance's effects with any reliability. Smaller samples would generate an unacceptably large number of false negatives (failure to detect an effect when it exists), a threat to internal validity which we discuss below. Given these factors and given the expense of animal studies, researchers assessing whether some substance is toxic may expose animals to relatively large doses to ascertain if there is any effect (and to guard against the potential for false negatives). If there is a positive result, toxicologists must then extrapolate a predicted incidence in humans at a more realistic lower dose rate. This is the purpose of calculating NOAELs and BMDs.

Unfortunately, precision is impossible. There are a number of ways in which high-dose toxicity testing differs from lower dose effects: there may be limits to the solubility of the compound; enzymes may become saturated at high doses, limiting absorption; detoxification mechanisms in the liver and elsewhere may be saturated; and metabolites may cause toxicity that would not occur with lower doses. All of these factors may produce non-linear effects and, therefore, extrapolation to dosages that reflect typical human exposure is problematical (which, one may recall, is a reason to prefer BMD calculations to NOAEL calculations in some circumstances). This is especially the case if the animals were exposed to relatively high doses.

As one FDA advisory guidance to industry notes:

For pharmaceuticals with low rodent toxicity, use of the MTD can result in the administration of very large doses in carcinogenicity studies, often representing high multiples of the clinical dose. This has led to the concern that exposures in rodents greatly in excess of the intended human exposures might not be relevant to human risk; because they so greatly alter the physiology of the test species, the findings might not reflect what would occur following human exposure.¹

Whether one employs a NOAEL or a BMD approach, the results of these studies permit the establishment of a reference dose (RfD) or reference concentration (RfC) of a toxin, that is, a dose or a concentration above which humans should not be exposed. These reference doses begin with the NOAEL or BMDL results and then lower the reference dose based on both the quality of the animal studies and uncertainty factors. These factors include, most importantly, a factor to adjust for the uncertainty concerning the differences between animal models and

¹ Food and Drug Administration, Guidance for Industry S1C(R2) Dose Selection for Carcinogenicity Studies, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm074919.pdf>.

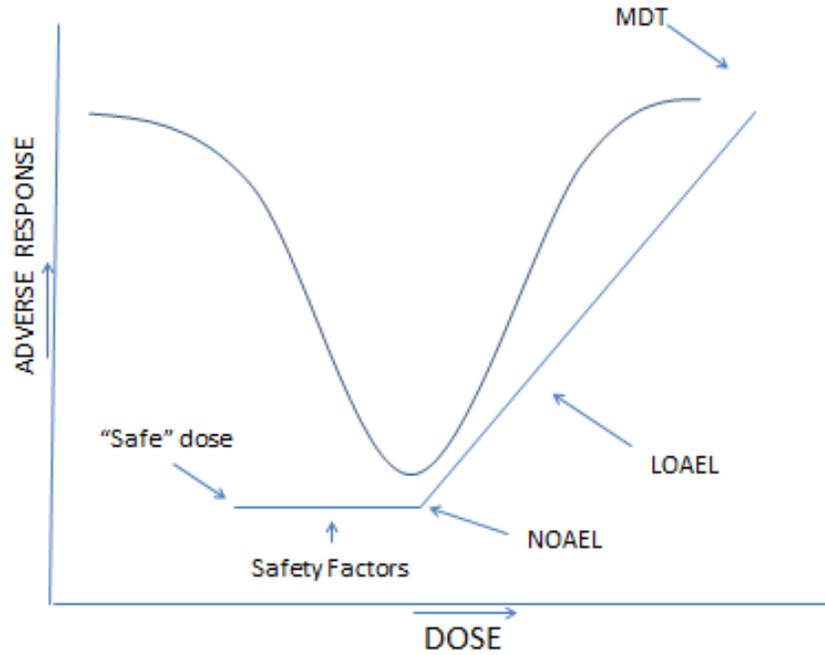
humans (inter-species variability), a factor to account for differences in individual susceptibility (intra-species variability), and in a NOAEL analysis uncertainty due to data base deficiencies. After these adjustments, the RfD or RfC may be a dose that is as low as one one-thousandths of the NOAEL.

RfD and RfC are not the only terms used to describe acceptable doses. The terms “acceptable daily intake” (ADI) and “tolerable daily intake” (TDI) are used by the Food and Drug Administration and the World Health Organization. One may encounter other terminology as well. However, the value of all of these terms begins with the calculation of a NOAEL, BMD or some similar concept and then divides this value by appropriate safety factors.

Note that even when we begin with a no-threshold model, that is, a model that assumes there is no harmless dose, RfDs often need to be established. For example, even if it is the case that exposure to any level of ionizing radiation is harmful, nevertheless we need to establish permissible exposures for those whose job requires them to come in contact with radiation. As in other situations, one derives a slope that measures risk per unit of dose and then selects an “acceptable” level of risk and thus an acceptable dose. Here, as is the case with other RfD and similar calculations the established limits do not mean that this level of exposure is risk free for all individuals.

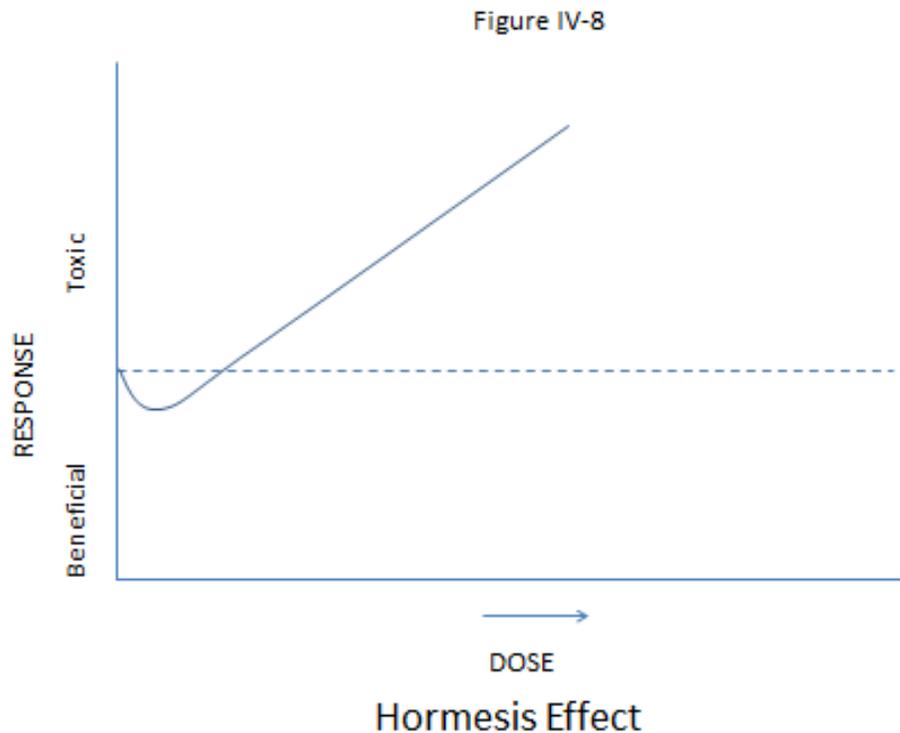
Both the NOAEL/BMD approaches with respect to noncancerous toxins and the no-threshold approach to toxins that cause cancer assume a monotonic effect. That is, the dose-response curve increases (or decreases) through the entire dose range as in Figures IV-4 and IV-5. Some toxicologists have suggested that with respect to some toxins the dose response curve is nonmonotonic. However, depending on the toxic substance, the argument comes in two different forms. In one form of the argument, very low doses may be more harmful than would be predicted by adopting an assumption of a monotonic curve. A number of toxicologists believe that endocrine-disrupting chemicals, one of which is DES, are harmful at very low doses; that is doses below those that would be established through the usual way in which regulatory toxicologists establish RfDs based on NOAEL or BMD analyses. Figure IV-7 illustrates this argument. The “safe dose” established by traditional methods may in fact set a dose that is more harmful than somewhat higher doses.

Figure IV-7



SOURCE: Laura N. Vandenberg et al., "Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses," 33(3) *Endocrine Reviews* 378 (2012). Copyright © 2012. The Endocrine Society. Reproduced with Permission.

A second form of the argument adopts the position that at very low doses a toxic substance may in fact be beneficial rather than harmful. This effect is called hormesis. Figure IV-8 illustrates this argument.



SOURCE: Adapted from Edward Calabrese, "Hormesis: Principles & Applications," *Homeopathy* (2015) 104, 69, Figure 1B. Copyright © 2015. The Faculty of Homeopathy.

In a short *Science* article, Kaiser notes, for example, that "dioxin and its chemical cousins are among the most deadly compounds on Earth. Spike a rat's water with 10 parts per billion—the equivalent of 7 teaspoons of dioxin dissolved in an Olympic-sized swimming pool—and there's a 50/50 chance that the rat will die of liver cancer. Yet even tinier concentrations of dioxins fed to rats inhibit tumors." Jocelyn Kaiser, *Sipping From a Poisoned Chalice*, 302 *Sci.* 376 (Oct. 17, 2003).

Whether or not one accepts the nonmonotonic hypotheses, they do point out the fact that linear models, especially with respect to potential carcinogens can be quite problematic in the regulatory arena where permissible thresholds are far from the experimental data. If regulators wish to set permissible exposures to one additional cancer per million over a 70-year lifetime, this extrapolation leads to doses many orders of magnitude below the experimental data. For example, the data used to establish a RfD or RfC may be based on bioassays that exposed rodents to hundreds of milligrams per unit of body weight per day, while the permitted dose to humans may be a fraction of a microgram per body weight per day. Recall that this was the situation in the Bendectin CD Rat study.

b. Extrapolation across Species

Setting aside dosage issues, as the thalidomide and gasoline examples and the discussion of failure rates in clinical trial indicate, extrapolation across species remains problematic because of species differences per se. Here it is important to distinguish exposure and dose. Even when exposure to some concentration of a substance in the environment is the same for animals and humans, the dosage to which each is exposed may differ based on factors such as the route of absorption. For example, because human dermis is more vascularized than that of most laboratory animals, finding a suitable animal model is more difficult with respect to dermal absorption.

The cross-species issue becomes particularly significant when adverse animal studies cast doubt on the safety of some substance for humans. The artificial sweetener saccharin provides a good example. Like a number of other artificial sweeteners, that is, aspartame and cyclamate, saccharin was under a cloud because animal studies suggested it might be a carcinogen. Based on these studies, Congress mandated that products containing saccharin contain the following warning: "Use of this product may be hazardous to your health. This product contains saccharin, which has been determined to cause cancer in laboratory animals." Subsequent animal research indicated that other artificial sweeteners were not animal carcinogens, but studies in rats do show an increased incidence of urinary bladder cancer at high doses of saccharin, especially in male rats. However, mechanistic studies have shown that these results are the result of the presence of the protein α_2 globulin, high pH, and high levels of calcium phosphate that result in the formation of precipitates in the bladder, causing irritation, cell proliferation, and ultimately tumors. Based on this conclusion and the failure to detect adverse effects in epidemiological research, saccharin was delisted in 2000 from the National Toxicology Program's Report on Carcinogens and Congress repealed the warning label requirement.

This and other example of differences in metabolic processing of a chemical in the animal species and in humans demonstrate that positive carcinogenicity bioassays in rodents certainly suggest further research but one must guard against a knee-jerk extrapolations to humans before there is evidence that the animal in question and humans share similar mechanisms. Of course the case for a human effect is much stronger if there are significant effects in several animal species and strains of laboratory animals.

c. Susceptibility

On top of the fact that animal species may be more or less susceptible to a substance than humans, there is the additional complication that there is substantial variation in human susceptibility. Some variation is the result of differential absorption. For example, iron deficiency, which is more likely among disadvantaged children, increases the rate of absorption of ingested lead. The rate of metabolic processes also may vary across individuals. And,

obviously existing conditions may affect the extent to which a toxin interferes with one's daily functions. A lung toxin may have little impact on an otherwise healthy individual but a substantial impact on one with an existing loss of lung function.

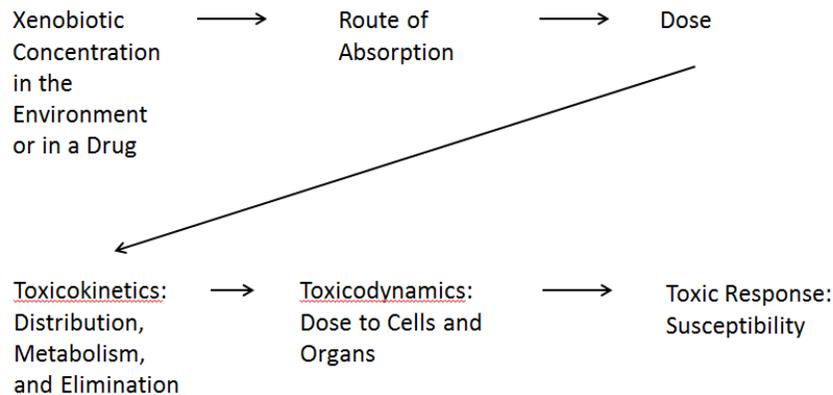
Even when dose is a constant and individuals appear to be in similar states of health, we observe substantial variation in outcomes. For example, the relative risk of getting lung cancer is quite high for smokers. One study of women smokers indicated a relative risk of 5.5 for those smoking 1 to 10 cigarettes a day to a relative risk of 20.0 for those smoking more than 31 cigarettes a day. Nevertheless, the great majority of average smokers (perhaps 85 to 90%) do not contract this disease. A full understanding of variations in susceptibility is still out of our reach, but some factors are understood. Obviously, genetics plays a role. Of course, this basic insight is as old as the observation that some illnesses run in families, but increasingly, we are gaining a better understanding the way in which specific genetic makeups predispose people to a disease. The best known current example is the existence of certain mutations in the BRCA1 or BRCA2 genes that greatly increase a person's risk of developing breast cancer and ovarian cancer. To offer another example, the way in which some ethnic groups metabolize alcohol may affect the incidence of alcohol dependence. Note, however, that this type of information tells us about predisposition to a certain illness, not how a particular toxin interacts with one's genetic makeup to produce an adverse outcome. We discuss this at greater length in the genetics materials in Section V.

Variations in susceptibility are not solely a matter of endogenous factors, however. They may be the result variations in the level of exposure to other exogenous factors. One of the best understood of these effects is the synergistic effect of tobacco use and asbestos exposure on the likelihood of contracting lung cancer, a topic discussed above in the epidemiology section. Recall that a history of smoking makes individuals much more susceptible to lung cancer when they are exposed to nonbackground levels of asbestos. At a more aggregate level, some communities—often poorer communities—expose their residents to a greater variety as well as a greater concentration of potentially hazardous substances whose cumulative effect is difficult to quantify. Nevertheless, toxicologists are devoting increasing attention to the problem of assessing the effects of cumulative exposures.

Figure IV-9 summarizes much of the preceding discussion in a graphical overview of exposure, dose, and injury.

Figure IV-9

Overview of Exposure, Dose, and Injury



SOURCE: Courtesy of the authors.

d. The Quality of In Vivo Studies

Before we turn to legal cases involving toxicological evidence we should make one more point about the importance of assessing the quality of a particular toxicological study before accepting its conclusions. As the preceding discussion suggests, *in vitro* and *in vivo* studies routinely raise questions of external validity. However, researchers must also be concerned with questions of internal validity. Internal validity addresses the question of whether it is correct to infer that an observed relationship between two variables is causal or that the absence of an observed relationship implies the absence of cause. Threats to internal validity usually can be thought of as specification errors. Specification errors occur when the researcher fails to consider a factor or factors (i.e., other variables) that mediates the observed effect between two variables, either because they explain changes in both the “cause” and the “effect” or intervene between the “cause” and the “effect” and act independently on the “effect.” As we discussed above, epidemiology research routinely confronts threats to internal validity because it cannot use randomization to control for all variables that may explain an observed relationship. On the other hand, by using randomization and blinding, clinical trials and well conducted animal studies can greatly increase internal validity.

To this point, we have proceeded as if all *in vivo* studies employ these methods. However, a fair amount of research now indicates that at least with respect to animal studies conducted to test new drugs, internal validity problems are more frequent than one would hope. This research was generated by the observation that positive therapeutic effects on animals often are not replicated in human clinical trials. Recall from the discussion of clinical

trials that only a small percentage of drugs that show promise in animal studies offer therapeutic benefits in humans. Apparently, one set of reasons this is that case is that a substantial number of animal studies are not conducted using randomized, blinded designs. For example, one study evaluated 290 animal study abstracts containing two or more experimental groups that were accepted by the Society for Academic Emergency Medicine. Of the 290 study abstracts, 194 were not randomized and 259 were not blinded. The nonrandomized and nonblinded studies had 3.4- and 3.2-fold higher odds, respectively, of claiming a statistically significant outcome than did those that were randomized and blinded. Other studies have produced similar results. Publication bias (the tendency to publish research only when the results are positive) also helps to explain the similar tendency to find therapeutic effects in animals that do not translate to humans. Note that the FDA's statements of good laboratory practices, to be found at 21 C.F.R. § 58 are designed to avoid problems such as failed randomization.

Similar censuses have not been done to assess what percentage of animal studies simply designed to test whether a substance is toxic are likewise flawed, but this result should give us additional reasons to be cautious in extrapolating animal study findings to humans. Again, however, good laboratory practices guidelines indicate what should be done. For example, some of the EPA's guidelines may be found at 40 C.F.R. § 160 (a) prescribing good laboratory practices for conducting studies that support or are intended to support applications for research or marketing permits for pesticide products and 40 C.F.R. 792, prescribing good laboratory practices for conducting studies relating to health effects, environmental effects, and chemical fate testing.

B. Toxicology and the Law

The law's relationship to toxicology may be usefully divided into several subcategories. First, one must distinguish between public law regulatory uses of toxicology information versus the use of toxicology in litigation concerning individual cases. Second, within this latter category it is worthwhile to distinguish criminal cases and civil cases.

1. The Use of Toxicology in Regulation

Multiple federal and state statutes govern exposure to chemicals and drugs. Many require the relevant governmental agency to conduct risk-benefit analyses and to set threshold exposures for various substances. For example, the Occupational Safety and Health Administration (OSHA) regulates exposures in the workplace under the Occupational Safety and Health Act, 29 U.S.C. §§ 650 et seq. The Food And Drug Administration (FDA) governs the testing of new prescription drugs and some other substances such as tobacco smoke, food, food additives and cosmetics under the Food Drug and Cosmetics Act (FDCA), 21 U.S.C. §§ 301 et seq. Most importantly, the Environmental Protection Agency (EPA) regulates emissions from

power plants and other stationary and mobile sources of air pollution under the Clean Air Act, 42 U.S.C., §§7401 et seq., as well as discharges into rivers and streams under the Clean Water Act, 33 U.S.C. §§ 1251 et seq. Under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), 7 U.S.C. § 136, it regulates and oversees substances that are commercially important but are designed to kill things. It also has responsibility for the handling of hazardous waste from its creation through its disposal under the Resource Conservation and Recovery Act (RCRA), 42 U.S.C. §§ 6901 et seq. and clean-up of hazardous substance releases under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), 42 U.S.C. §§ 9601 et seq.

One of the weakest aspects of the federal regulatory regime is oversight of nonpharmaceutical chemical agents not intended for use in humans and not yet known to have adverse effects on people or the environment. Under the Toxic substance Control Act (TSCA), 15 U.S.C. §§ 2601 et seq., all that is required is premarketing notice to the EPA. No prior testing is required. The European Union has a more comprehensive regulatory structure commonly known by its acronym REACH, which calls for the regulation, evaluation, authorization and restriction of industrial chemicals. It does appear that TSCA may be updated in some fashion in the near future. Plaguing both approaches is the fact that there are so many chemicals, mixtures of chemicals and other potential toxins, and far too many potential adverse health endpoints for it to be feasible to test all exogenous agents at environmentally appropriate doses for all potential health consequences. Full safety assessments exist for one percent of the estimated 60 to 75 thousand chemicals in commerce and for as many as 80% there is no significant toxicological data at all.

Many of these laws task the relevant agency to engage in a risk assessment and to establish RfCs and RfDs using in vitro, in vivo and, where available, epidemiology data. Some agencies have established quite low RfCs. For example, under CERCLA, the EPA may require responsible parties to remove or treat potential carcinogens to a level that will reduce risk to one in a million at Superfund sites.

With some regularity, agency risk-benefit and risk-risk decisions are challenged in court. At stake is the reasonableness of the types of decisions discussed above. Following is one case where the court rejected an EPA determination.

Chlorine Chemistry Council v. Environmental Protection Agency
United State Court of Appeals, District of Columbia, 2000.
206 F.3d 1286.

Williams, J.

The Safe Drinking Water Act (“SDWA” or the “Act”) directs the Environmental Protection Agency to set standards for the regulation of covered drinking water contaminants. For each EPA sets a “maximum contaminant level goal” (“MCLG”), defined as “the level at which no known or anticipated adverse effects on the health of persons occur and which allows an adequate margin of safety.” 42

U.S.C. § 300g-1(b)(4)(A). The MCLG is somewhat aspirational. After having set it, EPA is to promulgate an enforceable standard, known as a maximum contaminant level (“MCL”), which takes practical considerations into account while remaining “as close to the [MCLG] as is feasible.” § 300g-1(b)(4)(B).

In March 1998 EPA concluded that chloroform, a drinking water contaminant, exhibits a “nonlinear mode of carcinogenic action.”¹ In other words, exposures to chloroform below some threshold level pose no risk of cancer. But in promulgating the MCLG it retained the existing standard of zero, which was based on the previously held assumption that there was no safe threshold. * * * Petitioners, including the Chlorine Chemistry Council, a trade association comprised of chlorine and chlorine product manufacturers, petitioned this court for review, arguing that EPA violated its statutory mandate to use the “best available” evidence when implementing the provisions of the Safe Drinking Water Act. 42 U.S.C. § 300g-1(b)(3)(A). We agree.

Chloroform, a “nonflammable, colorless liquid,” is one of four compounds that together are classed as “Total Trihalomethanes” (“TTHMs”). These are byproducts of chlorination, the most widely used technique for ensuring the safety of drinking water. Chlorination plays a significant role in the control of microbial pathogens and in turn in the protection of public health; but on the basis of rodent tumor data the Agency has concluded that chloroform, a byproduct of this process, acts as a probable human carcinogen.

On July 29, 1994 EPA issued a proposed rule on disinfectants and disinfection byproducts in water. This included a zero MCLG for chloroform, based on EPA's finding of an *absence* of data to suggest a threshold level below which there would be no potential carcinogenic effects. *Id.* The Agency's default method of inferring risk at exposure levels for which it has no adequate data is linear extrapolation from cancer incidence inferred at exposures for which it does have data. Thus, either if the evidence supports linearity, *or* if there is “insufficient” evidence of nonlinearity, EPA assumes that if a substance causes cancer at *any* exposure it will do so at *every* non-zero exposure (though with cancer incidence declining with exposure). But EPA acknowledges its authority “to establish nonzero MCLGs for carcinogens if the scientific evidence” indicates that a “safe threshold” exists. And petitioners here assume the validity of the linear default assumption.

In 1996 Congress amended the SDWA, enshrining in the statute a timetable previously set by EPA for rules relating to disinfectants and disinfection byproducts associated with water treatment. 42 U.S.C. § 300g-1(b)(2)(C); Proposed Rule: National Primary Drinking Water Regulations: Monitoring Requirements for Public Drinking Water Supplies, 59 Fed.Reg. 6332, 6361 (1994). The relevant deadline here was November 1998. In preparation for the necessary rulemaking EPA formed an advisory group in 1997 whose purpose was “to collect, share, and analyze new information and data, as well as to build consensus on the regulatory implications of this new information.”

On the basis of the committee's findings and recommendations, EPA in November 1997 published a Notice of Data Availability (“NODA”), 62 Fed.Reg. 59,388 (1997), and in 1998 it published a second NODA specific to chloroform, 63 Fed.Reg. 15,674 (1998). Among the findings it discussed were

¹ [Ed. Note: The court's use of the term “nonlinear mode of carcinogenic action” is referring to what we have called a threshold assumption; that is, there is some threshold below which a substance is not harmful. On the other hand, when the court uses the terms “linear mode” or “linear extrapolation” it is referring to what we have called non-threshold models that assume that exposure to any amount of the substance has the potential to cause cancer. The use of this terminology is unfortunate because it potentially confuses two issues: (1) whether there is a threshold below which a substance is not harmful and (2) the shape of the dose-response curve which may or may not be linear regardless of whether there is a threshold.]

those arrived at by a panel of experts organized by the International Life Sciences Institute. The panel, whose work was subject to independent peer review and was convened under the auspices of the EPA, concluded on the basis of chloroform's mode of action that although it was "a likely carcinogen to humans above a certain dose range, [it was] unlikely to be carcinogenic below a certain dose range." The panel recommended "the nonlinear [] or margin of exposure approach [as] the preferred approach to quantifying the cancer risk associated with chloroform exposure."

EPA agreed. It said that "[a]lthough the precise mechanism of chloroform carcinogenicity is not established," nevertheless "the chloroform dose-response should be considered nonlinear." Rather than operating through effects on DNA, which is consistent with linearity, chloroform evidently works through "cytotoxicity" (i.e., damage to the cells) followed by regenerative cell proliferation. *Id.* Employing the threshold approach that it found was entailed by chloroform's mode of action, EPA then calculated an MCLG of 600 parts per billion ("ppb"), based solely on carcinogenicity. This level built in a 1000-fold margin of error in relation to the maximum safe dosage implied from the animal studies used by EPA. *Id.* But because even lower chlorine doses cause liver toxicity (a non-cancer effect), EPA proposed an MCLG of 300 ppb.

When EPA came to promulgate its final rule in December 1998, however, its MCLG was again zero. It stuck with 1994's zero level despite its explicit statement that it now "believe[d] that the underlying science for using a nonlinear extrapolation approach to evaluate the carcinogenic risk from chloroform is well founded." It justified the action on the basis that "additional deliberations with the Agency's SAB on the analytical approach used" and on the underlying scientific evidence were needed "prior to departing from a long-held EPA policy." It could not complete such additional deliberations by the November 1998 statutory deadline, and, moreover, the rulemaking would not affect the enforceable MCL for TTHMs.

After briefing on the petition for review at issue here, but before oral argument, EPA moved for a voluntary remand to consider the SAB report on chloroform that would soon be available. But EPA made no offer to vacate the rule; thus EPA's proposal would have left petitioners subject to a rule they claimed was invalid. We denied the motion.

On February 11, 2000, the day of oral argument, EPA released a draft report by the SAB on chloroform. The report concluded that chloroform exhibits a "cytotoxic" mode of action. Such a mode of action (unlike a "genotoxic" mechanism, which acts directly on a cell's DNA) involves no carcinogenic effects at low doses; thus a nonlinear approach is "scientifically reasonable." After consideration of the draft SAB report, EPA stated that it "no longer believes that it should continue to defend its original decision," and moved that this court vacate the MCLG.

[In an omitted part of the opinion, the court determined that the petitioners do have standing because in other arenas the EPA has set very low chloroform cleanup goals that are based on the assumption that chloroform poses a risk of cancer at any dose.]

* * *

On the merits petitioners argue that EPA's decision to adopt a zero MCLG in the face of scientific evidence establishing that chloroform is a threshold carcinogen was inconsistent with the Safe Drinking Water Act. Section 300g-1(b)(3)(A) of the Act states unequivocally that "to the degree that an Agency action is based on science, the Administrator shall use * * * the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices." In promulgating a zero MCLG for chloroform EPA openly overrode the "best available" scientific evidence, which suggested that chloroform is a threshold carcinogen.

EPA provides several arguments in defense of its action. First, it argues that to establish a non-zero MCLG would be a “precedential step,” that represents “a major change in the substance of regulatory decisions related to chloroform.” We do not doubt that adopting a nonzero MCLG is a significant step, one which departs from previous practice. But this is a change in result, not in policy. The change in outcome occurs simply as a result of steadfast application of the relevant rules: first, the statutory mandate to set MCLGs at “the level at which no known or anticipated adverse effect on the health of persons occur,” 42 U.S.C. § 300g-1(b)(4)(A), as determined on the basis of the “best available” evidence; and second, EPA’s Carcinogen Risk Assessment guidelines, stating that when “adequate data on mode of action show that linearity is not the most reasonable working judgment and provide sufficient evidence to support a nonlinear mode of action,” the default assumption of linearity drops out. Proposed Guidelines for Carcinogen Risk Assessment, 61 Fed. Reg. 17,969/1. The fact that EPA has arrived at a novel, even politically charged, outcome is of no significance either for its statutory obligation or for fulfillment of its adopted policy.

* * *

EPA justifies its decision not to adopt a nonzero MCLG on the basis that it had to reevaluate one of its underlying technical assumptions—that ingestion of chloroform in drinking water accounts for 80% of total exposure to chloroform. As it stated in its final rule, EPA is currently considering use of a 20% relative source contribution for drinking water, which would lower the MCLG to 70 ppb. Along these lines, EPA’s counsel conceded at oral argument that a science-based MCLG would fall into the interval between 70 and 300 ppb. The uncertainty on this issue may have provided support for choosing the lowest nonzero MCLG from within that interval, but none for choosing an MCLG *outside* the range of uncertainty.

* * *

Finally, EPA argues that its statements in the 1998 Notice of Data Availability do not represent its “ultimate conclusions” with respect to chloroform, and thus in adopting a zero MCLG it did not reject what it considered to be the “best available” evidence. In fact, the zero MCLG merely represented an “interim risk management decision” pending the final SAB report. We find these semantic somersaults pointless. First, whether EPA has adopted its 1998 NODA as its “ultimate conclusion” is irrelevant to whether it represented the “best available” evidence. All scientific conclusions are subject to some doubt; future, hypothetical findings always have the potential to resolve the doubt (the new resolution itself being subject, of course, to falsification by later findings). What is significant is Congress’s requirement that the action be taken on the basis of the best available evidence *at the time* of the rulemaking. The word “available” would be senseless if construed to mean “expected to be available at some future date.” Second, EPA cannot avoid this result by dubbing its action “interim.” The statute applies broadly to any “[a]gency action”; whether the action is interim is irrelevant.

* * *

Finding the Agency’s December 1998 rule adopting a zero MCLG for chloroform to be arbitrary and capricious and in excess of statutory authority, see 5 U.S.C. § 706(2)(A) & (C), we vacate the rule. A separate order on briefing additional remedies will issue shortly.

So ordered.

Notes and Questions

1. The EPA's current statement about the risks of chloroform may be found at <http://www3.epa.gov/airtoxics/hlthef/chlorofo.html> Among other things, it states on the web page that:

EPA has determined that although chloroform is likely to be carcinogenic to humans by all routes of exposure under high-exposure conditions that lead to cell death and regrowth in susceptible tissues, chloroform is not likely to cause cancer in humans by any route of exposure under exposure conditions that do not cause cell death and regrowth. Therefore, EPA has not derived either an oral carcinogenic potency slope or an inhalation unit risk for chloroform.

2. What is the difference between the MCL and the MCLG? How is each calculated?
3. As should be obvious from the opinion, the Clean Water Act requires the EPA to weigh the costs and benefits of any level of safety. With respect to this statute, the agency must examine whether the drinking water regulations are overly costly compare to the health benefits they would provide resulting in increased and unjustified costs for water suppliers. See 42 U.S.C. § 300g-1(b)(3)(C), (b)(6)(A).
4. Several years after the Chlorine Chemistry Council case, the EPA faced a similar challenge, this time with respect to radionuclide levels in water. *City of Waukesha, v Environmental Protection Agency*, 320 F.3d. 228 (D.C. Cir. 2003). This time the court found the agency had sufficient justification for setting the MCLG at 0 µg/L given that there was contradictory data on the question of whether a linear no-threshold model was appropriate.
5. The Safe Drinking Water Act requirement that the agency use the "best available" evidence is not replicated in every statute. Its absence provides the agency greater leeway.
6. For a very useful summary of cases in which the courts are asked to reverse agency decisions, see Caroline Cecot & W. Kip Viscusi, *Review of Agency Benefit-Cost Analysis*, 22 GEO. MASON L. REV. 575 (2015).

2. Toxicology in Litigation

a. Toxicology in Criminal Cases

In criminal law, one of the areas in which toxicology frequently plays a role is drug testing, for example, how long it takes to metabolize a drug and the physiological effects of drugs and alcohol in an individual's system. Samples of urine, blood, breath, and hair may be used to estimate exposure and dose. Toxicology plays an especially important role is alcohol testing, usually in the context of charges of drunk driving or vehicular homicide. The toxicology of the effects of alcohol consumption is well studied.

Alcohol, more specifically ethanol, affects all systems of the body to some degree due to

its distribution into aqueous (water) areas throughout the body, with major acute and chronic effects on the central nervous system (CNS), gastrointestinal system, the liver, the cardiovascular system, and the endocrine system.

However, alcohol has its greatest short term effect on the CNS, and it is these effects that are the most important in the law-science interface. As with other chemicals, there is a significant dose-response relationship. Factors that influence alcohol's effect are the rapidity and recency with which the alcohol was consumed and the amount of food in the gastrointestinal tract at the time of the drinking.

Alcohol consumption produces significant decrements of visual acuity, tracking, division of attention, and reaction time, all of which are involved in driving a motor vehicle. (Surprisingly, it appears that the mechanism by which ethanol produces these effects is not fully understood.) Reaction time is particularly adversely affected. In complicated tasks in which subjects perform more than one attentive task at a time, typical of most driving situations, alcohol levels averaging 0.11 g/dL, increase reaction times up to 200%. It is not surprising, therefore, that the law typically criminalizes driving while intoxicated, creating several issues at the law/science interface.

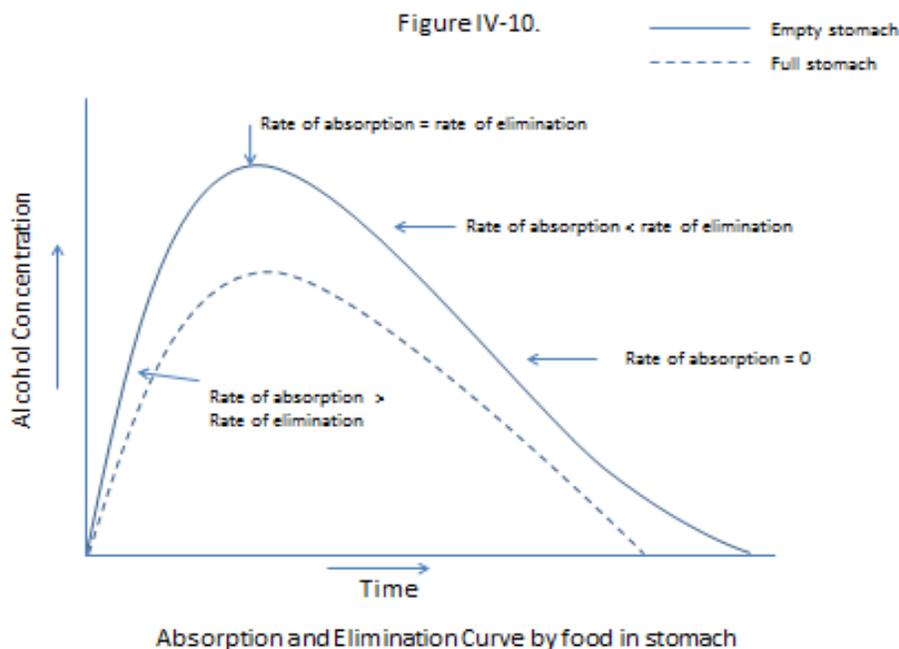
One of the more difficult legal issues with respect to alcohol consumption arises from the fact that many statutes contain a per se violation provision making it an offense to operate a motor vehicle when the driver's blood alcohol concentration is above some threshold, typically 0.08 g/dL. What does a test conducted at some later time tell us about the level of intoxication of the defendant while driving? The question, of course, is one of pharmacokinetics. Almost all states sidestep the extrapolation issue in most situations by providing that if a blood or breath test is done within 2 or 3 hours of the time of driving no extrapolation is necessary. However, under some circumstances, states may require the prosecution to relate the blood alcohol concentration (BAC) level at the time of the test back to the BAC level at the time of driving. This back-extrapolation is often called *retrograde extrapolation* in legal opinions.

One real advantage in the study of pharmacokinetics of alcohol is that we can study it in humans, rather than laboratory animals and thus no animal to human extrapolation is necessary. Not surprisingly, there is no dearth of individuals prepared to participate in alcohol research! Based on the results of this research, we know that an individual's BAC at any point in time is affected by when one began to drink, total consumption, time of last drink, and variables that influence the rate of absorption of alcohol in the stomach and small intestine and variables that influence the rate of elimination of alcohol from the body. With respect to rate of absorption, the most important variable appears to be whether one is drinking on an empty stomach or not. Food requires digestion and any alcohol trapped in food particles will take longer to be absorbed. In almost all situations, however, absorption is complete within 90 to 120 minutes after cessation of drinking.

Additional factors also influence the BAC in an individual at a given time. Because alcohol is soluble in water the alcohol content in the body is proportional to the total body water content. As a result, body weight and relative proportion of body fat affect BAC.

The elimination of alcohol primarily occurs through a process of oxidation. The great majority (over 90%) of the elimination of alcohol from the body occurs in the liver, with small amounts excreted in sweat, breath, and urine. (The fact that some alcohol is excreted in breath makes breath tests for alcohol concentration possible.) When alcohol is absorbed into the body it first passes through the liver. The process by which a substance is metabolized before entering the general circulation is called first-pass metabolism (FPM). Many toxic substances undergo hepatic FPM. In the first step of hepatic FPM of alcohol, an oxidative enzyme called alcohol dehydrogenase (ADH) converts alcohol into acetaldehyde. When the quantity of alcohol reaching the liver exceeds the metabolic capacity of the available ADH, the remainder passes into general circulation, raising the BAC. Because of the limited capacity of ADH, elimination of alcohol from the bloodstream occurs in a relatively straight line fashion (a zero-order process of elimination), although, as the court discusses in the following case, the rate of elimination varies from person to person.

Because food in the stomach slows absorption while the elimination rate remains relatively constant, given the same amount of alcohol intake, usually peak BAC will be higher when drinking on an empty stomach (See Figure IV-10).



SOURCE: Republished with permission of Taylor and Francis Group LLC Books, from Karen E. Stein and Thomas M. Brown, *Principles of Toxicology*, Third Edition, 2015; permission conveyed through Copyright Clearance Center, Inc.

United States v. Tsosie
United States District Court, District of New Mexico, 2011.
791 F.Supp. 2d 1099.

Browning, J.

THIS MATTER comes before the Court on: (i) the Plaintiff's Notice of Intent to Introduce Expert Witness Testimony Pursuant to Rules 702, 703 and 705, filed on October 6, 2010 (Doc. 37); and (ii) Defendant's Motion in *Limine* for Daubert¹ Ruling Regarding the Admissibility and Scope of Ms. Nancy's Drez's Expert Testimony, filed on April 20, 2011 (Doc. 86) ("Motion"). The Court held an evidentiary hearing on April 28, 2011. The primary issue is * * * whether Plaintiff United States of America's retrograde extrapolation is admissible under rule 702 of the Federal Rule of Evidence. The Court * * * concludes that the United States has met its burden of showing its retrograde extrapolation is reliable. The Court therefore denies the Motion.

FACTUAL BACKGROUND

The charges in this case arise from a fatal crash that occurred sometime before 5:11 a.m. on October 17, 2009. Defendant John Leonard Tsosie was driving one of the vehicles involved in the crash. He told law enforcement and medical personnel who treated him at the hospital that he fell asleep at the wheel, and a test the hospital administered following a blood draw at 6:15 a.m. revealed that Tsosie had a blood alcohol concentration ("BAC") of .07 mg/mL at that time. Tsosie stated that he had three beers the night before and had stopped drinking at 11:00 p.m. Manuel Johnson was driving the other vehicle, and his wife, Loretta, was the passenger. The crash occurred approximately seven minutes from their home. Neither of the Johnsons survived.

PROCEDURAL BACKGROUND

On March 24, 2010, a Federal grand jury returned a two-count indictment charging Tsosie with two counts of involuntary manslaughter for killing M. and L. Johnson while operating a motor vehicle while under the influence of alcohol.

On October 6, 2010, the United States filed its Notice of Intent to Introduce Expert Witness, notifying Tsosie, in part, that it plans to call Nancy Drez as an expert witness. * * * Dr. Drez is a forensic toxicologist. The United States intends to call Dr. Drez as an expert witness to testify about two opinions. First, she will offer opinion testimony regarding the impairment humans suffer as their BAC increases, which includes drowsiness and significant impairment of motor skills, reaction time, and other functions critical to safe driving. Tsosie does not challenge this testimony. Second, Dr. Drez will testify regarding the rates at which the human body absorbs and eliminates alcohol. Applying these principles to the evidence in this case through retrograde extrapolation, she will testify that Tsosie's BAC would have fallen within the range of .08 to .09 mg/mL at the time of the crash. Carrying the extrapolation back earlier into the night before the crash, she also will show that Tsosie's BAC would have been in the range of .12 to .17 mg/mL at the latest point he could have been expected to start strictly eliminating alcohol from his system. The United States contends Dr. Drez' testimony will show Tsosie failed to tell the truth when he stated that he drank only three beers the night before he took the wheel. According

to the United States' Notice:

3. Nancy G. Drez is the Implied Consent Supervisor, Toxicology Bureau, Scientific Laboratory Division of the New Mexico Office of the Medical Investigator. Her CV is attached as Government's Exhibit 3. As an expert in blood and breath analysis, alcohol impairment, blood alcohol content (BAC)/ breath alcohol content (BrAC) extrapolation and other issues related to chemical testing for alcohol, Drez will testify regarding the alcohol content of defendant's blood. A summary of her opinions is attached as Government's Exhibit 3. Drez will explain the nature of impairment at that level of alcohol concentration for an average individual and then for the defendant, given the characteristics known to her about the defendant.

4. Drez will also testify regarding BAC/BrAC extrapolation since the blood samples of the defendant were approximately two hours after the fatal crash with the Johnson vehicle. Based on her training and experience, Drez will opine regarding what range of alcohol levels were likely sustained by the defendant at the time of the collision. Given the facts that will be presented at trial, *Drez will present expert testimony that the defendant's BAC/BrAC was between .07 and .13 g/100ml at the time of the collision, depending on the time of the blood draw.*

5. The United States anticipates that Drez will testify regarding the effects of alcohol on the human ability to operate motor vehicles, including the impairment of motor skills, vision impairment and drowsiness. In this case, the defendant admitted on more than one occasion that after a night of drinking alcohol, he had fallen asleep at the wheel of his moving SUV. In addition, Drez can explain principles underlying alcohol absorption, metabolism, and elimination for forensic purposes. The United States anticipates that the expert opinion will include the conclusion that the defendant's BAC/BrAC level was in excess of the legal limit at the time of the fatal crash given the evidence in the case.

* * *

At the April 28, 2011 *Daubert* hearing, Tsosie * * * agreed that retrograde extrapolation is relevant and may be admissible. * * * Tsosie further stated that he does not contest that retrograde extrapolation is a valid methodology that is widely accepted in the scientific community.

Tsosie [argued however] * * * that Dr. Drez does not have sufficient information to reach her opinion regarding the range within which Tsosie's BAC fell at the time of the collision. * * *

Dr. Drez testified about the basis of her report and her retrograde extrapolation. * * * When a person stops consuming alcohol, his or her body eventually reaches an absorption point, where the body completes absorption of the alcohol he or she has ingested, and enters the elimination phase, where the body is only eliminating alcohol. Dr. Drez testified that the general population typically reaches the elimination phase of processing alcohol within a half-an-hour to an hour after consuming the last drink, with a statistically significant group reaching the elimination phase two hours after stopping, and only rare outliers taking more than two hours to reach the elimination phase. See P.M. Ganer & W.D. Bowthorpe, *Evaluation of Breath Alcohol Profiles Following a Period of Social Drinking*, 33 *Can. Soc. Forensic Sci. J.* 137, 142 (2000) ("On average, 69 minutes elapsed from the end of drinking until the start of the linear decline in BAC, with the longest taking 124 minutes."); A.W. Jones & L. Andersson, *Influence of Age, Gender, and Blood-Alcohol Concentration on the Disappearance Rate of Alcohol from*

Blood in Drinking Drivers, 40 J. Forensic Sci. 922, 924 (1995). Once a person completes absorption and enters the elimination phase, there is a linear decline of alcohol from the system at a typical rate of .01 to .02 mg/mL/h. Ganer & Bowthorpe, *supra*, at 143; Jones & Andersson, *supra*, at 922. Dr. Drez testified that heavy drinkers eliminate alcohol more quickly—as fast as .03 mg/mL/h according to one study. (citing Jones & Andersson, *supra*, at 922, 924). Ninety-five percent of more than 1000 drinking drivers in the Jones and Andersson study eliminated alcohol at a rate between .09 and .29 mg/mL/h, and only 2.2% eliminated alcohol at a rate slower than .01 mg/mL/h. See Jones & Andersson, *supra*, at 924. Jones and Anderssons suggested that the outliers that eliminated alcohol at a rate slower than .01 mg/mL/h were not truly in the elimination phase, but still in a “slow absorption phase.” Jones & Andersson, *supra*, at 924. The average person continues to eliminate alcohol at this rate until the person's BAC reaches .02 or .01 mg/mL, at which time the rate of decline tends to taper off until all the alcohol is eliminated.

In preparing her opinion on Tsosie's BAC at the time of the accident, Dr. Drez reviewed the medical records and police reports, including the dispatch report, in this case. Dr. Drez learned from the incident report that the accident was reported to the police at 5:11 a.m. Dr. Drez also spoke with Gayla Bias, the nurse who obtained Tsosie's blood sample the morning of the accident, and confirmed that the blood sample was taken at 6:15 a.m., as reflected in Tsosie's hospital record. Dr. Drez testified that tests on the blood sample revealed a BAC of 84 mg/dL, which, using accepted formula that account for distillation that occurs during processing, is equivalent to a BAC of .07 mg/mL. She also testified that she learned from the police report that Tsosie told investigators that he stopped drinking at 11:00 p.m. the night before the accident.

Dr. Drez applied the retrograde extrapolation principles to the facts of this case. Tsosie stated that he stopped drinking at 11:00 p.m. the night before the crash. The crash occurred sometime before the crash was reported to the police at 5:11 a.m. Tsosie's blood was drawn at the hospital at 6:15 a.m., and the results from that blood test show Tsosie had a BAC of .07 mg/mL at 6:15 a.m. Based on Tsosie's statement that he stopped drinking at 11:00 p.m., Dr. Drez assumed that, even if Tsosie were an outlier, he would have starting strictly eliminating by 1:15 a.m.—two hours and fifteen minutes later. At the low end of the range of strict elimination rates of .01 mg/mL/h—the circumstances most favorable to Tsosie—he would have eliminated at least .01 mg/mL in the more than one hour period between the accident and his blood draw, placing his BAC at .08 mg/mL or above at the time of the accident. At an elimination rate of .02 mg/mL/h, Tsosie would have eliminated more than .02 mg/mL, producing a BAC level of at least .09 mg/mL at the time of the accident. Dr. Drez extrapolated further back to conclude that Tsosie had a BAC of .12 to .17 mg/mL at 1:15 a.m., which is inconsistent with Tsosie's statement that he drank only three beers. . Dr. Drez stated that, if Tsosie consumed three beers almost instantaneously, the highest BAC she would expect him to achieve is .06 to .08 mg/mL, and Tsosie would have completely eliminated the alcohol from his system by 6:15 a.m.

On cross examination, Dr. Drez stated that there are no curves in her graph, because her “graph only is looking at the elimination phase.” * * * Dr. Drez testified that she did not know when Tsosie last ate, so she used an absorption period of more than two hours to give Tsosie the benefit of the doubt * * * She further testified that she is unfamiliar with Tsosie's drinking patterns, beyond his statement that he drank three beers the night of the accident and stopped drinking at 11:00 p.m., which is why she used a range to allow for a faster elimination rate if Tsosie is a heavy drinker and a slower rate if he is a light drinker. * * * Dr. Drez also stated that studies offer conflicting conclusions whether Native Americans eliminate alcohol faster, slower, or at the same rate as other races, but that using a range of elimination rates also accounts for this uncertainty, because none of the studies indicated that Native

Americans eliminate alcohol slower than .01 mg/mL/h. * * *

ANALYSIS

Tsosie * * * argues that Dr. Drez' opinion must be excluded, because sufficient facts do not support the opinion. In particular, Tsosie suggested that Dr. Drez cannot offer an opinion based on retrograde extrapolation, because she does not know what he ate while he was drinking the night before the crash, exactly how much he drank, or over what time period he drank it. * * *

Tsosie's argument goes to the reliability requirements rule 702 imposes. *See* Fed.R.Evid. 702 (requiring that expert opinions be "supported by sufficient facts or data"). The issue before the Court is therefore whether the United States has established by a preponderance of the evidence that Dr. Drez' opinion regarding the range of Tsosie's BAC at the time of the crash takes into account the facts that would impact her conclusion. * * * A qualified expert "may testify * * * in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case." Fed.R.Evid. 702. The Court concludes that Dr. Drez' retrograde extrapolation is admissible and will deny Tsosie's Motion. * * *

I. DR. DREZ' ANALYSIS IS ADMISSIBLE UNDER RULE 702.

The heart of Tsosie's argument is that Dr. Drez was required to know certain information * * * namely, the time period over which Tsosie consumed alcohol, and when and what he last ate, and that without knowing this information, her retrograde extrapolation is inadmissible. * * *

Dr. Drez' retrograde extrapolation satisfies rule 702 because it accounts for the known and unknown facts of this case, giving Tsosie the benefit of the doubt when facts are unknown. Dr. Drez used assumptions that favored Tsosie to account for the unknown variables of when he last ate and his drinking patterns.

Dr. Drez' use of reasonable assumptions is permissible under rule 702. * * *

Dr. Drez applied the retrograde extrapolation principles to the facts of this case. Tsosie stated that he stopped drinking at 11:00 p.m. the night before the crash. The crash occurred sometime before the crash was reported to the police at 5:11 a.m. Tsosie's blood was drawn at the hospital at 6:15 a.m., and the results from that blood test show Tsosie had a BAC of .07 mg/mL at 6:15 a.m. Based on Tsosie's statement that he stopped drinking at 11:00 p.m., Dr. Drez assumed that, even if Tsosie were an outlier, he would have started strictly eliminating by 1:15 a.m.—two hours and fifteen minutes after he stopped drinking. This allowance also accounts for the lack of data on Tsosie's last meal. ("[Winder:] You don't know what he ate, do you? [Dr. Drez:] I accounted for that by giving the benefit of the doubt and making it two hours."). Because she is unfamiliar with Tsosie's drinking patterns, she used a range to allow for a faster elimination rate if Tsosie is a heavy drinker and a slower rate if he is a light drinker. ("I address that by giving a range, which encompasses the different drinking patterns of individuals."). At the low end of the range of strict elimination rates of .01 mg/mL/h—the circumstances most favorable to Tsosie—he would have eliminated at least .01 mg/mL in the more than one hour period between the accident and his blood draw, placing his BAC at .08 mg/mL or above at the time of the accident. At an elimination rate of .02 mg/mL/h, Tsosie would have eliminated more than .02 mg/mL, producing a BAC level of at least .09 mg/mL at the time of the accident. Dr. Drez extrapolated further back to conclude

that Tsosie had a BAC of .12 to .17 mg/mL at 1:15 a.m., which is inconsistent with Tsosie's statement that he drank only three beers. Thus, Dr. Drez' retrograde extrapolation accounts for the unknown variables of Tsosie's last meal and drinking habits.

Dr. Drez thus relies on two fundamental assumptions: that Tsosie was in the elimination phase and that Tsosie eliminates alcohol at a rate that is within the range that a typical person eliminates alcohol. First, Dr. Drez assumed that Tsosie was in the elimination phase and not in the absorption phase when the accident occurred. Dr. Drez based this assumption on Tsosie's statement that he stopped drinking at 11:00 p.m. According to studies Dr. Drez cited, most people enter the elimination phase within one hour of finishing their last drink. Dr. Drez allowed that Tsosie could be an outlier who did not reach absorption until two hours and fifteen minutes after he finished drinking, but that her analysis with regard to his BAC at the time of the collision would not change even if he did not reach absorption until six hours after he stopped drinking. Because of this allowance, how much and how quickly Tsosie consumed alcohol—which could affect when he reached absorption—would not reasonably alter Dr. Drez' conclusions.

Dr. Drez' second assumption was that Tsosie is not an outlier in his elimination rate. The general population eliminates alcohol at a rate between .01 and .03 mg/mL/h, with the mean clustered more between .015 and .02 mg/mL/h. *See Jones & Andersson, supra*, at 922; *Jones & Andersson, supra*, at 924. * * * Dr. Drez prepared a chart that reflected that range of Tsosie's possible BAC based on elimination rates between .01 and .02 mg/mL/h. Dr. Drez stated that studies offer conflicting conclusions whether Native Americans eliminate alcohol faster, slower, or at the same rate as other races, but that using a range of elimination rates also accounts for this uncertainty, because none of the studies indicated that Native Americans eliminate alcohol slower than .01 mg/mL/h. The Court concludes that Dr. Drez' assumptions are reasonable, that her retrograde extrapolation “is based upon sufficient facts or data” and “is the product of reliable principles and methods,” and that Dr. Drez “has applied the principles and methods reliably to the facts of the case.” Fed.R.Evid. 702. Her retrograde extrapolation is scientifically valid and relevant to the facts of the case. * * * The Court will therefore deny Tsosie's Motion.

The Court denies Tsosie's request that the Court not allow Plaintiff United States of America to present expert testimony with regard to his BAC with the use of retrograde extrapolation.

Notes and Questions

1. In 1932, Swedish chemist E.M.P. Widmark first calculated absorption and elimination rates in the body, and his work still represents the benchmark for other scientists' studies today. Widmark created what we know today as the "BAC curve," which represents the rise and fall of an individual's BAC as his body absorbs and eliminates alcohol. The simplest version of the formula is below.

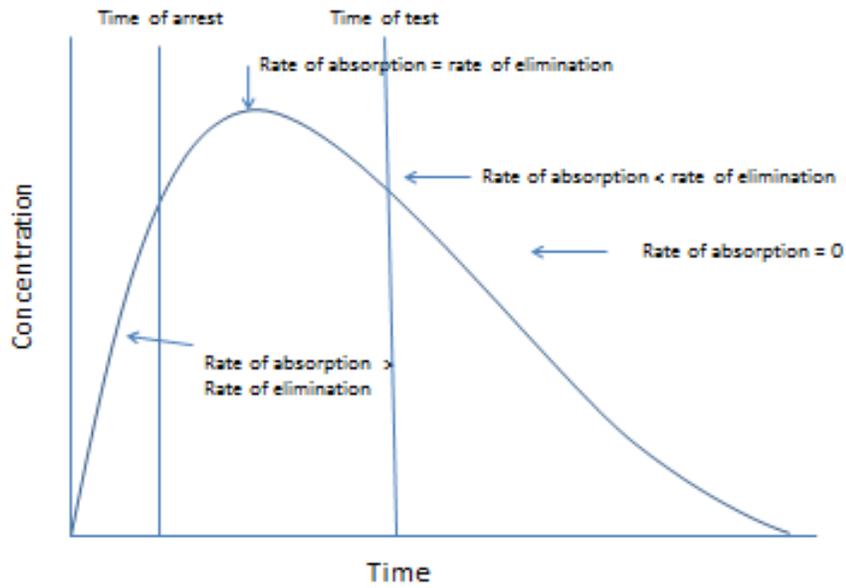
BAC calculations are based on the Widmark formula or some variation. The formula may be expressed as:

$$AC = - \frac{A}{Wr} (\beta t)$$

Where AC is alcohol content, A is the amount of pure alcohol consumed, W is the weight of the individual, r is the so-called Widmark factor that reflects an individual's percentage of body fat, gender, and age, and beta is the elimination rate. The Widmark formula has many variations, some of which are reviewed and discussed in <http://www.forensicmag.com/articles/2011/08/how-extrapolate-alcohol-certainty>, and in Posey, D. & Mozayani, A., *The Estimation of Blood Alcohol Concentration: Widmark Revisited*, 3 FORENSIC SCI., MED. & PATHOLOGY 33-39 (2007). See also <http://pubs.niaaa.nih.gov/publications/aa35.htm>.

2. The ability to make an accurate extrapolation requires some information about whether the individual was still in the absorption phase while driving. Figures IV-11 and IV-12 present two different scenarios with respect to this issue. What assumptions did the government's expert make, how did they address this issue, and why did the court conclude they were reasonable?

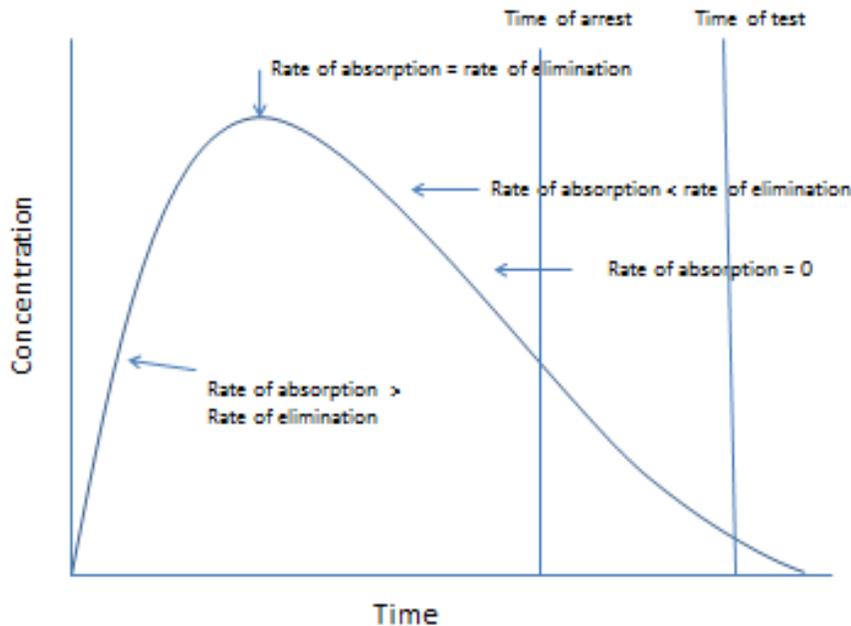
Figure IV-11.



Absorption and Elimination of an Oral Dose

SOURCE: Republished with permission of Taylor and Francis Group LLC Books, from Karen E. Stein and Thomas M. Brown, *Principles of Toxicology*, Third Edition, 2015; permission conveyed through Copyright Clearance Center, Inc.

Figure IV-12.



Absorption and Elimination of an Oral Dose

SOURCE: Republished with permission of Taylor and Francis Group LLC Books, adapted from Karen E. Stein and Thomas M. Brown, *Principles of Toxicology*, Third Edition, 2015; permission conveyed through Copyright Clearance Center, Inc.

3. Prosecutors and defense counsel have created a number of web pages explaining the pharmacokinetics of alcohol consumption. One of the better pages is published by the national district attorney association. See http://www.ndaa.org/pdf/toxicology_final.pdf. Some lawyers have actually created their own simplified BAC calculators. See, e.g., <http://www.impaired-driving-defence.com/>.
4. The *Tsosie* case focuses on acute effects of ethanol. Chronic exposure to this chemical produces several additional adverse effects. Ethanol is a teratogen and it leads to fibrosis and cirrhosis of the liver.

b. Toxicology in Civil Cases

Testimony by experts identified as toxicologists in appellate court civil litigation opinions was extremely rare in noncriminal cases prior to the 1940s. Two early cases are *Tindall v.*

American Furniture Co.,³⁶ (benzol exposure and anemia) and *Boal v. Electric Storage Battery Co.*³⁷ (tongue cancer and inhalation of sulphuric acid mist). Both cases involved job-related injuries.

Since that time, the frequency of toxicological testimony evidence has grown enormously. With the rise of toxic torts, toxicology has become an integral component of civil litigation. As is the case with respect to criminal cases, most case law addresses the question of the admissibility of expert opinions based on toxicological evidence. A complete discussion of all the issues raised in this arena is beyond the scope of this module. The following case addresses three related issues: (a) the admissibility and/or sufficiency of various types of toxicological evidence to prove causation, (b) the relationship between toxicological and epidemiological evidence, and (c) the relationship between regulatory standards for proof of causation and those in private civil actions.

Johnson v. Arkema, Inc.
United States Court of Appeals, Fifth Circuit, 2012.
685 F.3d 452.

Opinion

Per Curiam.

In this toxic tort case, we consider whether the district court erred in: (1) excluding the opinions of Gregory Johnson's expert witnesses on the element of causation; and (2) granting summary judgment in favor of Arkema, Inc. because Johnson was unable to prove causation without the opinions of his excluded causation experts. We AFFIRM the district court's judgment in all respects except as to Johnson's claims regarding his acute injuries, on which we REVERSE and REMAND for further proceedings.

I.

Johnson worked as a machine repairman at Owens Illinois Inc.'s glass bottling plant in Waco, Texas from May 1998 to the end of 2008. On two separate occasions, first in early June 2007 and again on July 15, 2007, Johnson was directed to perform work in close proximity to a device known as a C-4 Hood, which was designed, manufactured, and installed by Arkema. C-4 Hoods are utilized by Owens Illinois to apply a chemical known as Certincoat to the glass bottles it produces as the bottles are transported along a conveyor belt. Certincoat is composed mostly of monobutyltin trichloride (MBTC), an organometallic compound based on tin. Under the elevated temperatures of the C-4 hoods, MBTC vaporizes and then decomposes when it contacts the glass bottles on the conveyer belt. Hydrochloric acid (HCl) and tin oxide are byproducts of MBTC. Arkema's C-4 Hoods are designed to vacuum up and capture any vapors that are not deposited on the glass bottles, thus preventing the escape of MBTC, HCl and tin oxide into the workplace environment. According to Johnson, the C-4 Hood he worked near on those two occasions in the summer of 2007 failed to perform its proper preventative function, resulting in his exposure to Certincoat and its chemical byproducts.

Specifically, Johnson alleges that within fifteen minutes of first approaching the C-4 hood in

early June 2007 he: (1) smelled a sweet, unique chemical odor; (2) noticed chemical buildup on the conveyer belt; (3) developed a sore throat; (4) felt burning and watery eyes; and (5) experienced chest pain and breathing difficulty. Johnson nevertheless continued to work in these conditions for approximately four to five hours and, thereafter, neither reported the incident to his supervisor nor sought immediate medical attention. A few days later, on June 9, 2007, Johnson's family doctor diagnosed him with pneumonia. At his June 18, 2007 follow-up visit, Johnson reported that he "fe[lt] a lot better" and his doctor concluded that he could return to work the following day.

The next month, on July 15, 2007, Johnson was again instructed to work near the C-4 Hood. While doing so for approximately two to three hours, Johnson experienced the same symptoms that he felt during his first alleged instance of Certincoat exposure. This time, however, Johnson reported the incident to his supervisor and sought immediate medical attention at a local emergency room.

On August 8, 2007, upon Johnson's disclosure of the two exposure incidents to his treating physician, Dr. Camille Hinojosa, Johnson was diagnosed with chemical pneumonitis and advised to see a pulmonologist. According to Johnson, his lung condition progressively worsened over the course of the years following the exposure incidents, culminating in a diagnosis of severe restrictive lung disease and pulmonary fibrosis.

II.

On November 3, 2008, Johnson filed a personal injury lawsuit against Arkema, claiming that Arkema's C-4 Hood proximately caused his restrictive lung disease and pulmonary fibrosis. * * *¹

Arkema filed motions to exclude the opinions of Dr. Richard Schlesinger, Johnson's expert toxicologist, and Dr. Charles Grodzin, Johnson's expert pulmonologist, under Federal Rule of Evidence 702 and the Supreme Court's decision in *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993). Arkema also filed a motion for summary judgment, contending that Johnson was unable to present scientifically reliable evidence establishing that exposure to the chemicals in Certincoat can cause restrictive lung disease and pulmonary fibrosis. On December 16, 2010, the magistrate judge issued a report and recommendation to the district court regarding Arkema's *Daubert* motions. The magistrate judge recommended: (1) excluding Dr. Schlesinger's opinion, which only addressed causation, as unreliable and irrelevant; and (2) limiting Dr. Grodzin's opinion so that he could only opine on the nature and extent—but not the cause—of Johnson's illness. The district court adopted the report and recommendation and subsequently granted summary judgment in favor of Arkema. The district court reasoned that summary judgment was appropriate because—given the exclusion of Dr. Schlesinger's opinion and the limitation of Dr. Grodzin's opinion—Johnson "ha[d] no evidence that any lung injury he suffered [was] a result of his exposure to MBTC and/or HCl." In so doing, the district court rejected Johnson's claim that the similar symptoms experienced by his co-workers provided sufficient summary judgment evidence of causation. * * *

This appeal followed.

III.

¹ Johnson's brief defines pulmonary fibrosis as the "inflammation and progressive fibrosis of the pulmonary alveolar walls"; it is "one of a family of related diseases called interstitial lung diseases. All of these diseases can result in lung scarring."

* * *

A.

Johnson contends that the district court abused its discretion in excluding Dr. Schlesinger's expert opinion that MBTC and HCl¹ can cause restrictive lung disease and pulmonary fibrosis. The district court excluded Dr. Schlesinger's testimony after determining that: (1) Dr. Schlesinger could not cite to one epidemiological or controlled study of humans indicating that exposure to MBTC or HCl could cause restrictive lung disease and pulmonary fibrosis; (2) Dr. Schlesinger relied, in part, on two animal studies that were highly distinguishable from and not correlated to Johnson's two instances of MBTC and HCl exposure; and (3) the scientific literature is devoid of any data or peer-reviewed articles indicating that exposure to MBTC or HCl will result in chronic lung disease, and such a proposition is not generally accepted in the scientific community. Johnson argues that the district court erred in so ruling because: (1) MBTC and HCl are part of a toxicological class of chemicals labeled as irritants that are known to potentially cause pulmonary fibrosis; (2) Dr. Schlesinger based his opinion on reliable scientific data concerning MBTC and HCl exposure—including animal studies, * * * and guidelines from regulatory and advisory bodies—that support his conclusions; and (3) Dr. Schlesinger's opinion is buttressed by the temporal connection between Johnson's exposure and illness. As set forth below, because we are unable to conclude that the district court abused its broad discretion in performing its gatekeeping function under *Daubert*, we affirm the exclusion of Dr. Schlesinger's expert opinions.

1.

Johnson first claims that the district court erred in discounting Dr. Schlesinger's "class of chemicals" theory. Johnson asserts that Dr. Schlesinger's opinion is reliable because "MBTC and HCl are part of a group of chemicals labeled by toxicologists as 'strong irritants.'" According to Johnson, this classification is significant because "[a]ll 'strong irritants' have the same physiological effect when they contact biological tissue—production of inflammation." Moreover, numerous peer-reviewed studies of exposure to other chemicals labeled as irritants—including chlorine, ammonia, and nitric acid vapor—have reported lung scarring following acute exposure to those respective irritants. Thus, although Dr. Schlesinger only relied on one MBTC and one HCl study in forming his opinions, Johnson contends that Dr. Schlesinger's conclusions are reinforced by the more prevalent studies involving other irritants.

Our review of Supreme Court and this circuit's case law confirms that, in forming a reliable opinion regarding the effects of exposure to a particular chemical, an expert may extrapolate data from studies of similar chemicals. * * * However, "[t]o support a conclusion based on such reasoning, the extrapolation or leap from one chemical to another must be reasonable and scientifically valid." *Moore v. Ashland Chem., Inc.*, 151 F.3d 269, 279 (5th Cir.1998) (en banc) ("Thus, courts are free to reject a theory based on extrapolation when "there is simply too great an analytical gap between the data and the opinion proffered." *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997).

We applied the foregoing principles in our decision in *Wells v. SmithKline Beecham Corp.*, 601 F.3d 375, 380 (5th Cir.2010). In that case, three experts relied on a study of a class of drugs known as

¹ Although Certincoat also contains tin oxide, Dr. Schlesinger did not offer the opinion that tin oxide can cause restrictive lung disease and pulmonary fibrosis

“dopamine agonists” in support of their conclusion that a specific drug within the class, Requip, could have potentially caused the appellant's compulsive gambling problem.¹ We held that the district court did not abuse its discretion in excluding the experts, in part, because they failed to bridge the analytical gap between the generalized nature of the class-wide dopamine agonist study and the specific characteristics of Requip, “a drug that functions differently than other dopamine agonists.” *Id.*

* * *

In this case, we conclude that the district court did not abuse its discretion in excluding Dr. Schlesinger's “class of chemicals” theory. Dr. Schlesinger opined that MBTC and HCl can cause pulmonary fibrosis because they are part of a class of chemicals labeled as irritants:

It is generally accepted in the field of toxicology that both HCl and MBTC belong to a class of chemicals known as irritants. Toxicologically, all irritants have the same effect when they contact biological tissue, namely production of inflammation.

It is an accepted fact that acute inhalation of irritants can result in chronic diseases, including restrictive lung disease and pulmonary fibrosis.

* * *

While all irritants produce inflammation, as described above, *respiratory irritants are different in their specific chemical structure*. These differences relate to toxic potency (the exposure concentration needed to produce damage) and solubility (which affects the area of the lung an inhaled irritant would be expected to reach). Exhibit A # 6. However, while chemicals within a class may differ in toxic potency and solubility, the mechanism of toxicity is the same, as described above. Therefore, *if exposure to an irritant is of sufficient concentration to cause inflammation*, there are no other differences among irritants in the same class in terms of capability to cause a particular lung injury.

(Emphasis added). Dr. Schlesinger did not go further, however, and explain how, based on any of the specific properties and toxicities of similar irritants when compared with those of MBTC and HCl, Johnson's exposure to MBTC and HCl was at a sufficient concentration level to cause restrictive lung disease and pulmonary fibrosis. *See Moore*, 151 F.3d at 278–79 (“Dr. Jenkins made no attempt to explain his conclusion by asserting that the Toluene solution [to which plaintiff was exposed] had properties similar to another chemical exposure to which reactive airway dysfunction syndrome, or (RADS)) had been scientifically linked.”); *see also Mitchell v. Gencorp. Inc.*, 165 F.3d 778, 782 (10th Cir.1999) (although the “record contain[ed] some testimony about the similarities between benzene and [d]efendant's products,” there was no “additional testimony explaining exactly what these similarities [were] and how the similarities cause[d] the human body to respond to [d]efendant's chemicals in a manner similar to benzene”). Put differently, save for highlighting their shared classifications as irritants, Dr. Schlesinger did not attempt to explain any direct correlation or “fit” between the chemicals in Certincoat and the known scientific data concerning exposure to, for example, chlorine, ammonia, or nitric acid vapor. Accordingly, given the diverse chemical structures and toxicities of irritants, which Dr.

¹ A dopamine agonist is a drug “that stimulates the dopamine receptors in the brain to alleviate symptoms of Parkinson's [Disease].

Schlesinger acknowledged,¹ we hold that the district court did not abuse its discretion in concluding that Dr. Schlesinger's "class of chemicals" theory presented "too great an analytical gap between the data and the opinion proffered."² *Joiner*, 522 U.S. at 146, 118 S.Ct. 512.

2.

Johnson next asserts that reliable and relevant scientific data concerning exposure to HCl supports Dr. Schlesinger's conclusion that HCl causes scarring to lung tissue. * * *

Johnson cites a 1993 study of HCl's effect on nine baboons who were exposed "for fifteen minutes to three concentrations (500 ppm, 5,000 ppm, and 10,000 ppm) of HCl for a one year period." The study found that one of the nine baboons developed fibrosis after being exposed to a 10,000 ppm concentration of HCl. It ultimately concluded that HCl inhalation did not result in "the development of impaired respiratory/pulmonary function, except at the highest concentration." Although Johnson was only exposed to a ten to fifty ppm concentration of HCl, Johnson claims that the baboon study is reliable and relevant because: (1) Johnson was exposed to HCl for a much longer time period than the baboon who developed fibrosis; (2) baboons are considered to be an animal species that is a surrogate of man; and (3) the study shows that HCl is capable of causing fibrosis.

We have previously recognized the "very limited usefulness of animal studies when confronted with questions of toxicity." *Allen v. Pa. Eng'g Corp.*, 102 F.3d 194, 197 (5th Cir.1996) (quoting *Brock v. Merrell Dow Pharm.*, 874 F.2d 307, 313 (5th Cir.1989)). Accordingly, "studies of the effects of chemicals on animals must be carefully qualified in order to have explanatory potential for human beings."³ *Id.* Here, the district court found the baboon study unreliable and irrelevant because Dr. Schlesinger did not even attempt to show that there was a "correlation between the duration and length of the baboon exposure and Mr. Johnson's exposure." Likewise, Dr. Schlesinger admitted that the respiratory tracts of humans are "pretty unique," further diminishing the significance of the baboon study. Finally, Johnson's reliance upon the baboon study was weakened by the fact that there are no other studies of baboons or other animals that corroborate the baboon study's conclusions. In light of *Allen's* "careful qualification"

¹ One of the articles Dr. Schlesinger submitted with his report also implicitly addressed the diverse characteristics of irritants, providing that "[t]he health effects of an acute exposure to an irritant gas or vapor are dependent on the physiochemical properties of that particular gas or vapor, as well as specific host factors." (Emphasis added)

² This outcome may have been different had Dr. Schlesinger presented other reliable scientific evidence to support his causation opinion. For instance, if Dr. Schlesinger had other reliable evidence demonstrating that the concentration levels of MBTC and HCl were sufficiently high to impair respiratory function, then the analytical leap found in his "class of chemicals" theory could potentially have been reduced to a mere step, rendering Dr. Schlesinger's opinion reliable.

³ In *Allen*, we concluded that a study's finding that ethylene oxide (EtO) caused cancer in rats provided "at best speculative support" for the conclusion that EtO could cause cancer in humans because a different study of mice produced no such results. *Id.* In explaining our conclusion, we adopted the following logic of the appellee's expert: "Thus, the lack of capacity for the F-344 rat to predict how even the mouse model responds necessarily undercuts humans because a different study of mice produced no such results. *Id.* In explaining our conclusion, we adopted the following logic of the appellee's expert: "Thus, the lack of capacity for the F-344 rat to predict how even the mouse model responds necessarily undercuts confidence that the rat will predict accurately how other species including humans will respond [to EtO exposure]." *Id.*

requirement, we conclude that the district court did not abuse its discretion in rejecting the baboon study. *See also Joiner*, 522 U.S. at 144–45, 118 S.Ct. 512 (finding that the court did not abuse its discretion in rejecting the experts' reliance on animal studies—which involved the injection of “massive doses” of certain chemicals into infant mice—because the “studies were so dissimilar to the facts presented in th[e] litigation”). * * *

Finally, Johnson contends that he was exposed to amounts of HCl that were between two and ten times the permissible exposure levels set by the Occupational Safety and Health Administration (OSHA) and the National Institute for Occupational Safety and Health (NIOSH). Johnson also references the Acute Exposure Guideline Levels set by the National Research Council (NRC), which provide that Johnson could have been exposed to a “disabling” and possibly “lethal” dose of HCl.

In *Allen*, we addressed the significance of guidelines promulgated by regulatory and advisory bodies such as IARC, OSHA and EPA utilize a “weight of the evidence” method to assess the carcinogenicity of various substances in human beings and suggest or make prophylactic rules governing human exposure. This methodology results from the preventive perspective that the agencies adopt in order to reduce public exposure to harmful substances. *The agencies' threshold of proof is reasonably lower than that appropriate in tort law*, which “traditionally make[s] more particularized inquiries into cause and effect” and requires a plaintiff to prove “that it is more likely than not that another individual has caused him or her harm.” *Allen*, 102 F.3d at 198 (emphasis added) (quoting *Wright v. Willamette Industries, Inc.*, 91 F.3d 1105, 1107 (8th Cir.1996)). * * *

In sum, the Airgas MSDS, baboon study, and OSHA, NIOSH, and NRC guidelines do not sufficiently support Johnson's theory that HCl is known to cause scarring to lung tissue. The district court did not abuse its discretion in dismissing this data as irrelevant and unreliable under *Daubert*.

3.

Johnson also argues that reliable and relevant scientific data concerning exposure to MBTC supports Dr. Schlesinger's conclusion that MBTC causes scarring to lung tissue. Johnson first references Arkema's MSDS, which explains that MBTC “CAUSES RESPIRATORY TRACT IRRITATION” * * * [T]he district court did not abuse its discretion in rejecting the only evidence underlying Arkema's MSDS, namely, one unpublished study performed by Arkema in 1988 concerning MBTC's effect on rats. The study was designed to assess the toxic effects of MBTC when administered by inhalation to rats for six hours per day, five days per week, for four weeks at target concentrations of one, ten, and thirty milligrams per cubic meter. The study did not make any conclusions regarding restrictive lung disease and pulmonary fibrosis, and instead only found that exposure to MBTC had a discernable effect on the lung tissue of rats.¹ The district court determined that the rat study was irrelevant and unreliable

¹ Specifically, the study concluded that:

Grossly, the incidence of lung discoloration was increased in exposed males and females. Microscopically, amorphous material, (perhaps the test material or monobutyltin dihydroxy chloride, the hydrolysis product of monobutyltin trichloride) and alveolar edema were evident in the lungs of exposed males and females. Other lung changes which occurred with increased incidence and severity in the exposed groups included peribronchial lymphoid cell accumulation and perivascular lymphoid cell infiltrate, extravasated erythrocytes (males only), and accumulation of alveolar macrophages. Dose

because Dr. Schlesinger admitted that “there is no correlation between the durations of exposure” experienced by the rats, on the one hand, and Johnson, on the other. Based on *Allen*'s requirement that animal studies be “carefully qualified in order to have explanatory potential for human beings,” we conclude that the district court did not abuse its discretion in discounting this rat study.¹ *Allen*, 102 F.3d at 197. It follows that the district court did not abuse its discretion in discounting Arkema's MSDS because its warnings were founded on the rat study.

Johnson also raises the fact that he was exposed to a concentration level of MBTC that was between 100 and 500 times OSHA's permissible MBTC exposure limit of .1 milligrams per cubic meter. The district court was unpersuaded by the sheer magnitude of, according to OSHA's exposure limit, Johnson's over-exposure to MBTC. It found the maximum exposure limit misleading because OSHA set the .1 milligram per cubic meter threshold for all organotins, not just MBTC. Critically, Dr. Schlesinger conceded that this threshold for organotin exposure was “clear[ly]” not set based on data relating specifically to MBTC. Instead, according to Dr. Schlesinger, the OSHA threshold would be “based on whichever [organotin] they had the most data on in terms of inhalation.” Dr. Schlesinger also conceded that some organotin compounds are more toxic than others. Given Dr. Schlesinger's concessions, we conclude that the district court did not abuse its discretion in refusing to treat the OSHA exposure limit as reliable scientific evidence. *See Allen*, 102 F.3d at 198 (regulatory “agencies' threshold of proof is reasonably lower than that appropriate in tort law”).

Accordingly, we hold that Arkema's MSDS, the rat study, and OSHA's guidelines do not sufficiently support Johnson's theory that MBTC is known to cause scarring to lung tissue. The district court did not abuse his discretion in dismissing this data as irrelevant and unreliable under *Daubert*.

* * *

5.

In conclusion, we hold that the district court did not abuse its discretion in excluding Dr. Schlesinger's expert opinion under *Daubert*. Dr. Schlesinger could not cite to one epidemiological or controlled study of humans indicating that exposure to MBTC or HCl could cause restrictive lung disease and pulmonary fibrosis. *See Allen*, 102 F.3d at 197 (“Undoubtedly, the most useful and conclusive type of evidence in a case such as this is epidemiological studies.”). Also, Dr. Schlesinger neither extrapolated from existing data concerning chemicals similar to those in Certincoat nor correlated existing animal studies to Johnson's two exposure episodes. Instead, he relied on blanket statements from presumably credible sources—such as material safety data sheets and advisory guidelines—but failed to present the scientific evidence upon which those statements were founded. *Cf. Joiner*, 522 U.S. at 146, 118 S.Ct. 512 (“[N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.”). Finally, Dr. Schlesinger

related responses were shown only by alveolar edema in both sexes and by alveolar erythrocytes in males only.

¹ Johnson was exposed to between ten and fifty milligrams per cubic meter of MBTC. Johnson was, therefore, exposed to amounts of MBTC that were similar in concentration but not duration to the amounts of MBTC involved in the rat study. Nevertheless, Dr. Schlesinger could not adequately explain, as required by *Allen*, why MBTC's effect on the rats provides a reliable scientific basis for the conclusion that MBTC can cause restrictive lung disease and pulmonary fibrosis in human beings.

did not offer evidence that his theory has been generally accepted by the scientific community. The district court's exclusion of Dr. Schlesinger's expert opinion is affirmed.

* * *

IV.

After excluding the causation opinions of Dr. Schlesinger and Dr. Grodzin, the district court granted Arkema's motion for summary judgment because Johnson could not "prove the causation necessary to support a claim under Texas law." Johnson alleges that the district court erred in granting Arkema's motion for summary judgment because: (1) there is a strong temporal connection supporting causation; (2) the symptoms experienced by other Owens Illinois' employees provide additional circumstantial evidence of causation; and (3) Arkema's expert pulmonologist conceded that tin oxide is known to cause scarring of the lung tissues.

* * *

A.

Johnson first argues that the strong temporal connection between his exposure to Certincoat and the onset of his symptoms offsets the need to present expert testimony to establish causation. Johnson relies on the Supreme Court of Texas's decision in *Morgan v. Compugraphic Corporation*, which held that "[g]enerally, lay testimony establishing a sequence of events which provides a strong, logically traceable connection between the event and the condition is sufficient proof of causation." 675 S.W.2d 729, 733 (Tex.1984). Johnson argues that such a sequence exists in this case because "Johnson (1) had never smoked or had any history of asthma or lung disease prior to exposure, (2) worked within 2–3 feet of Arkema's machine that was leaking chemical fumes, (3) was exposed to chemical fumes at a level far above the OSHA limit, (4) could see, smell and feel the chemical burning his throat and lungs, (5) suffered classic symptoms of exposure to the chemical, (6) was administered oxygen and transported to the emergency room after 2–3 hours of constant exposure, and (7) despite continuous medical treatment to reduce lung inflammation, suffered permanent scarring to his lung tissue."

In its 2007 decision in *Guevara v. Ferrer*, the Texas Supreme Court summarized the meaning of *Morgan*. 247 S.W.3d 662 (Tex. 2007). The court first explained that "[t]he general rule has long been that expert testimony is necessary to establish causation as to medical conditions outside the common knowledge and experience of jurors." *Id.* at 665. The court reiterated, however, that "non-expert evidence alone is sufficient to support a finding of causation in limited circumstances where both the occurrence and conditions complained of are such that the general experience and common sense of laypersons are sufficient to evaluate the conditions and whether they were probably caused by the occurrence." *Id.* at 668–69. Such is generally the case when the lay testimony "establish[es] a sequence of events which provides a strong, logically traceable connection between the event and the condition."

In the underlying dispute in *Guevara*, the plaintiff had presented evidence at trial of: (1) the decedent's condition before an automobile accident; (2) the accident itself; and (3) the decedent's post-accident condition, including his numerous medical treatments. *Id.* at 667. The court found that such evidence could establish that the accident caused "basic physical conditions which (1) are within the common knowledge and experience of laypersons, (2) did not exist before the accident, (3) appeared

after and close in time to the accident, and (4) are within the common knowledge and experience of laypersons, caused by automobile accidents.” *Id.* The court nevertheless reversed because the evidence was legally insufficient to support a finding that the automobile accident caused all of the medical expenses awarded by the jury:

Non-expert evidence of circumstances surrounding the accident and Arturo's complaints is sufficient to allow a layperson of common knowledge and experience to determine that *Arturo's immediate post-accident condition which resulted in his being transported to an emergency room and examined in the emergency room were causally related to the accident.* Thus, the evidence is legally sufficient to support a finding that some of his medical expenses were causally related to the accident. On the other hand, the evidence is not legally sufficient to prove what the conditions were that generated all the medical expenses or that the accident caused all of the conditions and the expenses for their treatment.

Id. at 669–70 (emphasis added). * * *

Here, Johnson's alleged chronic injuries, the severe restrictive lung disease and pulmonary fibrosis, did not develop shortly after the Certincoat exposure incidents but instead manifested in the years following the incidents. In light of *Guevara*, we conclude that this significant gap in time renders the fact-finder unable to evaluate the cause of Johnson's chronic lung disease based solely on its common sense and general experience. We, therefore, agree with the district court's conclusion that Johnson needs the assistance of experts to prove that his Certincoat exposure caused his chronic injuries.

On the other hand, Johnson's acute injuries—which immediately followed his exposure to Certincoat and precipitated an emergency room visit and at least two other doctors' office visits during the summer of 2007—are within those limited circumstances where expert opinion is unnecessary. . . . Accordingly, the district court erred in granting summary judgment to Arkema regarding Johnson's alleged acute injuries. We therefore reverse and remand, in part, for further proceedings concerning Johnson's alleged acute injuries.

* * *

V.

For the foregoing reasons, we AFFIRM the district court's judgment in all respects except as to Johnson's claims regarding his acute injuries, on which we REVERSE and REMAND for further proceedings.

Notes and Questions

1. The court rejected the plaintiff's expert's "class of chemicals" analysis. What parts of the toxicology discussion above are relevant to this issue? Is the court correct to be wary of the generalization implicit in the expert's approach?
2. How persuasive is the baboon study? What are the issues one must address in answering this question?
3. Was the trial court correct in disregarding the "rat study" mentioned in the Arkema

- MSDS? Why? Is it relevant to the case at hand?
4. The court notes that Dr. Schlesinger could not cite to one epidemiological or controlled study of humans indicating that exposure to MBTC or HCl could cause restrictive lung disease and pulmonary fibrosis. What effect should this have on the assessment of the toxic effect of these chemicals? How should it affect how we assess the other evidence presented by the plaintiff?
 5. Is the court correct that OSHA guidelines are of little use to the plaintiff? Why?
 6. Part IV of the opinion distinguishes between acute and chronic injuries. Note this should be distinguished from acute versus chronic exposures. Acute exposures can lead to a chronic injury.
 7. Why did the court conclude in Part IV that the plaintiff needed an expert in order to prove his chronic injury but not his claim of an acute injury?

Templin et al. address the question raised in the *Chlorine Chemical* case: how does exposure to chlorine produce cancer in rats. [See Michael V. Templin et al., "Chloroform-Induced Cytotoxicity and Regenerative Cell Proliferation in the Kidneys and Liver of BDF₁ Mice," 108 *Cancer Letters* 225 (1996) (available at <http://www.sciencedirect.com/science/article/pii/0304383596042346>)]. As one can see, the article basically supports the court's position.

V. SPECIFIC CAUSATION

A. Introduction

Epidemiology and (to a somewhat lesser extent) toxicology are concerned with the incidence of disease in populations, and these researchers do not investigate the question of the cause of an individual's disease. This question, often referred to as specific causation, is beyond the domain of the science of epidemiology and toxicology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiologists investigate whether an agent can cause a disease or, equivalently is a risk factor for disease, not whether an agent did cause a specific plaintiff's disease.

Nevertheless, the specific causation issue is a necessary legal element in a toxic substance case. The plaintiff must establish not only that the defendant's agent is capable of causing disease but also that it did cause the plaintiff's disease. It is almost uniformly the case that courts require proof of general causation before they will permit an expert to testify as to specific causation.

When a court has concluded that there is sufficient evidence on general causation it

must confront the legal question of what is acceptable proof of specific causation. When epidemiologic evidence is available the courts must decide what role this evidence should play in addressing the specific causation question. As explained below, these studies play a critical role in providing evidence relevant to specific causation. The role has been worked out largely by courts rather than by epidemiologists.

Recall that the civil standard of proof is a preponderance of the evidence, that is the plaintiff must establish each prima facie element, including factual cause, is more likely than not or 50% + likely to have been the case. The relative risk from epidemiologic studies can be adapted to this 50% + standard to yield a probability or likelihood that an agent caused an individual's disease. An important caveat is necessary, however, before proceeding. The discussion below speaks in terms of the magnitude of the relative risk or association found in a study. As emphasized above, before an association or relative risk is used to make a statement about the probability of individual causation, the inferential judgment, described in Section III. F. supra, that the association is truly causal rather than spurious is required: an agent cannot have caused a specific individual's disease unless it is first recognized as a cause of that disease in general. The following discussion should be read with this caveat in mind.

B. The Logic of Relative Risks Greater than 2.0

Some courts have reasoned that when epidemiologic studies find that exposure to the agent causes an incidence in the exposed group that is more than twice the incidence in the unexposed group (i.e., a relative risk greater than 2.0), the probability that exposure to the agent caused a similarly situated individual's disease is greater than 50%. These courts, accordingly, hold that when there is group-based evidence finding that exposure to an agent causes an incidence of disease in the exposed group that is more than twice the incidence in the unexposed group, the evidence is sufficient to satisfy the plaintiff's burden of production and permit submission of specific causation to a jury. In such a case, the factfinder may find that it is more likely than not that the substance caused the particular plaintiff's disease. Courts, thus, have permitted expert witnesses to testify to specific causation based on the logic of the effect of a doubling of the risk.

This reasoning is a reasonable first step in adapting group-based epidemiologic evidence to an individual. But its validity requires several other steps and consideration of a number of assumptions on which this thinking is based.

C. Beyond the Basic Logic: The Appropriateness of Applying the Average of a Group Study to a Nonstudy Individual

Whether the results of a study can be applied to nonstudy individuals, known as "external validity," must be considered. If a study includes only women, whether the results of that study can be applied to men is a matter of external validity. Indeed, Judge Barbara

Rothstein, who presided over the multi-district litigation regarding PPA, an appetite suppressant, faced very similar questions. She was confronted, after a *Daubert* challenge by defendants, with whether plaintiffs' experts could testify, based on the results of an epidemiologic case-control study (the "HSP") that found a strong association (OR 16.58, 95% confidence interval 1.52 to 182.21) between PPA and hemorrhagic stroke in women between the ages of 18 and 49 (no men reported use of PPA, so the study could not directly assess risk to them), about causation in other adult women outside the age cohort of those studied, adult males, and children of both sexes. Although because of the multidistrict procedure, Judge Rothstein only addressed general causation, her analysis would be equally applicable in the specific causation context. We reproduce below the relevant portion of Judge Rothstein's opinion with regard to whether there was sufficient external validity to permit the plaintiffs' experts to testify about causation for these other groups.

In re Phenylpropanolamine (PPA) Products Liability Litigation
United States District Court for the Western District of Washington, 2003.
289 F. Supp. 2d 1230, 1244-46.

* * *

3. *Hemorrhagic Stroke in the Various "Sub-Populations":*

The HSP focused on men and women between the ages of eighteen and forty-nine. It did not offer any conclusions as to individuals outside of that age range, and the results were inconclusive as to men. The lack of epidemiological evidence directly associated with men, children, and individuals above the age of forty-nine is not fatal under *Daubert*. As discussed below, plaintiffs' experts demonstrate that it is scientifically acceptable to extrapolate the conclusions of the HSP to these sub-populations.

a. *Hemorrhagic stroke in individuals above the age of forty-nine:*

Defendants generally dispute whether extrapolation to a different age group is good science. However, in arguing against extrapolation to individuals above the age of forty-nine, defendants' experts primarily point to the fact that the risk of stroke increases as age increases. The court sees no reason why the increasing risk of stroke would render the HSP and the non-epidemiological lines of evidence unreliable as applied to this age group. *See* Dep. of Dr. Jerome Avorn, Defs.' Ex. E-1 at 363 ("[A]ll of the evidence we have is that risks only go up in the elderly. . . . [T]here are no drugs I'm aware of that get safer the older you get.") As such, the court finds testimony associating PPA with hemorrhagic stroke in individuals above the age of forty-nine reliable and, thus, admissible under *Daubert*.

b. *Hemorrhagic stroke in children and men:*

* * * [I]n disputing the propriety of extrapolating evidence from women to men, and from adults to children, defendants and their experts go to great lengths to highlight differences between these sub-populations.

Plaintiffs' experts assert that the weight of the evidence, including that obtained through extrapolation, supports the opinion that PPA can cause stroke in children and men. The court must address whether this extrapolation constitutes good science *See, e.g., Domingo*, 289 F.3d at 606 (“[S]tudies involving similar but not identical situations may be helpful, [so long as] an expert [] set[s] forth the steps used to reach the conclusion that the research is applicable.”).

It is axiomatic that children differ from adults in various ways, just as younger children differ from older children, and younger adults differ from the elderly. Men and women, likewise, differ in some respects. As might be expected, the incidence rates of stroke, types of stroke, and some of the risk factors for stroke vary between these groups. Plaintiffs' experts concede these differences, but maintain that these sub-populations share far more similarities than differences. After considering all possible differences, plaintiffs' experts find no basis for concluding that PPA poses a risk exclusive to adult females.¹

Because of the many barriers to including children in studies, scientists and medical practitioners routinely extrapolate study results and data on adults to children. This practice, despite its limitations, finds wide support in reputable sources. *See, e.g., Robert M. Ward, Adverse Effects of Drugs in the Newborn, in Rudolph's Pediatrics* 146 (Colin D. Rudolph et al. eds. 21st ed., 2001) (“Children continue to be excluded from studies of most new drugs, so that drug therapy of those patients is seldom guided by large controlled trials.”); George C. Rodgers, Jr. & Nancy J. Matyunas, *Oski's Pediatrics* 61–62 (Julia A. McMillan et al. eds. 3d ed., 1999) (“In the absence of controlled, randomized clinical trials in children, pediatricians must either extrapolate information from adult studies or use uncontrolled reports of clinical experience in children, both of which have major flaws.”). Plaintiffs' experts also point to the presumption in pediatric toxicology that toxic effects seen in adults will be as great, if not greater, in children.

Plaintiffs' experts attest to the equally commonplace practice of extrapolation between the genders, based on, in significant part, the historical exclusion of women from scientific studies. Defendants' experts note current studies accounting for the differences between men and women, but do not establish that this very recent shift has yet effectuated a change in the practice of extrapolation. Until such a change occurs, the court will not deem this practice scientifically unreliable.

Plaintiffs' experts clearly set forth the steps followed in extrapolating this evidence. While defendants demonstrate some of the problems posed by extrapolation and dispute the conclusions reached, they do not establish that plaintiffs' experts utilized scientifically unreliable methodologies. The court finds the direct and extrapolated evidence sufficiently reliable evidence upon which to base expert opinion. As such, it also finds opinions as to these sub-populations admissible under *Daubert*.

An even more serious problem of external validity arises when there are differences in dose between the studied population and another population to which the study results are applied. Thus, if in a study of the risk of lung cancer from smoking those exposed smoked half a pack of cigarettes a day for 20 years, the attributable proportion of risk in that study will not apply to a population that smoked 2 packs of cigarettes for 30 years without strong

¹ Similarly, the FDA did not differentiate between men and women, and found no reason to believe the risks posed by PPA were limited to individuals within the age range studied in the HSP. *See FDA Proposal to Withdraw Approval of New Drug Applications*, 66 Fed. Reg. 42670 (proposed Aug. 14, 2000) (“Although the Yale study focused on men and women 18 to 49 years of age, the agency has no reason to believe that the increased risk of hemorrhagic stroke is limited to this population.”).

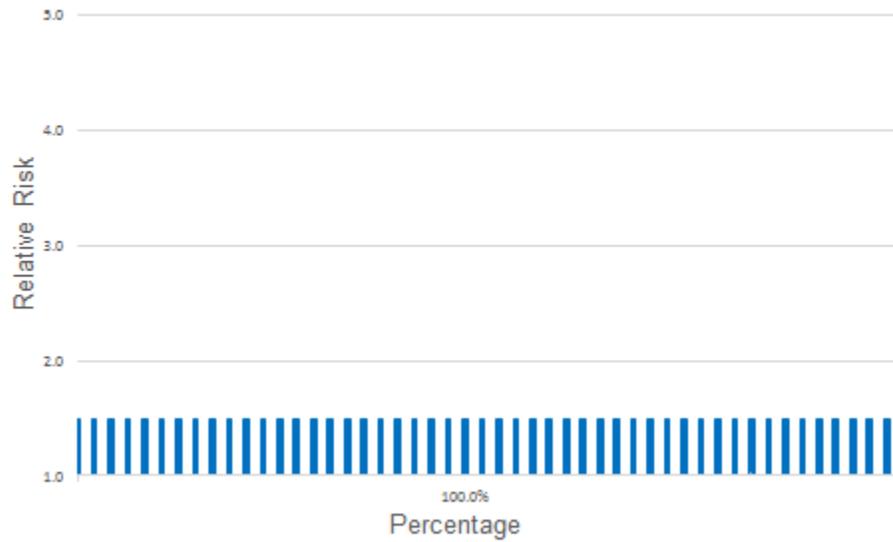
assumptions about the dose-response relationship. This is also applicable to risk factors for competing causes. Thus, if all of the subjects in a study are participating because they were identified as having a family history of heart disease, the magnitude of risk found in a study of smoking on the risk of heart disease cannot validly be applied to an individual without such a family history. Conversely, if an individual has been differentially exposed to other risk factors from the study subjects, the results of the study will not provide an accurate basis for the probability of causation for the individual. Consider once again a study of the effect of smoking on lung cancer among subjects who have no asbestos exposure. The relative risk of smoking in that study would not be applicable to an asbestos insulation worker.

D. Heterogeneity: The Effect on Employing the Average Outcome

Even when the conditions explained above are sufficient to give confidence in applying study results externally to others as Judge Rothstein did in the *PPA Products Liability Litigation*, heterogeneity must be considered. The results of a study reflect only the *average* among that population. If there is significant heterogeneity among the study population, say, with regard to other risk factors for the same disease being studied, then the average effect found in the study will not accurately reflect the probability of causation for most individuals generally, including those in the study. “A 1.5-fold relative risk may be composed of a 5-fold risk [RR of 5.0] in 10% of the population, and a 1.1-fold risk in the remaining 90%, or a 2-fold risk in 25% [RR of 2.0] and a 1.1-fold for 75%, or a 1.5-fold risk for the entire population.”³⁸ These alternatives are displayed graphically below in Figures V-1 – V-3. As should be evident, there are any number of distributions of relative risks among subpopulations that may exist for any study result. We address the implications of heterogeneity when there are sources of information that permit refinement of individual probability estimates in Section V. F. 1., *infra*.

Figure V-1. Homogeneous Population

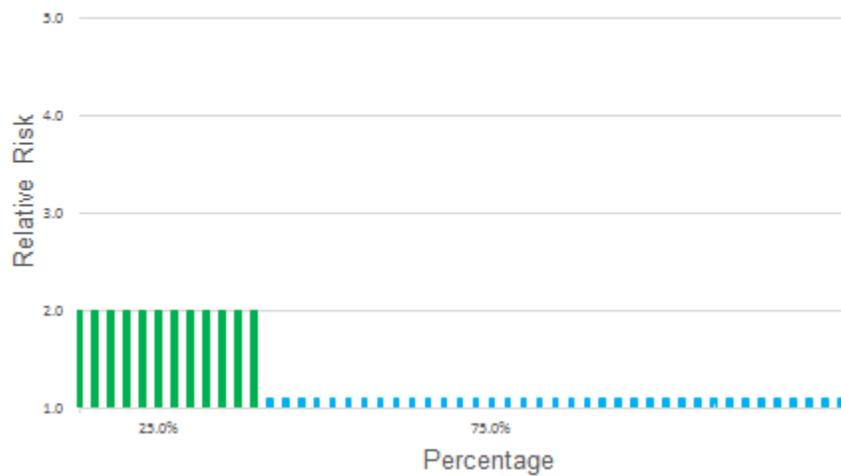
RR = 1.5 for 100%



SOURCE: Courtesy of the authors.

Figure V-2. Heterogeneous Population

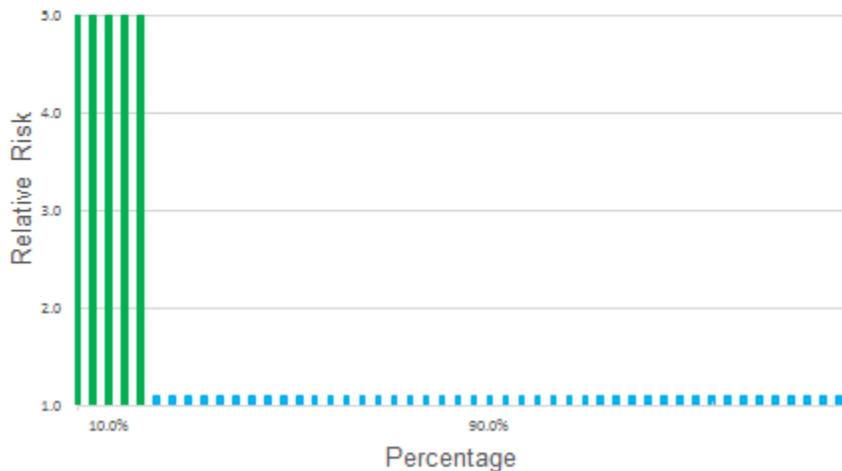
RR = 2 for 25%
RR = 1.1 for 75%



SOURCE: Courtesy of the authors.

Figure V-3. Heterogeneous Population

RR = 5.0 for 10%
RR = 1.1 for 90%



SOURCE: Courtesy of the authors.

E. The (Often Unarticulated) Assumptions in the > 2.0 Threshold

Scientific inquiry often rests on assumptions that enable or are foundational to that inquiry. A foundational assumption of science is that there are natural phenomena that cause the outcomes we observe.³⁹ Without such an assumption, which seems quite reasonable based on experience, inquiry into, for example, the causes of disease would be pointless. Often assumptions are less foundational, but nevertheless critical to conducting inquiry in various fields of science or even for specific studies.

While translating the results of an epidemiologic study to assess the probability of specific causation through the use of the APR statistic is not, as explained above, formally science, it does rely on several assumptions, which are rarely articulated and may not always exist. To remind you of one such assumption, the propriety of the “doubling reasoning” depends on group studies identifying a genuine causal relationship and a reasonably reliable measure of the increased risk.

Another assumption embedded in using the risk findings of a group study to determine the probability of causation in an individual is that the disease is one that never would have been contracted absent exposure. Put another way, the assumption is that the agent did not merely accelerate occurrence of the disease without affecting the lifetime risk of contracting the disease. In many instances, this assumption seems quite solid: Birth defects are an example of an outcome that is not accelerated, by definition. However, for many of the chronic diseases

of adulthood, it is not possible for epidemiologic studies to distinguish between acceleration of disease and causation of new disease. If, in fact, acceleration is involved, the relative risk from a study will understate the probability that exposure accelerated the occurrence of the disease.⁴⁰

See Abstract from Brad A. Racette, *Welding-Related Parkinsonism: Clinical Features, Treatments, and Pathophysiology*, 56 *Neurology* 8 (2001), available at: <http://www.neurology.org/content/56/1/8.full.pdf>.

Notes and Questions

1. Parkinsonism is an umbrella term that refers to the entire category of neurological diseases that cause rigidity and slowness of movement. Idiopathic Parkinson's disease (PD) is PD for which no particular cause can be determined. Welding-related parkinsonism is parkinsonism caused by exposure to welding fumes. In order to determine whether individuals with welding-related parkinsonism experience similar symptoms to individuals with idiopathic PD, Racette et al. examined a number of symptoms of parkinsonism: frequency of tremor, bradykinesia, rigidity, asymmetric onset and postural instability, clinical depression, dementia, and drug-induced psychosis.
2. What does the Racette study tell us about whether welding causes PD? About whether welding accelerates the occurrence of PD in welders? The abstract describes the study as of case-control design. Is it?

A third assumption underlying the > 2.0 threshold reasoning is that the alleged agent operates independently from other risk factors. Interaction means that when the agent and another risk factor are both present the combined increased risk in disease is greater than the sum of the increased incidence due to each agent separately. (There would be interaction, as well, if the increased risk was less than the sum of the increased risk for each, reflecting a protective effect of the combination of the two.) To take a prominent example, the relative risk of lung cancer due to smoking is around 10, while the relative risk for asbestos exposure is approximately 5. If the two operated independently, we would find a 15-fold (10 representing the effect of smoking and 5 representing the effect of asbestos exposure) increase in risk. But, the relative risk for someone exposed to both is not 15, but closer to 60, reflecting an interaction between the two. Neither of the individual agent's relative risks can be employed alone to estimate the probability of causation in someone exposed to both asbestos and cigarette smoke. In order to make such a calculation, one has to account for the synergistic effect produced by exposure to both, i.e., the 45-fold increase in risk above the arithmetic increase of 15.

Before concluding, there are two additional assumptions, ones that appear to be the case most of the time, that have been identified that undergird the > 2.0 reasoning: (1) The agent of interest is not responsible for fatal disease due to causes other than the disease of

interest. If that were the case in a study examining mortality, the relative risk found in the study would understate the role that the agent plays in causing death; and (2) The studied agent does not serve a protective effect in a subpopulation of the group being studied. If such an assumption were incorrect in a given study, then the relative risk found in the study would understate the real increased risk for the sub-group that was affected detrimentally by the agent. (Can you explain why that is so?)

F. Refining the Probability of Causation Through Subgrouping, Including Genetic Subgrouping and Molecular Epidemiology

We have seen that the methods of the disciplines principally involved in the scientific investigation of potentially toxic agents suspected of causing human disease—epidemiology and toxicology—are not really focused on discriminating among competing causes to determine “the” cause of any individual case of disease. Even if for legal purposes we accept the relative risk > 2.0 reasoning (see Section V. B., *supra*), we should appreciate that there is a subtle difference in interpreting a study finding a relative risk greater than 2.0 to mean: (1) it is more likely than not that any selected case of disease was caused by the exposure; and (2) that it is more likely than not that a particular person’s case of disease was caused by the exposure and would not have occurred without exposure. Can you, in light of the discussion in Section D. above, see why? Consider a jurisdiction that applies a strict relative risk > 2.0 rule. Suppose epidemiologists, after repeated study, concluded that a certain agent-disease association was causal and that the relative risk was approximately 3.0—equivalent to an attributable risk of two-thirds. If everybody in the exposed population sued, how many plaintiffs would prevail on the causation issue? On the other hand, suppose that—notwithstanding the difficulties of determining small increases in risk—valid epidemiologic studies coalesced around a causal association with relative risk of 1.8—equivalent to an attributable risk of 44%. If everybody in the exposed population sued, how many plaintiffs would prevail on the causation issue?

1. The Role of Information about the Individual Plaintiff

Intuition tells us that additional information about an individual case can increase our confidence in accepting or rejecting an inference of causation derived from group-based studies. If, for example, a particular disease were caused entirely (or overwhelmingly) by a small number of competing causes, ruling out some or all of the competing causes would make it more likely that the suspect exposure was the actual cause. If, on re-trial of his case, Mr. Stubbs (Section II. D., *supra*) were able to prove that he had consumed no raw fruits or vegetables, no shellfish, and no milk during the period within which he contracted typhoid fever, would his case to the jury have been stronger? Ruling out competing causes is the essence of the logic of differential etiology, discussed in Section V. H., *infra*.

From an epidemiologic perspective, the presence or absence of various competing

causes (or risk factors) among individuals in a study is an undesirable source of heterogeneity in the study sample. In particular, known competing causes are potential confounding factors in an epidemiologic study. As we have seen (Section III E. 3., supra), they can sometimes be taken into account by selection of subjects (e.g., excluding from the study persons known to have been exposed to a competing cause) or by stratification (computing separate relative risks for those exposed to and not exposed to the competing cause) or multivariate analysis. Stratification and multivariate analysis are discussed in Section III E. 3. c., supra.

But confounding factors are only one source of heterogeneity in a study sample. Another, particularly in retrospective studies of environmental or occupational exposure, is the amount of exposure. Often exposure must be treated as a simple yes-or-no condition, or in rough and not entirely accurate groupings (e.g., no, low, moderate, or high exposure). A plaintiff might wish to argue that even if the average relative risk for an exposed population was < 2.0, her relative risk was above average because of above-average exposure. Or a defendant might wish to argue that even if the average relative risk for an exposed population was > 2.0, the plaintiff's relative risk was below average because of below-average exposure. Such arguments were made in the following case.

Estate of George v. Vermont League of Cities and Towns
Supreme Court of Vermont, 2010.
2010 VT 1, 993 A.2d 367.

Skoglund, J.

Claimant[, who worked for the City of Burlington Fire Department for thirty-six years, first as a firefighter and later as assistant chief,] appeals from the superior court's order granting summary judgment to insurer in this workers' compensation case. * * *

* * *

In 2003, claimant died of non-Hodgkin's lymphoma (NHL). His estate brought a workers' compensation action, alleging that his work as a firefighter caused him to develop NHL. The Vermont Department of Labor denied his claim. The Commissioner ruled that although claimant proved that there was an "association" between NHL and firefighting, he failed to establish a "causal connection" between the general activity of firefighting and NHL. The Commissioner found no evidence as to the number of fires that claimant fought, the level of his participation in those fires, or the number of such fires that were industrial or commercial in nature, where known carcinogens might have been present. There was similarly no evidence as to the frequency of exposure or types of exposures that claimant may have had. Without this information, the Commissioner found that NHL was possibly, but not probably, related to his employment. The Commissioner thus concluded that claimant failed to meet his burden of proof and she denied the claim.

Claimant appealed this decision to the superior court, and the Commissioner certified the following question for determination: was claimant's NHL causally related to his work as a firefighter? In August 2007, insurer moved for summary judgment on this question. It asserted that the opinions of claimant's experts should be excluded under Vermont Rule of Evidence 702 as both irrelevant and scientifically unreliable, and that without any admissible evidence of causation, claimant was not

entitled to workers' compensation benefits.

* * *

In this case, claimant relied on the testimony of three experts -- Dr. Tee Guidotti, Dr. James Lockey, and Dr. Grace LeMasters -- to prove that his NHL was causally related to his employment. * * * All three experts relied upon epidemiological studies as the basis for their conclusions.

* * *

* * * The trial court here adopted a relative risk factor of 2.0 as a benchmark, finding that it easily tied into Vermont's "more likely than not" civil standard and that such a benchmark was helpful in this case because the eight epidemiological studies relied upon by claimant's experts reflected widely varying degrees of relative risk.

The trial court found that only two of the eight epidemiological studies relied upon by the experts in this case reflected a relative risk greater than 2.0 -- Figgs and Sama -- while the remaining six showed "little or no association" between firefighting and lymphomas. Notwithstanding the results of these studies, Dr. Guidotti opined that firefighting was in fact what caused claimant's lymphoma. Other than an undefined reference to "weight-of-the-evidence methodology," however, the court could not discern the scientific method that Dr. Guidotti used to reach his conclusion. The court also noted that the studies that Dr. Guidotti relied upon may have been overinclusive, reflecting associations between other types of lymphomas and generic cancers in firefighters.¹ For these reasons, the court could not find that Dr. Guidotti's testimony was based upon sufficient facts or data or that he applied the principles of epidemiological analysis reliably to this case. See V.R.E. 702(1), (3).

* * *

Claimant * * * asserts that the court should not have used a relative risk of 2.0 as a benchmark in evaluating whether the experts' testimony was based on sufficient facts or data. He also maintains that the court erred in stating that six of the epidemiological studies he offered showed "little or no association" between NHL and firefighting. In a related vein, claimant argues that, contrary to the trial court's finding, Dr. Guidotti adequately explained his methodology, and his reliance on a "weight of the evidence" methodology was scientifically acceptable. Claimant argues that the court should have credited Dr. Guidotti's explanation of why the "true risk" ratio for the type of cancer suffered by claimant "probably exceeds 2.0," notwithstanding the results in the majority of the epidemiological studies upon which he relied.

We find these arguments without merit. Claimant was required to show by a preponderance of the evidence that his NHL was causally related to his employment. * * * Given claimant's burden of proof * * * and the inherent limitations of epidemiological data in addressing specific causation, the trial court reasonably found the 2.0 standard to be a helpful benchmark in evaluating the epidemiological evidence underlying Dr. Guidotti's opinion.

* * *

* * * The [trial] court concluded * * * that Dr. Guidotti's opinion was not based on sufficient facts or data, and that Dr. Guidotti had not applied scientific principles and methods reliably to the facts of this case. The court did not abuse its discretion in reaching its conclusion.

* * *

* * * Dr. Guidotti did not specify the precise weight he gave to each study or how he reached his

¹ As Dr. Guidotti noted, NHL "is a collection of widely disparate diseases that are not commonly separated in epidemiological studies." He stated that NHLs consist of at least thirty recognized types, and he opined that new types will be identified as immunological and genomic methods become more sophisticated.

conclusion that the studies, taken together, demonstrated a statistically significant result, when seventy-five percent of the studies, individually, failed to reach that conclusion. Dr. Guidotti stated that his analysis was “based on the observation that improving the accuracy of cumulative exposure to combustion products in whatever data set is available results in an increased estimate of risk, which reflects the strength of association.” He opined that “[i]n the key studies available, a career of 40 years clearly places a firefighter at increased risk of NHL and is sufficient to conclude that the risk was in fact elevated to at least an approximate doubling.” How? Why? Dr. Guidotti failed to specifically account for the level of relative risk shown by each of the studies, describe what precise weight was given to each study, particularly in light of the different types of studies involved, or account specifically for showings such as that found in the Baris study that the level of excess risk of NHL was not associated with an increased number of lifetime runs, and that, in fact, the standardized mortality ratio was highest in those individuals who made the lowest number of firefighting runs.

* * *

* * * “[W]hen an expert opinion is based on data, a methodology, or studies that are simply inadequate to support the conclusions reached, *Daubert* and Rule 702 mandate the exclusion of that unreliable opinion testimony.” The trial court identified reasonable grounds for its decision, and as we have often repeated, it is for the trial court, not this Court, to weigh the evidence and assess the credibility of witnesses. We find no abuse of discretion in its exclusion of Dr. Guidotti’s testimony here.

We next consider the court’s evaluation of the meta-analysis conducted by Dr. Lockey and Dr. LeMasters. * * *

* * * [I]t does appear, as claimant argues, that the meta-analysis with respect to NHL was based on eight studies, apparently the same eight studies used by Dr. Guidotti in reaching his conclusion. The meta-analysis found the summary risk estimate for NHL to be 1.51, again a value less than 2.0. The study concluded that the findings of an association between firefighting and significant increased risk for specific types of cancer raised red flags and should encourage further development of innovative comfortable protective equipment, allowing firefighters to do their jobs without compromising their health. A conclusion that NHL is considered a “probable cancer risk” for firefighters is not sufficient to establish that claimant’s NHL was caused by firefighting, particularly given that this conclusion rests on a finding of relative risk of less than 2.0. As the trial court found, moreover, the study did not in fact assert that firefighting caused claimant’s NHL, and Dr. Lockey failed to adequately explain how this study showed that it was more likely than not that firefighting caused claimant’s cancer.

* * *

* * * Our law requires claimant to show, not merely that firefighting increased the likelihood of injury, but that it more likely than not caused his disease. Claimant failed to establish good grounds for such a conclusion here. * * *

* * *

* * * Without evidence of specific causation, summary judgment was properly granted to insurer.

Affirmed.

[The concurring opinion of Dooley, J., is omitted.]

Reiber, C.J., dissenting.

I cannot agree with the majority's decision to affirm the trial court's conclusion that summary judgment was appropriate after it improperly excluded claimant's expert opinions. * * * I would reverse and remand this case.

* * *

* * * [B]oth the trial court and the majority have exceeded their proper roles in this case and evaluated the evidence put forward by claimant to determine whether claimant should ultimately prevail on the merits. As the concurrence states, the trial court and the majority have concluded that "the evidence was inadequate." The problem is that this is a merits determination that should have been put to the jury. * * *

The only way that insurer could prevail on summary judgment is if the expert opinions of both of claimant's medical doctors are held to be inadmissible. Perhaps it is the foundation for the medical doctor's opinions that the majority and the concurrence find "inadequate." Regarding that question -- a question of admissibility -- the only way to dismiss the medical doctors' opinions here would be if there were "too great an analytical gap between the data and the opinion[s] proffered." *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997). * * * The trial court held that the gap was too great here because claimant did not have studies meeting the 2.0 relative risk standard. I agree that the 2.0 standard corresponds with the ultimate issue that must be decided on the merits: whether it is more likely than not that claimant's non-Hodgkin's lymphoma was caused by firefighting. The problem is that it is not the standard for admissibility.

The standard for admissibility is whether there is too great a gap between the studies offered and the medical doctor's opinions based in part on those studies. *Id.* But here, there is no gap at all: two of the studies relied upon by the doctors -- the Figgs study and the Sama study -- show statistically significant results that meet even the trial court's strict 2.0 admissibility standard. Those studies directly support the doctors' conclusions that it is more likely than not that claimant's non-Hodgkin's lymphoma was caused by firefighting. The doctors' opinions are therefore admissible. That should be the end of the admissibility analysis. But even where there is a gap between the studies and the doctors' opinions, as there is for those studies which show a relative risk of less than 2.0, that gap is more than filled here by specific knowledge about claimant that makes it more likely that claimant's non-Hodgkin's lymphoma was caused by firefighting. * * *

* * *

The opinions offered by claimant's experts were based on numerous statistically significant scientific studies with confidence intervals for relative risk entirely above 1.0. Those studies, published in peer-reviewed scientific journals, are routinely used by experts to determine the issue litigated here. * * *

The trial court's adoption of the 2.0 relative risk standard as the threshold for admitting evidence of epidemiological studies, with no consideration of a study's statistical significance, goes far enough in passing judgment on the evidence to amount to an evaluation of the merits of the case, rather than a proper inquiry into the methodology and reliability of the studies used by the experts. Whether it is more likely than not that claimant's firefighting caused his non-Hodgkin's lymphoma is the exact fact question that must be resolved on the merits.

The 2.0 standard for admissibility is also problematic because it sets a threshold that requires each study to prove that claimant should win on the merits. By definition, the 2.0 standard only admits each study if that study independently meets the more-likely-than-not standard for proving

causation. * * * By requiring each study to show -- on its own -- that it is more likely than not that claimant's cancer was caused by firefighting, the trial court failed to recognize that claimant is free to combine various pieces of evidence to make his case. * * *

The trial court's analysis appears to stem in part from a mistaken belief that an epidemiological study that fails to meet the 2.0 relative risk standard is not statistically significant. That is simply not true. Statistical significance and relative risk are two different concepts, and a doubling of the risk is not required for a study to be statistically significant. * * *

* * *

If a study has a [95%] confidence interval in a range that is entirely above 1.0, it is statistically significant, and any questions about the strength of the relationship shown by the study go to the study's weight, not its admissibility. If the trial court applied this standard here, the experts would be allowed to rely on the Burnett, Ma, Figgs, and Sama studies -- all of which had a confidence interval entirely above 1.0. The trial court abused its discretion in excluding those studies.

* * *

Further, although "epidemiology focuses on general causation rather than specific causation," [] epidemiological studies can be combined with specific information about an individual to show specific causation, as both of the medical doctors did here. * * * The 2.0 standard makes much more sense when a plaintiff is using epidemiological studies alone to prove specific causation. But here, as discussed in detail below, claimant's experts relied on more than just the epidemiological studies.

* * *

* * * The [trial] court apparently accepted insurer's erroneous position that all of claimant's experts looked only at the epidemiological studies and did nothing to relate those studies to anything particular about claimant. While it is true that the epidemiologist Dr. LeMasters appropriately limited her proposed testimony to the epidemiological studies, each of the medical doctors (Dr. Lockey and Dr. Guidotti) looked at several factors particular to claimant before concluding that it is more likely than not that claimant's disease was caused by firefighting. As Dr. Guidotti stated, the studies on general causation "inform[] our interpretation of the case, and then we try to bring it down to the particulars of that case, with as much knowledge as we have available."

* * *

First, both doctors considered claimant's extraordinarily long forty years of service as a firefighter. Dr. Lockey specifically looked at the fact that claimant "worked as a fireman for forty years." Similarly, Dr. Guidotti noted that claimant's forty years of exposure "places him in a high-risk category," specifically for non-Hodgkin's lymphoma "among other things." This deposition testimony in itself is sufficient to allow claimant to argue that the epidemiological studies underestimate the real risk that claimant faced through his firefighting and that even studies showing a relative risk of less than 2.0 can therefore support his claim that firefighting more than doubled his risk of getting non-Hodgkin's lymphoma. The trial court completely failed to address the fact that the experts in this case rendered opinions that this particular claimant was a firefighter for a much longer period of time than the average firefighter discussed in the studies.

Second, Dr. Lockey and Dr. Guidotti looked at the fact that claimant was likely exposed to more toxins than the average firefighter, since claimant's firefighting career covered a time when protective equipment was often not used. Dr. Lockey noted that claimant was a firefighter "during a timeframe back in the '60s and '70s when control measures more likely than not were not as good as they are currently." Dr. Guidotti similarly noted that it was not until the 1970s that a self-contained protective

breathing apparatus was widely introduced and that even then “relatively few” firefighters actually used such an apparatus. According to Dr. Guidotti, “there was a gap in the 1970s and the early ‘80s when firefighters very often were not using their personal protection when fighting fires.”

Third, Dr. Guidotti also looked at the particular type of non-Hodgkin’s lymphoma that claimant contracted. Non-Hodgkin’s lymphoma does not refer to just one disease; rather, it is a large category that includes at least thirty recognized types of lymphoma. Dr. Guidotti noted that only some of those types “are known to be associated with environmental exposures and occupations.” Claimant had small cell lymphoma, which Dr. Guidotti noted is associated with environmental exposures. In particular, it is associated with exposure to solvents, including some of the same chemicals that are “released during firefighting.” Thus, Dr. Guidotti concluded that “the chemicals that are known to be associated with small cell lymphocytic lymphoma seem to be more than likely the kinds of things that one would encounter on the job.” This information was a significant factor leading Dr. Guidotti to state that, although “[s]cientific certainty in the matter is unattainable,” it was his opinion that the evidence favored the conclusion that claimant’s “lymphoma arose from work as a firefighter.”

Finally, both doctors also ruled out other possible causes of claimant’s disease before reaching their ultimate conclusions. Dr. Lockey examined claimant’s medical records and looked at whether there were “any other potential factors as it applies to [claimant] that would be known to be associated with a risk for the occurrence of non-Hodgkin’s lymphoma.” Dr. Lockey concluded that he “could not identify any other known risk factors based on the information that was available to me.” Dr. Lockey specifically noted that to his knowledge claimant “apparently did not have an immune deficiency disorder, which is the primary risk. As far as I was aware, he was not HIV positive, which would put him at risk for non-Hodgkin’s lymphoma.” This was a major factor leading Dr. Lockey to conclude “with a reasonable medical probability that [claimant’s] work as a firefighter was the cause of his non-Hodgkin’s lymphoma.” Dr. Guidotti also examined claimant’s medical records and similarly noted that this main risk factor could be ruled out for claimant, since a “severe immune problem ... would have expressed itself by inability to work.” Because alternative explanations for contracting non-Hodgkin’s lymphoma were ruled out, the trial court should not have excluded the opinions concluding that it was more likely than not that claimant’s disease was caused by firefighting. * * *

The trial court failed to recognize that both medical experts relied on numerous factors specific to claimant. The trial court stated that “Dr. Guidotti’s testimony in particular relies solely upon these [epidemiological] studies.” This was clear error. * * *

* * *

The majority notes that the Baris study found that the level of excess risk of non-Hodgkin’s lymphoma “was not associated with an increased number of lifetime runs, and that, in fact, the standardized mortality ratio was highest in those individuals who made the lowest number of firefighting runs.” [] There are three problems with the majority’s approach here: (1) it does not address any of the other factors particular to this claimant that the doctors relied upon in making their conclusions, such as claimant’s lack of protective safety equipment, lack of other known risk factors, and contraction of a type of non-Hodgkin’s lymphoma that is linked to solvents released during fires; (2) the majority’s questioning of the experts’ opinions is precisely the type of issue that goes to the weight of those opinions, not to their admissibility; and (3) the majority’s foray into interpretation of the Baris study is misleading and contrary to how the experts interpret that study.

The Baris study itself noted that “[s]mall numbers of observed deaths in the subcategories of the . . . cumulative runs analyses resulted in imprecise risk estimates.” Dr. Guidotti noted that “Baris is very

clear ... that they don't consider that the runs analysis was particularly useful." Dr. Lockey, working with Dr. LeMasters, takes the same position and lists numerous possible alternative explanations, including "gross misclassification," "a chance finding," and "a healthy survivor effect." At a minimum, these observations raise issues of disputed fact.

Dr. Guidotti and Dr. Lockey unsurprisingly found these alternative explanations more convincing than insurer's counterintuitive claim (adopted by the majority today) that increased exposure to fires can decrease the likelihood of getting cancer. It is undisputed that firefighting exposes firefighters to known carcinogens: most of the studies presented to the trial court below state that as a given. For instance, the first sentence of the Baris study notes that "[f]irefighters are exposed under uncontrolled conditions to a wide variety of toxic chemicals including known and suspected carcinogens, such as benzene and formaldehyde in wood smoke, polycyclic aromatic hydrocarbons (PAHs) in soot and tars, arsenic in wood preservatives, asbestos in building insulation, diesel engine exhaust, and dioxins." A carcinogen is defined as "a substance or agent producing or inciting cancer." Webster's New Collegiate Dictionary 165 (1981). Thus, it is difficult to understand why the majority puts any stock in the claim that increased exposure to carcinogens decreases one's chance of cancer -- a claim that is inherently self-contradictory, is called into question by the Baris study itself, and is resoundingly rejected by all three of claimant's experts below.

The majority has to mention this strange finding from the Baris study because there is no other way to affirm the trial court's decision. Dr. LeMasters, Dr. Lockey, and Dr. Guidotti all put much more stock in the Baris study's finding that firefighters who are employed for more than twenty years are at a greater risk than other firefighters for contracting non-Hodgkin's lymphoma. If those three experts are correct -- or, rather, if a jury could conclude that they are correct -- that increased exposure to fires leads to increased risk of non-Hodgkin's lymphoma, claimant can argue that the generalized studies showing an association among firefighters underestimate the risk that he personally experienced. Then, even studies showing a relative risk of less than 2.0 help claimant make out a prima facie case that his forty years as a firefighter made it more likely than not that firefighting caused his disease.

* * *

These facts could easily lead a reasonable jury to conclude that because claimant fought fires for forty years, he was exposed to more carcinogens -- and was at greater risk for contracting non-Hodgkin's lymphoma -- than the average firefighter discussed in the epidemiological studies. * * * Thus, the trial court abused its discretion when it excluded the proposed expert testimony.

Although it is my view that adopting the 2.0 standard was a clear error of law here, even if that standard were acceptable the trial court abused its discretion by failing to provide any explanation as to why it excluded evidence based upon the Aronson, Figgs, and Sama studies -- all three of which exceeded the 2.0 standard. Granted, the Aronson study could properly be excluded because its 2.04 relative risk finding was not statistically significant, as it had a confidence interval that included the number 1.0. But the trial court never explains that as a reason for excluding the Aronson study. More importantly, the Figgs and Sama studies could not be excluded as statistically insignificant, because both of these studies had confidence intervals entirely above 1.0. The Figgs study found a relative risk of 5.6, and the Sama found a relative risk of 3.27. Both of these studies were statistically significant and met the trial court's strict 2.0 standard. Therefore, the court's failure to explain why these studies were not themselves sufficient support for the opinion evidence constitutes an abuse of discretion that requires reversal.

* * *

The trial court * * * appears to require some unspecified percentage (the majority? all?) of surveyed epidemiological studies to meet the 2.0 standard before the jury can even hear about any of the studies. * * *

Although the trial court found that the epidemiological studies “reflect widely varying degrees of relative risk,” that is not a reason to exclude all of the studies. Just because the studies had different results does not mean that they are all wrong, and claimant should be allowed to argue to the jury why the Figgs and Sama studies are the studies that arrived at the correct relative risk for claimant. That is particularly true here, where claimant's experts found that although the relative risks were different, they for the most part all pointed in the same direction. As Dr. Lockey stated, there was “consistency across the medical literature based on epidemiology studies of, in fact, a cause-effect relationship between this profession and the occurrence of non-Hodgkin's lymphoma.”

* * *

In summary, the trial court abused its discretion in numerous ways by summarily excluding all of claimant's evidence and granting insurer's motion for summary judgment * * * The trial court should have admitted the two medical doctors' expert testimony, which found specific causation based on four statistically significant studies (Burnett, Ma, Figgs, and Sama) and specific information about claimant. Two of those studies (Figgs and Sama) meet even the strict 2.0 relative risk standard and therefore directly support the experts' conclusions that it is more likely than not that claimant's injuries were caused by firefighting. The other two studies (Burnett and Ma) can be combined with specific information about claimant to bridge any “analytical gap between the data and the opinion proffered.” [] The trial court should have also admitted testimony from claimant's epidemiologist to help explain the underlying studies and how firefighting can cause non-Hodgkin's lymphoma.

For these reasons, I would reverse and remand to the trial court to apply the proper legal standard for the admission of evidence. I therefore dissent.

I am authorized to state that Justice Johnson joins this dissent.

Notes and Questions

1. Assess the persuasiveness of the majority and dissenting opinions.
2. One of the claimant's experts testified that his inference of causation was stronger because he “could not identify any other known risk factors based on the information that was available to me.” In particular, the expert noted that the claimant did not have an immune deficiency, a known risk factor for NHL. This is a version of the “differential etiology” approach to specific causation discussed in Section V. H. 3., *infra*.
3. Compare the analysis in *Estate of George* with the analysis in *Lindquist*, *supra* Section III. I. In both cases epidemiologic research (including some studies used in both cases) at least arguably supported the existence of an association between firefighting and the disease in question. In *Estate of George*, the claimant's experts could eliminate at least some other risk factors for the disease; in *Lindquist*, the claimant had exposed himself to the “most significant risk factor” (tobacco smoke) and had a family history that at least suggested the possibility of a genetic risk factor

- as well. (We discuss genetic risk factors *infra*, Section V. F. 3). Yet the claim in *Estate of George* failed because without proof of more than a doubling of risk the claimant could not establish that his occupation “more likely than not caused” his NHL, while the claim in *Lindquist* succeeded even without any quantitative risk evidence because the claimant’s evidence was sufficient to prove that on-the-job “exposure contributed in a material degree” to his emphysema. Suppose the legal standard applied in *Estate of George* had been applied in *Lindquist*, and vice versa. Would the outcome of either case have changed? If so, how?
4. Both the majority and the dissent noted the current medical view that NHL is not a single disease but comprises a large category of malignancies, but the two opinions assigned different significance to this fact. As medical science detects finer and finer distinctions among subtypes of what were once considered single diseases, disputes about the proper categorization, and about the implications of the proper categorization for inferences of causation, are likely to become increasingly common and increasingly important. For an example, see *Milward v. Acuity Specialty Products Group, Inc.*, 664 F. Supp. 2d 137 (D. Mass. 2009), *rev’d*, 639 F.3d 11 (1st Cir. 2011). For a scientific review of “weight of the evidence” in the context of research on toxic disease causation, see Douglas L. Weed, *Weight of Evidence: A Review of Concept and Methods*, 25 RISK ANALYSIS 1545 (2005).
 5. One of the claimant’s experts testified that “the weight of evidence favors the interpretation that [claimant’s] lymphoma arose from work as a firefighter.” The majority affirmed the trial court’s exclusion of this testimony in part because the expert could not quantify the weight to be given to each study in the body of scientific evidence the expert considered. The exclusion was based on Vermont Rule of Evidence 702 (which is patterned after Federal Rule of Evidence 702), and on *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), the leading federal case on the admissibility of expert testimony. *Daubert* and two subsequent decisions of the United States Supreme Court, discussed in Section II E., *supra*, instruct trial judges to act as gatekeepers for expert testimony, admitting such testimony only if it is based on scientifically “reliable” methodology and “fits” the facts of the case. The most frequently employed approach, perhaps best exemplified by *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), independently examines each piece of scientific evidence an expert relies on and asks if it supports the expert’s conclusion. Under this approach, testimony that the “weight of the evidence” supports the conclusion has generally fared badly. Yet scientists often accept “weight of the evidence” as sufficient support for regulatory decisions based on hypotheses of toxicity that cannot be directly tested experimentally. One federal court of appeals reversed a trial court’s decision excluding an expert’s “weight of the evidence” testimony as to general causation. *Milward v. Acuity Specialty Products Group, Inc.*, 639 F.3d 11 (1st Cir. 2011). On remand, a different district judge excluded the testimony of the plaintiff’s expert on specific causation. *Milward v.*

Acuity Specialty Products Group, Inc., 969 F. Supp.2d 101 (D. Mass. 2013), *aff'd*, 820 F.3d 469 (1st Cir. 2016).

6. Both the majority and the dissent discussed the finding of the Baris study that the standardized mortality ratio was highest in the group that had made the smallest number of firefighting runs. One of the plaintiff's experts testified that a "healthy survivor" effect could help explain this finding. The healthy survivor effect suggests that the firefighters who work the longest include those who are least likely to develop NHL in response to their on-the-job exposure to carcinogens. If the effect is real, why would some firefighters (as the dissent described it) be "naturally less susceptible to contracting non-Hodgkin's lymphoma?"

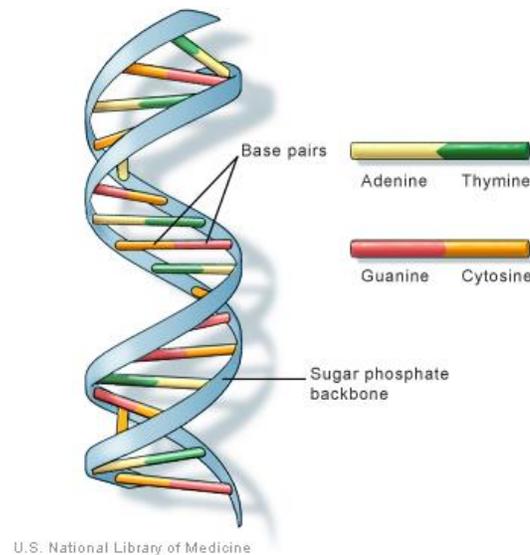
2. Using Genetics to Refine the Probability of Causation: Individual Susceptibility to Toxic Effects

We have seen that differences in exposure to other risk factors, or in the dose of the exposure of interest, might mean that the average relative risk determined in an epidemiologic study does not accurately reflect the degree of risk conferred by exposure on a particular individual. These exposures are exogenous to the individual. What if entirely endogenous features affect an individual's risk? That is, what if the risk posed by exposure depends in part upon an individual's genetic endowment?

Long experience has shown that the same exposure to a toxin—or a beneficial drug—does not affect different people identically. See Section V. H. 3., *infra*. For example, most people exposed to asbestos do not develop mesothelioma, but a few do. Most patients obtain pain relief from codeine and related drugs, but a small minority of patients do not. Researchers increasingly trace such differences to genetic variability in the population.

Humans ordinarily inherit their genetic material in DNA that is (in all body cells except sperm and egg cells) arranged in 23 pairs of chromosomes with one member of each pair inherited from each parent. DNA (deoxyribonucleic acid) is a molecule that consists of two long chains of subunits called nucleotides. The chains spiral around each other in a double helix. Each nucleotide includes one of four bases: adenine, thymine, cytosine, or guanine. These bases project into the space between the chains' "backbone" to form, in specific complementary pairings, the "rungs" of the DNA double helix "ladder" (see Figure V-4).

Figure V-4. Schematic of a Segment of DNA



SOURCE: U.S. National Library of Medicine. Image in the public domain.¹

DNA's ability to transmit information results from the fact that these base pairs can be arranged in many different sequences. More specifically, groups of three bases correspond to particular amino acids. For example, a group consisting of three adenine bases codes for the amino acid lysine; a group consisting of two adenine bases followed by a cytosine base codes for the amino acid asparagine. Coding for amino acids is important because amino acids are the chemical subunits of proteins, which have critical biochemical and structural roles in our bodies.

A gene is a segment of DNA, found at a particular location ("locus") on a chromosome, that codes for a particular sequence of amino acids. The amino acid sequence is generated when the segment of DNA is "transcribed" into a molecule of messenger RNA and then "translated" through a process that delivers and concatenates the specified amino acids. A protein forms when one or more of these amino acid sequences folds into three-dimensional shape under the influence of other biochemical constituents. The sequence of amino acids, in association with any other chemical controls on how the protein folds, determines the protein's structure and function. The proper structure and function of the proteins synthesized pursuant to these genetic instructions are essential to health. The human genome contains on the order of 20,000 protein-coding genes.

Different body parts need different proteins; one would not want a brain cell manufacturing the digestive enzymes secreted by stomach or intestinal cells. A gene is said to be "expressed" if the cell, tissue, or organism under study is actively manufacturing the amino

¹ <http://ghr.nlm.nih.gov/handbook/basics/dna>, visited Feb. 4, 2016.

acid sequence associated with that gene. How does a gene in a particular cell know whether to express itself? Some genes have a regulatory function: they code for proteins that affect the expression of other genes. But gene expression is also controlled by “epigenetic” factors, outside of the genes themselves, that influence gene activity. These factors include a number of biochemical constituents: histone proteins that form part of the structure of chromosomes and affect which sequences of DNA are exposed so they may be transcribed; several forms of RNA, and some of the vast regions of DNA that do not code for amino acid sequences. These epigenetic instructions are affected by a cell’s environment and may be enduring. For example, some epigenetic changes that are associated with differentiation of cells into various tissues and organs in a developing embryo may persist through fetal development and into childhood and adulthood.

As you know, genes vary from person to person. Because of inherited alterations—mutations—in the DNA sequence, a given gene may occur in a number of variable forms, which geneticists call alleles. If there are two or more alleles of a gene that occur at frequencies above those expected to arise by newly-occurring mutations, that gene is said to be polymorphic. For example, the familiar blood group types are determined by a gene that has three different alleles, commonly known as A, B, and O. Some genes are much more variable than that, with numerous alleles found in the human population. Others—presumably genes in which any new allele that might appear by mutation would be highly deleterious to survival or reproduction—display very little variability.

For any given gene, each individual ordinarily possesses two copies, one inherited from each parent, which may both be the same allele or may be different. Thus a person who inherits a blood type A allele from each parent will have blood type A; a person who inherits an A allele from one parent and a B allele from the other will have blood type AB. Biologists refer to the inherited pair of alleles as the “genotype” for a gene, and to the physical manifestation of the genotype—for example, the presence or absence of the A and B antigens on a person’s red blood cells—as the “phenotype.”

The relation between genotype and phenotype is not always simple. Phenotypic traits that are more complex than the ABO blood types may result from the action of multiple genes. Even for traits determined by a single gene, not every genotypic change necessarily leads to a change in phenotype. Both single-gene and polygenic traits may be affected by environmental factors operating through epigenetic controls on gene expression. Thus the same genotype might or might not affect the phenotype in different people. The degree to which different genotypes are reflected in different phenotypes (such as the presence or absence of a disease) is called penetrance.

Some polymorphisms affect the expression of the gene (whether – and to what extent – its protein manufacturing machinery is turned on or off) or the activity of the resulting protein (whether or how well it does its biochemical job, which can include the job of turning other genes on or off). Polymorphisms result from genetic mutations, which may have harmful, beneficial or neutral effects on survival and reproduction – or which may have effects that are harmful under some environmental conditions and beneficial under others. For example, an

alteration of a single specific DNA nucleotide in the gene that codes for a portion of the protein hemoglobin results in a substitution of just one particular amino acid for another at a specific location in the protein. That substitution affects the shape of the protein, producing a phenotype called sickle-cell to describe the shape of red blood cells with the altered protein. The altered hemoglobin is less effective at transporting oxygen but confers greater resistance to certain malaria parasites. Although inheriting two altered alleles leads to sickle disease, persons whose genotype includes just one altered allele are able to transport oxygen properly under most environmental conditions but also are less likely to die of a childhood infection with malaria.

The example of the sickle trait allele demonstrates that different genotypes may affect humans' interaction with external disease-causing agents—in that case, the parasitic plasmodia that cause malaria. The same is true of the types of environmental toxins frequently involved in toxic tort suits. As toxicologists and medical researchers have increasingly elucidated the molecular mechanisms of toxicity and disease, the potential role of genetic variability has become increasingly apparent.

Most toxins are metabolized—biochemically altered in one or more ways—by the body. A polymorphic gene might affect a protein that is involved in the metabolism of some potentially toxic substance. The variants could either interfere with or facilitate a biochemical reaction that either detoxifies the agent or creates a harmful metabolic by-product. In these ways genetic variation could alter individual responses to toxic exposure.

Toxic exposure might also affect disease processes more directly. Cancer, for example, is now understood to result from a series of changes to a cell's DNA that eventually override the molecular controls that normally regulate cell division and proliferation. Carcinogens may directly cause damage to DNA that controls these regulatory mechanisms or may disrupt other cellular mechanisms that limit or repair such damage to DNA.¹ Inherited variations in the genes that code for proteins involved in the processes that affect the relative vulnerability of DNA to accumulating damage may affect individuals' susceptibility to exposure-induced carcinogenesis.

Increasing mechanistic understanding of toxicity and disease helps scientists select genes for focused study aimed at determining whether a given exposure causes more increased risk to persons with certain genotypes than to others. Recent technological advances have made it practical for researchers investigating the relation between exposure and disease to determine the genotypes of the individuals in their studies. Doing so may allow researchers to assess the degree to which variability in genotype is reflected in variability in susceptibility to the toxic effect of the exposure of interest. That knowledge can then be applied to an individual plaintiff in a toxic tort suit, provided that the plaintiff's genotype is also known.

It is important to understand that genetic susceptibility to toxicity is not (at least in the overwhelming majority of, if not all, cases) an all-or-nothing phenomenon. It is not that a

¹ The "somatic" mutations that lead to cancer – acquired in particular cells during a person's lifetime—must be distinguished from the interindividual genetic diversity that results from inherited "germline" mutations found in all of a person's cells.

particular toxic exposure inevitably causes people with some genotypes to get sick while being absolutely harmless to people with other genotypes. Rather, genetic variability may place some people at greater risk than other people of developing disease as a result of the same exposure to a toxic agent. The methods for finding and quantifying these differences in risk are fundamentally epidemiologic—even though a person’s genotype is an individual property determined at a molecular level. Do you see why?

The key insight is that these studies test the hypothesis that genotype constitutes a risk factor for toxic effect. This is precisely analogous to the traditional epidemiologic hypothesis that a toxic exposure is a risk factor for illness. In traditional epidemiology, if some members of the study group have been exposed to an additional risk factor—say, tobacco smoke in a study of the effect of asbestos exposure on lung cancer risk—researchers may stratify the study group to assess the effect of each risk factor individually as well as the interaction of the risk factors. The same approach enables molecular epidemiologists to assess the interaction of genotype and exposure as risk factors for disease.

Investigations into the suspected link between cigarette smoking and breast cancer provide an example. Conventional epidemiologic research failed to demonstrate this link despite strong reasons to suspect its existence. Genomic investigations observed that variations in the *NAT2* gene, which codes for a carcinogen-neutralizing enzyme,¹ dramatically influenced the breast cancer danger from smoking. Women whose genes coded for the most protective form of the enzyme had no increased risk of breast cancer even if they smoked, but women smokers with less protective forms of the gene were eight times more likely to get breast cancer than were women with the same genotype who did not smoke. It remains true, however, that not all women smokers with the less protective genotype will develop breast cancer, some women smokers develop breast cancer even though they do not have that genotype, and some women develop breast cancer even though they neither smoke nor have that genotype. What does this tell you about the role of the *NAT2* gene and smoking with respect to breast cancer?

As another example, consider the still-unsettled question whether the commonly-used solvent trichloroethylene (TCE) causes a type of kidney cancer called renal cell carcinoma (RCC).² Before genomic techniques became available, classical epidemiologic studies associated high, long-term, occupational exposure to TCE with an increased risk of RCC. The relative risks found in these studies were generally greater than 1 but less than 2. Toxicologists figured out some of the chemical pathways that metabolize TCE in the body, which are thought to produce biologically active and potentially carcinogenic metabolites. One of these pathways involves enzymes that are coded by genes that are polymorphic, with some alleles that produce

¹ An enzyme is a protein that catalyzes a chemical reaction. For example, the process of chemical digestion—breaking down carbohydrates, proteins and fats into simpler components that can be used by the body—requires enzymes produced by the salivary glands, stomach, pancreas, and small intestine.

² The suspected causal connection of TCE exposure to RCC, and especially the dose-response curve for this suspected toxicity, are controversial because TCE has been very widely used and is ubiquitous in the environment, so a great many people have been exposed to at least small doses of TCE.

functional enzymes and some alleles that do not. Knowing this, a research group in Germany added a molecular epidemiology component to their work. They compared a group of workers who were occupationally exposed to TCE and had RCC to another group of workers who were similarly exposed but did not have cancer (a study that fits into which category of epidemiologic research?) and analyzed the workers' genotypes for two genes involved in the pertinent metabolic pathway.

For each of the two genes studied, the researchers found that the odds of having RCC were higher for workers who had at least one copy of an allele that coded for a functional protein. For one gene the odds ratio was 2.7 (95% confidence interval 1.18 to 6.33). For the second gene, the odds ratio was 4.2 (95% confidence interval 1.16 to 14.91). Were the higher odds ratios statistically significant? This study, published in 1997, illustrates the potential of molecular epidemiology to refine the risk estimates derived from classical epidemiology by identifying genetic variations that affect the degree of risk conferred by toxic exposures.

On the other hand, the link among TCE exposure, genotype, and RCC also illustrates how shifting to a molecular scale does not eliminate the methodologic issues of observational epidemiology. In 2003, the same research group performed a somewhat larger hospital-based case-control study of occupational TCE exposure and RCC risk. That study found a statistically significant association. After adjusting for age, gender, and smoking, workers whose longest-held job was in an industry with TCE exposure had an odds ratio for RCC of 1.80 (95% confidence interval 1.01 to 3.20) as compared to workers whose longest-held job involved no TCE exposure; workers who had ever worked in metal degreasing had an odds ratio of 5.57 (95% confidence interval 2.33 to 13.32). But when the researchers checked the study participants' genotypes, they did not observe a statistically significant association and concluded that their research "does not confirm the working hypothesis of an influence" of the studied genotypes "on renal cell cancer development due to high occupational exposures to trichloroethylene."¹ The authors noted that variable genetic susceptibility to TCE toxicity, if found, could have implications for whether the workers with RCC are eligible for compensation under German law. How would you interpret the results of these studies overall?

Another complication is that single genetic variations may not influence toxic susceptibility independently. Interactions with other genes or with other environmental factors may also play a role. A review of studies of genetic susceptibility to health effects of air pollution, for example, observed that the studies produced conflicting results. The review examined seven potentially relevant genes involved in antioxidant activity, but noted that many other potentially important genes exist and concluded that because antioxidant mechanisms are complex, it is unlikely that any one polymorphic gene has a large effect on susceptibility. Non-genetic factors also play a role, so four-way interactions (among a given gene, pollutants, other genes, and environmental factors other than pollution) are real possibilities. The

¹ Bernd Wiesenhütter et al., *Re-assessment of the Influence of Polymorphisms of Phase-II Metabolic Enzymes on Renal Cell Cancer Risk of Trichloroethylene-exposed Workers*, 81 INT'L ARCHIVES OCCUPATIONAL ENVTL. HEALTH 247, 247 (2007).

Environmental Genome Project has identified nearly 90,000 variations in more than 600 genes believed to be involved in responses to environmental exposures.

Nevertheless, studies of genetic variation in susceptibility to the harmful effects of toxic exposures have begun to be used in American toxic tort cases. Consider the following example, which is adapted from expert witness reports filed on behalf of the defendant in 2015 in a then-pending toxic tort suit.

Plaintiff v. Oil Co.: Expert Report by T. Toxicologist
State Superior Court.
Docket # YY-NNNN.

[The expert's qualifications are omitted].

Plaintiff, a man in his thirties, has been diagnosed with Acute Myeloid Leukemia (AML). Plaintiff alleges that he contracted AML as a result of occupational exposure to benzene in gasoline manufactured by Defendant, Oil Co. Exposure to benzene can cause AML, but Defendant disputes that exposure to benzene at the concentrations present in gasoline can cause AML. That dispute concerns general causation and is beyond the scope of this report. Defendant also disputes that exposure to benzene in gasoline caused Plaintiff's AML. That issue of specific causation is the subject of this expert report. In particular, I was retained by counsel for Defendant to evaluate and report on Plaintiff's genetic susceptibility to the leukemogenic effects of benzene. [Other issues addressed in the expert report are omitted].

To prepare this evaluation I asked a genomic sequencing laboratory to review the scientific literature for research that has associated genetic polymorphisms with susceptibility to benzene toxicity and then to analyze Plaintiff's genome to determine the Plaintiff's genotypes with respect to those genes. The availability of normal, non-cancerous tissue from the Plaintiff was essential for this study. We were able to obtain formalin-fixed, paraffin-embedded blocks of tissue taken from a colon biopsy of Plaintiff several years before Plaintiff's cancer diagnosis. These specimens were then analyzed using Next Generation Sequencing [NGS], a technology that allows rapid analysis of the whole genome even with small amounts of DNA. [The methodologic details of the analysis, including chain-of-custody documentation for the tissue samples, are omitted, although it is worth noting that DNA sequencing is a multi-step process and the laboratory expressed differing degrees of confidence in the analysis of different DNA segments, meaning it is possible that the analysis missed some variations in the Plaintiff's genes or that some detected variations might have been spurious].

Genomics is the generally accepted scientific study of the genome, a person's DNA and RNA. Epidemiology studies populations and attempts to extrapolate findings back to an affected person with a disease. Genomics starts with the affected person and reveals a personal molecular record of health, disease, and potential future quantifiable disease risks. Unlike epidemiology studies on populations that use the group to extrapolate to the person, NGS uses the molecular person of Plaintiff and compares to the group.

Benzene must be metabolized for it to become carcinogenic. The ability to metabolize benzene varies by individual. It remains unclear what role various metabolites play in the carcinogenicity of benzene, but investigators have suggested several likely pathways. Persons with inherited susceptibility to benzene hematotoxicity due to a polymorphism in benzene metabolizing genes may be at greater risk of AML compared to the population without such polymorphisms. In addition, variations in genes that

code for proteins involved in DNA repair and maintenance may affect susceptibility to benzene toxicity.

Upon manual curation of the automated sequencing results of Plaintiff's genome, a scientifically acceptable level of evidence was found for seven variants in six genes involved in benzene metabolism, detoxification, and repair pathways. Six published studies were used in a comparative analysis of Plaintiff's specific genotypes to determine if he had an inherited susceptibility to develop benzene related hematotoxicity at the claimed range of benzene exposure concentrations.

1. The first study involved the gene *NQO1*. Benzene is metabolized by the liver enzyme CYP2E1 to benzene oxide which spontaneously forms phenol and is itself further metabolized by CYP2E1 to hydroquinone. Hydroquinone and related hydroxy metabolites are converted in the bone marrow to benzoquinones which are hematotoxic and genotoxic compounds. The gene *NQO1* codes for an enzyme that converts these back to less toxic metabolites. There exists an allele of this gene in which a variation of one DNA base alters one amino acid in the encoded protein that renders the enzyme inactive. This study found that persons with two copies of the altered allele have a 2.4-fold risk of benzene poisoning, a risk factor for acute non-lymphocytic leukemia (a class of leukemia that includes AML), compared with persons who have one or no copy of the altered allele. Plaintiff has one copy of this genetic variant, indicating that he has no increased risk to benzene hematotoxicity.

2. The second study examined a group of Chinese workers occupationally exposed to benzene who suffered from benzene poisoning, which is considered a risk factor for benzene-induced lymphoma. This study focused on three genes involved in benzene metabolism: *NQO1*, *GSTT1*, and *GSTM1*. This study found a 2.82-fold increased risk of benzene poisoning (Odds Ratio 2.82, 95% confidence interval 1.42 to 5.58) for individuals with two copies of the same altered *NQO1* allele studied in #1 above, as compared to individuals with one or no copies of the altered allele. The study also found that individuals with a "null" genotype for the *GSTT1* gene had a 1.91-fold increased risk of benzene poisoning (Odds Ratio 1.91, 95% confidence interval 1.05 to 3.45) as compared to individuals with a non-null genotype. Finally, this study found a 20.41-fold increased risk of benzene poisoning (Odds Ratio 20.41, 95% confidence interval 3.79 to 111.11) for individuals who had all of the following genetic variations: two copies of the variant allele of *NQO1*, the null genotype for *GSTT1*, and the null genotype for *GSTM1*. Plaintiff has one copy of the *NQO1* variant allele and the non-null genotype for both *GSTT1* and *GSTM1*. Thus Plaintiff has no increased risk for benzene poisoning based on this study.

3. The third study (really a group of studies) examined a gene called *MPO*. *MPO* codes for an amino acid sequence that, after further biochemical processing, is incorporated into a protein that may have antimicrobial benefits but also can catalyze the conversion of pre-carcinogens like benzene to carcinogenic byproducts. A variant allele of *MPO*, however, results in reduced expression of the *MPO* gene and therefore possibly reduces the production of carcinogenic compounds. Several studies have examined this effect for benzene specifically. One study found no effects for the variant allele and benzene poisoning. A second study, involving subjects with fairly low benzene exposure, found that people with one or two copies of the variant *MPO* allele had more white blood cells than people with no copies of the variant allele (reduced white blood cell count being one effect of benzene exposure). A third study found no effect of the variant *MPO* allele on the frequency of chromosome breakage in benzene-exposed individuals (chromosome breakage being a strongly suspected mechanism of benzene-induced leukemogenesis). Plaintiff has one copy of the variant *MPO* allele, indicating a reduced risk to catalyze pre-carcinogens such as benzene into toxic metabolites.

I conclude that these studies are scientifically convincing that Plaintiff did not have any gene-gene polymorphism interaction that would increase his risk for AML if exposed to benzene.

[Other genomic issues considered in the expert's report, as well as portions of the expert's ultimate conclusions related to those issues, are omitted.]

Notes and Questions

1. This excerpt from an expert's report must be considered in its litigation context. The report was submitted by an expert retained by counsel for the defendant. The plaintiff, of course, bore the burden of proving that his exposure to defendant's benzene caused his AML. The defendant therefore perceived value in undermining any inference that the plaintiff was particularly genetically susceptible to benzene's carcinogenic effects. (Do you see why that is the case?) The authors of a 2010 article assessing research needs in environmental carcinogenesis concluded that "[a]ddressing the role of genetic susceptibility to carcinogenic exposures is . . . important; however, the stable and reproducible associations are few."¹
2. In another portion of the expert's report, the expert also concluded that the plaintiff had an "inherited predisposition" to AML because of a variation in another gene that codes for a protein not involved in benzene metabolism. That variation, the expert opined, places the plaintiff at "high associated risk for AML unrelated to benzene." We explore this type of argument in the next section.

3. Using Genetics to Refine the Probability of Causation: Individual Inherited Susceptibility to Disease as a Competing Cause

In the preceding section we considered the possibility that a person's inherited genome could affect her or his probability of contracting a disease after a given exposure to a toxin. Toxic exposure, however, is not necessary for genes to affect health. Some genotypes create health consequences on their own. For example, as described in the prior section, a person who inherits two copies of the allele for the sickle trait in the gene that codes for hemoglobin will have the symptoms of sickle cell disease.

Inheritance of a genetic disease can be considerably more complicated, even for diseases caused by changes in a single protein coded by a single gene. Cystic fibrosis is a case in point. In 1989, researchers announced the sequencing of "the cystic fibrosis gene," which codes for a protein involved in transporting molecules across cell membranes. A person who inherits from each parent a copy of a variant allele that results in the omission of one amino acid from the protein will have cystic fibrosis. But the gene in question is highly polymorphic. Researchers have identified more than 1,600 different mutations that can produce cystic fibrosis. The mutations are located in various parts of the gene, affect the protein in various ways, and

¹ Elizabeth A. Ward et al., *Research Recommendations for Selected IARC-Classified Agents*, 118 ENVTL. HEALTH PERSP. 1355, 1356 (2010).

produce disease of varying severity. Some of the altered versions of the protein work less well (and cause worse symptoms) than would be anticipated based on their structural changes alone, because “helper proteins” coded by other genes recognize the altered protein as incorrect and destroy it before it can be put in position to do its job.

Disease-causing genetic variations with extremely high penetrance, such as the sickle cell allele or the various cystic fibrosis alleles, are relatively rare. For the types of complex, chronic diseases such as cancer that often figure in toxic tort cases, inherited genetic variations typically modify risk to varying degrees rather than leading invariably to disease or providing absolute protection against disease. The so-called “breast cancer susceptibility genes,” *BRCA1* and *BRCA2*, which code for proteins involved in DNA repair, are examples that have been known since the mid-1990’s. Certain alleles of these highly polymorphic genes are considered to have high penetrance—that is, they confer relatively high degrees of risk of breast cancer (as well as some other cancers). Attempts to estimate exactly how much risk have produced varying results, but a meta-analysis of multiple studies published in 2007 estimated mean lifetime breast cancer risks of 57% (95% confidence interval 47% to 66%) for *BRCA1* mutation carriers and 49% (95% confidence interval 40% to 57%) for *BRCA2* mutation carriers. These risks compare to estimates that 12.3% of all women will be diagnosed with breast cancer during their lifetime. Exactly how much extra risk results from these genotypes seems to be influenced by other factors, including a person’s genotype for genes other than *BRCA1* and *BRCA2*.

Suppose a plaintiff with breast cancer alleged that an environmental exposure—say, hormone replacement therapy drugs, tobacco smoke, or childhood exposure to the pesticide DDT (all of which are suspected, based on evidence of varying quality, to increase the risk of breast cancer)—caused her illness. Would knowing the plaintiff’s genotype for *BRCA1* and *BRCA2* be relevant to the causation element of her case? Suppose it turned out that the plaintiff had a high-risk mutation in *BRCA1*. How should that affect the outcome?

The defense in our hypothetical would argue that the plaintiff’s own genetics, rather than the plaintiff’s toxic exposure, caused the breast cancer—that even absent the plaintiff’s exposure, she still would have developed breast cancer because of her genetic predisposition to the disease. This argument treats the plaintiff’s genes as a competing cause of, or an alternative explanation for, the plaintiff’s illness. Can you identify the critical assumption underlying this argument? How is the role of the breast cancer plaintiff’s *BRCA1* genotype different from the role of the AML plaintiff’s *NQO1* genotype? What about the gene for the protein alpha-1-antitrypsin (AAT), discussed in *Lindquist, supra*, that is implicated in one to three percent of emphysema cases?

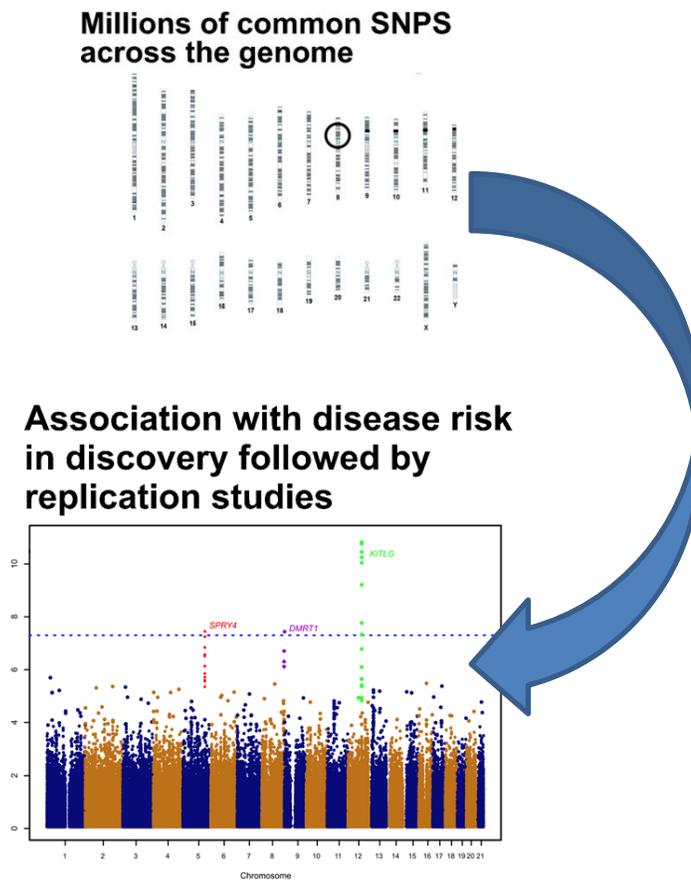
Until recently, studies of inherited susceptibility of disease almost always began by identifying families with unusually high incidence of the disease. Comparing the genes of family members with the disease to the genes of family members without the disease allowed researchers to zero in on certain DNA regions for further investigation. This is how the *BRCA1* gene was discovered. As a practical matter, this research design is useful only to identify high-penetrance genes, which account for only a small fraction of the cases of complex disease. As genome sequencing has become faster and more affordable, researchers have increasingly

shifted to genome-wide association studies (GWAS), which obtain data on polymorphisms throughout the genome of large numbers of subjects and assess which DNA regions contain polymorphisms that are associated with an increased risk of disease. Can you see how this method is similar to traditional epidemiologic research?

In a GWAS, literally millions of polymorphisms are checked for association with disease. Recall that epidemiologists normally consider an observed association to be statistically significant if there is no more than a 5% chance that an association at least that large would have occurred by chance. What would happen if the same standard ($\alpha=.05$) were used for statistical significance in a GWAS? To avoid generating large numbers of false positives by random chance, GWAS researchers typically use much more stringent standards of statistical significance, requiring associations with a probability of being generated by random error that is on the order of 1 in 1 million to 1 in 100 million.

Notwithstanding strict tests of statistical significance, GWAS often identifies numerous DNA regions (“loci”) in which polymorphisms are associated with increased risk of disease. These associations may occur in genes (DNA segments involved in coding for proteins) or they may occur in DNA regions that control gene expression. Many of these polymorphisms are relatively common and are associated with risk increments that are relatively small. After they are identified, further research is required to identify with precision the gene or noncoding DNA segment involved and to estimate (using conventional statistical testing) the magnitude of the effects of individual genotypes. Figure V-5 illustrates, in greatly simplified form, the basic concept of GWAS research. In the “Manhattan plot” on the bottom, locations on the chromosomes are spread along the x axis; each dot represents a single-nucleotide polymorphism (SNP), and a SNP’s position on the y axis shows the statistical strength of the association between that polymorphism and the risk of the disease being studied.

Figure V-5. The Basic Concept of GWAS Research



SOURCE: Adapted from Roelof Koster and Stephen J. Chanock, "Hard Work Ahead: Fine Mapping and Functional Follow-Up of Susceptibility Alleles in Cancer GWAS," *2 Current Epidemiology Rep.* 205, 207 (2015). © Springer International Publishing AG (outside the USA) 2015. With permission of Springer.

As with pregenomic epidemiologic studies, the associations identified by GWAS must be assessed for causality. Some researchers suspect that many of these associations do not reflect truly causal genetic contributions to disease risk. Mechanistic understanding of the molecular pathways involved can provide important evidence of causation, but new associations are being reported much more quickly than their underlying biology can be elucidated.

Questions remain even if researchers reach the conclusion that an observed association between genotype and disease risk indicates a true causal relation. Often, more than one gene is implicated. For example, a review of breast cancer genetics published in 2008 listed risk-conferring alleles of some half-dozen genes. Each allele was fairly common and individually was associated with only a small increase in risk. The authors estimated, however, that someone with two copies of all of these higher-risk alleles would face 6 times the lifetime risk of breast

cancer as somebody with none of these alleles. In this and other cases, it is not clear whether the higher-risk alleles of each gene independently add to the likelihood of disease or whether there is interaction that makes combinations of higher-risk alleles confer more than the sum of the individual risks. (For more on interaction, see Section V. D., *supra*).

Genetic risk modification may go both ways. If some alleles confer relatively higher risk, others may be associated with reduced risk of disease. For apparently protective genes too, the questions are whether observed associations are truly causal and whether biological mechanisms can be worked out.

To think about how the results of GWAS research might affect a toxic tort case, imagine a plaintiff who alleges that her or his disease was caused by a toxic exposure. Assume that the plaintiff can prove, with sufficient epidemiologic evidence, that the exposure in question more than doubles the risk of the disease, and that sufficient confirmatory evidence exists for the association to be considered causal. Now suppose the plaintiff is found to carry several alleles that have been associated with modest increases in risk for the plaintiff's disease. What issues does this finding present for the plaintiff's case? In what way are those issues similar to or different from the issues presented by genotypes with high penetrance?

In this hypothetical, the epidemiologic and GWAS research tend to show that both the toxic exposure and the risk alleles are risk factors for the plaintiff's disease. It would be important to know whether the risks they confer are independent or whether they interact. One way that this might be evaluated would be to compute how the toxic exposure affects risk for people of the various genotypes and to compute how the risk conferred by the various genotypes varies with toxic exposure. It would also be useful to understand the mechanism by which the different alleles increase risk—and whether those mechanisms are mediated (always or sometimes) by the toxic exposure. These types of information are rarely available in the epidemiologic or genomic research, however.

Risk alleles in general, and low-penetrance risk alleles in particular, have played a limited role in toxic tort cases to date. Some defendants have been able to obtain plaintiffs' genotype information and have argued that various genes predisposed the plaintiff to the disease in question. More typical (at least so far) are claims that the plaintiff's condition really has a genetic rather than a toxic origin. For example, in many cases alleging that vaccines cause autism or other neurologic conditions, courts have accepted a growing scientific consensus that these conditions result from genetically regulated events during brain development, even if the causative genetic defect cannot be identified in a particular plaintiff. Such arguments tend to conflate general and specific causation. Consider how they work in the following case excerpt.

Bowen v. E.I. Du Pont de Nemours & Co., Inc.
Superior Court of Delaware, 2005.
2005 WL 1952859, *aff'd*, 906 A.2d 787.

Toliver, J.

Factual Background

[T]he plaintiffs are eight minor children and their parents who have alleged that the children suffered injuries manifested at birth as a result of the exposure of the children's mothers to an agricultural product sold under the trade name of Benlate. Benlate was manufactured by the defendant, the DuPont Company. More specifically, the plaintiffs contend that the mothers of the children were dermally exposed to Benlate during the early stages of their pregnancies. Once deposited, the Benlate was alleged to have passed thru the skin to the developing fetus via the placenta where it acted to retard fetal growth and cell development. The product, which the plaintiffs allege is a human teratogen, was being used as directed at the time of the exposure.

The exposure and births in question are alleged to have taken place between 1984 and 1995. * * *
* The injuries the children suffered which the plaintiffs attribute to Benlate include anophthalmia and microphthalmia¹ as well as other forms of arrested development, physical, emotional and intellectual.

The defendant denies that Benlate is a human teratogen or that it otherwise was responsible for the problems experienced by the plaintiffs. Those problems, the defendant contends, were caused by factors independent of the defendant and Benlate. * * *

Benlate is described as a fungicide developed by the defendant primarily for commercial agricultural use and is designed to prevent and cure fungal infections in plants and crops. The defendant first placed the product on the market for sale in 1970. Although it was only sold commercially in the United States, the product was available for purchase for home use outside the United States, and in particular, in the United Kingdom and New Zealand where the exposures complained about herein took place. The sale of Benlate was halted and it was withdrawn from all markets in 1995.

* * * On April 27, 2004, this Court * * * ordered the cases grouped in pairs, resulting in four trials. The claims made by and on behalf of Emily Bowen and Darren Griffin were to be tried first * * *.

B. Motion to Exclude Plaintiffs' Expert Witnesses Based Upon DRE 702

As was to be expected, both sides retained numerous experts to provide assistance in preparing the case for trial generally as well as for purposes of testifying at trial concerning general and specific causation. * * * The plaintiffs engaged * * * experts in the fields of genetics, teratology, toxicology, dermal exposure and dermal absorption * * *. They are Dr. Charles V. Howard, Dr. David L. MacIntosh, [and] Dr. Michael A. Patton. * * *

* * *

The defendant has contended from the start of this litigation that Emily Bowen's injuries and condition constitute CHARGE Syndrome, which is generally thought to be genetic, as opposed to

¹ Children afflicted with anophthalmia are born with no eyes and those suffering from microphthalmia are born with very small eyes.

environmental, in origin.¹ The plaintiffs disputed this contention and initially offered the testimony and opinions of Dr. Patton. Dr. Patton's qualifications as an expert in the field of genetics in this case are not questioned by the defense.

Based upon his initial examinations and review of her medical records and related information, Dr. Patton concluded in 2002 and in 2003 that Emily Bowen's features did not constitute CHARGE Syndrome. Dr. Patton agreed with two other physicians that had seen her during this period of time, that Emily Bowen did not meet enough of the criteria that would make such a diagnosis appropriate. As a result and given the state of the science at that time, he concluded that her problems did not have any recognizable root in genetics. However, he acknowledged that if his findings relative to her physical condition or the state of the science changed, his opinion could change.

* * * Dr. Howard is a medical doctor and lecturer at the University of Liverpool in Liverpool, England, where he received his medical training from 1965 to 1970. He began at that institution in 1971 and assumed his current position as a senior lecturer in 1991 in the Department of Human Anatomy and Cell Biology. * * * Dr. Howard belongs to several professional organizations, including the British Society of Toxicological Pathologists and the Society for Developmental Pathology. He considers himself a toxicologist and a fetal pathologist, and is not, by his own admission, an expert in genetics.

Dr. Howard, relying on the initial opinions of Dr. Patton, i.e., that Emily Bowen's birth defects did not constitute the "CHARGE Syndrome", ruled out genetics as a cause. Given that conclusion and Dr. McIntosh's [sic] findings relative to the amount of Benlate that was dermally absorbed, Dr. Howard, based upon his education, training, research and experience regarding Benlate, concluded that Benlate was a human teratogen to which Emily Bowen was exposed while being carried in her mother's uterus. It was that exposure, he opined, that proximately caused the birth defects experienced by Emily Bowen.

The defendant, based upon DRE 702 in light of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, and its Delaware progeny, moved, on March 23, 2003, to exclude the testimony of Drs. Howard [and] MacIntosh * * *. The motions were taken under advisement.

C. Further Genetic Testing

[B]ased upon newly developed genetic testing methodologies and the results of related testing in the six remanded cases, the defendant moved, on July 12, 2004, to subject Emily Bowen and Darren Griffin to testing for gene mutations that had been cast as causes of conditions similar to those suffered by the instant plaintiffs. That motion was initially denied and the defendant, after supplementing the record, moved the Court to reconsider. Over the plaintiffs' objections, the Court, on October 15, 2004, ordered that the testing take place * * *

In January 2005, the parties became aware of the results of the additional testing. The tests revealed that Emily Bowen's genetic profile contained a gene, CHD7, which had mutated. The geneticists who discovered that mutation as well as those who confirmed its existence, now believe it is the cause

¹ "CHARGE" is an acronym which stands for Coloboma (absence of or defect in ocular tissue), heart defect, atresia of choanae (blockage between back of nose and mouth), retarded growth and development, genital hypoplasia (arrested development) and ear anomalies. Lalani SR, Safiullah AM, Molinari LM, Fernbach SD, Martin DM, Belmont JW. *SEMA3E Mutation in a Patient with CHARGE Syndrome*. J. Med. Genet. 41:99, 2004. According to Dr. Patton's declaration, CHARGE is defined as an association of features or pattern of malformations which occur together more commonly than by happenstance. Dr. Patton also stated that the principal debate seems to have been whether there is a common underlying cause or causes.

of CHARGE Syndrome. While not all individuals with CHARGE Syndrome tested up to that point in time had the aforementioned mutation, it appears that each individual with the CHD7 mutation was diagnosed with CHARGE Syndrome.¹ The defense contends as a result that Emily Bowen not only has CHARGE Syndrome, but that it was caused by the CHD7 mutation which is genetic in origin only. Stated differently, there were no environmental or external causes.²

Two of the plaintiffs' experts, Dr. Howard and Dr. Patton, have responded to the additional test results with conclusions that are different than those originally offered. Dr. Patton, notwithstanding his previous conclusion that Emily Bowen did not exhibit CHARGE Syndrome and that he could rule out genetics as a cause of her afflictions, now believes that the CHARGE Syndrome diagnosis is correct. He further opines that the mutated CHD7 gene played a substantial role in bringing about that condition. However, he could not rule out a teratogenic cause in general or Benlate specifically, because as he conceded, he is not qualified to do so in that he is not a teratologist, a toxicologist or an expert in either field.

By contrast, Dr. Howard, continues to argue that Benlate is somehow the cause of Emily Bowen's problems and now believes that the CHD7 acted together with Benlate to bring about those injuries. In spite of that position, he does concede that it is very likely that Emily Bowen has CHARGE Syndrome. That concession is based upon Dr. Patton's supplemental findings upon which Dr. Howard relied since he has no expertise in the field of genetics. He further acknowledged that Benlate is not responsible for the mutation in question and that he knows nothing about the CHD7 gene other than what he read in one article on the subject, i.e., the Vissers Study.

Although he is able to maintain his view of Benlate as a human teratogen, Dr. Howard is not able to state how or in what percentage or proportion Benlate and the CHD7 mutation act together to produce CHARGE Syndrome in Emily Bowen. Nor is he aware of any testing or studies which confirm or support his theory regarding the interaction between Benlate and the CHD7 mutation.

D. Supplemental and Renewed DRE 702 Motions

On April 11, 2005, the defendant filed several supplemental motions based upon the recent genetic test results and the expert opinions filed in response by the plaintiffs' expert witnesses. * * *

* * * At the conclusion of [oral argument], the Court granted the defendant's motions as to Dr. Patton, Dr. McIntosh [sic] and Dr. Howard. Given those findings, the defendant's motion for summary judgment was also granted as to both plaintiffs. This Court reasoned that without the testimony of those witnesses the plaintiffs could not establish that Benlate was a human teratogen or that it was the specific cause of the injuries being complained of by either plaintiff.

The motion as to Dr. Patton was granted limiting his testimony as requested on grounds of relevance and competency based upon his admitted lack of expertise in teratology and toxicology. * * *

* * *

* * * Dr. Howard was excluded as an expert witness in Emily Bowen's case based upon Dr.

¹ The study first identifying the CHD7 mutation as a cause of CHARGE Syndrome, was presented in the medical journal "Nature Genetics" in its August 2004 edition (hereinafter the Vissers Study"). Vissers, L., Brunner, H., et. al., *Mutations in a New Member of the Chromodomain Gene Family Cause CHARGE Syndrome*, Nature Genetics 36(9): 955, 2004.

² The results of the testing were negative as to Darren Griffin.

Patton's amended opinion that Emily Bowen's injuries could be deemed genetic in origin and Dr. Howard's reliance on Dr. Patton as an expert in that area. Since he could not, given his lack of expertise and/or qualification as a geneticist, provide an opinion resting in genetics or otherwise supporting his post-CHD7 discovery theory that the CHD7 mutation and Benlate acted together, Dr. Howard could not testify as an expert witness as to Emily Bowen via DRE 702.

* * *

Dr. Howard's Testimony Regarding Emily Bowen

In order to establish the cause of a condition, an expert must not only be able to state the cause of a condition, the witness, or the party offering the testimony, must also be able to exclude other possible/putative causes. In scientific circles, this is known as performing a differential diagnosis. It is a commonly accepted method of addressing the issue of the origin or cause of a medical condition. As the Fourth Circuit Court of Appeals stated in *Westberry*, such a diagnosis:

. . . is a standard scientific technique of identifying the cause of a medical problem by eliminating the likely causes until the most probable one is isolated. A reliable differential diagnosis typically, though not invariably, is performed after "physical examinations, the taking of medical histories, and the review of clinical tests, including laboratory tests," and generally is accomplished by determining the possible causes for the patient's symptom and then eliminating each of these potential causes until reaching one that cannot be ruled out or determining which of those that cannot be excluded is the most likely . . . (Citations omitted.)

[*Westberry v. Gislaved Gummi AB*, 178 F.3d 267, 262-63 (4th Cir.1999).]

In the instant case, both sides have referenced this method of addressing the question of causation. The defense argues that the plaintiffs must not only be able to attribute responsibility for Emily Bowen's injuries to Benlate, they must also be able to exclude the most likely cause of Emily Bowen's problems, genetics and CHARGE Syndrome. The plaintiffs state that they did perform a differential diagnosis via the testimony of Dr. Patton and Dr. Howard and were able to establish Benlate as the cause of her problems. That conclusion was based upon the negative results of prior chromosomal based genetic testing. Two years later, as indicated above, dramatic advances had been made thus allowing the more precise testing of Emily Bowen and Darren Griffin ordered here.

When Dr. Patton changed his diagnosis following the CHD7 test results, Dr. Howard could no longer exclude genetics as, in the words of Dr. Patton, a "substantial cause" of the injuries in question. Dr. Howard then amended his opinion that Benlate was the sole cause of Emily Bowen's injuries to conclude that Benlate interacted with the CHD7 mutation to proximately bring about the problems visited upon her. Dr. Howard did so without any expertise in genetics, having very little knowledge about CHD7 or how, when, and to what degree it combined with Benlate to cause the injuries complained about. Moreover, he admitted that his theory has never been tested, peer reviewed or otherwise subjected to professional scrutiny.

The Court's decision to exclude Dr. Howard as a witness in Emily Bowen's case was based in the first instance on DRE 702's requirement that the witness be "qualified." * * * Given the fact that Dr. Howard admits that he is not a geneticist and has no training, education or experience generally, or

specifically, as to CHD7, he is not qualified via DRE 702 to opine relative to any interaction between CHD7 mutation and Benlate. Nor can he perform a valid differential diagnosis excluding CHD7 or genetics as a cause of the injuries visited upon Emily Bowen under the circumstances.

Dr. Howard's amended opinion and proposed testimony was further excluded because it was not reliable and therefore runs afoul of DRE 402 and 702. His theory regarding the interaction between the CHD7 mutation and Benlate as the cause of Emily Bowen's injuries has not been validated by any scientific discipline, study or entity. It has not been the subject of any peer review nor has it been accepted by any relevant scientific community. There was no testing or publication of this theory prior to the discovery of the CHD7 mutation and its link to CHARGE Syndrome. It is readily apparent as a result, that the theory did not arise out of research or testing but was a product of the instant litigation, a factor which supports its rejection.

Lastly, there is no evidence of any cause other than the CHD7 mutation. Dr. Howard is unable to explain how, why, or where the CHD7/Benlate combination works, nor have the plaintiffs been able to otherwise produce any testimony, at least from those qualified to provide it, that there exists a disease or disability producing gene, in this case CHD7, which requires the presence of an environmental agent to manifest itself. The position advocated by the defense is clear—the mutated CHD7 gene was the sole and proximate cause of Emily Bowen's CHARGE Syndrome. That theory has substantial support in the record in that it has been tested, peer reviewed and published, apparently without consequential dissent.

The Court must further conclude that Dr. Howard's revised opinion is not sufficiently tied to the facts of the case so as to assist the jury in resolving any of the issues involved in this case. It is not the product of reliable scientific principles and methods. In short, while it does relate to causation, the proposed testimony is nothing more than an unsupported theory, or "ipse dixit."

[The court also excluded Dr. MacIntosh's proffered expert testimony. Dr. MacIntosh had "calculated the amount of Benlate that would have been absorbed through the skin of the mothers of Emily Bowen and Darren Griffin * * * based upon the testimony provided by the Bowen and Griffin mothers concerning the uncovered areas of their bodies that came into contact with the Benlate spray. He did not attempt to estimate the amount or quantity of the spray * * * and relied completely on the EPA [dermal absorption] model and formula in reaching his conclusions." The court held, however, that Dr. MacIntosh was not qualified to give an expert opinion on such matters, because Dr. MacIntosh admitted that "while he might be an expert in dermal exposure, dermal absorption is a specialized area in which he was not an expert but had only a working knowledge of the subject. Unfortunately for the plaintiffs, there is no authority in support of the proposition that a 'working knowledge' is the equivalent of 'expertise' for purposes of DRE 702 * * *." The court further held that Dr. MacIntosh's "testimony is not reliable, i.e., it was not based upon a relevant methodology. It had not been tested, subjected to peer reviewed publication or been accepted within any recognized scientific community relating to dermal absorption prior to its use here."]

CONCLUSION

For the foregoing reasons, the Court entered the orders relative to Drs. Patton, Howard and McIntosh [sic] on May 9, 2005. It was based upon the May 9 orders that the defendant's motion for summary judgment was granted on that same date. There was no need as a result to proceed to a trial on the merits.

Notes and Questions

1. The Delaware Supreme Court affirmed the Superior Court summary judgment, holding that “the record supports the trial judge’s conclusion that Dr. MacIntosh was not qualified to give a dermal absorption opinion and that the opinion he did proffer was not the product of a reliable methodology. * * * Because Dr. MacIntosh’s opinion is critical to establishing Dr. Howard’s contentions that Benlate specifically caused the children’s birth defects, we need not reach or address the issue of Dr. Howard’s qualifications or methodology.” *Bowen v. E.I. DuPont de Nemours & Co., Inc.*, 906 A.2d 787, 797-98 (Del.2006).
2. Other cases where the plaintiffs have attempted (unsuccessfully) to use biomarker information include *Young v. Burton*, 667 F.Supp.2d 121 (D.D.C.2008); *City of San Antonio v. Pollock*, 284 S.W.3d 809 (Tex.2009); *Snyder v. Sec’y of Dept. of HHS*, 2009 WL 332044 (Fed.Cl.2009).

G. Using Toxicogenomics and Biomarkers as Proof of Specific Causation

The science called toxicogenomics marries the experimental techniques of toxicology to the analytical techniques of genomics. Using laboratory animals or cells or tissues cultured in vitro, researchers can expose genetic material containing many variations of many genes to a suspected toxin and observe any variations in response, or they can compare exposed and nonexposed genetic material and observe any differences. Researchers may study polymorphisms in the DNA sequence of particular genes or non-coding DNA regions, rearrangements and other variations of the structure of genes along a chromosome, perturbations in gene expression, and other biochemical features.

One goal of such studies is to identify biomarkers—biochemical characteristics that reveal a toxic relation of interest. We have already described one type of biomarker in discussing genetic variations that affect a person’s susceptibility to some agent’s toxic effects. The genetic variation is an observable characteristic that provides information about the likelihood that exposure to the agent will produce an illness. It is a biomarker of susceptibility.

Biomarkers of susceptibility may be relevant evidence of causation (or lack of causation) in a toxic tort case, as we have seen. But to answer the most difficult causation questions in toxic torts—whether the plaintiff was exposed to a medically significant dose of the alleged toxic agent, and whether the plaintiff’s disease resulted from that exposure as opposed to some other cause—biomarkers of exposure or effect potentially could be even more powerful evidence. A biomarker of exposure is an observable change that occurs with exposure but is otherwise absent. If an observable change that occurs with exposure but is otherwise absent also is medically significant and harmful, it is a biomarker of effect.

Consider, for example, the claim, discussed in the *Expert Report of T. Toxicologist* at

Section V. F. 2., *supra*, that the plaintiff's exposure to benzene caused acute myelogenous leukemia (AML). AML begins at some point in the process through which certain stem cells divide and differentiate into various types of white blood cells. During that process, metabolites of benzene may interact with DNA to cause errors during the copying of chromosomes. Research has revealed that blood cells of individuals occupationally exposed to benzene are more likely to have particular chromosomal aberrations associated with the development of AML. The observation suggests a biological mechanism of benzene carcinogenicity, tending to confirm classical epidemiologic studies that detected an association between exposure to benzene and incidence of AML. The chromosome aberrations also could provide a marker of benzene exposure or effect.

Chromosome aberrations are but one type of potential biomarker. Particular gene alterations could indicate exposure to or the effect of particular toxins. Patterns of gene expression—the level of protein-making activity of many genes at once—are often influenced by toxins and could be useful biomarkers. The relative abundance or characteristics of other biochemical constituents are potential biomarkers as well. Researchers are investigating all of these possibilities with respect to numerous known or suspected toxic agents.

You can readily see how biomarkers could lead to better fact-finding on causation in toxic tort cases. Litigators and courts may hope that this research will discover that each cause of a given disease leaves a distinctive biological trace, like a tiny molecular flag claiming the disease as its own—which would provide persuasive scientific evidence about specific causation issue in each individual case. To assess the likelihood that this hope will become reality requires some understanding of the methods and output of toxicogenomic research.

As with traditional toxicology, toxicogenomic research may proceed *in vitro* or *in vivo*. *In vitro* research may investigate potential markers by exposing cells or tissues to a toxin and comparing the response, such as gene expression patterns, to that of control cells or tissues that are not exposed to the toxin. *In vivo* toxicogenomic experimentation measures the potential biomarker of interest in live animal models either exposed or not exposed to the substance. Toxicogenomic experimentation on humans would be unethical, but researchers can compare exposed and unexposed groups of persons—perhaps only persons who have been diagnosed with the disease under study—to see if any biomarkers differentiate the groups. What questions would you want answered about these types of studies if a party tried to use them in a toxic tort case?

Scientists searching for biomarkers wish to ensure that the biomarkers are valid. Biomarker validity, from the scientific perspective, entails a number of technical requirements related to the marker's intended use. Despite the large amount of research into potential biomarkers, validation of new markers remains frustrating.

Several validation issues loom particularly large for the use of biomarkers as evidence on the specific causation issue in toxic tort cases. In a general sense, of course, the markers must be analytically valid so the results of a search for them can be trusted. To meet the legal system's needs, however, markers must reliably help to distinguish between "true" and "false" causation claims in people who have already become ill. This implies that a marker must persist

long enough to be detected when the claim is litigated, that a marker's presence demonstrates the alleged causal relation, and that a marker's absence demonstrates that the disease was caused by something other than the suspected cause.

How well the presence of a biomarker demonstrates causation depends on the biomarker's specificity: the probability that the biomarker will *not* be present in cases that are not instances of true causation. A biomarker's presence alone will provide conclusive, deterministic proof of causation only if the marker is perfectly specific to the exposure-disease combination.

Conversely, how well the absence of a biomarker demonstrates the absence of causation depends on the biomarker's sensitivity: the probability that the biomarker *will* be present in cases that are instances of true causation. A biomarker's absence alone will provide conclusive, deterministic disproof of causation only if the marker is perfectly sensitive to the exposure-disease combination.

A perfectly specific and perfectly sensitive biomarker would be a "signature" connecting harm to exposure in the same way that the handful of currently known signature diseases do. No such biomarker is presently known; some may be discovered but it seems unlikely that such high levels of specificity and sensitivity will be typical. In general, biomarker studies accept a trade-off between sensitivity and specificity. Perfect specificity would require that an exposure results in a given harm via a metabolic pathway not shared by other causes and produces a marker that is unique to that pathway and detectable after disease manifestation. Perfect sensitivity would require that exposure results in a given harm via only one biochemical pathway and always produces a marker that can always be detected after the disease has manifested. Many toxins or their metabolites, however, are thought to produce illness by multiple pathways, and many pathways are thought to be shared by multiple toxins.

To illustrate the difficulty in finding biomarkers that are highly specific and highly sensitive, consider again the connection between benzene exposure and leukemia. Studies have found some chromosomal aberrations that occur at much higher frequencies in leukemias of patients with known occupational benzene exposure, but they also occur in the control groups of these studies. Furthermore, a number of different aberrations have been associated with benzene exposure. Despite the progress in chromosomal and genetic study of leukemia in people exposed to benzene, the search for a signature biomarker continues. More generally, at the very least we can say it remains to be seen whether large numbers of biomarkers of signature specificity and sensitivity are prevalent in our cells just waiting to be found.

On the other hand, the lack of signature biomarkers does not mean that biomarker studies will be useless for the resolution of toxic tort causation disputes. Valid biomarkers of exposure or effect would provide probabilistic rather than deterministic evidence, but they would still be relevant. For example, it would be relevant to know if a plaintiff's leukemia exhibited chromosome aberrations associated with benzene exposure even if there were some probability that a benzene-caused leukemia would not exhibit those aberrations and even if there were some probability that a leukemia exhibiting those aberrations would not have been caused by benzene exposure.

Another example already seen in court opinions involves claims that certain types of breast cancer were caused or accelerated by certain kinds of hormone replacement therapy prescribed to relieve symptoms of menopause. To assist in treatment physicians routinely test these tumors to determine if they are positive for chemical receptors for estrogen and progesterone as well as for excess expression of the gene for another receptor called HER2. Expert witnesses often testify that tumors positive for these receptors are hormone-dependent and therefore more likely to have been caused by the patient's exposure to administered hormones. Breast cancers associated with variant *BRCA1* genes, by contrast, are more likely to be negative for all three of these receptors.

Even biomarkers that do not allow discrimination of particular toxins may be of some use in toxic tort cases. Some research suggests, for example, that for certain toxin-induced cancers, the pattern of gene expression is discernibly different depending on whether the toxin was of a type that causes cancer by damaging DNA directly or of a type that causes cancer by interfering with other cellular repair and control mechanisms. More generally, some research also suggests that certain chemically-induced tumors have gene expression profiles that differ from examples of the same type of cancer that do not result from chemical exposure.

If valid biomarkers of causation are found, they will be probative in the sense that they alter the probability that an element of the plaintiff's case—causation in fact—is true. But, absent perfect specificity and sensitivity, they will provide probabilistic evidence in a population-based way similar to the probabilistic evidence provided by epidemiologic and molecular epidemiologic studies. Consider what you have learned about biomarkers as you read the abstract of Brauch, H., Weirich G., Hornauer M.A., Störkel, S., Wöhl, T., and Brüning, T.J., "Trichloroethylene exposure and specific somatic mutations in patients with renal cell carcinoma," *Natl Cancer Inst.* 1999 May 19; 91(10):854-61, available at <http://jnci.oxfordjournals.org/content/91/10/854.full.pdf>. This early research study involved a suspected toxic relation we have already considered (see Section IV. B. 2. B., supra): exposure to trichloroethylene (TCE or, in the excerpt that follows, TRI) and kidney cancer.

Notes and Questions

1. The excerpted study was small: it had only 44 subjects in the exposed group and 204 subjects in the two control groups. A subsequent study on a different group of subjects by a different research group failed to find the mutational "signature" that the excerpted study's results suggested might exist. Barbara Charbotel et al., *Trichloroethylene Exposure and Somatic Mutations of the VHL Gene in Patients with Renal Cell Carcinoma*, 2 J. Occupational Med. & Toxicology 13, at *6 (2007), <http://www.occup-med.com/content/pdf/1745-6673-2-13.pdf>.
2. Even if valid biomarkers of sufficient sensitivity and specificity provide relevant evidence of whether a plaintiff's exposure to a particular toxic agent caused that individual plaintiff's disease, other difficult questions concerning cause-in-fact may remain. Suppose a plaintiff with renal cell carcinoma had been exposed to TCE

- manufactured by several different companies. Even if genomic analysis could help prove that TCE exposure caused the plaintiff's cancer, would the genomic analysis help to prove which manufacturer's product caused the plaintiff's cancer? This issue, sometime dubbed the "indeterminate defendant" problem, has already figured prominently in toxic tort cases involving signature diseases, such as exposure to asbestos from multiple sources followed by development of asbestosis or mesothelioma. Courts have taken diverse approaches to this difficult problem. For some examples, see *Borel v. Fibreboard Paper Prods. Corp.*, 493 F.2d 1076 (5th Cir. 1973); *Lohrmann v. Pittsburgh Corning Corp.*, 782 F.2d 1156 (4th Cir. 1986); *Thacker v. UNR Inds., Inc.*, 603 N.E.2d 449 (Ill. 1992); *Rutherford v. Owens-Illinois, Inc.*, 941 P.2d 1203 (Cal. 1997); *Gregg v. A-J Auto Parts, Co.*, 943 A.2d 216 (Pa. 2007); *Sienkiewicz v. Greif (UK), Ltd.*, [2011] UKSC 10, [2011] 2 A.C. 229 (appeal taken from Eng.); *Bostic v. Georgia-Pacific Corp.*, 439 S.W.3d 332 (Tex. 2014). These issues are discussed in, for example, Michael D. Green, *Second Thoughts About Apportionment in Asbestos Litigation*, 37 Sw. U. L. REV. 531 (2008); Joseph Sanders, *The "Every Exposure" Cases and the Beginning of the Asbestos Endgame*, 88 TUL. L. REV. 1153 (2014); Steve C. Gold, *Drywall Mud and Muddy Doctrine: How Not to Decide a Multiple-Exposure Mesothelioma Case*, 49 IND. L. REV. 117 (2015).
3. For discussions of the potential role of biomarkers in toxic tort litigation, see Gary E. Marchant, *Genetic Data in Toxic Tort Litigation*, 45 BRIEF 22 (Winter 2016); Steve C. Gold, *When Certainty Dissolves Into Probability: A Legal Vision of Toxic Causation for the Post-Genomic Era*, 70 WASH. & LEE L. REV. 237 (2013); Andrew R. Klein, *Causation and Uncertainty: Making Connections in a Time of Change*, 49 JURIMETRICS J. 5 (2008); Jamie A. Grodsky, *Genomics and Toxic Torts: Dismantling the Risk-Injury Divide*, 159 STAN. L. REV. 1671, 1672 (2007).

H. The Sufficiency of Relative Risks ≤ 2.0

1. The Basic Logic Recalled

Based on the reasoning explained in subsection B supra, if the outcome of a study finds a relative risk of 2.0 or less, the study would not support finding that specific causation exists. The probability of causation would be 50% or less and therefore not satisfy the standard of proof. Some courts have adopted a rule that a relative risk must exceed 2.0 for specific causation to be established.⁴¹

2. Adjusting the Probability for Specific Individuals

As explained in subsection C above, the results of a study may not accurately reflect the probability of causation in any given individual. If a non-study person has been exposed to a

greater dose (and if there is a positive dose-response curve), then the probability of causation for that individual will be greater than that found in the study with a lower dose. Similarly, if an individual does not have other risk factors to which the study population has been exposed, then the probability of causation for that individual is greater than the study result.

3. Adjustments Based on Heterogeneity and Subgrouping

Subsection D, above, addressed heterogeneity—that different subgroups within a study population (as well as the general population) may differ in their susceptibility to disease (recall Figures 5–7). Some heterogeneity may be due to a difference in risk factors to which those in the study have been exposed. Recall the *Lindquist* case in which the plaintiff seeking workers compensation for his emphysema due to exposure to fire smoke had also been a tobacco smoker. When information is available that permits identification of those subgroups and to which a given individual belongs, that information permits a more refined assessment of the probability of specific causation for that individual. One common method for this adjustment involves an assessment of a specific plaintiff’s risk factors for known competing causes of disease, referred to as a differential etiology (or differential diagnosis), a subject that arose in *Estate of George*, supra, and which we address in the next subsection. Another emerging basis for subgrouping is genetic susceptibility, covered in Section V. F. 2., supra. This subgrouping permits more refined risk assessments both for a subgroup with an identified genetic risk factor and for the subgroup without it (Figure V-3 reflects one possible distribution of disease reflecting different susceptibilities because of genetics or otherwise).

4. Adjustments through a Differential Etiology

The most common effort to employ evidence about a specific individual to adjust the likelihood of causation derived from a study is by elimination of risk factors that could be responsible for the individual’s disease. This is similar to the process of differential diagnoses that physicians regularly use to determine the illness or disease causing a patient’s symptoms. If a physician can rule out an infection as a cause of a person’s springtime nasal congestion, for example, the probability increases that seasonal allergy is the correct diagnosis. The differential diagnosis identifying disease or illness may help guide the physician’s treatment decisions.

The same reasoning process may be applied to inferring the cause of a disease rather than the cause of symptoms: eliminating the possibility that other competing causes could have been responsible for that individual’s disease increases the probability that the environmental agent was the cause. In toxic tort cases, physicians or scientists frequently testify to an opinion of specific causation based on such reasoning, which often is labeled as a “differential diagnosis,” although it might better be called a “differential etiology,” as the objective is not to diagnose a disease but to infer its cause.”

Thus, we might reason, as some courts have, that for some individuals a study result less

than a doubling may still justify a conclusion that the study agent was more likely than not the cause of that individual's disease. For example, imagine that there are three independent risk factors for a disease: agent X, genetic mutation Y, and unknown causes (UC). Let us also assume that the attributable proportion of risk for each of these risk factors is:

$$\begin{aligned} X &= .3 \\ Y &= .6 \\ UC &= .1 \end{aligned}$$

If, through a differential etiology, it can be determined that an individual does not have the genetic mutation Y, then the probability that Agent X caused that individual's disease is greater than 50%. Can you calculate the precise probability? What formula would express the revised probability of causation when a risk representing an APR of Z can be eliminated through a differential etiology? How would unknown causes (and risks) be accounted for?

A differential diagnosis alone cannot, as a logical matter, establish general causation: absent some proof of agent-disease causation, all that a differential etiology demonstrates is what didn't cause the plaintiff's disease, not what did. Even if all known causes of a plaintiff's disease can be ruled out, it does not follow that the toxic exposure caused the plaintiff's illness, that is, that the toxic exposure has been "ruled in." Thus courts have routinely refused to admit differential diagnosis testimony if a plaintiff lacks sufficient admissible evidence of general causation.

5. Accounting for Idiopathic Disease

In addition to eliminating known risk factors where the facts support such, there is the matter of unknown ("idiopathic") causes. For example, the majority of birth defects are idiopathic. Do you see why idiopathic causes are sometimes described as the "soft underbelly" of differential etiologies? At what point would such causes preclude use of a differential etiology alone to ascribe a probability in excess of 50% to the agent alleged to have caused the victim's disease?

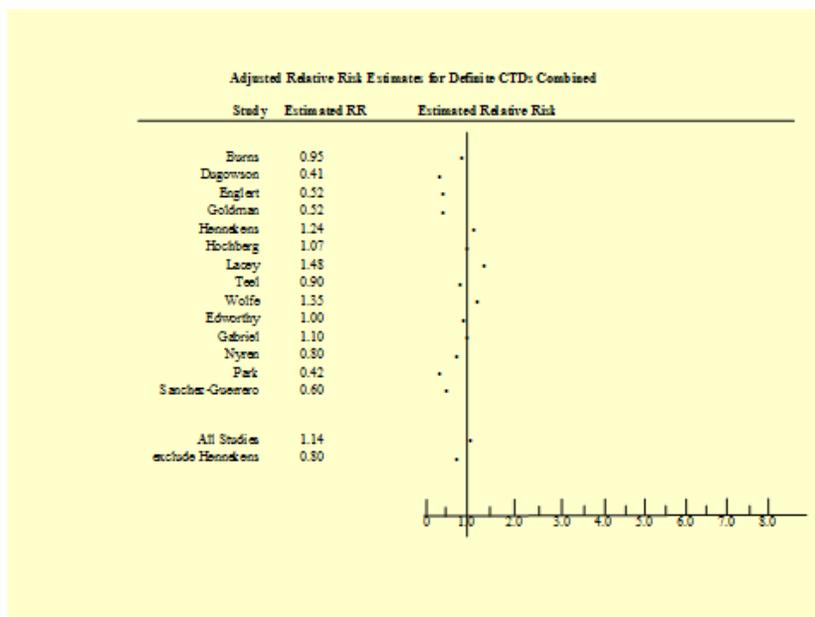
Medical researchers, of course, continue to try to identify the causes of currently unexplained disease. Some of those causes are likely to be genetic, some environmental, and some the result of gene-environment interactions. As we write this, the "ability to sequence genes has gotten ahead of [the] ability to know what it means," as one physician put it. Gina Kolata, *Genetics Often Muddle Options in Cancer Care*, NEW YORK TIMES, March 12, 2016 at A1, A11 (quoting Eric P. Winer). It may not be possible ever to understand all risk factors for all diseases. Nevertheless, genomic and other research will surely, over time, reduce the proportion of disease considered idiopathic—thereby improving public health by permitting new methods of disease prevention and treatment and, incidentally, reducing causal uncertainty in toxic tort cases.

6. Small Relative Risks and General Causation

Above we have explained the logic of refining the probability of specific causation based on individualized or subgroup information when study results find relative risks below 2.0. Notwithstanding that logic, there remains a serious question about these efforts because of general causation. Recall that the results of an observational epidemiologic study can be affected by random error, bias, and confounding. Those sources of error result in a fair amount of noise rather than true causal relationships, and this problem is most acute with small increases in the relative risk. That is why several epidemiologists have stated that unless a study reaches a threshold relative risk of 2.0 to 3.0, it should not be taken seriously⁴² unless repeatedly replicated or other confirmatory evidence exists, such as studies of the agent at higher doses that find a greater risk of the same disease. Do you see why the latter addresses this concern about using studies with relative risks below 2.0?

An illustration of the scattershot, yet almost surely spurious, outcomes that one might find due to noise when there is no real causal relationship is revealed in Figure 9. The data for this figure is adapted from the report prepared by the court-appointed experts in the silicone gel breast implant litigation.⁴³ It reflects all epidemiology studies that were located on the connection between silicone gel breast implants and connective tissue disease.

Figure V-6. Silicone Gel Breast Implants Study Results



SOURCE: Courtesy of the authors.

VI. THE ROLE OF CONSENSUS ORGANIZATIONS

There are a number of organizations that assess the evidence bearing on whether a chemical or other agent is a toxin and present their conclusion and the evidence bearing on the matter to the public.

The best known such agency is the International Agency for Research on Cancer (IARC), a well-regarded international public health agency established in 1965 by the World Health Organization. It evaluates the human carcinogenicity of various agents and other exposures, such as occupational ones, and its target audience is international and national public health agencies, although others rely on IARC work.

In conducting its carcinogenicity assessments, IARC obtains all of the published studies on the matter, including animal studies as well as any human studies. An interdisciplinary group of scientists synthesize and evaluate that evidence (employing a weight of the evidence methodology) to reach an assessment on the potential human carcinogenicity of the studied agent., IARC publishes a monograph containing that evidence and its analysis of the evidence and provides a categorical assessment of the likelihood the agent is carcinogenic. IARC may also provide dose-response information if the available evidence supports such.

In a preamble to each of its monographs, IARC explains its principles and procedures and its use of human epidemiology, animal toxicology, and mechanism and other evidence for its assessments. In its monograph conclusions, IARC employs four categories to summarize the

evidence on carcinogenicity based human (epidemiologic) and animal (toxicologic) studies: (1) sufficient evidence of carcinogenicity; (2) limited evidence of carcinogenicity; (3) inadequate evidence of carcinogenicity; and (4) evidence suggesting a lack of carcinogenicity.⁴⁴ An overall assessment is also provided as human carcinogenicity, which range from the substance is carcinogenic in humans or is probably or possibly carcinogenic through to probably not carcinogenic, and unclassifiable based on the available evidence. When an IARC monograph for a given agent is available, it is generally recognized as authoritative. A graphic with the agents studied by IARC, its conclusion and the year of the study can be found at <http://www.bloomberg.com/graphics/2015-red-meat-cancer/>. Unfortunately, IARC has conducted evaluations of only a fraction of potentially carcinogenic agents, and many suspected toxic agents cause effects other than cancer.

Other consensus organizations also exist. The International Labour Organization, among its activities, publishes a list of agents that cause occupational diseases, including cancer.⁴⁵ In the United States, the National Institute on Occupational Safety and Health, a division of the Centers for Disease Control and Prevention has developed a research agenda regarding occupational safety and health and conducts research in this field as well as funding extramural research.

¹ See generally Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 29 (ALI 2010).

² H.L.A. Hart & T. Honoré, *Causation in the Law* (2d ed. 1985); Richard Wright, *Causation in Tort Law*, 73 Cal L. Rev. 1735 (1985).

³ [T]he actor's negligent conduct is not a substantial factor in bringing about harm to another if the harm would have been sustained even if the actor had not been negligent.

Restatement (Second) of Torts § 432(1) (1965).

⁴ Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 27 (2010).

⁵ See David W. Robertson, *The Common Sense of Cause in Fact*, 75 Tex. L. Rev. 1765 (1997).

⁶ The classic case is *Dillon v. Twin State Gas & Elec. Co.*, 163 A. 111 (N.H. 1932).

⁷ David Hume is credited with first expressing this insight. See David Hume, *A Treatise of Human Nature* bk. I, pt. III, §§ 14–15 (A. Selby-Bigge ed., rev'd P. Nidditch 1978). See also Hart & Honoré, *supra* note 2, at 10–11, 14–15, 44–49.

⁸ [T]he generalizations which are needed to defend [T] particular causal statements are, for the most part, truisms derived from common experience. They concern the effects of impacts, blows, gross mechanical movements, and are often so deeply embedded in our whole outlook on nature that we scarcely think of them as separate elements in causal statements."

Hart & Honoré 15.

⁹ "Scientists know very little about how, in a mechanistic sense, toxic substances cause disease such as cancer or birth defects. Nonetheless, they may know a considerable amount about whether toxic substances cause disease or injury through inferences drawn from statistical associations and other indirect means." Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?*, 7 High Tech. L.J. 189, 209-210 (1992) (emphasis and footnotes omitted).

¹⁰ W. Winkelstein, *A new perspective on John Snow's communicable disease theory*, 142 Am. J. Epidemiol. S3-9. (1995)

¹¹ 509 U.S. 579 (1993).

¹² Peter Huber, *Galileo's Revenge: Junk Science in the Courtroom* (1991).

¹³ 293 F. 1013, 1014 (1923).

¹⁴ 959 F.2d 1349, 1360 (6th Cir. 1992).

¹⁵ Arthur L. Herbst et al., *Adenocarcinoma of the Vagina: Association of Maternal Stilbestrol Therapy with Tumor Appearance*, 284 New Eng. J. Med. 878 (1971).

¹⁶ *Baker v. Chevron USA, Inc.*, 680 F. Supp. 2d 865, 880 (S.D. Ohio 2010) ("[R]egulatory agencies are charged with protecting public health and thus reasonably employ a lower threshold of proof in promulgating their regulations than is used in tort cases."), *aff'd sub nom.* *Baker v. Chevron U.S.A. Inc.*, 533 F. App'x 509 (6th Cir. 2013).

¹⁷ See Thomas O. Mcgarity & Sidney A. Shapiro, *Regulatory Science in Rulemaking and Tort: Unifying the Weight of the Evidence Approach*, 3 Wake Forest J. L. & Pol'y 65, 72, 98 (2013).

¹⁸ Alfredo Morabia, *History of Epidemiologic Methods and Concepts* (2004).

¹⁹ 911 F.2d 941 (3d Cir. 1990).

²⁰ *Id.* 954.

²¹ See, e.g., Leon Gordis, *Epidemiology* (5th ed. 2014); Dona Schneider & David E. Lilienfeld, *Foundations of Epidemiology* (4th ed. 2015).

²² On the lack of pure experiment in clinical trials and the potential for bias in the results of those studies, see Jeremy Howick, *The Philosophy of Evidence-Based Medicine* (2011)

²³ See Michael D. Green, *Bendectin and Birth Defects* 230 (1996).

²⁴ See, e.g., Nathan Mantel, *The Detection of Disease Clustering and a Generalized Regression Approach*, 27 Cancer Research 209 (1967).

²⁵ Herbst et al., *supra* note 15.

²⁶ See Duncan C. Thomas, *Statistical Methods in Genetic Epidemiology* (2004).

-
- ²⁷ See Nyi Naing, *Easy Way to Learn Standardization: Direct and Indirect Methods*, 7 Malaysian J. Med. Sci. 10 (2000).
- ²⁸ See *id.*
- ²⁹ 874 F.2d 307, 312 (5th Cir. 1989).
- ³⁰ The technique of multivariate analysis is beyond the scope of this module. For a discussion designed for those who are unschooled in mathematics, see Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in Federal Judicial Center & National Research Council, *Reference Manual on Scientific Evidence* 303 (3d ed. 2011).
- ³¹ See Mohamad A. Pourhoseingholi et al., *How to Control Confounding Effects by Statistical Analysis*, 5 Gastroenterol. & Hepatol. Bed Bench 70 (2012).
- ³² See Douglas L. Weed, *Epidemiologic Evidence and Causal Inference*, 14 Hematology/Oncology Clinics N. Am. 797 (2000).
- ³³ 533 F. Supp. 581, 584 (N.D. Okla. 1981).
- ³⁴ See Douglas L. Weed & Stephen D. Hursting, *Biologic Plausibility in Causal Inference: Current Methods and Practice*, 147 Am. J. Epidem. 415 (1998) (examining use of this criterion in contemporary epidemiologic research and distinguishing between alternative explanations of what constitutes biologic plausibility, ranging from mere hypotheses to “sufficient evidence to show how the factor influences a known disease mechanism”).
- ³⁵ Letter from Honorable Carl B. Rubin to Michael D. Green (May 9, 1994) (on file with a co-author).
- ³⁶ 306 F.2d 894 (N.C. 1939).
- ³⁷ 98 F.2d 815 (3d Cir. 1938).
- ³⁸ Ofer Shpilberg et al., *The Next Stage: Molecular Epidemiology*, 50 J. Clinical Epidem. 633, 637 (1997).
- ³⁹ See Understanding Science, http://undsci.berkeley.edu/article/basic_assumptions.
- ⁴⁰ See Sander Greenland & James M. Robins, *Epidemiology, Justice, and the Probability of Causation*, 40 *Jurimetrics* J.321 (2000); Sander Greenland, *Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodologic Error That Has Become a Social Problem*, 89 Am. J. Pub. Health 1166 (1999).
- ⁴¹ See, e.g., *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1320 (9th Cir. 1995).
- ⁴² See Samuel L. Lesko & Allen A. Mitchell, *The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology* 599, 601 (Wiley Brian L. Strom ed., 4th ed. 2005) (“it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies”); Gary Taubes, *Epidemiology Faces its Limits*, 269 Sci. 164 (1995) (explaining views of several epidemiologists about a threshold relative risk of 3.0 to seriously consider a causal relationship); N. E. Breslow & N. E. Day, *Statistical Methods in Cancer Research, in The Analysis of Case-Control Studies* 36 (IARC Pub. No. 32, Lyon, France 1980) (“[r]elative risks of less than 2.0 may readily reflect some unperceived bias or confounding factor”); David A. Freedman & Philip B. Stark, *The Swine Flu Vaccine and Guillain-barré Syndrome: A Case Study in Relative Risk and Specific Causation*, 64 *Law & Contemp. Probs.* 49, 60 (2001) (“If the relative risk is near 2.0, problems of bias and confounding in the underlying epidemiologic studies may be serious, perhaps intractable.”).
- ⁴³ See Barbara A. Diamond et al., *Silicone Breast Implants in Relation to Connective Tissue Disease and Immunologic Dysfunction: A Report by a National Science Panel to the Honorable Sam C. Pointer Jr., Coordinating Judge for the Federal Breast Implant Multi-District Litigation* (undated), available at <http://www.fjc.gov/BREIMLIT/SCIENCE/report.htm>; see also Esther C. Janowsky et al., *Meta-Analysis of the Relation Between Silicone Breast Implants and the Risk of Connective-Tissue Diseases*, 342 N. Eng. J. Med. 781 (2004).
- ⁴⁴ E.g., 90 International Agency for Research on Cancer, *Monographs on the Evaluation of Carcinogenic Risks to Humans: Human Papillomaviruses 9-* (2007), available at <http://monographs.iarc.fr/ENG/Monographs/vol90/index.php>.
- ⁴⁵ http://www.ilo.org/safework/info/publications/WCMS_125137/lang--en/index.htm.